VAASAN YLIOPISTO

TEKNIIKAN JA INNOVAATIOJOHTAMISEN YKSIKKÖ

OHJELMISTOTEKNIIKKA

Jannika Rintamäki

SISÄISEN UHAN HAVAITSEMINEN TERVEYDENHUOLLON KÄYTTÖ-LOKEISTA

Diplomityö, joka on jätetty tarkastettavaksi diplomi-insinöörin tutkintoa varten Vaasassa 13.12.2020

Työn valvoja Professori Jouni Lampinen
Työn ohjaaja Professori Jouni Lampinen

SISÄLLYSLUETTELO

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen yksikkö
Tekijä: Jannika Rintamäki
Diplomityön nimi: Sisäisen uhan havaitseminen terveydenhuollon käyttölokeista
Valvoja: Professori Jouni Lampinen
Ohjaaja: Professori Jouni Lampinen
Tutkinto: Diplomi-insinööri
Oppiaine: Ohjelmistotekniikka
Opintojen aloitusvuosi: 2012
Diplomityön valmistumisvuosi: 2020 **Sivumäärä:** 74

TIIVISTELMÄ

Sosiaali- ja terveydenhuollossa on siirrytty käyttämään sähköisiä potilastietoja. Potilasturvallisuuden takaamiseksi laki edellyttää keräämään lokitietoja niiden käytöstä. Käyttölokeista voidaan havaita käyttäjien suorittamaa potilastietojen väärinkäyttöä auditoimalla, mutta tietojen suuri määrä vaikeuttaa niiden manuaalista läpikäyntiä.

Kun suurista tietomääristä yritetään löytää oleellista tietoa, samankaltaisuuksia ja poikkeavuuksia, voidaan hyödyntää tiedonlouhinta- ja koneoppimistekniikoita. Tekniikat ovat tärkeä osa väärinkäytön ja sisäisen uhan havaitsemiseksi kutsuttuja tutkimusaloja. Tutkielmassa etsittiin terveydenhuoltoon sopivia sisäisen uhan havaitsemismenetelmiä, jotka hyödyntävät käyttölokeja.

Tutkimusmenetelmänä havaitsemismenetelmien etsintään käytettiin integroivaa kirjallisuuskatsausta, jonka aineistoon valikoitui 19 laatuarvioitua tieteellistä julkaisua. Sisällytetyt julkaisut vuosilta 2009–2019 kerättiin tietotekniikan alan tietokannoista. Tutkielman keskeisin tulos on itse kirjallisuuskatsaus, jossa esitellään aihealueen aiempia tutkimuksia ja muodostetaan synteesi. Synteesi sisältää tiivistetyn nykytilannekuvauksen sisäisen uhan havaitsemisratkaisuista terveydenhuoltoympäristössä. Toimiva järjestelmä selvittää, viittaako käyttölokitietue, käyttäjä tai potilas väärinkäyttöön. Järjestelmän havaitsemisstrategia hyödyntää yksinkertaisia sääntöjä, hälytysten priorisointia ja vähentämistä, suosittelua, normaalikäytön selitysmallinteita tai läheisyysmittoja. Järjestelmän kannalta tärkeitä tietoja ovat käyttölokit, organisaatio- ja hoitotiedot.

Terveydenhuoltoon sopivien havaitsemismenetelmien löytäminen on mahdollista kirjallisuuskatsauksen avulla, vaikka yhtenäisten hakusanojen muodostaminen tuo haasteita. Katsaus osoitti, että soveltuvien menetelmien kokonaisuus on monipuolinen, ja että niiden avulla havaitsemistyötä on todennäköisesti mahdollista tehostaa. Lisäksi sisäisen uhan havaitsemisen tutkimusala on aktiivinen, joten uusia havaitsemisstrategioita voi löytyä lisää lähitulevaisuudessa. On todennäköistä, että terveydenhuoltoympäristön erityispiirteiden vuoksi tulevaisuudenkin ratkaisut nojaavat vahvasti käyttölokeihin. Jatkotutkimuksissa olisi syytä selvittää menetelmien käytännön soveltuvuutta suomalaisessa terveydenhuollossa olemassa olevien järjestelmien rinnalla.

AVAINSANAT: Käyttölokit, sisäinen uhka, sähköiset potilastiedot, tiedonlouhinta, koneoppiminen

UNIVERSITY OF VAASA
School of technology and innovation
Author: Jannika Rintamäki
Topic of the Thesis: Insider threat detection from health care access logs
Supervisor: Professor Jouni Lampinen
Instructor: Professor Jouni Lampinen
Degree: Master of Science in Technology
Major of Subject: Software Engineering
Year of Entering the University: 2012
Year of Completing the Thesis: 2020 **Pages:** 74

ABSTRACT

Social and health care are using electronic health records. To guarantee the patient safety the law demands they collect log data of the accessed records. Access logs can be used in audit process to detect users who misuse patient records but manual detection is difficult due to large number of data.

Data mining and machine learning techniques are used to discover relevant information, patterns and anomalies from large amounts of data. These techniques are an important part of misuse and insider threat detection research areas. This thesis searched access logs utilizing methods to detect insider threats in health care environment.

Research method to find the detection methods was integrative literature review which resulted in 19 quality assessed research papers. Included papers published during 2009–2019 were found from information technology databases. The most essential result of the study was the review itself, which showcases the former studies of the field and forms a synthesis. The synthesis summarizes the status quo of insider threat detection solutions in health care environment. Functional system detects whether log message, user or patient points to misuse. Approach of the system utilizes simple rules, alert prioritization and reduction, recommendation, normal use explanation templates or similarity measures. Important information for the system are access logs, organization and treatment data.

The literature review is able to find methods suitable for health care even if the formation of common search words is challenging. The study showed that the list of suitable methods is diverse and that improvement of the detection work is plausible. Also, insider threat detection research field is active which makes it possible that new detection strategies are found in near future. Due to the uniqueness of health care environment it is probable that the future solutions also utilize access logs. In future studies the applicability of the methods to the Finnish health care should be investigated parallel to the existing methods.

KEYWORDS: Access logs, insider threat, electronic health records, data mining, machine learning

1 JOHDANTO

Sosiaali- ja terveydenhuollossa potilastietoja tallennetaan ja käytetään sähköisesti potilastietojärjestelmien avulla. Potilastietojärjestelmissä on kyse yhteistoiminnallisista tietojärjestelmistä (eng. collaborative information systems, CIS), joissa käyttäjät kommunikoivat ja työskentelevät yhteisten tehtävien parissa. Yhteinen tehtävä on potilaan hoito, jota käsittelee dynaaminen työryhmä, johon kuuluu hoitajia ja lääkäreitä eri osastoilta. (Chen, Nyemba & Malin 2012a: 332.) Sähköisten potilastietojen (eng. electronic health records) tallentaminen tietokantoihin, tietojen siirtäminen eri järjestelmien välillä ja tiedonhaku on tuonut mukanaan paljon etuja. Esimerkiksi potilastietojen säilyttäminen ja hoitojakson kulku on tehokkaampaa ja tietoihin pääsy useissa sairaalaympäristöissä mahdollistaa laadukkaamman hoidon. Kuitenkin potilastietojen siirtyminen sähköiseen muotoon on lisännyt potilaiden henkilökohtaisten tietojen luvattoman ja laittoman käytön uhkaa. (Ontario 2015: 3–4.)

Tietojärjestelmät sairaalaympäristöissä on suojattu ulkoisten uhkien varalta muun muassa virustorjunnalla, palomuurilla ja käyttäjien identiteetin todennuksella eli autentikoinnilla. Ulkoisten uhkien torjunta ei kuitenkaan poista potilastietojärjestelmien valtuutettujen käyttäjien väärinkäytön uhkaa. Paljastunut potilastietojen väärinkäyttö voi saada potilaan välttelemään jatkossa hoitoon hakeutumista tai jättämään tarvittavien arkaluonteisten tietojen kertomisen hoitotilanteessa. Lisäksi yksikin potilastietojen väärinkäyttötapaus voi huonontaa sosiaali- ja terveydenhuollon palveluntarjoajan mainetta ja heikentää potilaiden luottamusta sosiaali- ja terveydenhuollon ammattimaisuutta kohtaan (Ontario 2015: 5). Yleisinä keinoina väärinkäyttöä vastaan käyttäjille voidaan pitää koulutusta potilaiden yksityisyyden suojasta, solmia salassapitosopimuksia ja lisätä henkilöstön tietoisuutta mahdollisista seuraamuksista väärinkäyttötapauksissa (Ontario 2015: 12). Koulutus, sopimukset ja tietoisuuden lisääminen eivät silti estä väärinkäyttöä eikä tilanteita, joissa toinen henkilö pääsee potilastietoihin käsiksi valtuutetun käyttäjän käyttäjätunnuksilla.

Suomessa sähköisten potilastietojen käsittely on säädetty laissa. Sosiaali- ja terveysministeriön asetus potilasasiakirjojen laatimisesta sekä niiden ja muun hoitoon

liittyvän materiaalin säilyttämisestä (99/2001: 4 §) vaatii, että potilastietoja saa käsitellä ainoastaan potilaan hoitoon liittyvät henkilöt omien työtehtäviensä ja vastuidensa mukaisesti. Lisäksi laki sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä (159/2007: 2 luku 4 §) edellyttää, että potilastietojen saatavuus, käytettävyys ja muuttumattomuus tulee turvata koko niiden säilytysajan. Lain (159/2007: 2 luku 5 §) mukaan sosiaali- ja terveydenhuollon palveluntarjoajan täytyy pitää lokirekisteriä, johon se tallentaa lokitiedot potilastietojen käytöstä ja luovutuksesta. Käyttö- ja luovutuslokirekisterien tiedoista tulee ilmetä mitä ja minkä palveluntarjoajan tietoja on käytetty/luovutettu, kuka tietoja on käyttänyt/luovuttanut, mihin käyttötarkoitukseen ja milloin tietoja on käytetty/luovutettu.

Tietojen mahdollinen väärinkäyttö olisi syytä havaita, koska potilaiden yksityisyys on merkittävä osa potilasturvallisuutta. Lain edellyttämä lokitietojen kerääminen mahdollistaa väärinkäyttötapausten havaitsemisen, mutta se on kuitenkin haastavaa, koska lokitietoja kertyy huomattavan suuria määriä. Lokitietojen suureen määrään on syynä potilastietojen ja niiden käyttäjien paljous. Tehokas tietojen manuaalinen seuranta ja läpikäynti on lähes mahdotonta, joten tietojen väärinkäytön havaitsemiseen olisikin hyvä löytää koneellisia menetelmiä. Tehokkaalla väärinkäytön havaitsemisella taas voidaan ennaltaehkäistä valtuutettujen käyttäjien suorittamaa luvatonta potilastietojen käyttöä.

1.1 Tutkimuksen taustat

Työn aihe on saatu tietotekniikan alan yritykseltä X, joka on erikoistunut toteuttamaan tietokanta- ja verkkopohjaisia ratkaisuja. Yksi yrityksen tuotteista on lokitietojen keruu ja seurantajärjestelmä. Järjestelmällä on useita ominaisuuksia ja käyttötarkoituksia, kuten seurata sovellusten ja palveluiden käyttöä ja varmistaa tietojärjestelmän tietoturvan toteutumista. Tämän työn kannalta oleellinen käyttötarkoitus on ennaltaehkäistä tietojärjestelmän ja sen tietojen väärinkäyttöä.

Järjestelmä kerää lokitietoja useista eri järjestelmistä, jotka sisältävät potilas- ja käyttäjätietoja. Yhdistämällä eri lähdejärjestelmien tiedot, saadaan kokonaiskuva esimerkiksi yksittäisen käyttäjän toiminnasta. Lokeissa on saatavilla tietoja kuten tapahtuman ajankohta, kuka tietoja katsoi, kenen tietoja on katsottu, käytetty sovellus, työasema ja käyttötarkoitus. Lokitietojen lisäksi saatavilla on myös hoito- ja työaikatietoja. Järjestelmä kykenee havaitsemaan potilaiden väärinkäyttöön viittaavia poikkeamia toimintaympäristön asiantuntijoiden kanssa muodostettujen sääntöjen avulla. Poikkeamia ovat esimerkiksi potilastietojen haut potilaan hoitosuhteen tai käyttäjän työajan ulkopuolella.

1.2 Tutkimuskysymys ja tutkielman tavoitteet

Millaisilla terveydenhuollon käyttölokeja hyödyntävillä menetelmillä voidaan havaita valtuutettujen käyttäjien suorittama potilastietojen väärinkäyttö?

Tämän tutkielman tutkimuskysymyksessä pohditaan, millaisia terveydenhuollon käyttölokeja hyödyntäviä koneellisia menetelmiä voidaan käyttää sähköisten potilastietojen väärinkäytön havaitsemiseen. Tutkimusmenetelmänä kysymykseen vastaamiseksi käytetään kirjallisuuskatsausta. Kirjallisuuskatsaustulokseen sisällytetään pääasiassa julkaistuja tieteellisistä tutkimuksista, joiden menetelmillä pystytään havaitsemaan valtuutettujen käyttäjien eli sisäisten uhkien suorittama väärinkäyttö terveydenhuoltoympäristössä. Työn tavoitteena on kartoittaa manuaalisten ja yksinkertaisten havaitsemisjärjestelmien tulevaisuuden kehityssuuntia, jotta väärinkäytösten havaitsemistoiminnoista saadaan vieläkin monipuolisempia ja tehokkaampia.

1.3 Työn rakenne

Kappaleessa kaksi käsitellään aiheen kannalta oleellista teoriaa tutkielman ymmärtämisen helpottamiseksi. Kappaleessa esitellään yleisesti mitä ovat lokit,

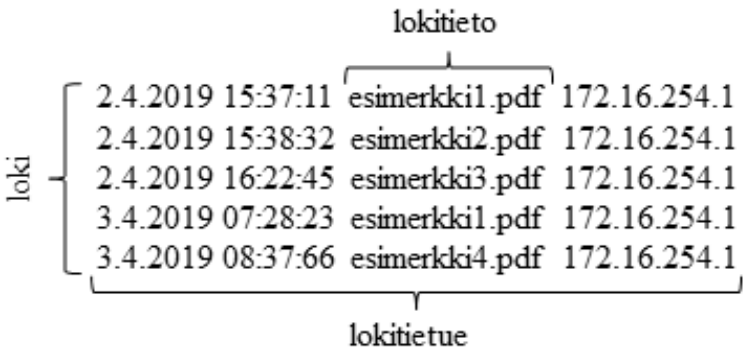
väärinkäytön havaitseminen tutkimusalana ja koneoppimis- ja tiedonlouhinta-menetelmät. Kolmannessa kappaleessa käydään läpi kirjallisuuskatsauksen rakenne ja sen yksityiskohtainen toteutus. Neljännessä kappaleessa esitellään kirjallisuus-katsauksen tulokset jäsenneltynä. Viidennessä luvussa analysoidaan tutkimuksen tuloksia ja esitetään synteesi järjestelmästä, joka soveltuu potilastietojen sisäisen uhan havaitsemiseen käyttölokeista. Tutkielman viimeisessä luvussa esitellään tutkimuksen johtopäätökset sekä pohditaan mahdollisia jatkotutkimuksia.

2 LOKIANALYYSI, SISÄINEN UHKA JA KONEELLISET MENETELMÄT

Tässä kappaleessa käsitellään työn ymmärtämisen kannalta oleellista teoriaa. Aluksi selvennetään, mitä tarkoitetaan lokitiedoilla ja lokeilla. Lisäksi kappaleessa käydään läpi kerätyille lokitiedoille suoritettavaa lokianalyysia, joka on muutakin kuin lokitietojen manuaalista läpikäyntiä. Kappaleessa esitellään myös väärinkäytön havaitsemisen tutkimusalaa, joka pohjautuu vahvasti tilastotieteen, hahmontunnistuksen, koneoppimisen ja tiedonlouhinnan tekniikoihin.

2.1 Lokitietue, lokitiedot, loki ja lokianalyysi

Tietokoneohjelmisto luo ja tallentaa automaattisesti tekstiä, kun se saa tietyn herätteen. Kun esimerkiksi ohjelmiston käyttäjä X avaa tiedoston Y, avaus toimii herätteenä. Ohjelmistossa esiintyneestä tapahtumasta (eng. event) luodaan tapahtumaa ilmaiseva merkkijono. Tällaista automaattisesti luotua merkkijonoa kutsutaan lokitietueeksi (eng. log message). Lokitietueet luodaan ja tallennetaan yleensä yhteen tai useampaan lokitiedostoon (eng. log file). Lokitiedosto siis pitää rekisteriä tapahtumia kuvaavista lokitietueista, ja tätä lokitietueiden kokoelmaa kutsutaan lokiksi tai lokirekisteriksi. Lokitietue pitää sisällään attribuutteja eli lokitietoja (eng. log data), jotka kertovat, miksi tietue on luotu. Jos käyttäjä X on kirjautunut hänelle luoduilla tunnuksilla ohjelmistoon ja avannut tiedoston Y, luotu lokitietue sisältää todennäköisesti käyttäjätunnuksen ja avatun tiedoston nimen. Tunnus ja tiedostonimi ovat lokitietoja ja ne kertovat, että lokitietue on luotu, koska käyttäjä X avasi tiedoston Y. (Chuvakin, Schmidt & Phillips 2013: 2–3.)



Kuva 1. Loki, lokitietue ja lokitieto.

Yksittäinen lokitietue pitää sisällään runsaasti tapahtumaa kuvaavia lokitietoja. Tietue sisältää tyypillisesti aikaleiman, lähteen ja dataa. Aikaleima on ajanhetki, jolloin lokitietue luotiin. Lähde taas on lokitietueen luoneen järjestelmän IP-osoite tai verkkoaseman nimi. Datalle ei ole standardoitua esitystapaa, mutta tyypillisesti se sisältää useita tietoalkioita, kuten käyttäjän, ohjelmiston tai avatun tiedoston nimen tai vaikka siirrettyjen tavujen määrän. (Chuvakin ja muut 2013: 6.)

Lokianalyysi (eng. log analysis) on menetelmä, jolla pyritään löytämään merkittävää sisältöä analysoimalla lokitietoja (Chuvakin ja muut 2013: 14). Lokianalyysin kannalta on tärkeää kerätä kaikki tarvittavat lokitietueet yhteen paikkaan, jotta tietojen vertailu ja johtopäätösten tekeminen helpottuu. Lokitietueet voidaan kerätä suoraan lokipalvelimelle (eng. log server) tai erilliselle lokienkeruupalvelimelle (eng. log collector server), josta ne siirretään lokipalvelimelle (Chuvakin ja muut 2013: 12). Jos lokitietoja täytyy säilyttää pitkiä aikoja, tiedot tallennetaan tyypillisesti relaatiotietokantaan. Relaatiotietokannassa yksi rivi vastaa yhtä lokitietuetta. Relaatiotietokantojen käyttö on yleistä, koska lokitietoja on nopea etsiä ja hakea yksinkertaisten SQL-lauseiden avulla. (Chuvakin ja muut 2013: 78.) Oleellinen osa lokianalyysia on myös lokitietojen suodatus ja normalisointi. Analyysissa kaikista kerätyistä lokitietueista suodatetaan epäolennaiset lokitiedot pois. Normalisointivaiheessa eri palvelimilta kerättyjen lokitietueiden lokitiedot muunnetaan vastaamaan yhtä valittua formaattia. (Chuvakin ja muut 2013: 145–146.)

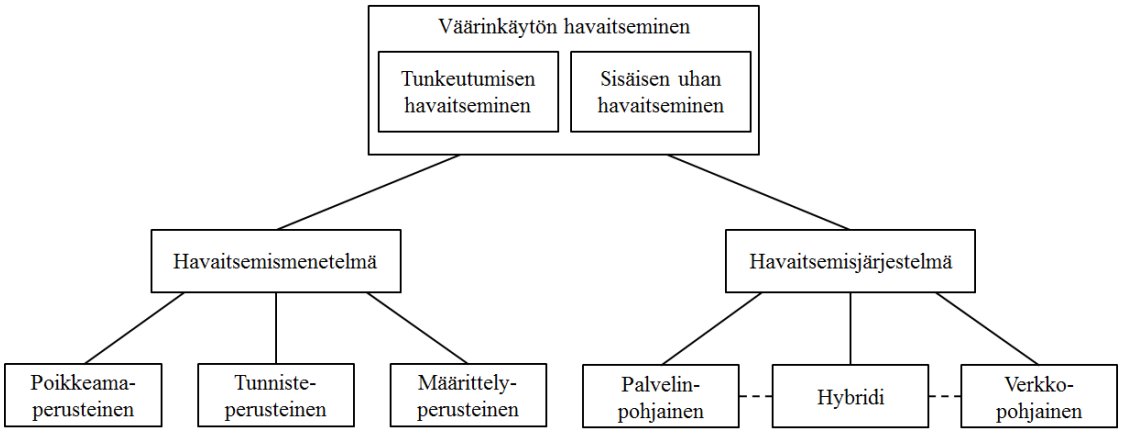
Lokianalyysin avulla voidaan havaita esimerkiksi tietojen väärinkäyttöön viittaavia epäilyttäviä lokitietueita. Analyysin ja havaintojen teko voi olla manuaalista tai hyödyntää korrelaatiota. Korrelaatio tai tapahtumakorrelaatio (eng. event correlation) toteutetaan tyypillisesti korrelaatiosääntöihin tai tilastotieteeseen perustuvilla tekniikoilla. Suoraviivaisten korrelaatiosääntöjen avulla etsitään yhteyksiä kahden tai useamman tapahtuman väliltä vertailemalla lokitietueita. (Chuvakin ja muut 2013: 146; Kent & Souppaya 2006: 3-4–3-5.) Tilastotieteeseen perustuva korrelaatio voi laskea esimerkiksi tapahtumien esiintymistiheyttä ja auttaa havaitsemaan tästä poikkeavien lokitietueiden löytämisessä. Tilastotieteellinen korrelaatio voi perustua myös koneoppimistekniikoihin, joita käsitellään tarkemmin kappaleessa 2.3. (Chuvakin ja muut 2013: 182, 187.) Lokianalyysissa hyödynnetään yhä useammin myös tiedonlouhintatekniikoita, kun ei tiedetä tarkalleen, millaista sisältöä ollaan etsimässä (Chuvakin ja muut 2013: 193). Tiedonlouhintatekniikoita on ehdotettu käytettäväksi myös terveydenhuoltoympäristön käyttölokien analysointiin (Asaro & Ries 2001: 855).

Lokianalyysi on tyypillinen osa järjestelmälle tai prosessille suoritettavaa auditointia. Auditointi (eng. auditing) on prosessi, jolla varmistetaan, että järjestelmä tai prosessi toimii odotetusti. Käyttölokit muodostavat osan auditoinnissa käytettävästä jäljityslokista (eng. audit trail). Sairaalaympäristössä auditoinnilla varmistetaan, että käyttäjien suorittama potilastietojen käyttö on asiallista. Potilastietojen käytöistä kerätyt käyttölokit ovat sovellustason lokeja (eng. application level access logs). (Chuvakin ja muut 2013: 22.)

Edellä esitetyistä määritelmistä on syytä huomata, että lokeihin liittyvissä julkaisuissa esiintyy vaihtelua käytetyssä terminologiassa. Esimerkiksi lokitietueesta voidaan käyttää nimitystä tapahtumatallenne (eng. event record). Lisäksi lokitietueen synonyymejä ovat tarkastustallenne (eng. audit record) ja lokimerkintä (eng. log entry). Myös lokitietueen sisältämiä lokitietoja voidaan kutsua tapahtumakentiksi (eng. event field) tai tarkastustiedoiksi (eng. audit data). Tapahtumatallenteiden kokoelmaa kutsutaan lokiksi, mutta myös lokiin viittaavia termejä ovat esimerkiksi jäljitysloki (eng audit log) ja tapahtumaloki (eng. event log). (MITRE 2010: 3; Chuvakin ja muut 2013: 30.)

2.2 Väärinkäytön havaitseminen ja sisäinen uhka

Väärinkäytön havaitseminen (eng. misuse detection) on tutkimusala, joka tutkii erilaisten tietoturvauhkien havaitsemista erilaisten järjestelmien avulla. Väärinkäytön havaitseminen on ylätason käsite ulkoiselle tunkeutumisen havaitsemiselle (eng. intrusion detection) ja sisäisen uhan havaitsemiselle (eng. insider threat detection). Tunkeutumisen havaitsemiseen perustuva järjestelmä pyrkii havaitsemaan luvattomasti sisään pyrkiviä ulkopuolisia uhkia. Vastaavasti sisäisiä uhkia havaitsevan järjestelmän tarkoituksena on havaita valtuutettujen käyttäjien luvatonta toimintaa ja käyttöoikeuksien ylityksiä. Valtuutetut käyttäjät ovat usein jopa suurempi uhka kuin ulkoiset tunkeutujat. (Chung, Gertz & Levitt 1998: 1–2.) Kuvassa 2 on kuvattu edellä mainittua väärinkäytön havaitsemisen jakoa alatasoihin uhkatyypin mukaan sekä esitetty havaitsemismenetelmien- ja järjestelmien yleistä terminologiaa.



Kuva 2. Väärinkäytön havaitsemisen terminologia.

Väärinkäytön havaitsemisjärjestelmät (eng. misuse detection systems, MDS) jaetaan pääsääntöisesti kolmeen luokkaan riippuen niiden asennustavasta ja käyttämästä tiedosta. Havaitsemisjärjestelmä voi olla palvelin- (eng. host-based) tai verkkopohjainen (eng. network-based) tai niiden yhdistelmä eli hybridi. Palvelinpohjainen havaitsemisjärjestelmä on tyypillisesti tietokoneohjelma, joka asennetaan jokaiselle palvelimelle erikseen. Palvelinpohjainen järjestelmä voi seurata palvelimelle tallennettuja lokitietoja, vastaanotettua tietoliikennettä ja käyttöjärjestelmän käyttöä.

Verkkopohjaiset järjestelmät taas hyödyntävät eri puolille verkkoa asennettuja antureita, joilla pyritään havaitsemaan poikkeavaa tietoliikennettä. (Larson & Cockcroft 2003.)

Väärinkäytön havaitsemismenetelmät jaetaan kolmeen luokkaan. Havaitseminen voi olla poikkeama- (eng. anomaly-based), tunniste- (eng. signature-based) tai määrittelyperusteista (eng. specification-based). Poikkeamaperusteisessa väärinkäytön havaitsemisjärjestelmässä järjestelmään luodaan profiileja, jotka jäljittelevät käyttäjän tyypillistä käyttäytymistä ja mikäli järjestelmä havaitsee siitä poikkeavaa käytöstä, se hälyttää mahdollisesta väärinkäytöksestä. Käyttäytymisprofiilit luodaan seuraamalla oikeiden käyttäjien toimintaa tietyn ajanjakson ajan. Tunnisteperusteisessa havaitsemisjärjestelmässä esimerkiksi lokitietoja verrataan ennalta määriteltyihin väärinkäyttöön viittaaviin tunnisteisiin eli sääntöihin, ja jos tiedoista löytyy vastaavuus, järjestelmä ilmoittaa tietojen mahdollisesta väärinkäytöksestä. Näiden lisäksi on mahdollista käyttää tilalliseen protokolla-analyysiin (eng. stateful protocol analysis) pohjautuvaa määrittelyperusteista järjestelmää. Järjestelmä muodostaa protokollastandardien määritelmien avulla kuvan, miten protokollia tulisi käyttää ja havaitsee jos niiden käyttö on poikkeavaa. (Liao, Lin, Lin & Tung 2012: 17; Juniper Networks 2016: 5.)

Väärinkäytön havaitsemiseen tarkoitettuja järjestelmiä on kehitetty lukuisia. Erään tulkintatavan mukaan järjestelmien lähestymistavat väärinkäytön havaitsemiseen voidaan jakaa viiteen luokkaan. Lähestymistapa voi olla tilastotieteeseen (eng. statistics), hahmoon (eng. pattern), sääntöön (eng. rule), tilaan (eng. state) tai heurestiikkaan (eng. heurestic) perustuva. Jokaisessa luokassa on havaitsemismenetelmiä, jotka ovat joko poikkeama- tai tunnisteperusteisia tai niiden yhdistelmiä. (Liao, Lin, Lin & Tung 2012: 18–19.)

Tilastotieteeseen perustuvan luokan lähestymistavat pohjautuvat tilastotieteeseen (eng. statistics), etäisyyteen (eng. distance), bayesiläisyyteen (eng. Bayesian) ja peliteoriaan (eng. game theory). Hahmoihin perustuvan luokan tavat taas voidaan jakaa hahmonsovitukseen (eng. pattern matching), Petri-verkkoihin (eng. Petri net), näppäinlyöntien valvontaan (eng. keystroke monitoring) ja tiedostojärjestelmän

tarkastukseen (eng. file system checking). Lisäksi sääntöihin perustuvia tapoja ovat sääntöpohjaiset (eng. rule-based) menetelmät, tiedonlouhinta (eng. data mining), malli/profiili -pohjaiset (eng. model/profile-based) menetelmät sekä tukivektorikone-menetelmä (eng. support vector machine). Tilaan perustuvia lähestymistapoja ovat tila-siirtymä analyysi (eng. state-transition analysis), käyttäjän aikeiden tunnistaminen (eng. user intention identification), Markovin prosessia (eng. Markov process) hyödyntävä menetelmä ja protokolla-analyysi (eng. protocol analysis). Viimeisenä heurestiikkaan perustuvat tavat voidaan jakaa menetelmiin, jotka hyödyntävät hermoverkkoja (eng. neural networks), sumeaa logiikkaa (eng. fuzzy logic), geneettisiä algoritmeja (eng. genetic algorithm), immuunijärjestelmää (eng. immune system) ja parviälyä (eng. swarm intelligence). (Liao ja muut 2012: 18–19.)

Edellä kirjoitetuista määritelmistä on syytä huomioida, ettei väärinkäytön havaitsemisen tutkimusalan terminologia ole täysin vakiintunut. Useissa tutkimuksissa ja kirjoissa keskitytään tunkeutumisen havaitsemisjärjestelmien (eng. intrusion detection systems, IDS) tutkimiseen ja juurikin ulkoisten uhkien havaitsemiseen. Tällaisen järjestelmän havaitsemistoiminta voi olla käytetyn menetelmän perusteella poikkeamien havaitsemista (eng. anomaly detection) tai väärinkäytön havaitsemista (eng. misuse detection). Näillä termeillä kuitenkin viitataan edellä määriteltyihin poikkeama- ja tunnisteperusteiseen havaitsemiseen. (Tsai & Yu 2010: 2.) Tunnisteperusteisesta havaitsemisesta käytetään toisinaan myös nimitystä tieto- tai sääntöperusteinen (eng. knowledge/rule-based) havaitseminen. Väärinkäytön havaitsemisen alalla on tunnistettu ulkoisten uhkien lisäksi sisäisten uhkien olemassaolo. Väärinkäytössä ei siis ole läheskään aina kyse tunkeutumisesta, vaan myös sisäisten uhkien, kuten valtuutettujen käyttäjien luvaton toiminta on osa sitä. (Chung ja muut 1998: 2.) Mikäli tutkitaan sisäisten uhkien havaitsemista, tunkeutumisen havaitsemisesta ja tunkeutumisen havaitsemisjärjestelmistä (IDS) puhuminen on ontuvaa. Näin ollen väärinkäytön havaitsemisesta on syytä puhua tunkeutumisen ja sisäisen uhan havaitsemisen osalta ylätason käsitteenä ja käyttää IDS:n sijaan yleisempää nimitystä väärinkäytön havaitsemisjärjestelmä (MDS). Lisäksi havaitsemismenetelmää kuvaava väärinkäytön havaitseminen tulisi korvata yleisellä tasolla termillä tunnisteperusteinen väärinkäytön havaitseminen. Koska sisäisten uhkien havaitsemiseen voidaan käyttää tunkeutumisen

havaitsemisen menetelmiin ja järjestelmiin perustuvia periaatteita, tunkeutumisen havaitsemista tutkivien julkaisujen sisällyttäminen tähän työhön ei tuota ristiriitoja.

Tässä tutkielmassa keskitytään sisäisen uhan havaitsemiseen, ja koska käytössä on palvelimelta lokitietoja keräävä ohjelma, palvelinpohjainen havaitsemisjärjestelmä on kiinnostavampi. Havaitsemismenetelmistä keskitytään poikkeama- ja tunnisteperusteisiin menetelmiin.

2.3 Koneoppiminen, hahmontunnistus ja tiedonlouhinta

Kun suuresta määrästä kerättyä dataa pyritään löytämään koneellisesti säännönmukaisuuksia ja hyödyllistä tietoa, puhutaan yleensä tutkimusaloista, kuten koneoppiminen (eng. machine learning), hahmontunnistus (eng. pattern recognition) ja tiedonlouhinta (eng. data mining). Näillä aloilla on paljon yhtäläisyyksiä, kuten tilastotieteen hyödyntäminen, ja lisäksi paljon päällekäisyyksiä, koska alat käyttävät menetelmiä toinen toisiltaan. Hahmontunnistus on saanut alkunsa erilaisia koneita kehittävästä insinööritekniikasta (eng. engineering), kun taas koneoppiminen kehittyi tietojenkäsittelytieteen (eng. computer science) lähtökohdista. Hahmontunnistus ja koneoppiminen voidaankin nähdä saman tutkimusalan erilaisina näkökulmina. (Bishop 2006: vii.)

Koneoppiminen on yksi tekoälyn osa-alue, jossa on tyypillistä käyttää ongelmien ratkaisuun automaattisesti oppivia algoritmeja. Tavallisesti ongelmaa ratkaistaessa ohjelmoija luo esimerkiksi funktion Summa(), joka hyväksyy kaksi kokonaislukua syötteenä, kuten luvut 1 ja 2. Funktio tulostaa ulostulona kokonaisluvun 3. Koneoppimisessa ohjelma taas tietää syötteinä olevat luvut 1 ja 2 sekä ulostulon 3. Koneoppimisalgoritmien keskeinen tehtävä on luoda funktio tai malli, jolla syötteistä saadaan haluttu ulostulo. Koneoppimisessa oppiminen (eng. learning) on puhtaasti matemaattista, eikä algoritmi ymmärrä, mitä se on oppinut. Oppiminen on enemmänkin opetusta (eng. training), koska algoritmi opetetaan löytämään oikeat vastaukset (ulostulot) jokaiseen sille esitettyyn kysymykseen (syöte). Algoritmi paranee, mitä

enemmän se saa kokemusta useista syötteistä. (Alpaydin 2014: 2; Mueller & Massaron 2018; 126.) Koneoppimisalgoritmit jaetaan pääsääntöisesti kolmeen ryhmään niiden käyttötarkoituksen mukaan. Algoritmi voi perustua ohjattuun (eng. supervised), ohjaamattomaan (eng. unsupervised) tai vahvistettuun (eng. reinforcement) oppimiseen. (Mueller ja muut 2018; 133–134.)

Ohjatussa oppimisessa algoritmi oppii esimerkkisyötteistä ja niihin liitetyistä ulostuloista. Ohjatun oppimisen algoritmille annetaan siis esimerkkitapauksia eli niin sanottu opetusjoukko, josta algoritmi muodostaa yleisiä sääntöjä. Sääntöjen avulla koneoppimisalgoritmi voi ennustaa oikean ulostulon, kun ohjelmaan syötetään ilman ulostuloa olevia syötteitä. Yleensä ohjatun oppimisen algoritmeilla ratkaistaan regressio- ja luokitteluongelmia. Regressio-ongelman ratkaisu on jatkuva numeerinen arvo ja luokitteluongelmassa ratkaisuksi ehdotetaan epäjatkuvia arvoja, kuten 1 ja 0 tai merkkijonoja, kuten ennalta määriteltyjä luokkia. (Mueller ja muut 2018; 132.)

Ohjaamattomaan oppimiseen perustuva algoritmi oppii ilman opetusjoukkoa. Näin ollen syötteisiin ei ole liitetty tietoa ulostuloista. Algoritmin tavoitteena on löytää syötteistä säännönmukaisuuksia ja havaita, mikäli jotkin rakenteet toistuvat useammin kuin muut. Esimerkki ohjaamattomaan oppimiseen perustuvasta menetelmästä on klusterointi, jossa syötteistä löytyy samankaltaisuuksia, joiden perusteella voidaan muodostaa toisistaan eroavia klustereita tai ryhmiä. (Alpaydin 2014: 11; Mueller ja muut 2018; 134.)

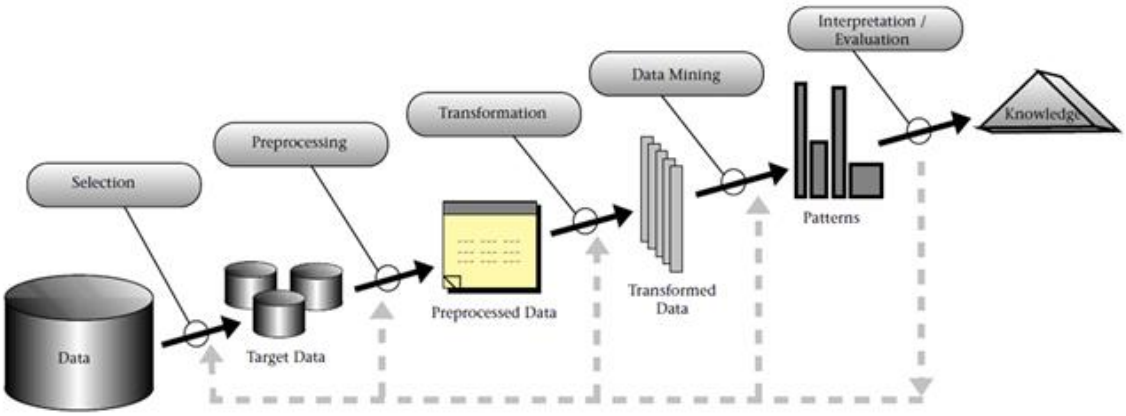
Ohjatun ja ohjaamattoman oppimisen lisäksi puhutaan vahvistetusta oppimisesta. Siinä algoritmin saamilla syötteillä ei ole ohjaamattoman oppimisen tapaan tietoa syötteiden luokasta. Vahvistetun oppimisen algoritmin antava ratkaisu voi olla esimerkiksi toimintojen sarja (eng. sequence of actions), jolla päästään ratkaisuun. Algoritmin ehdottamalle ratkaisulle voidaan antaa positiivista tai negatiivista palautetta, jonka avulla se oppii tekemään parempia päätöksiä. (Alpaydin 2014: 13; Mueller ja muut 2018; 134.)

Hahmontunnistus (eng. pattern recognition) on yksi tietotekniikan tutkimusala, joka keskittyy säännönmukaisuuksien (eng. pattern) automaattiseen tunnistamiseen datasta.

Tunnistaminen tapahtuu tietokonealgoritmien avulla ja säännönmukaisuuksien perusteella datalle suoritetaan toimenpide, kuten luokittelu erilaisiin kategorioihin. Hahmontunnistusongelma voidaan ratkaista itse keksittyjen sääntöjen tai heurestiikan avulla. Lähestymistapa voi perustua myös koneoppimisen ohjatun, ohjaamattoman ja vahvistetun oppimisen tekniikoihin. Esimerkki hahmontunnistuksesta voi olla esimerkiksi käsinkirjoitettujen numeroiden tunnistamista ja luokittelua kuvista. (Bishop 2006: 1-2.)

Tiedonlouhinta (eng. data mining) ja ”tiedon löytäminen tietokannoista” eli KDD (eng. knowledge discovery in databases) ovat tutkimusaloja, joissa algoritmit pyrkivät löytämään tietokantojen datasta toistuvia rakenteita ja tietoa (eng. knowledge). (Theodoridis & Koutroumbas 2008: 1–3.) Vaikka tiedonlouhinta esiintyy varsinkin liikemaailman sovelluksissa omana prosessina, tietojenkäsittelytieteen tulkinnan mukaisesti se on kuitenkin yksi KDD-prosessin vaihe. KDD on monivaiheinen prosessi, jolla pyritään löytämään hyödyllinen tieto tietokantojen datasta. Ennen hyödyllisen tiedon saamista KDD-prosessissa datan käsittelyyn sisältyy viisi päävaihetta (Fayyad, Piatetsky-Shapiro & Smyth 1996: 39–41):

- Valinta (eng. selection)
- Esikäsittely (eng. preprocessing)
- Muunnos (eng. transformation)
- Tiedonlouhinta
- Tulkinta ja arviointi (eng. interpretation & evaluation).



Kuva 3. KDD-prosessin rakenne (Fayyad ja muut 1996: 41).

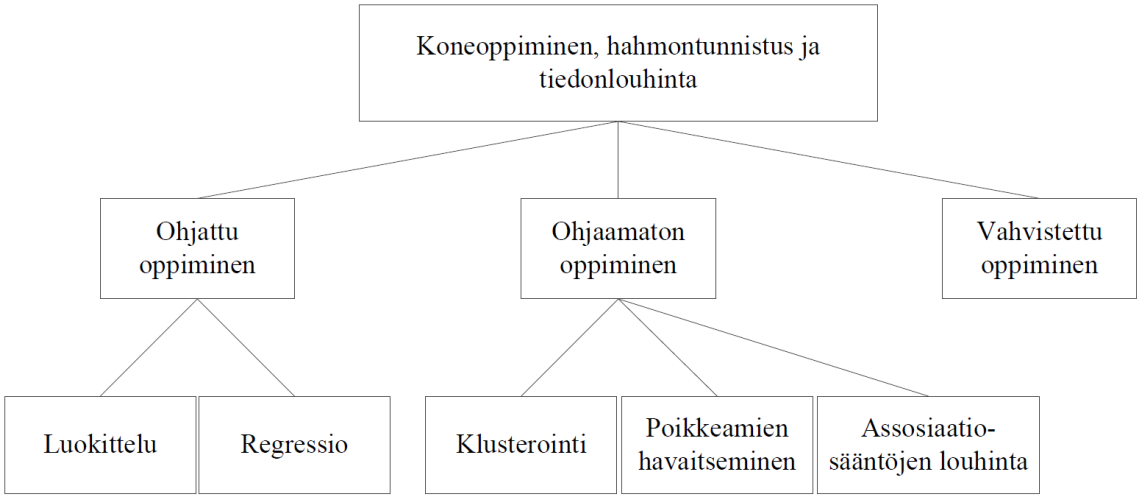
Valinta-vaiheessa alkuperäistä dataa pyritään rajaamaan eli varsinainen hyödyllisen tiedon etsintä suoritetaan vain pienemmälle kohdeaineistolle (eng. target data). Seuraavaksi kohdeaineisto esikäsitellään eli siitä poistetaan kaikki ylimääräinen ja sitä täydennetään, mikäli siinä havaitaan puutteita. Kohdeaineistosta voi esimerkiksi puuttua joitain tietokenttiä, joiden käsittelyyn täytyy muodostaa strategia. Seuraavassa vaiheessa esikäsitelty data muunnetaan tiedonlouhintatekniikoille sopivaksi, minkä jälkeen muunnetusta datasta voidaan alkaa louhia hahmoja. Kun tiedonlouhinta on suoritettu ja tulokset saatu tarkoituksenmukaiseen esitysmuotoon, käyttäjä tulkitsee ja arvioi, ovatko saadut tulokset hyödyllistä tietoa. (Gullo 2015: 18–19.)

Tämän tutkielman kannalta oleellista ovat erilaiset tiedonlouhintatekniikat. Tekniikoita on useita ja ne voidaan jakaa toimintaperiaatteiden mukaan kuuteen luokkaan (Fayyad ja muut 1996: 44–45):

- Luokittelu (eng. classification)
- Klusterointi (eng. clustering)
- Poikkeamien havaitseminen (eng. anomaly/outlier detection)
- Assosiaatiosääntöjen louhinta (eng. association rule discovery)
- Regressio (eng. regression)
- Tiivistäminen (eng. summarization)

2.2.4 Koneoppimis- ja tiedonlouhintatekniikat

Koneoppimisen, hahmontunnistuksen ja tiedonlouhinnan käsitteet yhdistämällä niiden tekniikoiden ja termien jako on kuvan 4 kaltainen.



Kuva 4. Koneoppimisen, hahmontunnistuksen ja tiedonlouhinnan terminologia.

Luokittelulla viitataan yleensä ohjattuun luokitteluun, jonka tarkoituksena on luokitella aiemmin tuntematon tietue kuuluvaksi johonkin ennalta määritettyyn luokkaan (Fayyad ja muut 1996: 44; Gullo 2015: 20). Luokittelualgoritmin syötteenä on olemassa ennalta tunnettu opetusjoukoksi (eng. training set) kutsuttu joukko, joka pitää sisällään useamman attribuutin sisältäviä tietueita. Esimerkiksi jos oletetaan, että on olemassa tietokanta, jonka yksi tietue sisältää neljä attribuuttia: (1) yliopistossa työskentelevän henkilön nimi, (2) toimenkuva (esim. yliopistonlehtori, apulaisprofessori tai professori), (3) tyovuodet yliopistossa ja (4) Boolean tyyppinen luokka, joka kertoo, onko henkilöllä vakituinen virka. Opetusjoukko voi sisältää esimerkiksi neljä tietuetta:

```
{Jukka, apulaisprofessori, 3, ei}
{Maria, apulaisprofessori, 7, kyllä}
{Martti, professori, 2, kyllä}
{Anne, yliopistonlehtori, 7, kyllä}.
```

Opetusjoukon perusteella luokittelualgoritmi loisi todennäköisesti mallin, joka koostuu säännöistä:

```
IF toimenkuva = professori OR tyovuodet > 3
THEN vakituinen = kyllä.
```

Näin ollen algoritmi luokittelisi uuden ennalta tuntemattoman tietueen

```
{Sofia, professori, 4, ?}
```

luokan arvoksi ”kyllä” eli ennustaisi kyseinen henkilön olevan yliopiston vakituinen työntekijä. (Gullo 2015: 20.)

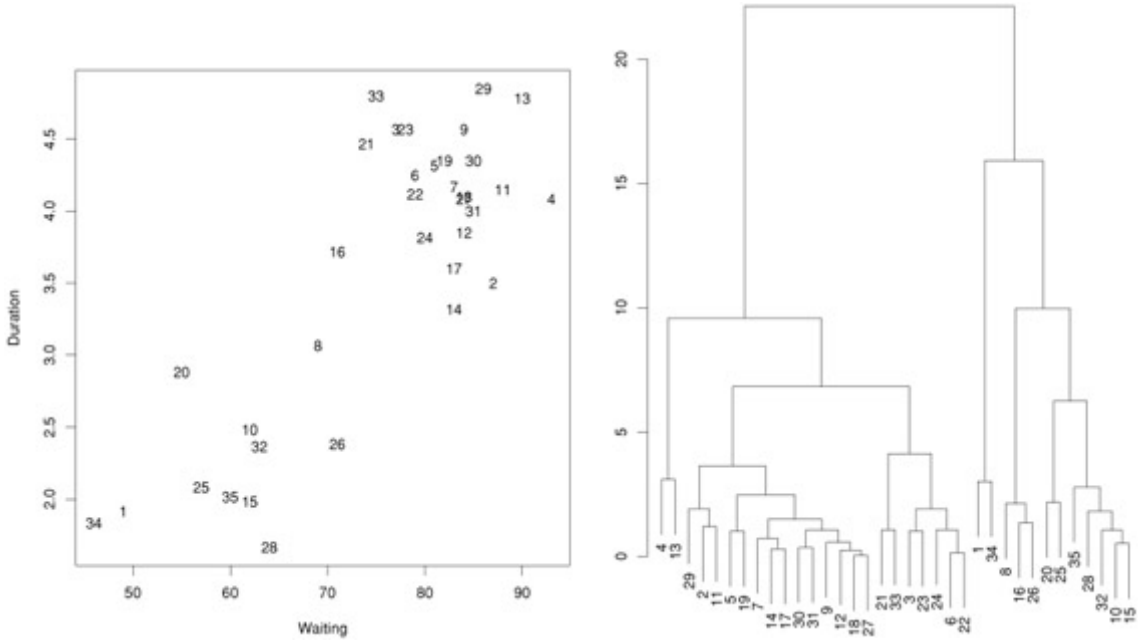
Klusterointi pyrkii jakamaan havaintojen joukon (pistejoukko) ryhmiin, joissa yhden ryhmän havainnot (pisteet) muistuttavat toisiaan. Täten eri ryhmiin kuuluvien havaintojen tulisi vastaavasti erota toisistaan. Havaintojen samankaltaisuuden tai eroavaisuuden vahvuus voidaan määritellä erilaisten läheisyysmittojen (eng. proximity/similarity measure) avulla. Klusteroinnilla voidaan paljastaa havaintojen joukosta rakenteita, joita on muuten vaikea havaita. Klusteroinnilla muodostettuja ryhmiä voidaan myös käsitellä helpommin, kun käsiteltävänä on tuhannen yksittäisen havainnon sijaan vain kymmeniä ryhmiä. Toisin kuin luokittelussa, klusteroinnissa ei käytetä opetusjoukkoa, jossa havaintojen luokka olisi valmiiksi määritelty. Opetusjoukon puuttumisen vuoksi klusterointia kutsutaan usein myös ohjaamattomaksi luokitteluksi. Yleisesti klusterointimenetelmät voidaan jakaa osittaviin (eng. partitional) ja hierarkkisiin (eng. hierarchical) menetelmiin. (Fayyad ja muut 1996: 44–45; Gullo 2015: 20.)

Osittavan klusteroinnin tehtävänä on jakaa pistejoukko toisistaan irrallisiksi ryhmiksi, joiden lukumäärä K on ennalta määritelty. Klusteroinnin aluksi lukumäärän mukainen määrä pisteitä arvotaan alkuklustereiksi. Klusterien samankaltaisuutta mitataan tyypillisesti läheisyysmitalla, joka on usein euklidinen etäisyys. Pisteiden etäisyydet mitataan alkuklustereihin nähden ja piste lisätään lähimpään klusteriin. Klusterin määrittämistä varten pistejoukon pistettä verrataan yleensä joko jokaisen klusterin

keskiarvoon (centroid), lähimpään pisteeseen (single-linkage) tai kauimmaiseen pisteeseen (complete-linkage). Kaikkien pisteiden etäisyydet mitataan vuorollaan ja lopuksi arvioidaan, kuinka lähellä klusterin sisältämät pisteet ovat toisiaan ja kuinka kaukana eri klusterit ovat toisistaan. Klusterointi suoritetaan samalle pistejoukolle useaan kertaan, jotta löydetään optimaaliset klusterit, joissa klusterien sisäisten pisteiden etäisyys toisistaan on mahdollisimman lyhyt ja klusterien väliset etäisyydet mahdollisimman pitkät. (Hand, Mannila & Smyth 2001: 178.)

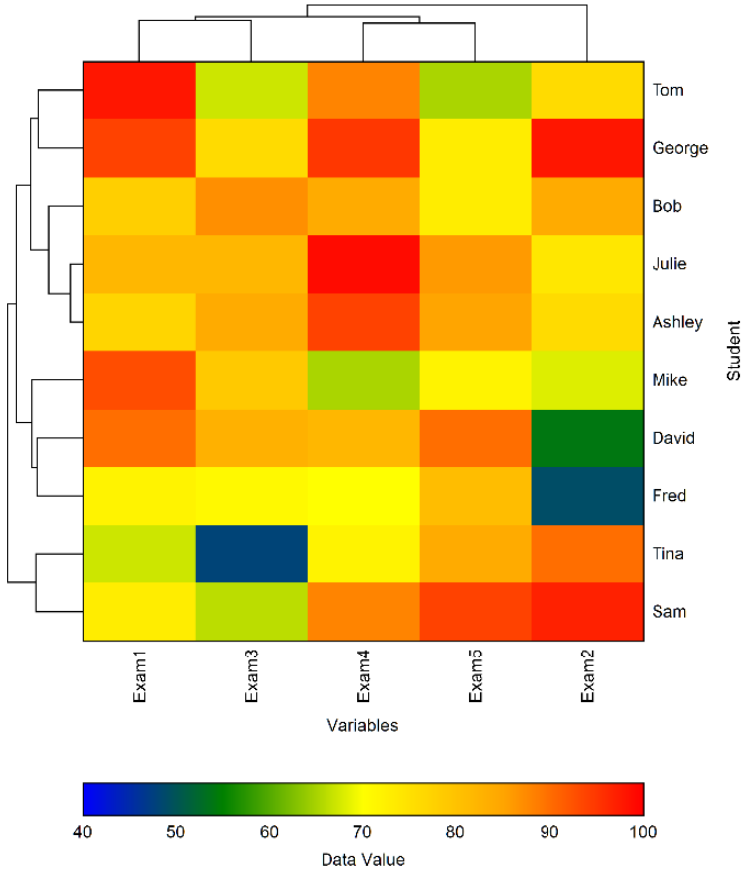
Hierarkkisessa klusteroinnissa pistejoukko jaetaan ryhmiin samankaltaisuuden perusteella, joko kokoavalla (eng. agglomerative) tai jakavalla (eng. divisive) menetelmällä. Kokoavassa tavassa jokainen piste on aluksi oma klusterinsa ja klustereita sulautetaan uusiksi suuremmiksi klustereiksi, kunnes lopulta on jäljellä vain yksi klusteri. Jakavassa menettelyssä aluksi on yksi suuri klusteri, josta erotellaan useita pienempiä klustereita. Kokoava klusterointi on näistä tavoista useammin käytetty. (Hand ja muut 2001: 185.)

Hierarkkista klusterointia visualisoidaan dendrogrammin avulla. Dendrogrammi on puurakenne, jonka avulla havainnollistetaan, missä järjestyksessä klusterit on luotu. Puurakenteen haarojen korkeus kertoo, kuinka samankaltaisia klusterit ovat. Haarojen yhdistyminen puun alaosassa viittaa klusterien samankaltaisuuteen ja yläosassa yhdistyminen kertoo klusterien olevan erilaisia. (Hand ja muut 2001: 185–186.)



Kuva 5. Pistejoukko ja hierarkkisen klusteroinnin dendrogrammi (Hand ja muut 2001: 185–186.)

Klusterointituloksia esitetään usein myös lämpökartan avulla. Lämpökarttaa kutsutaan joskus myös tupladendrogrammiksi. (NCSS 2020: 450: 1.) Esimerkki lämpökartasta on kuvassa 6. Värit kertovat arvosanoista, jotka oppilas on saanut kokeista. Mitä punaisempi väri sitä parempi arvosana. Dendrogrammit kuvaavat kokeiden ja oppilaiden samankaltaisuutta. Arvosanojen perusteella oppilaat Julie ja Ashley ja kokeet 4 ja 5 ovat samankaltaisimpia.



Kuva 6. Lämpökartta eli tupladendrogrammi (NCSS 2020: 450-1).

Poikkeamien havaitseminen on tekniikka, jossa havaintojoukosta etsitään havaintoja, jotka ovat merkittävästi erilaisia kuin joukon muut havainnot (Fayyad ja muut 1996: 45). Lisäksi tiedonlouhintamenetelmiä voi kuulua myös assosiaatiosääntöjen louhintaan, jossa havaintojoukosta muodostetaan normaalia toimintaa kuvaavia sääntöjä eli toistuvia rakenteita. Rakenteet esitetään usein graafisessa muodossa, josta ilmenee useiden muuttujien välisiä suhteita, ja suhteiden vahvuudet (Fayyad ja muut 1996: 45). Jos poikkeamien havaitsemisella pyritään normalisoimaan kaikki havainnot ja jäljelle jäävät tulkitaan poikkeamiksi, assosiaatiosäännöillä pyritään etsimään sellaiset säännöt, joilla havainnot pystytään normalisoimaan. Tekniikoina voidaan käyttää myös regressiota ja tiivistämistä, mutta näitä luokkia ei käsitellä tarkemmin tässä tutkielmassa.

3 KIRJALLISUUSKATSAUS

Tässä tutkielmassa pyritään saamaan tietoa koneellisesta sisäisen uhan havaitsemisesta terveydenhuollossa, ja vastaus tutkimuskysymykseen kirjallisuuskatsauksen avulla. Vaikka tutkielman aihe on rajattu yhteen spesifiseen alueeseen, tutkimuskysymys on jätetty melko laajaksi, mikä tekee kirjallisuuskatsauksen hyödyntämisestä luontevaa. Yksi tyypillinen perustelu kirjallisuuskatsauksen käytölle on sen kyky rakentaa laajasta asiakokonaisuudesta helpommin ymmärrettävä kokonaiskuva (Salminen 2011: 3). Tässä työssä katsauksen avulla voidaan rakentaa kattava kuva sisäisen uhan havaitsemiseen käytettävistä koneellisista menetelmistä ja pohtia niiden soveltuvuutta tietynlaisten lokitietojen (käyttölokit) tarkasteluun tietynlaisessa ympäristössä (terveydenhuolto). Tuloksena saadaan suuntaviivat olemassa olevien manuaalisten ja yksinkertaisten havaitsemisjärjestelmien kehittämiseen.

3.1 Kirjallisuuskatsaustyypin valinta

Kirjallisuuskatsauksia on erään ryhmittelyn mukaan kolmenlaisia. Katsausten perustyypit ovat kuvaileva kirjallisuuskatsaus, systemaattinen kirjallisuuskatsaus ja meta-analyysi. Kuvaileva kirjallisuuskatsaus on yksi eniten käytetty katsaustyyppi. Kuvailevassa kirjallisuuskatsauksessa ei ole tiukkoja sääntöjä, joten käytettävän tutkimusaineiston valinta on vapaampaa. (Salminen 2011: 6.) Systemaattinen katsaus pyrkii nimensä mukaisesti järjestelmälliseen tutkimuskirjallisuuden löytämiseen, aineiston laadun arviointiin ja yhteensovittamiseen (Stolt, Axelin & Suhonen 2015: 14). Kolmas katsaustyyppi on meta-analyysi, joka voi olla määrällinen tai laadullinen riippuen kerättävästä tutkimusaineistosta.

Tässä tutkielmassa tutkittavaa aineistoa ei haluttu rajata liikaa, jotta tutkimuskysymykseen saadaan riittävän laaja vastaus. Tästä syystä tutkielmassa päädyttiin kuvailevan katsauksen tekemiseen. Kuvaileva katsaus voidaan jakaa edelleen kahteen tyyppiin; narratiiviseen ja integroivaan katsaukseen. (Salminen 2011: 6.) Narratiivinen kirjallisuuskatsaus ei sisällä paljon metodeja ja laajin tapa toteuttaa se on

yleiskatsaus. Narratiivisen yleiskatsauksen avulla aiempaa tutkimustietoa pyritään tiivistämään ja sille on tyypillistä, ettei tutkimusaineiston valinta ole kovin systemaattista. (Salminen 2011: 7.) Toinen kuvaileva katsaustyyppi on integroiva kirjallisuuskatsaus, joka mahdollistaa tutkittavan ilmiön monipuolisen kuvauksen. Integroivaa katsausta pidetään hyvänä tapana luoda uutta tietoa aiemmin tutkitusta aiheesta ja lisäksi se auttaa tutkittavan aineiston kriittisessä arvioinnissa ja tiedon yhteensovittamisessa. (Salminen 2011: 8.) Integroivan katsauksen eteneminen on prosessimaista ja se sisältää viisi vaihetta. Tyypillisesti vaiheet on nimetty alla olevan luettelon mukaisesti (Stolt ja muut 2015: 13; Cooper 1989: 14).

1) Tutkimuskysymyksen muodostaminen
2) Analysoitavan aineiston keruu
3) Aineiston laadun arviointi
4) Aineiston analysointi ja tulkinta
5) Tulosten esittäminen

Tämän tutkielman kirjallisuuskatsaukseksi valittiin integroiva katsaus, koska sen prosessimaisuuden avulla katsaus pysyy johdonmukaisena ja etenee järjestelmällisesti tulosten esittämiseen.

3.2 Tutkimuskysymyksen muodostaminen

Integroivan kirjallisuuskatsauksen ensimmäinen vaihe eli tutkimuskysymyksen nimeäminen on käsitelty jo aiemmin osiossa 1.2.

3.3 Analysoitavan aineiston keruu

Integroivan kirjallisuuskatsauksen toisessa vaiheessa analysoitavan aineiston keruuprosessi käydään tarkasti läpi. Tässä vaiheessa katsausta vastataan kysymyksiin, mitä hakusanoja on käytetty, ja mistä tietoa on etsitty. Näiden lisäksi aineiston valintaan

käytetyt kriteerit esitetään ja keruuprosessista tehdään kooste, josta käy ilmi katsauksessa käytettävän aineiston koko.

3.3.1 Ennen aineiston keruuta

Ennen varsinaisen aineiston keruun toteuttamista, tutkimusongelmaan liittyvää aihealuetta ja sopivia hakusanoja kartoitettiin yksinkertaisten Google-hakujen avulla. Jo hakulausekkeen, kuten ”automatic misuse detection of EHR from access logs” avulla löytyneiden artikkelien perusteella voi todeta, että aihetta ei ole tutkittu erityisen kattavasti. Hakujen avulla kuitenkin ilmeni, että käyttölokeja hyödyntäviä sähköisten potilastietojen väärinkäyttöä havaitsevia järjestelmiä on olemassa. Kartoitushakujen perusteella voi myös todeta, että tyypillinen tapa havaita epäilyttävää potilastietojen käyttöä on hyödyntää tunnettuja ja yksinkertaisia sääntöjä. Tällaisia sääntöjä voi olla esimerkiksi julkisuuden henkilöksi tunnistetun potilaan tietojen katselu tai potilaan ja potilasta hoitavan henkilön jakama sama osoite tai sukunimi. Vaikkakin yksinkertaisten sääntöjen pohjalta voi havaita väärinkäyttöä, väärien hälytysten (eng. false positive) määrä voi olla korkea, koska säännöt eivät arvioi, kuinka todennäköisesti kyseessä on väärinkäyttötapaus. Näin ollen jokainen hälytys näyttäytyy havaitsemisjärjestelmän käyttäjälle väärinkäyttönä.

Kartoituksen perusteella ilmeni myös, että tehokas keino vähentää sisäisen väärinkäytön uhkaa on rajata järjestelmän käyttäjien pääsy vain niihin tietoihin, joihin heillä on tarve. Potilastietojärjestelmän kohdalla käyttäjien (lääkärit, hoitajat jne.) pääsy rajataan vain niihin potilaisiin, joihin heillä on hoitosuhde tai käyttö rajataan käyttäjän roolin mukaan. Sisäisen uhan estämisessä käytettävän pääsyn rajoituksen (eng. access control) sisällyttämistä kirjallisuuskatsaukseen harkittiin, mutta tietoihin pääsyn rajaus terveydenhuoltoympäristöissä on erityisen haastavaa vaarantamatta potilaiden turvallisuutta ja hoitojakson sujuvuutta. Potilaiden hoitojaksot voivat olla vaikeasti ennustettavia, koska potilas voi vain poiketa ennalta sovitulla klinikkakäynnillä, viettää useita päiviä sairaalahoidossa esimerkiksi leikkauksen jäljiltä tai käyttää kiireellisissä tilanteissa päivystyspoliklinikan palveluita. Sairaalahenkilökunta saattaa myös liikkua työssään, jolloin pääsy potilastietoihin on tarpeellista vaihtelevissa ympäristöissä ja

ajankohdissa. Näiden lisäksi klinikkatyössä tehdään paljon yhteistyötä ja potilastietoja saatetaan tarvita myös opetustilanteissa. Myös henkilökunnan rooleja ja työnimikkeitä voi olla satoja, eikä nimike välttämättä kerro, millaisiin potilastietoihin käyttäjä tarvitsee pääsyn. (Fabbri & LeFevre 2011a: 2; Boxwala, Kim, Grillo & Ohno-Machado 2011: 499.)

Monimutkaisissa ympäristöissä käyttörajoitusjärjestelmiin on ohjelmoitu niin sanottu ”riko lasi” -ominaisuus, jolla rajoituksen pystyy ohittamaan. Tyypillisesti tietoturva-asiantuntija tutkii jokaisen ohituksen jälkikäteen. Sairaalaympäristössä pilotoidun järjestelmän testauksen perusteella ominaisuus ei sovellu terveydenhuoltoon. Kuukauden testijakson aikana kirjatuista käyttölokeista yli puolet käytöistä hyödynsivät ”riko lasi” -ominaisuutta. Käytännössä tämä olisi tarkoittanut satojen tuhansien käyttölokitietueiden jälkikäteistä tarkastustyötä. (Rostad & Edsberg 2006: 4; Chen ja muut 2012a: 342.) Edellä mainituista syistä potilastietojärjestelmän käyttäjillä on tyypillisesti melko laajat oikeudet ja mahdollisia väärinkäytöksiä pyritään havaitsemaan auditoimalla jälkikäteen järjestelmän keräämiä tietoja (Fabbri ja muut 2011a: 2). Näin ollen tietoihin pääsyn rajausta ei pidetty tässä tutkielmassa oleellisena, eikä sitä sisällytetty kirjallisuuskatsauksen hakusanoihin. Sisäisten uhkien estämisen sijaan tutkielma ja katsauksen hakusanat keskittyvät niiden havaitsemiseen.

3.3.2 Perushakulauseke

Katsauksen toinen vaihe aloitetaan muodostamalla perushakulauseke, joka käsittelee tutkimuskysymystä. Hakulauseke muodostetaan Boolean operaattoreiden (AND/OR), sulkeiden, sanankatkaisujen (*) ja fraasien (" ") avulla. Perushakua varten valitut sanat on esitetty taulukossa 1. Osion 1 sanoilla haku pyritään keskittämään väärinkäytön havaitsemisen tutkimusalalle ja osiolla 2 koneellisiin menetelmiin. Osio 3 ohjaa hakua käyttölokien hyödyntämiseen ja osio 4 terveydenhuoltoympäristöön.

Taulukko 1. Kirjallisuuskatsauksen perushakusanat englanniksi ja suomeksi.

Osio 1	**Osio 2**	**Osio 3**	**Osio 4**
misuse detection (väärinkäytön havaitseminen)	machine learning (koneoppiminen)	log analysis (lokianalyysi)	electronic health records (sähköiset potilastiedot)
insider threat detection (sisäisen uhan havaitseminen)	pattern recognition (hahmontunnistus)	auditing (auditointi)	electronic medical records (sähköiset potilastiedot)
intrusion detection (tunkeutumisen havaitseminen)	data mining (tiedonlouhinta)	access logs (käyttölokit)	electronic medical information (sähköiset potilastiedot)

Sanojen pohjalta muodostettu yksittäinen perushakulauseke on

```
("misuse detection" OR "insider threat detection" OR "intrusion detection") AND ("machine learning" OR "pattern recognition" OR "data mining") AND ("log analysis" OR "auditing" OR "access logs") AND ("electronic health records" OR "electronic medical records" OR "electronic medical information").
```

Taulukossa on esitetty perushakusanat suomeksi ja englanniksi, mutta perushakulausekkeeseen valitut sanat ovat ainoastaan englanninkielisiä. Valintaan päädyttiin, koska aiheeseen sopivia suomenkielisiä tieteellisiä julkaisuja tai tietokantoja, jotka käsittelevät työn aihetta, ei ole juurikaan olemassa. Toteutetussa kirjallisuuskatsauksessa ei käytetty kaikkia perushakulausekkeen sanoja. Katsaukseen valittu aineisto löydettiin seuraavien hakusanojen avulla:

```
”insider threat detection”
"auditing" AND "electronic health records"
("machine learning" OR "data mining") AND "access logs" AND "electronic health records"
```

3.3.3 Tietokannat

Perushakulausekkeen muodostamisen jälkeen valitaan aineiston keruussa käytettävät tietokannat. Tämän tutkielman aihe sisältyy tietotekniikan alalle, joten myös tietokannat valitaan sen mukaisesti. Soveltuvia tietokantoja etsittiin Finna-palvelun kautta ja tähän tutkielmaan soveltuviksi valittiin tietokannat ACM Digital Library, IEEE Xplore Digital Library, ScienceDirect (Elsevier) ja SpringerLink. Koska aihe liittyy myös terveydenhuoltoon, näiden tietokantojen lisäksi valittiin sosiaali- ja terveyshallinto-tieteisiin keskittyvä tietokanta MEDLINE (PubMed).

3.3.4 Sisäänotto- ja ulosheittokriteerit

Perushakulausekkeen ja tietokantojen valinnan lisäksi aineiston keruuseen asetetaan erilliset valintakriteerit eli niin sanotut mukaanotto- ja poissulkukriteerit. Valintakriteerit on esitetty taulukossa 2.

Taulukko 2. Kirjallisuuskatsaukseen valittavan aineiston valintakriteerit.

Kriteerityyppi	**Mukaanottokriteerit**	**Poissulkukriteerit**
Julkaisukieli	englanti	≠ englanti
Julkaisuvuosi	2009–2019	< 2009
Aineistolaji	Tutkimusartikkeli Konferenssijulkaisu Muu artikkeli Väitöskirja Muu kirja	Korkeakoulututkielma
Saatavuus	Maksuton Koko teksti saatavilla	Maksullinen Rajattu teksti saatavilla

Kirjallisuuskatsaukseen päädyttiin ottamaan mukaan vain englanninkieliset aineistot, koska suomenkielisiä julkaisuja ja tietokantoja ei löydetty katsauksen valmisteluvaiheessa. Hyväksyttyjen aineistojen julkaisuvuodet on rajattu vuosien 2009–2019 välille. Tietokantahaut suoritettiin alkuvuodesta 2020, joten julkaisuvuosien ylärajaksi valittiin 2019. Alarajaksi valikoitui 2009, koska kerättävän aineiston halutaan

olevan suhteellisen uutta ja ajan tasalla olevaa. Julkaisuvuosirajauksella voidaan myös rajata aineiston koko sopivaksi. Tieteellisten julkaisujen maksuttomuus ja koko tekstin saatavuus päätettiin ottaa yhdeksi valintakriteeriksi, koska työhön käytetään ainoastaan yksittäisen opinnäytetyöntekijän resursseja, ja kyseessä on myös joustava integroiva katsaus. (Stolt ja muut 2015: 25–26.) Resurssien ja katsaustyypin vuoksi mukaanotettaviksi aineistolajeiksi hyväksytään laajasti erityyppisiä aineistoja, vaikka pääasiassa haut tehdäänkin tietokannoista, joissa aineistot ovat tutkimusartikkeleita ja konferenssijulkaisuja.

3.3.5 Aineistohaun tulokset

Tietokantahaut suoritettiin vuoden 2020 alussa. Tietokannoissa käytetyt hakusanat ja rajaukset, sekä saadut tulokset on esitetty seuraavaksi. Osio päättyy yhteenvetoon hakuprosessista.

ACM Digital Library -tietokantaan tehtiin haut sanoilla “insider threat detection”. Hakua rajattiin julkaisuajankohdalla vuosina 2009–2019 välillä julkaistut artikkelit. Tuloksena saatiin 58 aineistoa, joista luettiin 12 tiivistelmää. Lopulliseen aineistoon valikoitui yksi artikkeli. Vastaavalla rajauksella tehtiin myös haku sanoilla ”auditing” AND ”electronic health records”. Hakutulos oli 49 kappaleen aineisto, josta kaksi valittiin luettavaksi ja lopulliseen aineistoon. Hakusanoilla ("machine learning" OR "data mining") AND "access logs" AND "electronic health records" saatiin kuusi tulosta, joista yksi luettiin ja valittiin aineistoon. ACM -tietokannasta valittiin lopulliseen aineistoon neljä artikkelia.

IEEE Xplore Digital Library -kannan haussa käytettiin hakusanoja “insider threat detection”. Rajauksena käytettiin julkaisuvuosia 2009–2019. Aineistoja kertyi haussa 90, joista 20 artikkelia valittiin tiivistelmän lukuun. Koko tekstin tarkasteluun valittiin kolme julkaisua. IEEE -tietokannan tuloksista kaksi artikkelia sisällytettiin lopulliseen aineistoon.

ScienceDirect (Elsevier) -tietokantahaussa käytettiin sanoja “insider threat detection”. Haku rajattiin vuosina 2009–2019 julkaistuihin artikkeleihin ja kirjakappaleisiin. Tulosten määrä oli 38 julkaisua, joista otsikko ohjasi tiivistelmän tarkasteluun viisi tutkimusartikkelia. Kolme julkaisua luettiin kokonaan, ja yksi artikkeli hyväksyttiin aineistoon. Tuloksena hakusanoilla ("machine learning" OR "data mining") AND "access logs" AND "electronic health records" saatiin 16 julkaisua, joista kolmesta luettiin tiivistelmä. Näistä yksi luettiin kokonaan ja lisättiin lopulliseen aineistoon. ScienceDirect (Elsevier) -tietokannasta valittiin lopulliseen aineistoon kaksi julkaisua.

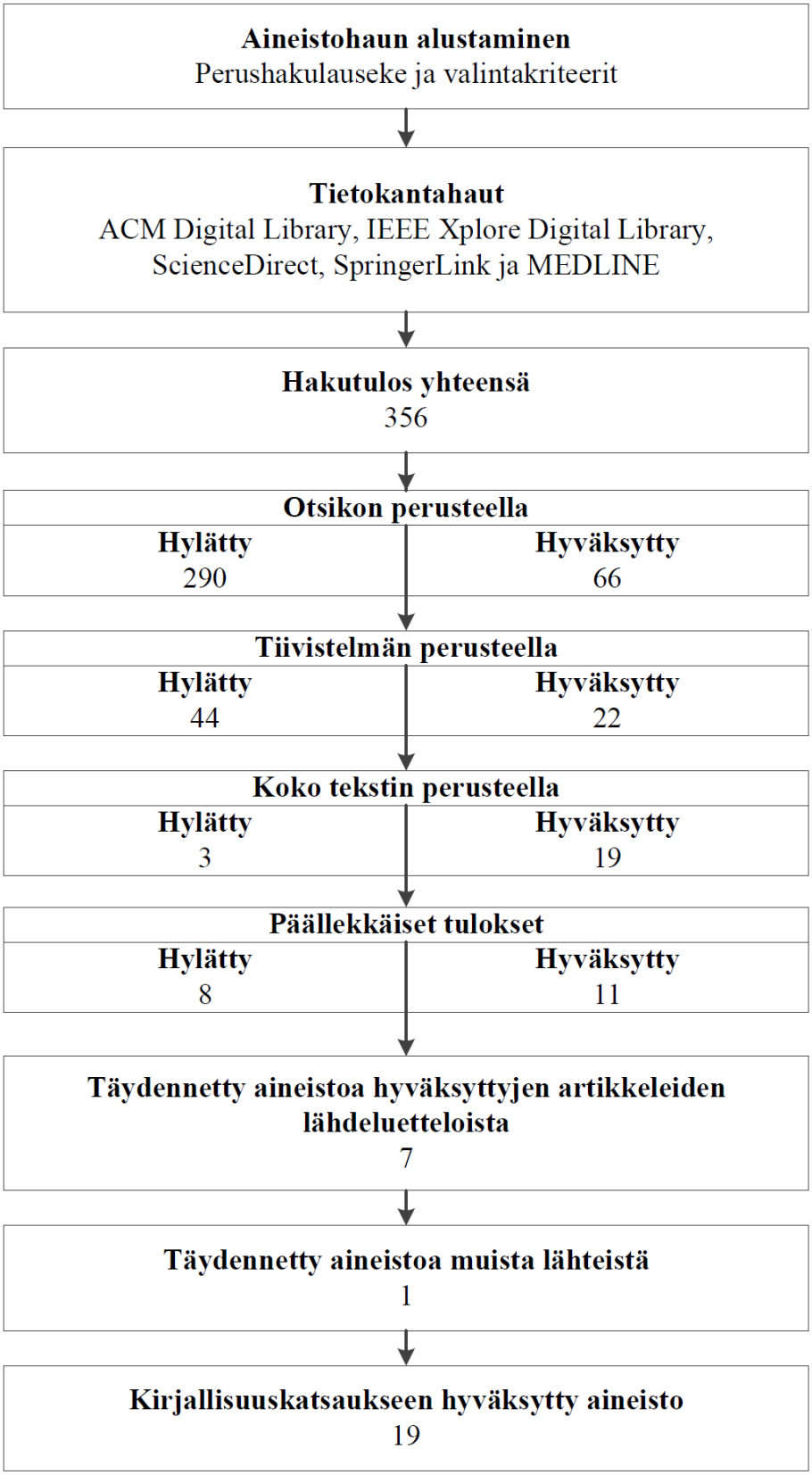
SpringerLink -kannasta tehtiin haut vuosilta 2009–2019 sanoilla “insider threat detection”. Hakutulos oli 28 artikkelia, joista 15 hylättiin otsikon ja 11 tiivistelmän perusteella. Koko tekstin tarkasteluun jäi yksi artikkeli ja lopulliseen aineistoon valittiin yksi julkaisu. Haku sanoilla ("machine learning" OR "data mining") AND "access logs" AND "electronic health records" löysi 18 tulosta, joista kahdesta luettiin tiivistelmä ja koko teksti, ja molemmat lisättiin aineistoon. Lopullinen valittujen artikkelien määrä SpringerLink -tietokannasta oli kolme.

MEDLINE (PubMed) -kannassa hakusanoiksi valittiin aluksi “insider threat detection”. Haun avulla löytyi neljä artikkelia, joista kaikki hyväksyttiin tarkempaan tarkasteluun otsikon perusteella. Tiivistelmän perusteella hyväksyttiin edelleen kaikki neljä. Koko tekstin perusteella jokainen neljästä artikkelista hyväksyttiin katsauksen lopulliseen aineistoon. Hakusanoilla ”auditing” AND ”electronic health records” saatiin 45 tulosta, joista tarkasteluun otettiin kaksi artikkelia ja valittiin molemmat lopulliseen aineistoon. Lisäksi tehtiin haku sanoilla ("machine learning" OR "data mining") AND "access logs" AND "electronic health records", joilla tulokseksi saatiin kolme artikkelia. Näistä tiivistelmä luettiin kaikista ja lopulliseen aineistoon valikoitui kaksi artikkelia. Yhteensä MEDLINE -tietokantahakujen avulla aineisto kasvoi kahdeksalla artikkelilla.

Tietokantahakujen jälkeen aineistoon sisällytettiin 19 artikkelia, joista kahdeksan poistettiin päällekkäisten tulosten vuoksi. Aineistoa kuitenkin täydennettiin valittujen artikkelien lähdeluettelojen ja Google-hakujen avulla. Täydennyksellä aineisto kasvoi kahdeksalla artikkelilla. Kirjallisuuskatsauksen lopullinen aineisto oli 19 artikkelia.

Lopulliseen aineistoon valikoitui ainoastaan tutkimuksia, jotka keskittyvät terveydenhuoltoympäristöön. Vaikka myös muiden ympäristöjen menetelmät saattaisivat soveltua potilastietojen väärinkäytön havaitsemiseen, terveydenhuolto-ympäristön erityispiirteet ohjasivat valintaa tähän suuntaan. Lisäksi erilaisia havaitsemismenetelmiä löytyi useita, joten lopullinen teknisten artikkelien määrä on varsin riittävä yhteen opinnäytetyöhön ja synteesin muodostamiseen.

Liitteessä 1 on aineistohaun tuloksena saadut artikkelit jaoteltuna hakulausekkeiden mukaan. Yhteenveto hakutuloksien määrästä ja valintaprosessista on kuvan 7 kaavion mukainen.



Kuva 7. Kirjallisuuskatsauksen aineistohaku.

3.4 Aineiston laadun arviointi

Katsaukseen sisällytettyjen tutkimusaineistojen laatua voidaan arvioida niiden merkityksellisyyden, tuoreuden ja aihealueen kohdistuksen perusteella. Aineisto voidaan luokitella merkityksellisyyden mukaan korkeaksi tai matalaksi (Whittemore & Knafl 2005: 549). Aineiston tuoreus määräytyy julkaisuajankohdan perusteella. Ajankohdista voidaan muodostaa jakauma, jonka perusteella tuoreudelle annetaan luokka 1, 2 tai 3, joista tuoreimmat aineistot saavat luokan 1. Lisäksi laatuarvioon lisättiin aihealueeseen liittyvä arviointikriteeri. Jos aineiston havaitsemismenetelmä hyödynsi oikeista potilaista kerättyjä käyttölokeja, laatu tulkittiin katsauksen kannalta paremmaksi. Aineiston laadun arviointi on esitetty liitteessä 2.

3.5 Aineiston analysointi ja tulkinta

Aineiston analysointivaiheessa pyrittiin sovittamaan aineistojen havaitsemismenetelmät teoriakappaleessa esiteltyihin yleisiin määritelmiin, vaikka aineistossa olisikin käytetty hieman poikkeavia ilmaisuja. Esimerkiksi jos artikkelissa pyritään havaitsemaan valtuutettujen käyttäjien suorittamaa potilastietojen ”epäilyttävää tai epäasiallista käyttöä”, järjestelmän voidaan olettaa etsivän sisäisiä uhkia, vaikka tällaista nimitystä ei käytetä. Vaikka epäilyttävä käyttö ei olisikaan aina oikeasti väärinkäyttöä, voidaan silti puhua sisäisen uhan havaitsemisesta, jossa osa hälytyksistä on vääriä. Toinen esimerkki on tutkimuksessa käytetty nimitys ”kyberhyökkäyksen havaitsemisen luottamuksellisuusskenaario”. Tässä skenaariossa ollaan kiinnostuneita käyttäjistä, jotka katselevat potilastietoja, jotka eivät heille kuulu tai käyttäjistä, jotka ovat saaneet valtuutetun käyttäjän tunnukset haltuunsa. Nimitys viittaa selvästi sisäisen uhan havaitsemiseen. Edellisten esimerkkien tavoin myös ”yksityisyyden loukkaamisen havaitseminen” valtuutettujen käyttäjien tekemänä on sisäisen uhan havaitsemista.

4 HAVAITSEMISMENETELMÄT

Kappaleessa 4 käydään läpi kirjallisuuskatsausaineiston sisäisen uhan havaitsemismenetelmiä, ja miten ne hyödyntävät terveydenhuollon käyttölokeja. Osion 2.2 esitetyistä lähestymistavoista sisäisten uhkien havaitsemiseen soveltuvia menetelmiä löytyi toteutetun kirjallisuuskatsauksen avulla useita. Katsaukseen sisälletyissä terveydenhuoltoon soveltuvissa artikkeleissa korostui jaon mukaisesti erityisesti tilastotieteeseen ja sääntöihin perustuvien luokkien lähestymistavat.

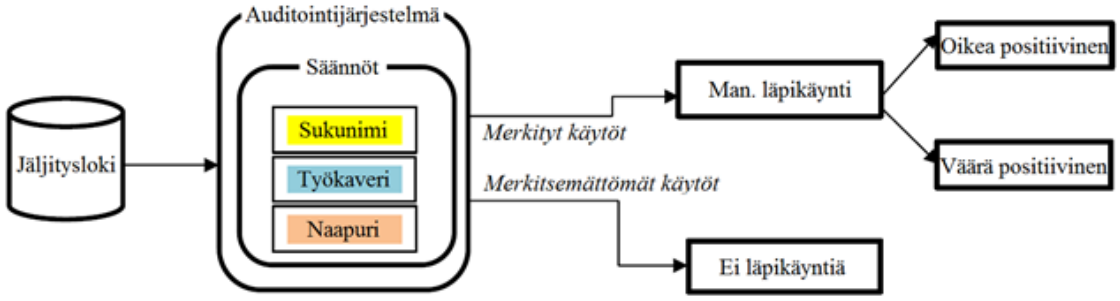
Terveydenhuollon sisäisen uhan havaitsemisessa on hyödynnetty tilastollista merkittävyyttä selvittävää khiin neliön (khi squared, χ^2) testiä ja peliteoriasta tuttua Stackelberg-pelimallia. Tutkimuksissa on tutkittu myös ohjatun oppimisen luokittelumenetelmiä, kuten lähimmän naapurin menetelmää (eng. k nearest neighbor, KNN), tukivektorikonetta (eng. support vector machine, SVM), logistista regressiota (eng. logistic regression, LR) ja naiivi bayes -luokitinta (eng. naive Bayes, NB. Ohjaamattoman oppimisen menetelmistä työhön sisällytettiin klusteroinnista hierarkkista klusterointia ja poikkeamien havaitsemisesta pääkomponenttianalyysin (principal component analysis, PCA) ja KNN-menetelmän yhdistelmää hyödyntävät menetelmät. Lisäksi löydetyissä poikkeamien havaitsemismenetelmissä, hyödynnettiin erilaisia läheisyysmittoja, joilla erotellaan selvästi poikkeavia havaintoja normaalien joukosta. Tuloksista löytyi myös menetelmiä havaintoja normalisoivien assosiaatiosääntöjen louhintaan.

Modernit potilastietojärjestelmät keräävät lokitietoja ja pitävät yllä käyttölokirekisteriä. Käyttölokitietueesta selviää ajankohta, jolloin järjestelmän käyttäjä, kuten lääkäri, on käyttänyt (esimerkiksi katsellut tai muokannut) potilaan tietoja. Useissa tutkimuksissa on havaittu, että potilastietojen käyttölokirekisterin ja terveydenhuolto-organisaatiosta tietokantaan kerättyjä tietoja voidaan käyttää sisäisten uhkien havaitsemiseen. Lokitietojen hyödyntämistä eli lokianalyysin tekoa on lähestytty eri tavoin.

4.1 Tunnisteperusteinen sisäisen uhan havaitseminen

4.1.1 Yksinkertaiset säännöt

Perinteisesti terveydenhuoltoympäristöissä potilastietojen väärinkäytön havaitsemiseen on käytetty lokianalyysijärjestelmää (auditointijärjestelmä), joka hyödyntää yksinkertaisia sääntöjä. Tällainen järjestelmä on tunnisteperusteinen sisäisen uhan havaitsemismenetelmä. Menetelmä on kuvan 8 kaavion mukainen.



Kuva 8. Sisäisten uhkien havaitseminen yksinkertaisten sääntöjen avulla (suomennettu englanninkielisestä lähteestä Hedda, Malin, Yan & Fabbri 2017: 867).

Järjestelmää varten voidaan luoda taulu, johon lisätään terveydenhuoltoon sopivia hälytysluokkia (eng. alert type) tai ominaispiirteitä (eng. feature), jotka viittaavat väärinkäyttöön. Hälytysluokkien taakse luodaan sääntöjä tutkimalla oikeita raportoituja tapauksia, joissa potilaan yksityisyyttä on rikottu. Jokaisesta säännöstä luodaan SQL-kysely, jolla on monta ehtoa. Tutkimuksissa käytettyjä sääntöjä ovat esimerkiksi (Boxwala ja muut 2011: 500; Kim, Grillo, Boxwala, Jiang, Mandelbaum, Patel, Mikels, Vinterbo & Ohno-Machado 2011: 726):

- Potilastietojen käyttöajankohta poikkeava
- Potilastietojen käyttömäärä korkea
- Potilas on VIP-rekisteriin merkitty henkilö
- Potilas työskentelee samalla osastolla kuin käyttäjä (työkaveri)
- Potilas kuuluu terveydenhuoltohenkilökuntaan (työkaveri)

- Potilas asuu samalla kadulla kuin käyttäjä (naapuri)
- Potilaan katuosoite maantieteellisesti lähellä käyttäjän osoitetta (naapuri)
- Potilas on käyttäjä itse
- Potilas on käyttäjän lapsi (sukulainen)
- Potilaalla sama sukunimi kuin käyttäjällä (sukulainen)
- Potilas on ollut kauan aikaa kuolleena
- Potilaalla ei ole ollut käyntiä lähiaikoina

Järjestelmä ilmoittaa lokitietueen käytön olevan väärinkäyttöä, kun jokin ehdoista täyttyy. Jos käytetään ainoastaan edellä esitettyjä sääntöjä, järjestelmän ilmoittamia tapauksia tutkiva tietoturva-asiantuntija (eng. privacy officer) ei tiedä, mikä tapauksista on tarkemman tutkimisen arvoinen. Järjestelmän ilmoittamien hälytysten ja niihin sisältyvien väärien hälytysten määrä voi olla suuri, paitsi jos tapaukset pystytään pisteyttämään ja priorisoimaan väärinkäytön todennäköisyyden mukaan (Boxwala ja muut 2011: 499.)

4.1.2 Hälytysten priorisointi

Hälytyksiä voi järjestää yksinkertaisten sääntöjen merkityksellisyyden mukaan. Sääntöjen merkitystä on arvioitu yhteensopivuutta (eng. goodness of fit) mittaavalla khiin neliön (χ^2) testillä. Sääntö on hyödyllinen, jos havaittujen hälytysten (eng. observed alerts) määrä on selvästi suurempi kuin satunnaisesti esiintyvien odotettujen hälytysten (eng. expected alerts) määrä. Khiin neliön testillä varmistetaan, että havaittujen ja odotettujen hälytysten määrän poikkeavuus on tilastollisesti merkittävä. (Hedda ja muut 2017: 868–869.) Havaittujen ja odotettujen hälytysten muodostaminen on selitetty liitteessä 3.

Säännön hyödyllisyyden todennuksen lisäksi hälytykset voidaan järjestää poikkeavuuksien suuruuden mukaiseen järjestykseen. Suuri poikkeavuus viittaa tärkeämpään sääntöön, joten tietoturva-asiantuntija voi priorisoida näiden sääntöjen hälytykset muiden edelle. Väärinkäyttöön viittaavien sääntöjen arvioinnin lisäksi on syytä tutkia merkityksettömiä sääntöjä, jotta voidaan varmistaa, ettei aineisto tee

kaikista säännöistä hyödyllisiä. Jos tutkitaan hälytysten määrää merkityksettömän säännön, kuten “potilaan etunimi vastaa käyttäjän etunimeä”, hälytysten havaittava määrä on lähellä oletettua määrää ja χ^2-testin tuloksen arvo on matala. (Hedda ja muut 2017: 873-874.)

Terveydenhuollon auditointiprosessin lokianalyysin tekoon on ehdotettu myös peliteorian hyödyntämistä. Auditointiprosessia mallinnetaan Stackelberg pelimallilla, joka on peli kahden pelaajan, puolustajan (sairaala) ja vihollisen (potilastietojärjestelmän käyttäjä), välillä. Puolustaja auditoi kohteiden joukkoa (potilastietojen käyttölokit) ja vihollinen valitsee kohteen, johon se hyökkää eli käyttää tietoa ilman lupaa. (Blocki, Christin, Datta, Procaccia & Sinha 2013: 1; Blocki, Christin, Datta, Procaccia & Sinha 2015: 1). Kun järjestelmän tekemiä hälytyksiä pyritään priorisoimaan, tietyntyyliset hälytykset saattavat jäädä vähemmälle huomiolle. Jos priorisointisäännöt ovat käyttäjän tiedossa tai ne ovat järjestelmän tuntevalle käyttäjälle pääteltävissä, käyttäjä saattaa hyödyntää järjestelmää ja laukaista ainoastaan hälytyksiä, jotka kiinnostavat vain vähän tietoturva-asiantuntijoita. (Laszka, Vorobeychik, Fabbri, Yan & Malin 2017: 196.)

Pelissä yksinkertaisten sääntöjen avulla luoduista ominaispiirteistä luodaan joukko hälytysluokkia (T), jossa samaan hälytysluokkaan (t) kuuluvat hälytykset ovat samanarvoisia. Esimerkiksi “potilas on VIP-rekisteriin merkitty henkilö” voi olla yksi hälytysluokista. Pelimallin avulla on tarkoitus määrittää optimaalinen strategia hälytysluokkien hälytysten läpikäymiseen. Luokat laitetaan priorisointijärjestykseen (o), joka on osa priorisointijärjestysten joukkoa (O). Puolustajalla on käytössään kiinteä budjetti (B), joka sillä on yhteensä varaa käyttää kaikkien hälytysten tutkintaan. Budjetti voi olla esimerkiksi tietoturva-asiantuntijalla käytettävissä olevat työtunnit tutkintoihin. Puolustaja määrää itselleen myös resurssit (C_t), kuten tietyn määrän tunteja budjetista, jotka se on valmis käyttämään kutakin hälytysluokkaa kohden. Vihollinen voi käynnistää hyökkäyksen (a), joka on jokin hyökkäysten joukosta (A). Todennäköisyyttä, jossa hyökkäys $a \in A$ saa aikaan hälytysluokan $t \in T$, merkitään symbolilla $R_{a,t}$. (Laszka ja muut 2017: 196.) Priorisoinnin algoritmi on esitetty liitteessä 4.

4.1.3 Ohjattu oppiminen ja luokittelu

Lokianalyysia on tehostettu ohjatun oppimisen luokittelualgoritmien avulla priorisoimalla hälytyksiä ja vähentämällä väärien hälytysten määrää. Näistä molempiin on hyödynnetty logistista regressiota (LR) ja tukivektorikonetta (SVM). Lisäksi väärien hälytysten määrää on pienennetty käyttämällä lähimmän naapurin -menetelmää (KNN) ja Naiivi Bayes -luokitinta (NB). Näillä ohjatun oppimisen malleilla hyödynnetään opetusjoukon havaintojen välisiä euklidisia etäisyyksiä tai havaintojen luokkien avulla laskettuja todennäköisyyksiä. Näistä tekniikoista logistiseen regressioon ja tukivektorikoneeseen pohjautuvia malleja on käytetty ennen potilastietoihin soveltamista esimerkiksi petosten havaitsemiseen luottokorttitransaktioista ja roskapostin havaitsevien filtterien luomiseen. (Boxwala ja muut 2011: 500; McGlade & Scott-Hayward 2018: 10)

Lokitietueiden väärinkäytön arviointiin ohjatun oppimisen avulla tarvitaan luokitin, ja luokitinta varten opetusjoukko. Jotta opetusjoukosta saadaan toimiva, siihen lisätään myös normaalia käytöstä kuvaavia sääntöjä. Tällaisia sääntöjä tai ominaispiirteitä ovat esimerkiksi (Boxwala ja muut 2011: 500):

- Potilas on hiljattain käynyt hoidettavana
- Käyttäjä on potilaan omalääkäri
- Käyttäjä on sama kuin potilaan tapaamiseen merkitty lääkäri
- Käyttäjä työskentelee samalla osastolla kuin potilaan omalääkäri
- Käyttäjä työskentelee samalla osastolla kuin tapaamiseen merkitty lääkäri
- Käyttäjä työskentelee osastolla, johon potilas on kirjattu sisäpotilaaksi
- Käyttäjä työskentelee päivystysosastolla

Opetusjoukkoon valitaan väärinkäyttöön ja normaaliin käyttöön viittaavia lokitietueita käyttölokirekisteristä esimerkiksi tietoturva-asiantuntijan avustuksella. Jokainen opetusjoukon lokitietue merkitään joko väärinkäyttöön viittaavaksi tai normaaliksi käytöksi. Väärinkäytöksi luetaan tapaukset, joissa potilastietojen käyttöön ei löydy potilaan hoitoon, laskutukseen tai terveydenhuoltotoimintoon liittyviä syitä.

Terveydenhuoltotoiminto voi olla esimerkiksi terveydenhuollon ammattilaisten koulutus. Opetusjoukon tietueet sisältävät valitun määrän sääntöjä. Taulukkoon 3 on muodostettu esimerkkiä varten supistettu opetusjoukko, joka sisältää seitsemän käyttölokitietuetta (ID) ja neljä ominaispiirrettä (OP).

Taulukko 3. Terveydenhuoltoympäristön supistettu opetusjoukko.

ID	**OP1**	**OP2**	**OP3**	**OP4**	**Luokka**
1	1	1	0	0	Väärinkäyttö
2	1	0	0	1	Väärinkäyttö
3	0	0	1	0	Normaalikäyttö
4	0	0	1	1	Väärinkäyttö
5	0	0	0	1	Normaalikäyttö
6	1	0	0	0	Normaalikäyttö
7	0	1	1	0	Normaalikäyttö

Opetusjoukon merkityistä lokitietueista voidaan luoda esimerkiksi LR-koneoppimismalli. LR-mallilla lasketaan opetusjoukon tapausten avulla todennäköisyyksiä. Selittyvä muuttuja on potilastiedon käyttö (lokitietue) ja muuttujan saava arvo tarkoittaa tässä esimerkissä väärinkäyttöä (1) tai normaalia käyttöä (0). Valinta on tehty näin päin, koska LR-mallissa selittyvän muuttujan arvolle, josta ollaan kiinnostuneita, annetaan arvoksi 1. Sen lisäksi, että käyttölokin tietueet voidaan ennustaa kuuluvaksi väärinkäytön ja normaalikäytön luokkiin opetusjoukosta muodostetun mallin avulla, LR määrittelee priorisointia varten jokaiselle ominaispiirteelle riskin (eng. odds). Riskiarvo on potilastiedon väärinkäytön todennäköisyyden suhde normaalin käytön todennäköisyyteen

$$\text{Riski} = \frac{p}{1-p} = \frac{P(y=1|x)}{1-P(y=1|x)} \quad (1)$$

Yhtälössä *p* tarkoittaa todennäköisyyttä sille, että potilastiedon käyttö on väärinkäyttöä, ja *x* on selittävä muuttuja (eng. predictor variable), jonka vaikutusta todennäköisyyksiin ja sitä kautta riskiarvoon tutkitaan. Selittävä muuttuja voi olla vaikka ”potilastietojen liiallinen käyttömäärä”. Logistisessa regressiossa todennäköisyydelle *p* tehdään niin kutsuttu logit-muunnos. Muunnos vastaa riskistä otettua luonnollista logaritmia. Kun tarkastellaan samanaikaisesti useampia selittäviä muuttujia, muuttujille valitaan suurimman uskottavuuden (eng. maximum likelihood) -menetelmällä kertoimet β_i. Logistisen regression malli usealla selittävällä muuttujalla on

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k . \quad (2)$$

Taulukossa 4 on erään tutkimuksen saamia LR-mallin kertoimia erilaisille ominaispiirteille.

Taulukko 4. Logistisen regression riskiarvot ominaispiirteille (muokattu lähteestä Boxwala ja muut 2011: 503).

Indeksi *i*	**Ominaispiirre**	**Kerroin (β_i)**	**Riskien suhde**
1	Työkaveri	3,158	23,524
2	Naapuri	2,599	13,450
3	Sama sukunimi	2,340	10,381
4	Potilastietojen liiallinen käyttömäärä	1,300	3,669
5	VIP	1,175	3,238
6	Potilas hiljattain hoidettavana	-0,616	0,540
7	Käyttäjä potilaan omalääkäri	-2.324	0,098

Taulukon 4 mukainen kerroin β_1 piirteelle ”työkaveri” on 3,158. Jos luonnollisen logaritmin kantaluku e korotetaan kertoimen suuruiseen potenssiin saadaan kahden riskin suhdeluku $e^{3,158}$ = 23,534. Riskien suhteen arvon mukaan työkaverin potilastietojen katselu on riskiltään 2253,4 % suurempi kuin käyttäjän, joka ei ole työkaveri. Taulukon 4 esimerkin mukaan tietoturva-asiantuntijan kannattaisi aloittaa väärinkäytösepäilyjen tutkiminen hälytyksistä, joissa käyttäjä on potilaan työkaveri.

Tukivektorikone (SVM) hyödyntää luokittelussa euklidisia etäisyyksiä. SVM-mallin avulla muodostetaan opetusjoukon luokiteltujen pisteiden välille taso. Tason molemmin puolin määritellään tukivektorit, jotka ovat tason kanssa samansuuntaiset ja yhtä kaukana siitä. Optimaalisin taso saadaan valittua, kun tukivektorien välinen etäisyys (marginaali) on suurimmillaan. Väärinkäytöksi luokitellut pisteet, jotka sijaitsevat kauimpana luokat jakavasta tasosta, priorisoidaan tärkeimmiksi. SVM-mallissa opetusjoukon pistejoukko muunnetaan kernelfunktion avulla usean ulottuvuuden avaruuteen, jossa data on mahdollista jakaa lineaarisesti joko väärinkäyttöön (1) tai normaaliin käyttöön (-1) (McGlade ja muut 2018: 16).

Lähimmän naapurin -menetelmällä (KNN) tutkitaan myös opetusjoukon havaintojen välisiä euklidisia etäisyyksiä. KNN-malli luokittelee uudet havainnot joko

väärinkäytöksi tai normaaliksi käytöksi lähimpien havaintojen etäisyyden avulla. KNN-mallin K on uuden havainnon läheisten pisteiden eli naapurien lukumäärä, ja naapurien luokista eniten esiintyvä ennustetaan luokaksi uudelle havainnolle. Jos K:n arvo on esimerkiksi 5, ja viiden lähimmän naapurin luokista neljä on väärinkäyttöä ja yksi normaalia käyttöä, uuden havainnon luokaksi valitaan väärinkäyttö.

NB-mallilla määritetään suhteellisia todennäköisyyksiä (eng. conditional probability) eri hypoteeseille. Oletetaan, että opetusjoukkoon kerätyt tiedot ovat taulukon 4 mukaisia usean muuttujan tietueita, joilla on luokka väärinkäyttö (1) tai normaali käyttö (0). Suhteellinen todennäköisyys NB-mallin mukaan on

$$P(H \mid X) = P(X \mid H)P(H). \tag{3}$$

joissa H_1 tarkoittaa hypoteesia 1, H_2 hypoteesia 2 ja muuttuja X sisältää ominaispiirteet. Kun hypoteesi H_1 tarkoittaa väärinkäyttöä ja hypoteesi H_2 normaalia käyttöä, uusi luokittelematon käyttölokitietue tutkituilla ominaispiirteillä X={OP1=1, OP2=1, OP3=0, OP4=1, Luokka=?} voisi saada NB-mallin mukaisiksi todennäköisyyksiksi $P(H_1|X)$=0,0211 ja $P(H_2|X)$ =0,0045. Koska hypoteesi H_1 on todennäköisempi, valitaan tietueen X luokaksi väärinkäyttö.

4.1.4 Ohjaamaton oppiminen, klusterointi ja suosittelu

Potilastietojen väärinkäytön havaitsemiseen on ehdotettu myös käyttäjien ja potilaiden historiatietoja hyödyntävää järjestelmää. Potilaalla, jonka potilastietoja on katseltu luvatta aiemmin, voi olla suurempi riski joutua uudelleen väärinkäytön kohteeksi. Myös käyttäjä, joka on aiemmin jäänyt kiinni väärinkäytöstä saattaa uusia tekonsa. (Menon, Jiang, Kim, Vaidya & Ohno-Machado 2014: 2.) Tällaisen järjestelmän toteutuksessa on käytetty suosittelujärjestelmistä tuttua yhteistoiminnalliseen suodatukseen (eng. collaborative filtering, CF) pohjautuvaa menetelmää. Yhteistoiminnalliseen suodatukseen perustuvaa mallia on testattu oikeasta terveydenhuoltoympäristöstä kerättyihin potilastietojen käyttölokitietoihin ja käyttäjistä kerättyihin organisaatiotietoihin.

Potilastietojen yhteydessä CF-perusteinen menetelmä yhdistää useiden samankaltaisten käyttäjien ja potilaiden tietoja ennustaakseen potilastietojen käytön väärinkäytöksi. Koska käyttäjien ja potilaiden määrä on suuri, CF-menetelmää voidaan vahvistaa hyödyntämällä ohjaamattomasta oppimisesta tuttua hierarkkista klusterointia. Klusteroinnilla käyttäjistä ja potilaista muodostetaan ryhmiä, joiden sisällä käyttäjät tai potilaat ovat samankaltaisia. Esimerkiksi kaikki hoitajat, jotka työskentelevät tutkimustyössä, voidaan kategorisoida yhteen ryhmään. Jos yksittäisistä käyttäjistä ja potilaista muodostetaan matriisi, matriisiin jää paljon tyhjiä rivejä ja sarakkeita, jolloin piilevien ominaispiirteiden arviointi on epäluotettavaa. Tällaista aineistoa pidetään harvana (eng. sparse) ja aineistoon liittyy niin sanottu kylmäkäynnistysongelma (eng. cold-start problem). Jotta tyhjiä soluja saadaan vähennettyä, käyttäjien ja potilaiden ryhmittely hierarkkisen klusteroinnin avulla on tarpeen. Uudet käyttäjät ja potilaat lisätään aina muodostettuihin klustereihin ja kylmäkäynnistykseltä vältytään. (Menon ja muut 2014: 4–5.)

CF-perusteinen ratkaisu yhdistää potilastietojärjestelmän käyttäjiin ja potilaisiin liittyviä selkeitä (eng. explicit) ja piileviä (eng. latent) ominaispiirteitä (sääntöjä). Piilevät säännöt luodaan potilastietojen käyttöhistorian perusteella ja ne ovat jokaiselle käyttäjälle tai potilaalle yksilöity tunniste kuten sormenjälki ihmiselle. (Menon ja muut 2014: 2.) Klusteroinnilla vahvistetussa järjestelmässä yksilöity tunniste on muodostettu käyttäjä- tai potilasryhmälle, eikä yksittäisille käyttäjille tai potilaille.

Oletetaan, että meillä on käyttölokista muodostettu lokitietue (Menon ja muut 2014: 2)

$$a' = (u', p', x_u', x_p', x_r'), \tag{4}$$

jossa $u \in \{1, ..., m\}$ on käyttäjäryhmän id ja m käyttäjäryhmien määrä, $p \in \{1, ..., n\}$ on potilasryhmän id ja n potilasryhmien määrä, x_u on käyttäjäryhmiin liittyvä ominaispiirre (esim. käyttäjä on lääkäri), x_p on potilasryhmiin liittyvä ominaispiirre (esim. potilaalla äskettäinen sairaalakäynti) ja x_r on käyttäjä- ja potilasryhmän välisiin suhteisiin liittyvä ominaispiirre (esim. käyttäjällä ja potilaalla sama sukunimi). Lokitietue määritetään, joko väärinkäytöksi tai normaaliksi käytöksi. Määrittelyä varten koneoppimismalli sovitetaan käyttämällä opetusjoukkoa (Menon ja muut 2014: 2)

$$T = \left\{\left(a^{(i)}, y^{(i)}\right)\right\}_{i=1}^{N}, \text{ jossa} \quad (5)$$

$a^{(i)}$ sisältää ominaispiirteet, jotka tekevät yhteenvedon potilastiedon käytöstä, $y^{(i)} \in \{0,1\}$ on merkintä, joka kertoo, onko kyseessä potilastiedon väärinkäyttöä (1) vai normaalia käyttöä (0) ja N on opetusjoukon koko (luokiteltujen rivien määrä). Opetusjoukko voidaan esittää matriisimuodossa.

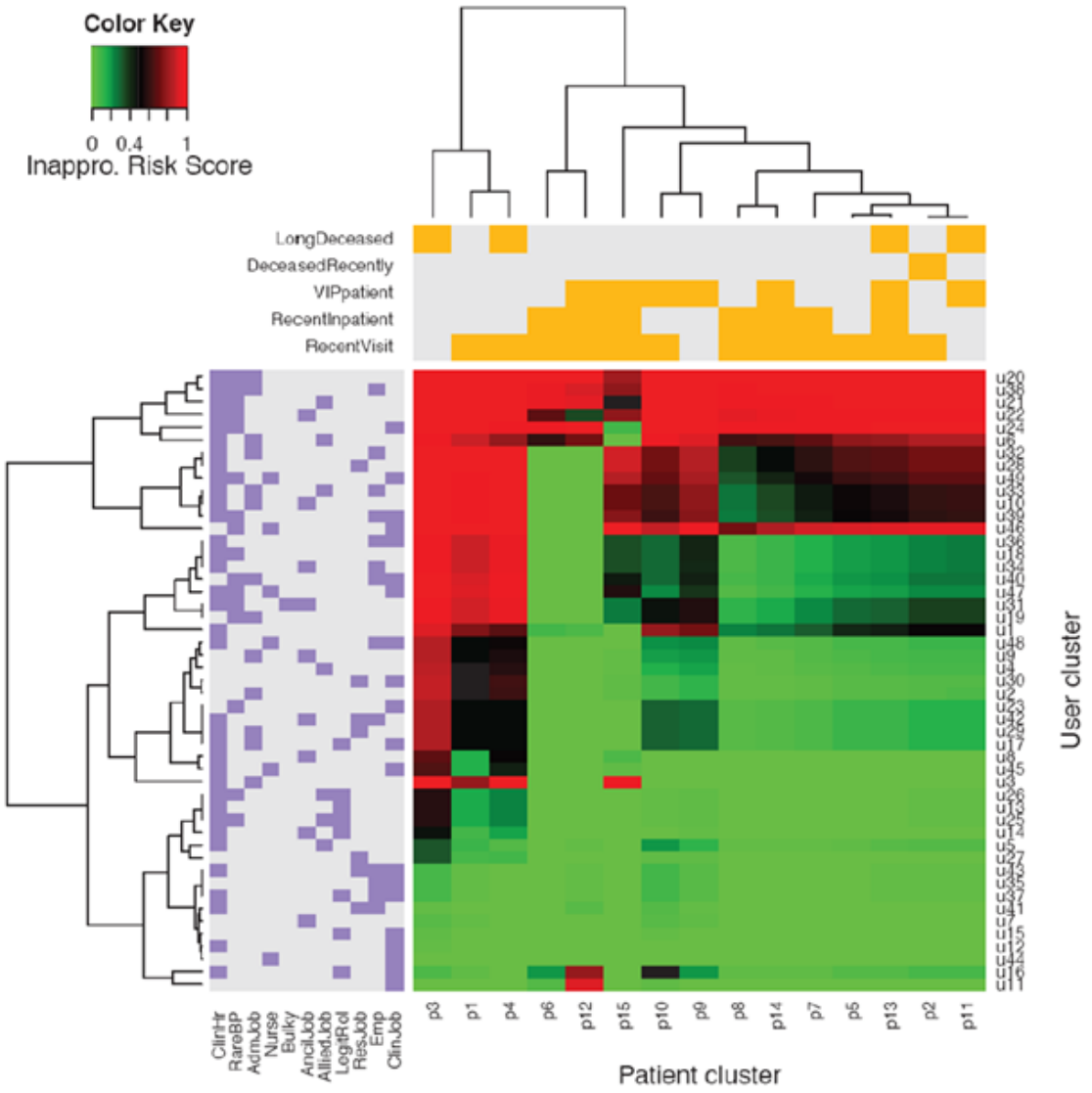
CF-perusteisella mallilla pyritään ennustamaan, miten vahvasti käyttäjä-potilas-pari viittaa väärinkäyttöön. Kun perinteisessä suosittelujärjestelmässä käyttäjä kertoo arvosanalla, paljonko hän pitää tuotteesta, kuten elokuvasta, väärinkäyttötapauksessa mitataan numeerisen arvon avulla käyttäjän ja potilaan välisen vuorovaikutuksen asiallisuutta. Käyttäjä-tuote-parille annetaan vain yksi arvosana, kun taas käyttäjä-potilas-parin solu voi sisältää useita arvoja, koska sama käyttäjä voi käyttää yhden potilaan tietoja useammin kuin kerran. Lokitietueen väärinkäytön todennäköisyyttä arvioidaan mallilla (Menon ja muut 2014: 4)

$$\hat{y}(a;\theta) = f\left(w^T \varphi(a) + \alpha_u^T \beta_p + \gamma_u + \delta_p + \mu\right), \quad (6)$$

jossa a on ominaispiirteistä muodostettu lokitietue, α_u on käyttäjäryhmän piilevä ominaispiirre, β_p on potilasryhmän piilevä ominaispiirre, $\varphi(a)$ on selkeä ominaispiirre ja $w^T\varphi(a)$ on termi, joka hyödyntää selkeästä ominaispiirteestä saatavilla olevaa tietoa. Termeistä $\gamma_u \in R$ on käyttäjäryhmään liittyvä painoarvo, $\delta_p \in R$ on potilasryhmään liittyvä painoarvo ja μ on globaali painoarvo.

CF-perusteisen mallin tulokset voidaan esittää esimerkiksi lämpökartan avulla, jossa käyttäjä- ja potilasklusterit on järjestetty hierarkkisen klusteroinnin avulla selkeään järjestykseen. Kuvan 9 esimerkin potilaista muodostettiin 15 klusteria ja käyttäjistä 49 klusteria. Testien perusteella käyttäjä- ja potilasklusterien välillä on informatiivisia tunnisteita. Kuvassa vasempaan reunaan on listattu käyttäjiin liittyviä ja ylhäälle potilaisiin liittyviä sääntöjä. Vaakarivillä on listattu myös potilasklusterit ja pystysuuntaan käyttäjäklusterit. Käyttäjäklusterin ja potilasklusterin säännöistä lasketaan mallilla käyttäjä- ja potilasklusterin välinen ”arvosana” eli riskiarvo. Suuri

riskiarvo on lähellä arvoa 1, jolloin potilastiedon käyttö on todennäköisesti väärinkäyttöä. Kun arvo on lähellä arvoa 0, käyttöä pidetään normaalina. CF-perusteisesta mallista muodostetusta lämpökartasta voidaan esimerkiksi tulkita, että kun käyttäjäklusteriin u10 kuuluvat käyttäjät käyttävät potilasklusteriin p3 kuuluvien potilaiden tietoja, kyseessä on todennäköisesti väärinkäyttö. Esimerkki normaalista käytöstä on käyttäjä-potilas-klusteripari p13 ja u15. Kuvasta ilmenee myös, mitkä potilas tai käyttäjäklusterit ovat samankaltaisia keskenään ja millaisten potilaiden tietojen käyttö on sallittua erilaisille käyttäjäryhmille. Esimerkiksi käyttäjäryhmän u31, jolle on tyypillistä käyttää yli 200 potilastietoa päivittäin (Bulky), toiminta viittaa väärinkäyttöön, jos käyttäjät käyttävät kauan kuolleina olleiden potilaiden (LongDeceased), kuten ryhmien p3 ja p4, potilastietoja (Menon ja muut 2014: 10).



Kuva 9. Käyttäjien ja potilaiden välisistä suhteista muodostettu lämpökartta (Menon ja muut 2014: 15).

4.2 Poikkeamaperusteinen sisäisen uhan havaitseminen

4.2.1 Ohjaamaton oppiminen ja assosiaatiosääntöjen louhinta

Terveydenhuollon sisäisen uhan havaitsemista on tutkittu järjestelmillä, joiden lokianalyysistrategia perustuu ohjaamattomaan oppimiseen ja on menetelmältään poikkeamaperusteinen. Sisäisen uhan selvittämistä voi lähestyä tekemällä oletuksen, että jokaiselle potilastiedon käytölle on aina olemassa syy (Fabbri & LeFevre 2011a: 1).

Järjestelmä luo malleja normaalista käytöstä ja kaikki normaalin ulkopuolelle jäävä on mahdollista väärinkäyttöä. Jos järjestelmä toimii tällä periaatteella ja saa syyt selville tehokkaasti, se pystyy rajaamaan suurimman osan lokitietueista väärinkäytösepäilyjen ulkopuolelle. Jotta järjestelmä pystyy pitämään lokitietueita normaaleina, sen täytyy kyetä vastaamaan kysymykseen, miksi potilastietoja on käytetty. Kuvitellaan esimerkki Anna nimisen potilaan käyttölokirekisteriin kerätyistä lokitiedoista. Syyn selittävän järjestelmän käyttölokirekisteri voisi olla esimerkiksi taulukon 5 mukainen.

Taulukko 5. Käyttölokirekisteriin kerätyt lokitiedot potilaasta Anna (taulukko suomennettu englanninkielisestä lähteestä Fabbri ja muut 2011a: 1).

Syykoodi	**Ajankohta**	**Käyttäjä**	**Potilas**
L100	03.08.2019 10:16:57	Hanna Hoitaja	Anna
L116	03.08.2019 11:22:43	Tiina Tohtori	Anna
L127	03.08.2019 17:09:03	Riina Radiologi	Anna
L900	12.10.2019 14:29:08	Katja Kirurgi	Anna

Ideaalisessa järjestelmässä jokaisesta yksittäisestä potilastiedon käytöstä tallentuisi lokitietue, josta selviää syy tietojen käytölle. Lokitietuetta klikkaamalla syykoodin mukainen selitys näkyisi potilaalle tai terveydenhuollon tietoturva-asiantuntijalle. Lyhyet tekstit olisivat esimerkiksi alla olevien mukaisia (Fabbri ja muut 2011a: 2.):

- L100: Potilaalla Anna oli tapaaminen lääkärin Tiina kanssa 01.08.2019
- L116: Hoitaja Hanna työskentelee lääkärin Tiina kanssa, ja potilaalla Anna oli tapaaminen lääkärin Tiina kanssa 01.08.2019.
- L127: Radiologi Riina arvioi potilaan Anna röntgen-kuvat lääkärille Tiina
- L900: Kirurgi Katja suoritti potilaalle Anna leikkauksen, kun lääkäri Tiina ohjasi potilaan Anna kirurgille Katja.

Edellä kuvatut selitykset ovat eri tyyppisiä. Syykoodin L100 takana oleva selitys on niin sanottu suora selitys (eng. direct explanation), jossa potilaan käyttölokitietue voidaan yhdistää suoraan lääkärin kanssa sovittuun tapaamiseen. Tapaaminen on kirjattu erilliseen tapaamiset-tauluun. Syykoodin L116 tapahtuma liittyy ryhmäselitykseen (eng. group explanation), jossa esimerkiksi potilaan potilastietoja käyttänyt työskentelee tapaamiseen merkityn lääkärin kanssa. Syykoodi L127 viittaa tapahtumaan, jossa selityksenä on konsultaatioselitys (eng. consultation explanation). Lääkäri määräsi potilaan röntgen-kuviin ja on yleistä käyttää radiologin antamia lausuntoja tukena hoitosuosituksissa. (Fabbri, LeFevre & Hanauer 2011b: 12–13.)

Jotta lokitietoja keräävä järjestelmä ei muuttuisi liian raskaaksi, sen pitäisi kyetä löytämään selityksiä ja luomaan tekstiä syykoodien taakse automaattisesti. Tällainen järjestelmä säästää tietoturva-asiantuntijan aikaa, eikä vaadi täydellistä ymmärrystä osastojen ja hoitohenkilökunnan yhteistyökuvioista. Lisäksi järjestelmä voisi mahdollistaa käyttäjäkeskeisen auditoinnin (eng. user-centric auditing), jossa potilas voi tarkastaa omien potilastietojensa käyttäjiä. Jos lokitiedot on esitetty taulukon 5 mukaisessa yksinkertaisessa muodossa, potilaat voisivat ilmoittaa epäilyksistä. (Fabbri ja muut 2011a: 1)

Kokeellisesti on pystytty osoittamaan, että terveydenhuoltoympäristössä valtaosalle käyttölokitietueista pystytään muodostamaan edellä esitellyn mukainen syy automaattisesti rajatulla joukolla selitysmallinteita (eng. explanation template). (Fabbri ja muut 2011a: 3.) Selitysmalline on SQL-kysely tietokantaan ja sen sisältämään käyttölokirekisteriin, ja sillä voidaan selittää useita yksittäisiä potilastietojen käyttöjä. Oletetaan, että kysely Q saa muodon (Fabbri ja muut 2011a: 3)

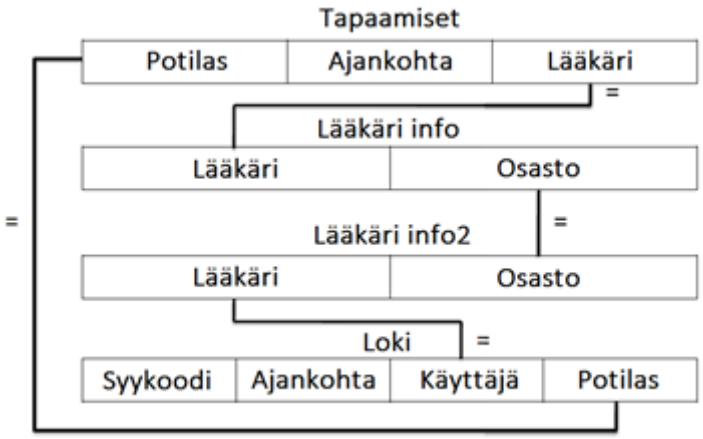
```
SELECT A_1, ..., A_m
FROM T_1, ..., T_n
WHERE E_1 AND ... AND E_i,
```

jossa $T_1,...,T_n$ ovat tauluja tietokannassa, $A_1,...,\ A_m$ ovat taulukon rivin alkioita eli attribuutteja, ja jokainen E_i on ehtolauseke, jossa vertaillaan attribuutteja esim. $A_1 \theta A_2$, kun $\theta \in \{<, \leq, =, \geq, >\}$.

Selitysmallinteen toimintaa voidaan havainnollistaa selitysgraafin ja siinä kulkevan polun (eng. path) avulla. Potilastiedon käyttö on normaalia, kun polku on hyväksyttävä. Polku taas on hyväksyttävä, jos se alkaa lokitietueen attribuutista ”Potilas” ja päättyy saman lokitietueen attribuuttiin ”Käyttäjä”. Oletetaan, että G on graafi, jossa jokainen taulujen T attribuutti on graafin solmukohta. Solmukohtien A_1 ja A_2 välillä on yhdistävä jana, jos A_1 ja A_2 ovat samassa tuplemuuttujassa tai jos kyselyn Q ehto attribuuttien A_1 ja A_2 välillä toteutuu. Esimerkkitapaus graafista ja sen toiminnasta voisi olla seuraavanlainen. Oletetaan, että järjestelmä muodostaa lokitietueesta tiedon (Fabbri ja muut 2011a: 3):

”Lääkäri Tiina käytti potilaan Birgitta potilastietoja, koska potilaalla Birgitta oli tapaaminen lääkärin Leena kanssa 01.09.2019, ja lääkäri Leena työskentelee lääkärin Tiina kanssa osastolla Lastentaudit.”

Esimerkin selitysgraafi polkuineen on kuvan 10 mukainen ja tietokannan taulut esimerkiksi kuvan 11 mukaiset.



Kuva 10. Selitysgraafi, jossa polku lokitietueen attribuutista muiden taulujen kautta lokitietueeseen (suomennettu englanninkielisestä lähteestä Fabbri ja muut 2011a: 3).

Potilas	Ajankohta	Lääkäri
Anni	01.08.2019	Tiina
Birgitta	01.09.2019	Leena

a) Tapaamiset

Lääkäri	Osasto
Leena	Lastentaudit
Tiina	Lastentaudit

b) Lääkäri info

Syykoodi	Ajankohta	Käyttäjä	Potilas
L1	03.08.2019	Tiina	Anni
L2	02.09.2019	Tiina	Birgitta

c) Loki

Kuva 11. Tietokannan taulut (suomennettu englanninkielisestä lähteestä Fabbri ja muut 2011a: 3).

Esimerkkiä vastaava selitysmalline eli SQL-kysely on (Fabbri ja muut 2011: 3)

```
SELECT L.Syykoodi, L.Potilas, L.Kayttaja, T.Laakari,
      T.Ajankohta, I1.Osasto

FROM Loki L, Tapaamiset T, Laakari_Info I1, Laakari_Info2 I2

WHERE L.Potilas = T.Potilas
      AND T.Laakari = I1.Laakari
      AND I1.Osasto = I2.Osasto
      AND I2.Laakari = L.Kayttaja.
```

Esimerkkinä muodostettu teksti on siis muodostettu attribuuttien avulla seuraavasti:

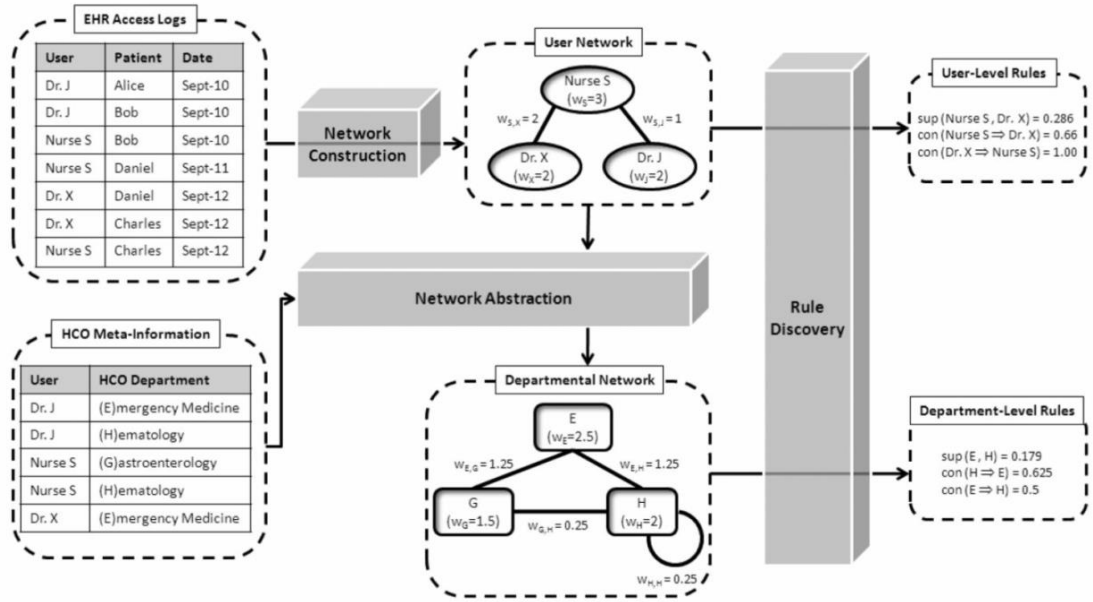
”Lääkäri [L.Kayttaja] käytti potilaan [L.Potilas] potilastietoja, koska potilaalla [T.Potilas] oli tapaaminen lääkärin [T.Laakari] kanssa [T.Ajankohta], ja lääkäri [L.Kayttaja] työskentelee lääkärin [T.Laakari] kanssa osastolla [I1.Osasto].”

Vaikka selitysmallinteita hyödyntävässä järjestelmässä mallinteita voidaan luoda manuaalisesti, järjestelmä toimii myös ohjaamattoman oppimisen avulla. Selitysmallinteen muodostamiseen voidaan käyttää esimerkiksi yksisuuntaista tiedonlouhinta-algoritmia, joka etsii hyväksyttäviä polkuja ja luo niiden pohjalta assosiaatiosäännöt eli selitysmallinteet toiminnoista, jotka kuvaavat normaalia käytöstä.

Tietoturva-asiantuntijan tehtäväksi jää arvioida, ovatko mallinteet hyödyllisiä. Algoritmin rakenteen kuvaus on tehty liitteessä 5.

Selitysmallinemenetelmän toimivuutta on testattu myös yhdessä yksinkertaisia sääntöjä hyödyntävän menetelmän kanssa. Selitysmallinteilla pystyy selittämään osan sääntöjen pohjalta tehdyistä hälytyksistä. Joidenkin sääntöjen kohdalla jopa lähes puolet hälytyksistä on selitetty normaaliksi käytöksi. Menetelmä auttaa myös hälytysten priorisoinnissa. Jos mallinteilla ei löydetä lokitietueelle normaalia käyttöselitystä ja sama tietue hälyttää myös sääntöjen kautta, nostetaan lokitietue priorisointilistan kärkipäähän. Jäljelle jäävät selitetyt hälytykset voidaan järjestää hälytystyypin ja selitysten määrän mukaan. Toisinaan useampi malline selittää tietyn lokitietueen käytön normaaliksi, jolloin selitys on yhtä mallinnetta uskottavampi. (Hedda ja muut 2017: 874)

Assosiaatiosääntöjä on louhittu myös toisella tapaa. Terveydenhuollon toiminnoista voidaan rakentaa verkosto, josta ilmenee, miten potilastietojen käyttäjät ovat yhteydessä toisiinsa potilastietojen kautta. Verkostosta voidaan taas louhia assosiaatiosääntöjä, joiden perusteella voidaan määrittää suhteelliset todennäköisyydet, joilla käyttäjä on katsellut potilastietoja suhteessa toiseen käyttäjään. Menetelmällä on mahdollista louhia sääntöjä myös hoito-osastojen välille, kun käyttölokitietoja yhdistetään organisaatiotietoihin. (Malin, Nyemba & Paulett 2011: 4.) Menetelmä on kuvan 12 kaltainen.

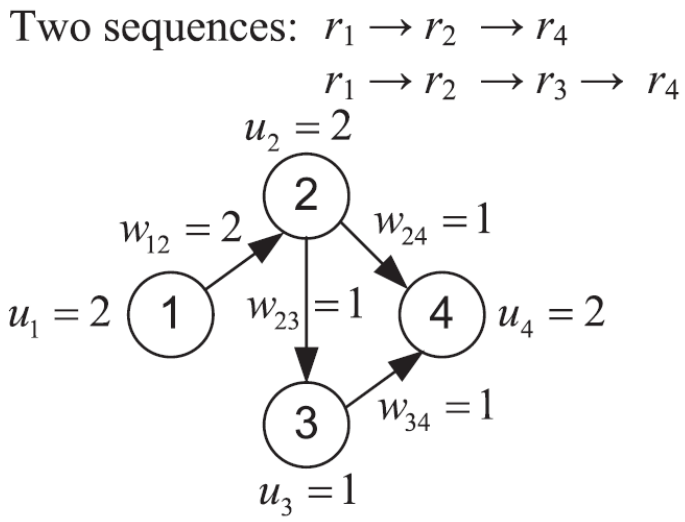


Kuva 12. Assosiaatiosääntöjen louhinta (Malin ja muut 2011: 17).

Käyttölokiaineistosta löytyvistä käyttäjistä piirretään solmupisteitä, joiden välille voidaan piirtää jana, kun kaksi käyttäjää on käyttänyt saman potilaan tietoja. Jokaiselle solmupisteelle ja janalle annetaan painoarvo *w*. Kuvan 12 esimerkissä painoarvo on suoraan käytettyjen potilastietomerkintöjen määrä. Esimerkissä lääkäri X on käyttänyt kahden potilaan tietoja, jolloin painoarvo w_X on kaksi. Lääkärin X ja hoitajan S välinen painoarvo $w_{S,X}$ on myös kaksi, koska he ovat molemmat käyttäneet potilaiden Daniel ja Charles tietoja. Assosiaatiosääntöjen perusteella voidaan laskea esimerkiksi todennäköisyys sille, että hoitaja S käyttää potilaan tietoja, kun saman potilaan tietoja on käyttänyt lääkäri X. (Malin ja muut 2011: 5.)

Verkoston ei tarvitse muodostua käyttäjien määristä, vaan siihen voidaan käyttää myös muita ominaisuuksia, jotka testaavat potilastietojen käytön tai hoitojakson asiallisuutta. Tällaisia ovat esimerkiksi käyttäjälle määrätty rooli, potilaalle annettava hoito tai potilaan hoito-osasto. (Zhang, Mehotra, Liebovitz, Gunter & Malin 2013: 5.) Hoitopolkua voisi tarkastella esimerkiksi käyttäjien roolien mukaan ja muodostaa kolmen käyttäjän polku. Käyttölokin aikaleimojen perusteella kävisi ilmi, että ensimmäiseksi potilaan tietoja on katsonut hoitaja, jonka jälkeen lääkäri ja viimeisenä laskutuksesta vastaava henkilö. Jos jokainen tarkasteltava käyttäjärooli (r_1, r_2, r_3 ja r_4) vastaa solmua ja polut $r_1 \rightarrow r_2 \rightarrow r_4$ ja $r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_4$ kahden potilaan hoitojaksoa,

kahdesta potilaasta ja neljästä erilaisesta käyttäjäroolista voitaisiin muodostaa kuvan 13 mukainen graafi.



Kuva 13. Hoitojakson eteneminen aikaleimojen ja ominaisuuksien perusteella. (Zhang ja muut 2013: 6).

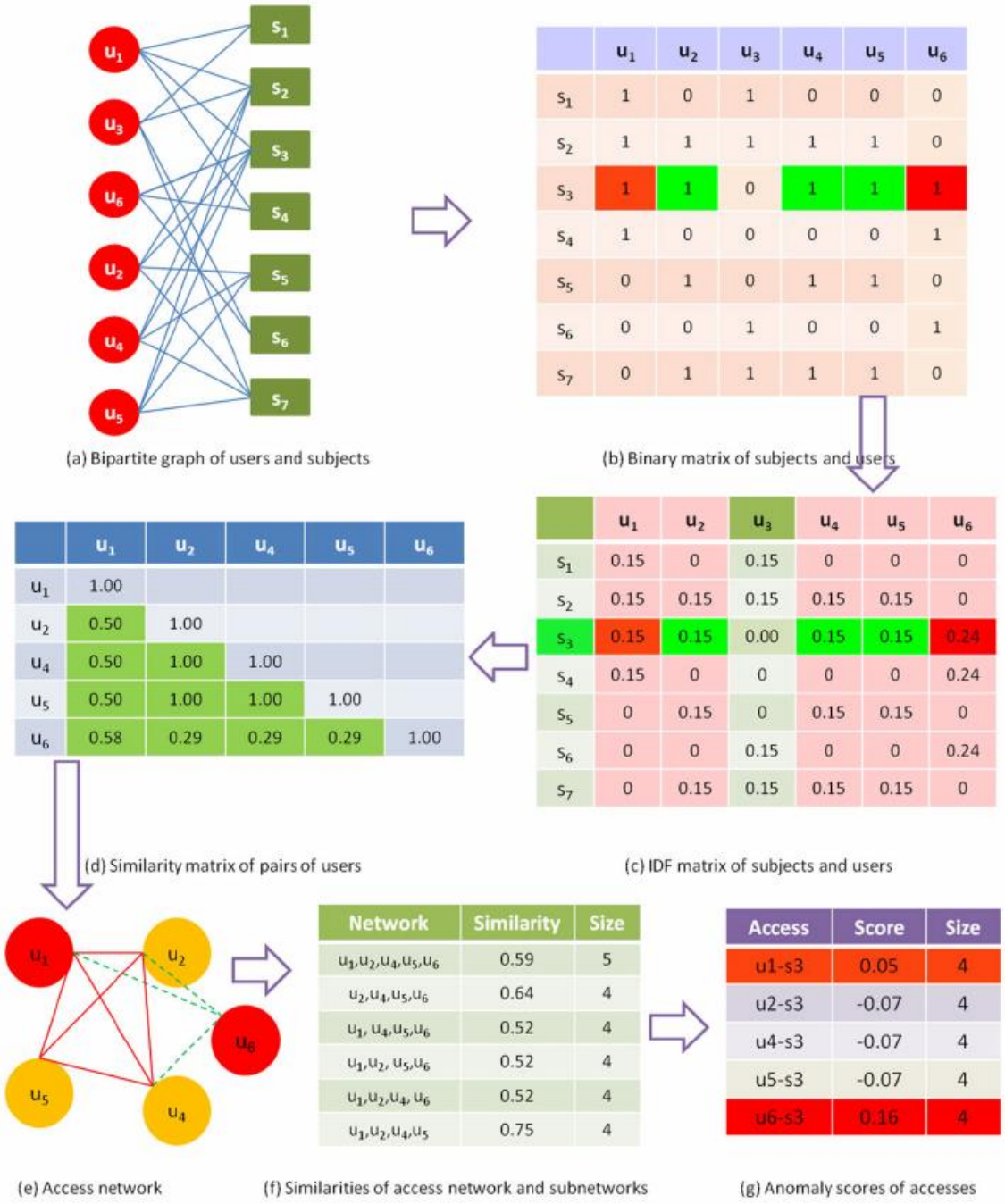
4.2.2 Ohjaamaton oppiminen ja poikkeamien havaitseminen

Assosiaatiosääntöjen lisäksi poikkeamaperusteinen ja ohjaamattomaan oppimiseen perustuva järjestelmä voi hyödyntää poikkeamien havaitsemista. Jos assosiaatiosäännöillä etsitään tietynlaisia järjestyksiä noudattavia polkuja, poikkeamien havaitsemisella etsitään samankaltaisten havaintojen joukosta esiin pistäviä poikkeavia havaintoja. Kuten osiossa 4.2.1 on todettu, käyttäjistä voidaan muodostaa verkostoja. Järjestelmän perusoletus on, että terveydenhuollon työympäristöt vaativat yhteistyötä. Siispä normaalit käyttäjät muodostavat normaaleja käyttäjäyhteisöjä, joiden rakenteet näkyvät käyttölokeista ilman organisaation rakenteen tarkempaa tuntemusta. Järjestelmä tekee näkyvistä rakenteista poikkeavista käyttäjistä (eng. anomalous users) hälytyksen. (Chen, Nyemba, Zhang & Malin 2011b: 2.)

Potilastiedon käyttöä pidetään epäilyttävänä, jos käyttäjiä useista erilaisista osastoista käyttää sitä. Jokaisen potilaan potilastiedoille muodostetaan yhteisö, joka koostuu kaikista tietojen käyttäjistä. Jokaista käyttäjää verrataan toiseen käyttäjään erillisellä

läheisyysmitalla. Läheisyysmitta muodostetaan molempien käyttäjien käyttämien potilastietojen potilaiden lukumäärien avulla. Myös yhteisölle määritellään läheisyys-arvo jokaisen käyttäjäparin etäisyyksien avulla. Algoritmi arvioi jokaisen käyttäjän vaikutusta potilastiedon käyttöön poistamalla käyttäjän yhteisöstä ja laskemalla, kuinka yhteisön läheisyysarvo muuttuu käyttäjän poiston jälkeen. Suuri muutos viittaa käyttäjän käyttävän tietoja, jotka eivät hänen alueelleen kuulu. Tällainen käyttäjä saattaa tarkastella myös muita hänelle kuulumattomia tietoja, mikä on usein merkki informaatiovarkaudesta. (Chen ja muut 2011b: 2; Ko, Divakaran, Liau & Thing 2016: 5.) Sisäisten uhkien havaitsemisprosessi käyttäjien läheisyysmittojen avulla voi olla esimerkiksi kuvan 14 mukainen.

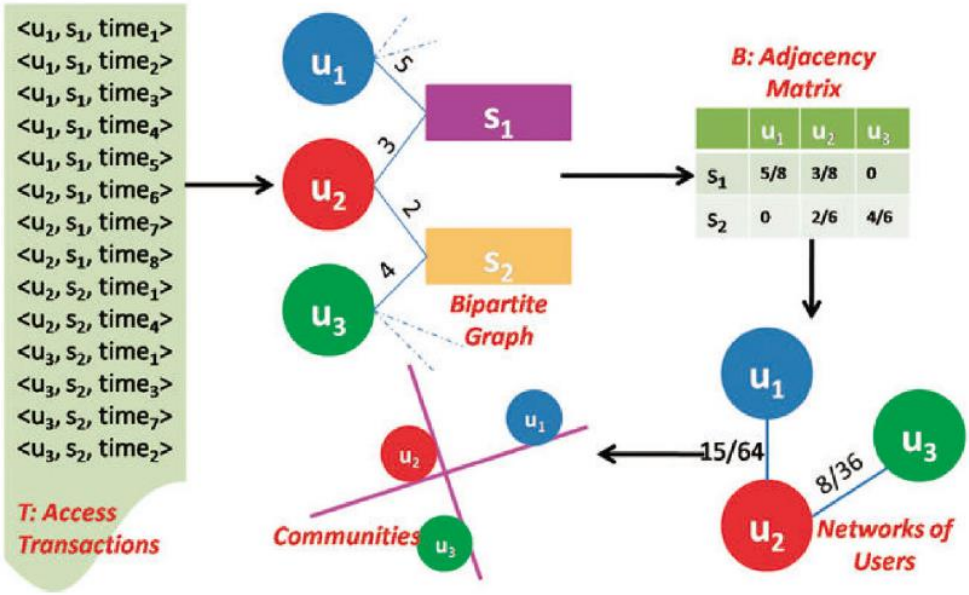
Kuvan 14 esimerkissä tarkastellaan kuutta käyttäjää ja seitsemää potilasta. Käyttäjä u_i on käyttänyt potilaan s_i potilastietoja, jos käyttäjän ja potilaan välillä on jana (a). Käytöistä muodostetaan matriisi, jossa käyttöä on merkitty numerolla 1 (b). Binäärimatriisi muutetaan matriisiksi, jossa käyttäjäkohtainen potilasmäärä vaikuttaa matriisiin merkittyyn painoarvoon (c). Painoarvojen avulla käyttäjiä verrataan toisiinsa ja käyttäjäparien välille lasketaan läheisyysmitat (d). Läheisyysmittojen perusteella muodostetaan käyttäjäyhteisö esimerkiksi potilaan s3 potilastietojen käyttäjistä (e). Käyttäjäyhteisölle lasketaan oma läheisyysmitta ja katsotaan, miten se muuttuu, kun yksi käyttäjistä poistetaan kerrallaan yhteisöstä (f). Yhteisön käyttäjille lasketaan läheisyysmittojen muutosten perusteella poikkeavuusarvo (g). Oletus on, että käyttäjäyhteisön läheisyysmitta laskee, kun poikkeava käyttäjä eli sisäinen uhka kuuluu siihen. Mitä suurempi positiivinen poikkeavuusarvo on, sitä todennäköisemmin käyttäjä on sisäinen uhka. (Chen ja muut 2012c: 6–7; Chen ja muut 2011b: 2–3.) Menetelmän matemaattiset kaavat ovat liitteessä 6.



Kuva 14. Sisäisen uhan havaitseminen käyttäjien läheisyysmittojen avulla (Chen ja muut 2011b: 12).

Poikkeavia käyttäjiä on pyritty havaitsemaan myös kaksiosaisella ohjaamattomaan oppimiseen pohjautuvalla järjestelmällä, joka hyödyntää poikkeamien havaitsemista. Ensimmäinen osa järjestelmää poistaa datasta kohinaa ja havaitsee siitä piilossa olevia kiinnostavia rakenteita eli oletettuja yhteisöjä. Tällainen osa järjestelmää voi hyödyntää

esimerkiksi PCA-menetelmää. PCA-osassa usean muuttujan lokiaineistoa yksinkertaistetaan vähentämällä siitä muuttujia, mutta samaan aikaan pyritään säilyttämään oleellinen informaatio. Toinen osa havaitsee käyttäjiä, jotka poikkeavat näistä monimutkaisista yhteisörakenteista. Toiseen osaan on hyödynnetty KNN-menetelmää, mutta ilman opetusjoukkoa. (Chen ja muut 2011a: 66) Periaatteellinen kuvaus yhteisöjen muodostamisesta on esitetty kuvassa 15.



Kuva 15. Yhteisöjen muodostaminen PCA-menetelmällä (Chen ja muut 2011a: 66).

Kuvassa 15 käyttölokitietueista muodostetaan käyttäjä-potilas -verkosto. Tämän pohjalta muodostetaan käyttäjä-käyttäjä -verkosto, josta päätellään PCA:n avulla mahdolliset yhteisöt. Kuvan esimerkissä potilaan s_1 tiedoista on luotu kahdeksan lokitietuetta, joista käyttäjä u_1 on käyttänyt viittä ja käyttäjä u_2 kolmea. Käyttäjän u_1 ja potilaan s_1 välinen suhde merkitään matriisiin arvolla 5/8. Käyttäjien u_1 ja u_2 välinen suhde taas on potilaan s_1 arvojen tulo 15/64.

Poikkeamaperusteista läheisyysmitan periaatetta on hyödynnetty käyttäjäyhteisöjen muodostamisen lisäksi myös potilastietojen käyttöön sallittujen terveydenhuollon osastojen määrittämiseen. Jos potilaan diagnoosi tunnetaan ja se on saatavilla järjestelmän tiedoista, voidaan laskea käyttötodennäköisyys, joka kertoo, onko tietyn

osaston työntekijälle tavanomaista käyttää tietyn diagnoosin saaneen potilaan tietoja (Fabbri & LeFevre 2013: 55):

$$P(\text{osasto } d \mid \text{diagnoosi } c) = \frac{\text{Diagnoosin } c \text{ potilaiden lkm, joilla käyttäjä osastolla } d}{\text{Diagnoosin } c \text{ potilaiden lkm}}.$$

Terveydenhuollossa voi kuitenkin olla työntekijöitä, jotka työssään käyttävät useiden eri diagnoosien saaneiden potilaiden tietoja. Siksi on tärkeää määrittää myös hoitotodennäköisyys, jolla määritetään, kuinka todennäköisesti potilaalla on tietty diagnoosi, jos tietyn osaston työntekijä käyttää hänen potilastietojaan (Fabbri ja muut 2013: 55):

$$P(\text{osasto } c \mid \text{diagnoosi } d) = \frac{\text{Diagnoosin } c \text{ potilaiden lkm, joilla käyttäjä osastolla } d}{\text{Osaston } d \text{ käyttämien potilaiden lkm}}.$$

Jos molempien käyttö- ja hoitotodennäköisyyksien arvot ylittävät niille asetetut kynnysarvot, tarkasteltava osasto vastaa tietyn diagnoosin saaneen potilaan hoidosta, ja käyttö on siten normaalia ja sallittua. Lopuista käyttölokitietueista jää epäilys ja ne täytyy luokitella normaaliksi muilla keinoin, kuten selitysmallinteilla. Esimerkki potilaiden diagnoosin ja osastojen välisistä todennäköisyyksistä on taulukossa 6.

Taulukko 6. Käyttö- ja hoitotodennäköisyyksiä eri osastoille, kun diagnoosina on munuaissiirre (muokattu lähteestä Fabbri ja muut 2013: 58).

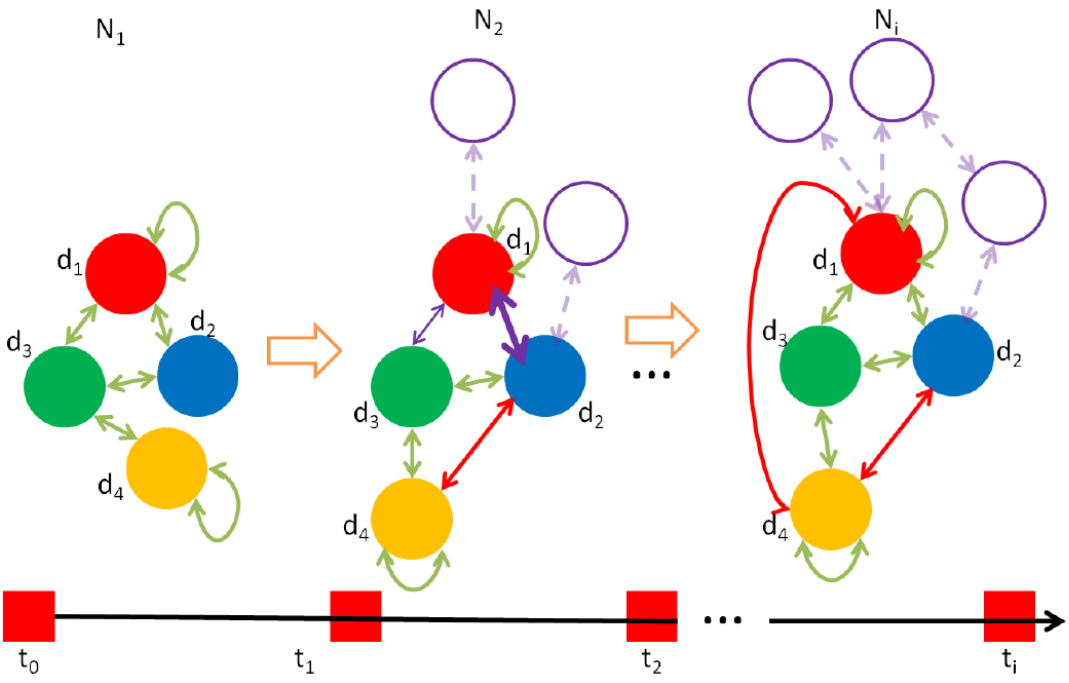
Osasto	**Käyttötodennäköisyys**	**Hoitotodennäköisyys**
Elinsiirto	0,56	0,20
Nefrologia	0,43	0,23
Farmasia	0,21	0,02
Dialyysi	0,10	0,11
Munuaishoitajat	0,09	0,08

Sisäisen uhan havaitsemisessa on hyödynnetty myös terveydenhuolto-organisaation osastojen välisiä yhteyksiä. Yhteyksistä ilmenee, onko potilaan hoito on ollut organisaation normaalin kulun mukaista. Väärinkäyttöön viittaa poikkeavien käyttäjien sijaan siis epäilyttävä potilas. Järjestelmä ilmoittaa poikkeavista potilaista tietoturva-asiantuntijalle. (Chen, Nyemba & Malin 2012b: 94.) Järjestelmä tutkii kahta osastoa kerrallaan ja laskee potilaiden määrän, joiden potilastietoja kumpikin osasto on käyttänyt. Osastojen välisiä yhteyksiä kuvataan kahdella läheisyysmitalla: varmuudella (eng. certainty) ja vastavuoroisuudella (eng. reciprocity). Varmuus kuvaa osastojen välisten yhteyksien vahvuutta ajan kuluessa. Varmuus mittaa, miten muutokset yhteisössä vaikuttavat osastojen välisiin yhteyksiin. Vastavuoroisuudella mitataan, miten samankaltaisia osastot ovat. Kaksi osastoa ovat vastavuoroisia, jos ne esiintyvät potilaiden tiedoissa yhdessä useammin kuin muut osastot. (Chen ja muut 2012b: 95–96.) Liitteessä 7 on esitetty varmuuden ja vastavuoroisuuden matemaattinen määritys.

Mittareiden muodostamista varten potilaista ja käyttäjistä muodostetaan matriisi, jonka alkiot saavat arvon 1, jos kyseisen alkion käyttäjä on käyttänyt alkion potilaan tietoja. Jos käyttöä ei ole käyttölokeissa, alkio saa arvoksi 0. Mittareita varten muodostetaan myös toinen matriisi käyttäjistä ja osastoista. Jos alkion käyttäjä on alkion osaston kirjoilla, alkion arvoksi merkitään 1. Muussa tapauksessa alkiota merkitään arvolla 0. Jos käyttäjä on kirjoilla useammalla osastolla arvo 1 jaetaan osastojen määrällä. Kolmella osastolla oleva käyttäjä saa arvoksi 0,33. (Chen ja muut 2012b: 95.)

Esimerkissä tarkastellaan yhden potilaan ympärille muodostuvaa yhteisöä eli lokaalia verkostoa. Ensimmäisen viikon aikana (t_0-t_1) potilaan tietojen käyttäjät ovat yhteydessä neljään osastoon (d_1, d_2, d_3, d_4). Osastojen välille piirretyt nuolet kertovat, että niillä on yhteisiä potilaita. Toisella viikolla (t_1-t_2) osaston d_1 ja d_2 välinen varmuus kasvaa ja osastojen d_1 ja d_3 välillä varmuus laskee. Lisäksi potilaan verkostoon ilmestyy käyttäjiä uusilta osastoilta. Osastojen muodostamia verkostoja ja niiden muutoksia on kuvattu kuvassa 16.

Potilaaseen liittyvien osastojen välille lasketaan myös niin sanotun globaalin verkoston arvot. Osastojen kaikkia käyttäjiä tarkastellaan ja saadaan laskettua globaalit varmuuden ja vastavuoroisuuden arvot. Jos potilaan lokaalin verkoston arvot eroavat vain vähän globaalin verkoston mittareiden arvoista, potilaan tietojen käyttö on todennäköisesti normaalia. Jos muutokset potilaan yhteisön mittareiden arvoissa ovat suuria, potilaan tiedot saattavat olla sisäisen uhan käytössä, jolloin järjestelmä tekee hälytyksen. (Chen ja muut 2012b: 101.)



Kuva 16. Osastojen muodostama verkosto (Chen ja muut 2012b: 97).

5 TULOSTEN JA HAVAINTOJEN ANALYSOINTI

Kappaleessa 5 kuvataan lyhyesti tuloksena saatujen tutkimusten puutteita ja muodostetaan kappaleen 4 menetelmistä tiivistetty synteesi aihealueen kokonaiskuvan ymmärtämisen helpottamiseksi.

5.1 Tutkimusten ja havaitsemismenetelmien puutteita

Tuloksena saaduissa tutkimuksissa on puutteita, jotka vaikeuttavat lopullisten suorien johtopäätösten tekoa. Tutkittujen menetelmien ja testattujen järjestelmien toimivuudesta kaikissa terveydenhuolto-ympäristöissä ei voida tehdä suoria yleistyksiä, koska testattuja aineistoja ei ole julkaistu eikä menetelmiä ole testattu suomalaisissa terveydenhuoltoympäristöissä.

Tilastollisen khiin neliön -testiä hyödyntävää järjestelmää testattiin oikealla viikon aikana kerätyllä yhdysvaltalaisella potilastietojen käyttölokidatalla ja saadut tulokset viittaavat yksinkertaisten sääntöjen merkittävyyden välillä olevan eroja. Testattavan aineiston koko oli 710000 käyttölokitietuetta (Hedda ja muut 2017: 870). Tutkimuksessa ei kuitenkaan tehty tietoturva-asiantuntijan avulla tarkempia selvityksiä, ovatko hälytyksen aiheuttaneet lokitietueet oikeasti väärinkäyttöä.

Logistisen regression ja tukivektorikoneen avulla voidaan automatisoida hälytysten priorisointia, mutta tutkimuksen perusteella sopivan opetusjoukon muodostaminen on työlästä ja manuaalista. Sopivan opetusjoukon muodostaminen vaatii myös ymmärrystä koneoppimismenetelmistä, mitä ei välttämättä ole ennalta tietoturva-asiantuntijoilla. Manuaalisesta työstä ja osaamisen puutteesta huolimatta luokittelumenetelmä osoittautui hyödylliseksi, koska perinteinen yksinkertaisia sääntöjä hyödyntävä järjestelmä teki kuukauden ajalta kerätyistä käyttölokeista kymmeniä tuhansia hälytyksiä. Hälytysten määrä on niin suuri, että priorisointi on välttämätöntä. Tutkimuksessa käytettyä 10,5 miljoonan käyttölokitietueen aineistoa tai 1291 tietueen opetusjoukkoa ei ole julkaistu.

Yhteistoiminnalliseen suodatukseen pohjautuva suosittelumenetelmä osoittaa, että täysin eri toimialalle ja tarkoitukseen kehitetty menetelmä soveltuu pienin muokkauksin sisäisen uhan havaitsemiseen. Tuloksissa menetelmä on sijoitettu ohjaamattoman oppimisen ja tunnisteperusteisen havaitsemisen alle, koska siinä hyödynnetään klusterointia, mutta samalla kuitenkin yksinkertaisten sääntöjen avulla muodostettuja tunnisteita. Mallin sovituksessa käytetään opetusjoukkoa, mutta yhteistoiminnallista suodatusta ei varsinaisesti sisällytetä puhtaasti ohjattuun oppimiseen, vaikka kyseessä onkin koneoppimistekniikka. Artikkelissa käytössä oli puolen vuoden ajalta kerätty 34,1 miljoonan käyttölokitietueen aineisto, josta oli tietoturva-asiantuntijoiden avulla muodostettu 1504 tietueen suuruinen opetusjoukko. Aineistoja ei ole julkaistu.

Assosiaatiosääntöjä hyödyntäviä menetelmiä on sovellettu oikeaan käyttölokiaineistoon, mutta tutkimusten mukaan tietoturva-asiantuntijat eivät ole arvioineet niillä saatuja tuloksia. Sääntöjen toimivuutta laajasti ja eri terveydenhuoltoyksiköissä ei voida täysin varmistaa, koska kausittaiset muutokset terveydenhuollossa ja eri osastojen käyttäjien toiminnoissa olevat eroavaisuudet saattavat vaikuttaa menetelmien toimintakykyyn. Assosiaatiosäännöillä on kuitenkin mahdollista osoittaa poikkeavuuksia, joita ei muilla tavoin löydy, ja mahdollisuus esittää tuloksia graafisesti helpottaa havaitsemistyötä. Erityisesti selitysmallinemenetelmällä on saatu aikaan lupaavia tuloksia potilastietojen normaalin käytön syiden selvittämisessä ja sitä on myös testattu muiden menetelmien rinnalla. Käytössä olevaa 4,5 miljoonan käyttölokitietueen aineistoa ei ole kuitenkaan julkaistu, eikä tutkimustulosten oikeellisuutta ole varmennettu terveydenhuollon tietoturva-asiantuntijoiden avulla.

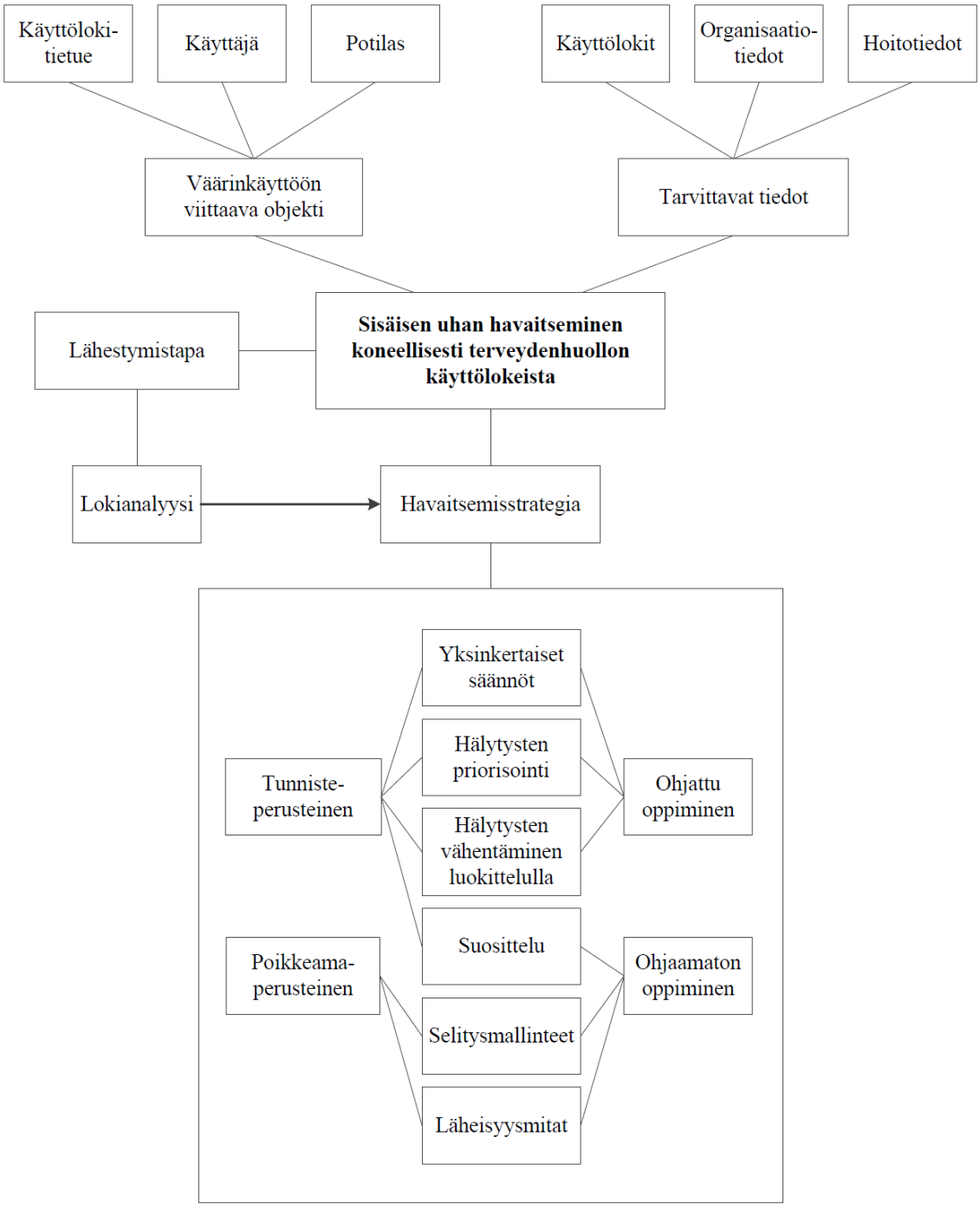
Poikkeamien havaitsemiseen keskittyviä menetelmiä on erilaisia ja niille on yhteistä muodostaa eri objektien välille läheisyysmittoja, joista selvästi poikkeavat objektit erottuvat mahdollisena väärinkäyttönä. Menetelmillä löytyy poikkeavuuksia, mutta usein poikkeama erottuu vain, jos se on riittävän selkeä, kuten jos käyttäjä on sattumanvaraisesti valinnut tarkasteltavaksi potilaan tiedot osastolta, jonka tietoja hän ei yleensä käytä. Menetelmien tarkkuus vaikuttaisi myös heikkenevän, kun esimerkiksi tarkasteltavia osastoja on paljon ja niiden välisissä toiminnoissa on suuria eroja. Menetelmien hyötynä on kuitenkin niiden ohjaamattoman oppimisen luonne, jolloin

terveydenhuoltoympäristön täydellinen tuntemus ei ole tarvittavaa. Menetelmien kohdalla käyttölokiaineistoja ei ole julkaistu, eikä terveydenhuollon tietoturva-asiantuntijoita ole ollut mukana tutkimustulosten todentamisessa.

5.2 Synteesin muodostus

Kirjallisuuskatsauksen tulosten perusteella on olemassa erilaisia koneellisia menetelmiä sisäisten uhkien havaitsemiseen terveydenhuollon käyttölokien avulla. Menetelmiä on myös mahdollista yhdistää tutkielman teoriakappaleessa esiteltyihin käsitteisiin. Kirjallisuuskatsaustuloksista voidaan koota havainnollistava yhteenveto sisäisen uhan havaitsemisesta terveydenhuollossa, kun käytössä on koneellisia menetelmiä, jotka analysoivat käyttölokeja.

Tuloksena saatujen havaitsemismenetelmien pohjalta voi todeta, että uhan havaitsemisessa väärinkäyttöön viittaa joko epäilyttävä käyttölokitietue, käyttäjä tai potilas. Tulosten kaikki menetelmät käyttävät havaitsemisessa käyttölokeja, mutta tehokas havaitsemiskyky vaatii myös muiden hyödyllisten tietojen käyttöä. Tällaisia tietoja ovat esimerkiksi organisaatiotiedot, kuten käyttäjätiedot ja hoito-osastot, sekä hoitotiedot, kuten ajanvaraustiedot, käyntitiedot ja diagnoosikoodit. Käyttölokeja hyödyntävät menetelmät lähestyvät sisäisen uhan havaitsemista lokianalyysin keinoin. Analyysiin on useita havaitsemisstrategioita. Strategia voi hyödyntää yksinkertaisia sääntöjä, hälytysten priorisointia, hälytysten vähentämistä luokittelemalla, suosittelua (käyttäjä-potilas-klusteriparit), normaalikäytön assosiaatiosääntöjä (selitysmallinteet) ja käyttäjiin ja osastoihin liittyviä läheisyysmittoja poikkeamien havaitsemiseen. Aihealueen synteesi on kuvan 17 mukainen.



Kuva 17. Sisäisen uhan havaitseminen koneellisesti terveydenhuollon käyttölokeista.

6 JOHTOPÄÄTÖKSET

Sisäisen uhan havaitsemista on tutkittu selvästi vähemmän kuin tunkeutumisen havaitsemista, vaikka sisäisiä uhkia pidetään useiden lähteiden mukaan vaarallisempana kuin ulkoisia uhkia. Kuitenkin sisäisen uhan havaitsemista käsittelevien artikkelien määrä on viime vuosina kasvanut. Valtuutettujen käyttäjien suorittama potilastietojen väärinkäyttö ja sen havaitseminen sopii hyvin sisäisen uhan havaitsemiseksi nimetyn tutkimusalan alle. Tutkimusala on aktiivinen, sillä aihepiiristä löytyy tuoreita julkaisuja. Siksi voidaan pitää todennäköisenä, että uusia menetelmiä sisäisten uhkien havaitsemiseen kehitetään lähivuosina lisää myös terveydenhuoltoympäristöihin sopiviksi. Tehtävää vaikeuttaa julkisten esimerkkikäyttölokien puuttuminen niiden arkaluonteisen sisällön vuoksi. Vaikka tässä tutkielmassa esitettiin toimiviksi testattuja lähestymistapoja sisäisten uhkien havaitsemiseen, puuttuvat aineistot vaikeuttavat johtopäätösten tekoa niiden toimivuudesta suomalaisessa terveydenhuoltoympäristössä. Uuden katsauksen tekeminen 5–10 vuoden kuluttua todennäköisesti löytäisi lisää käyttökelpoisia ratkaisuja.

Potilastietojen käyttölokeja hyödyntäviä sisäisen uhan havaitsemista tutkivia artikkeleja on olemassa useita, ja niitä on mahdollista löytää kirjallisuuskatsauksella. Yhtenäisten hakusanojen muodostaminen on haastavaa, koska väärinkäytön havaitsemisen ja suuriin tietomääriin käytettävien algoritmien (koneoppiminen, tiedonlouhinta jne.) tutkimusalojen terminologia ei ole ainakaan terveydenhuoltoympäristössä vakiintunut yksiselitteiseksi. Kirjallisuuskatsauksen hakusanoilla löydettyjen artikkelien lisäksi aineistoon lisättiin tutkielman kannalta tärkeitä tutkimusartikkeleita lähdeluetteloiden avulla. Katsauksen täydennys vaikuttaisi olevan tärkeä vaihe tietotekniikan katsauksissa erityisesti kun tutkitaan uusia aihealueita. Katsauksesta jätettiin ulkopuolelle tieteelliset julkaisut, joiden koko tekstiin ei ollut pääsyä. Katsauksen laajentaminen uusien tietokantojen ja käytettyjen tietokantojen koko aineistoon saattaisi kasvattaa aineistoa ja tuottaa uusia ideoita terveydenhuollon sisäisen uhan havaitsemiseen.

Terveydenhuoltoympäristö, potilastietojärjestelmät ja niistä kerätyt lokitiedot ovat niin poikkeuksellinen kokonaisuus, ettei luotettavan automaattisen sisäisen uhan

havaitsemisjärjestelmän toteutus sinne ole helppoa. Yksinkertaisten auditointisääntöjen rinnalla ei voida käyttää perinteisiä käyttäjäkohtaisia rooli- ja sisältöperusteisia pääsyrajoituksia, kuten monissa muissa ympäristöissä. Näin ollen sisäisten uhkien havaitseminen perustuu pääosin auditointiin ja sen sisällä tyypillisesti käyttölokeista tehtyyn lokianalyysiin. Toimintaympäristön erityisyys on syytä huomioida, kun kehitetään tulevaisuuden järjestelmiä. On todennäköistä, että käyttölokien hyödyntäminen tulee olemaan tärkeä osa tehokasta havaitsemisjärjestelmää terveydenhuollossa myös tulevaisuudessa.

Katsaustulosten perusteella perinteistä yksinkertaisia sääntöjä hyödyntävää järjestelmää voidaan vahvistaa koneellisesti synteesin mukaisesti monella tapaa. Toimiessaan varsinkin yksinkertaiset menetelmät, kuten selitysmallinteiden käyttö, voivat tarjota potilaille helpomman väylän tarkastella omien potilastietojensa käytön asiallisuutta, ja samalla järjestelmä ulkoistaa manuaalisen työn määrää tietoturva-asiantuntijoilta tuhansille aktiivisille potilaille. Lisäksi hälytysten priorisointi auttaa asiantuntijoita kohdentamaan resurssit todennäköisimpiin väärinkäyttötapauksiin. Terveydenhuolto-ympäristön väärinkäytön ja sisäisen uhan havaitsemisen ei siis tarvitse olla välttämättä paljon koneellisempaa ja täysin automatisoitua jatkossakaan ollakseen tehokasta. Manuaalista työtä ainoastaan jaetaan ja kohdennetaan uudella tapaa.

Tutkielman keskeisin tulos on integroivalla kirjallisuuskatsauksella saatu ajantasainen selvitys tutkimuksen nykytilasta. Synteesistä ilmenee, että haastavaan käyttöympäristöön soveltuu monipuolinen koneellisten menetelmien kokonaisuus. Katsaus osoitti myös todennäköiseksi, että menetelmien avulla on mahdollista kehittää havaitsemisjärjestelmiä tehokkaammiksi. Tutkielman tavoitteena oli antaa suuntaviivoja perinteisten havaitsemisjärjestelmien kehittämiseen ja tehostamiseen, mutta jää nähtäväksi, saadaanko menetelmistä käytännöllistä hyötyä todellisessa sisäisten uhkien havaitsemistyössä. Jatkotutkimuksissa olisi syytä selvittää tarkemmin synteesin lähestymistapojen soveltuvuutta oikeisiin suomalaisiin potilastietojen käyttöloki-aineistoihin. Käytännön järjestelmien kehittäjät voivat tehostaa havaitsemista rakentamalla olemassa olevan menetelmän rinnalle erillisiä komponentteja, jotka etsivät

sisäisiä uhkia toisiaan täydentävin strategioin. Uusien komponenttien avulla on mahdollista löytää aiemmin havaitsematta jääneitä potentiaalisia väärinkäytöksiä.

LÄHDELUETTELO

ACM Digital Library [online]. Saatavissa: https://dl.acm.org/

Alpaydin, E. (2014). *Introduction to machine learning*. 3. painos. The MIT Press. 640 s. ISBN 978-0262-02818-9.

Asaro, P. V. & J. E. Ries (2001). Data mining in medical record access logs. *Proceedings of the American medical informatics association (AMIA) annual symposium 2001* [online], 855–855.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. 1. painos. New York: Springer-Verlag. 738 s. ISBN 978-0387-31073-2.

Blocki, J., N. Christin, A. Datta, A. D. Procaccia & A. Sinha (2013). Audit games. *Proceedings of the 23rd international joint conference on artificial intelligence* [online], 41–47. AAAI Press. https://dl.acm.org/doi/10.5555/2540128.2540137

Blocki, J., N. Christin, A. Datta, A. D. Procaccia & A. Sinha (2015). Audit games with multiple defender resources. *Proceedings of the 29th AAAI conference on artificial intelligence* [online], 791–797. AAAI Press. https://dl.acm.org/doi/10.5555/2887007.2887117

Boxwala, A. A., J. Kim, J. M. Grillo & L. Ohno-Machado (2011). Using statistical and machine learning to help institutions detect suspicious access to electronic health records. *Journal of the American medical informatics association (AMIA)* [online] 18 (4), 498–505. https://doi.org/10.1136/amiajnl-2011-000217

Chen, Y. & B. Malin (2011a). Detection of anomalous insiders in collaborative environments via relational analysis of access logs. *Proceedings of the 1st ACM conference on data and application security and privacy* [online], 63–74. https://doi.org/10.1145/1943513.1943524

Chen, Y., S. Nyemba & B. Malin (2012a). Detecting anomalous insiders in collaborative information systems. *IEEE transactions on dependable and secure computing* [online] 9, 332–344. https://doi.org/10.1109/tdsc.2012.11

Chen, Y., S. Nyemba & B. Malin (2012b). Auditing medical records accesses via healthcare interaction networks. *Proceedings of the American medical informatics association (AMIA) annual symposium 2012,* 93–102.

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2011b). Leveraging social networks to detect anomalous insider actions in collaborative environments. *Proceedings of 2011 IEEE international conference on intelligence and security informatics* [online], 119–124. https://doi.org/10.1109/isi.2011.5984061

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2012c). Specializing network analysis to detect anomalous insider actions. *Security informatics* [online] 1 (5): 5, 1–24. https://doi.org/10.1186/2190-8532-1-5

Chung, C. Y., M. Gertz & K. Levitt (1998). DEMIDS: A Misuse Detection System for Database Systems. *Working conference on integrity and internal control in information systems* [online], 159-178. https://doi.org/10.1007/978-0-387-35501-6_12

Chuvakin, A. A., K. J. Schmidt & C. Phillips (2013). *Logging and Log Management. The Authoritative Guide To Understanding the Concepts Surrounding Logging and Log Management*. Syngress. 431 s. ISBN 978-1-59749-635-3.

Cooper, H. M. (1989). *Integrating Research. A Guide for Literature Reviews*. 2. painos. Sage Publications. 157 s. ISBN 0-8039-3430-0.

Fabbri, D. & K. LeFevre (2011a). Explanation-based auditing. *Proceedings of the very large data bases (VLDB) endowment* [online] 5, 1–12. https://doi.org/10.14778/2047485.2047486

Fabbri, D. & K. LeFevre (2013). Explaining accesses to electronic medical records using diagnosis information. *Journal of the American medical informatics association (AMIA)* [online] 20 (1), 52–60. https://doi.org/10.1136/amiajnl-2012-001018

Fabbri, D., K. LeFevre & D. A. Hanauer (2011b). Explaining accesses to electronic health records. *Proceedings of the ACM workshop on data mining for medicine and healthcare* [online], 10–17. https://doi.org/10.1145/2023582.2023585

Fayyad, U., G. Piatetsky-Sharipo & P. Smyth (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine* [online] 17: 3, 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Gullo, F. (2015). From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. *Physics Procedia* [online] 62, 18–22. https://doi.org/10.1016/j.phpro.2015.02.005

Hand, D., H. Mannila & P. Smyth (2001). *Principles of data mining*. The MIT Press. 546 s. ISBN 0-262-08290-X.

Hedda, M., B. A. Malin, C. Yan & D. Fabbri (2017). Evaluating the Effectiveness of Auditing Rules for Electronic Health Record Systems. *Proceedings of the American medical informatics association (AMIA) annual symposium 2017* [online], 866–875.

IEEE Xplore [online]. Saatavissa: https://ieeexplore.ieee.org/Xplore/home.jsp

Juniper Networks (2016). *Learn About Intrusion Detection and Prevention* [online]. Saatavissa: https://www.juniper.net/documentation/en_US/learn-about/LA_IntrusionDetectionandPrevention.pdf.

Kent, K. & M. Souppaya (2006). *Guide to computer security log management. Recommendations of the national institute of standards and technology (NIST)*. 72 s. https://doi.org/10.6028/nist.sp.800-92

Kim, J., J. M. Grillo, A. A. Boxwala, X. Jiang, R. B. Mandelbaum, B. A. Patel, D. Mikels, S. A. Vinterbo & L. Ohno-Machado (2011). Anomaly and signature filtering improve classifier performance for detection of suspicious access to EHRs. *Proceedings of the American medical informatics association (AMIA) annual symposium 2011* [online], 723–731.

Ko, L. L., D. M. Divakaran, Y. S. Liau & V. L. L. Thing (2016). Insider threat detection and its future directions. *International journal of security and networks*. https://doi.org/10.1504/ijsn.2017.084391

Laki sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä 2007. 159/9.2.2007.

Larson, R. E. & L. Cockcroft (2003). *CCSP: Cisco certified security professional certification exam guide*. Osborne: McGraw-Hill. 998 s. ISBN 978-0072-22691-1.

Laszka, A., Y. Vorobeychik, D. Fabbri, C. Yan & B. Malin (2017). A game-theoretic approach for alert prioritization. *Proceedings of the AAAI workshop on artificial intelligence for cyber security* [online].

Liao, H.-J., C.-H. R. Lin, Y.-C. Lin & K.-Y. Tung (2012). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications* [online] 36, 16-24. https://doi.org/10.1016/j.jnca.2012.09.004

Malin, B., S. Nyemba & J. Paulett (2011). Learning relational policies from electronic health record access logs. *Journal of biomedical informatics* 44 (2), 333–342. https://doi.org/10.1016/j.jbi.2011.01.007

McGlade D. & S. Scott-Hayward (2018). ML-based cyber incident detection for electronic medical record (EMR) systems. *Smart Health* [online] 12, 3–23. https://doi.org/10.1016/j.smhl.2018.05.001

MEDLINE (PubMed) [online]. Saatavissa: https://pubmed.ncbi.nlm.nih.gov/

Menon, A. K., X. Jiang, J. Kim, J. Vaidya & L. Ohno-Machado (2014). Detecting inappropriate access to electronic health records using collaborative filtering. *Machine Learning* 95, 87–101. https://doi.org/10.1007/s10994-013-5376-1

MITRE Corporation, The (2010). *Common event expression*. Saatavissa: https://cee.mitre.org/docs/CEE_Architecture_Overview-v0.5.pdf.

Mueller, J. P. & L. Massaron (2018). *Artificial intelligence for dummies*. John Wiley & Sons. 336 s. ISBN 978-1119-46765-6.

Ontario, Information and privacy commissioner of (2015). *Detecting and deterring unauthorized access to personal health information* [online]. Saatavissa: https://www.ipc.on.ca/wp-content/uploads/resources/detect_deter.pdf

Røstad, L. & O. Edsberg (2006). A study of access control requirements for healthcare systems based on audit trails from access logs. *Proceedings of the 22nd computer security applications conference*, 175–186. https://doi.org/10.1109/acsac.2006.8

Salminen, A. (2011). *Mikä kirjallisuuskatsaus? Johdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteellisiin sovelluksiin*. Vaasan yliopiston julkaisuja. ISBN 978-952-476-349-3.

ScienceDirect (Elsevier) [online]. Saatavissa: https://www.sciencedirect.com/

Sosiaali- ja terveysministeriön asetus potilasasiakirjojen laatimisesta sekä niiden ja muun hoitoon liittyvän materiaalin säilyttämisestä 2001. 99/19.1.2001.

SpringerLink [online]. Saatavissa: https://link.springer.com/

Stolt, M., A. Axelin & R. Suhonen (2015). *Kirjallisuuskatsaus hoitotieteessä*. Juvenes Print. 131 s. ISBN 978-951-29-6276-1.

Theodoridis, S. & K. Koutroumbas (2008). *Pattern recognition*. 4. painos. Academic Press. 984 s. ISBN 978-1597-49272-0.

Tsai, J. J. P. & Z. Yu (2010). *Intrusion detection. A machine learning approach*. World Scientific. 171 s. ISBN 978-1848-16447-5.

Zhang, H., S. Mehotra, D. Liebovitz, C. A. Gunter & B. Malin (2013). Mining deviations from patient care pathways via electronic medical record system audits. *ACM transactions on management information systems* [online] 4 (4), 17. https://doi.org/10.1145/2544102

LIITTEET

LIITE 1. Aineistohaun tulokset hakusanojen mukaan

"insider threat detection"

Chen, Y. & B. Malin (2011). Detection of anomalous insiders in collaborative environments via relational analysis of access logs. (ACM, PubMed)

Chen, Y., S. Nyemba & B. Malin (2012). Detecting anomalous insiders in collaborative information systems. (IEEE, PubMed)

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2011). Leveraging social networks to detect anomalous insider actions in collaborative environments. (IEEE, PubMed)

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2012). Specializing network analysis to detect anomalous insider actions. (PubMed, SpringerLink)

McGlade D. & S. Scott-Hayward (2018). ML-based cyber incident detection for electronic medical record (EMR) systems. (ScienceDirect)

"auditing" AND "electronic health records"

Fabbri, D. & K. LeFevre (2011). Explanation-based auditing. (ACM)

Fabbri, D. & K. LeFevre (2013). Explaining accesses to electronic medical records using diagnosis information. (PubMed)

Fabbri, D., K. LeFevre & D. A. Hanauer (2011). Explaining accesses to electronic health records. (ACM)

Menon, A. K., X. Jiang, J. Kim, J. Vaidya & L. Ohno-Machado (2014). Detecting inappropriate access to electronic health records using collaborative filtering. (PubMed)

("machine learning" OR "data mining") AND "access logs" AND "electronic health records"

Boxwala, A. A., J. Kim, J. M. Grillo & L. Ohno-Machado (2011). Using statistical and machine learning to help institutions detect suspicious access to electronic health records. (PubMed, SpringerLink)

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2012). Specializing network analysis to detect anomalous insider actions. (SpringerLink)

Fabbri, D., K. LeFevre & D. A. Hanauer (2011). Explaining accesses to electronic health records. (ACM)

Malin, B., S. Nyemba & J. Paulett (2011). Learning relational policies from electronic health record access logs. (PubMed, ScienceDirect)

LIITE 2. Aineiston laatuarviointi

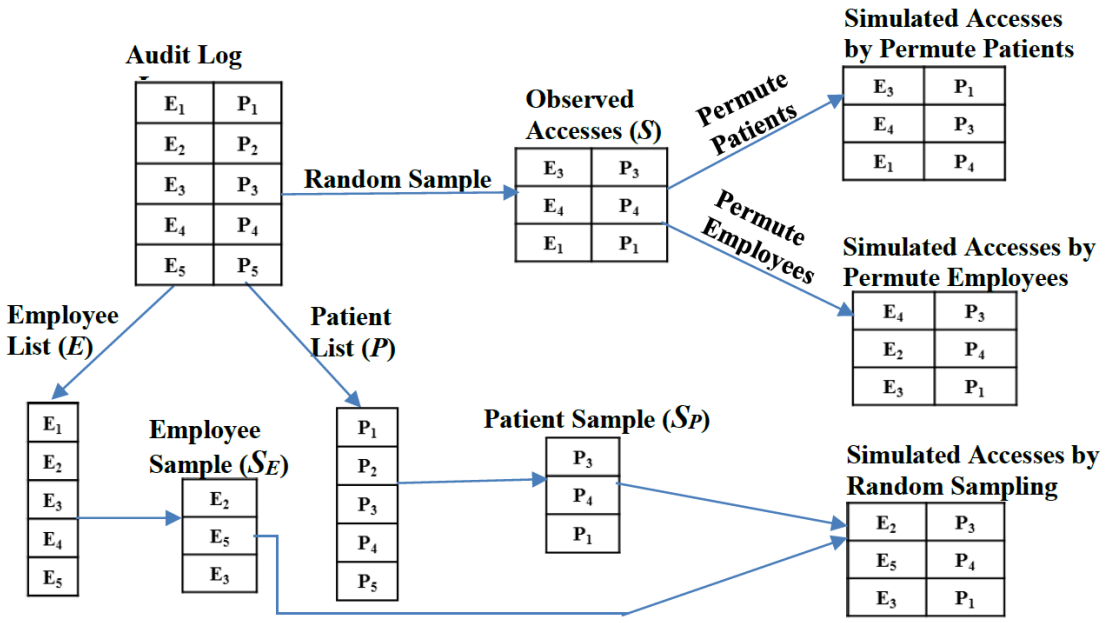
Julkaisu	**Merkitys**	**Tuoreus**	**Todellinen käyttöloki**
Blocki, J., N. Christin, A. Datta, A. D. Procaccia & A. Sinha (2013). Audit games.	Matala	2	0
Blocki, J., N. Christin, A. Datta, A. D. Procaccia & A. Sinha (2015). Audit games with multiple defender resources.	Matala	1	0
Boxwala, A. A., J. Kim, J. M. Grillo & L. Ohno-Machado (2011). Using statistical and machine learning to help institutions detect suspicious access to electronic health records.	Korkea	3	1
Chen, Y. & B. Malin (2011). Detection of anomalous insiders in collaborative environments via relational analysis of access logs.	Korkea	3	1
Chen, Y., S. Nyemba & B. Malin (2012). Auditing medical records accesses via healthcare interaction networks.	Korkea	2	1
Chen, Y., S. Nyemba & B. Malin (2012). Detecting anomalous insiders in collaborative information systems.	Korkea	2	1
Chen, Y., S. Nyemba, W. Zhang & B. Malin (2011). Leveraging social networks to detect anomalous insider actions in collaborative environments.	Korkea	3	1
Chen, Y., S. Nyemba, W. Zhang & B. Malin (2012). Specializing network analysis to detect anomalous insider actions.	Korkea	2	1
Fabbri, D. & K. LeFevre (2011). Explanation-based auditing.	Korkea	3	1
Fabbri, D. & K. LeFevre (2013). Explaining accesses to electronic medical records using diagnosis information.	Korkea	2	1
Fabbri, D., K. LeFevre & D. A. Hanauer (2011). Explaining accesses to electronic health records.	Korkea	3	1
Hedda, M., B. A. Malin, C. Yan & D. Fabbri (2017). Evaluating the Effectiveness of Auditing Rules for Electronic Health Record Systems.	Korkea	1	1
Kim, J., J. M. Grillo, A. A. Boxwala, X. Jiang, R. B. Mandelbaum, B. A. Patel, D. Mikels, S. A. Vinterbo & L. Ohno-Machado (2011). Anomaly and signature filtering improve classifier performance for detection of suspicious access to EHRs.	Korkea	3	1

Ko, L. L., D. M. Divakaran, Y. S. Liau & V. L. L. Thing (2016). Insider threat detection and its future directions.	Matala	1	0
Laszka, A., Y. Vorobeychik, D. Fabbri, C. Yan & B. Malin (2017). A game-theoretic approach for alert prioritization.	Matala	1	1
Malin, B., S. Nyemba & J. Paulett (2011). Learning relational policies from electronic health record access logs.	Korkea	3	1
McGlade D. & S. Scott-Hayward (2018). ML-based cyber incident detection for electronic medical record (EMR) systems.	Korkea	1	1
Menon, A. K., X. Jiang, J. Kim, J. Vaidya & L. Ohno-Machado (2014). Detecting inappropriate access to electronic health records using collaborative filtering.	Korkea	2	1
Zhang, H., S. Mehotra, D. Liebovitz, C. A. Gunter & B. Malin (2013). Mining deviations from patient care pathways via electronic medical record system audits.	Korkea	2	1

LIITE 3. Havaittujen ja odotettujen hälytystaajuuksien muodostaminen

Hedda, M., B. A. Malin, C. Yan & D. Fabbri (2017). Evaluating the Effectiveness of Auditing Rules for Electronic Health Record Systems. *Proceedings of the American medical informatics association (AMIA) annual symposium 2017* [online], 866–875.

Kerätystä käyttölokista voidaan muodostaa lista jokaisesta käyttäjä-potilas-parista, joista valitaan satunnaisesti näyte *S*. Esimerkiksi 710000 parista voidaan ottaa 100000 parin näyte. Näyteaineistosta etsitään uhkia yksinkertaisen sääntöjen avulla. Saatujen hälytysten ja näytteen koon suhteesta saadaan näytteen havaittu hälytystiheys. Seuraavaksi näytteen potilaat ja käyttäjät satunnaistetaan permutaation avulla kumpikin ryhmä erikseen. Tällä tavalla saadaan permutoitujen aineistojen hälytyksistä laskettua oletetut hälytystiheydet. Jotta vältytään aineiston vinoutumiselta tiettyyn suuntaan, käytetään permutaatiomenetelmän rinnalla toistakin satunnaistamismenetelmää. Yksinkertaisella satunnaisotannalla käyttäjä-potilas-parien listasta valitaan 100000 käyttäjää (S_E) ja 100000 potilasta (S_P) kumpikin erikseen, ja yhdistetään aineistoksi. Myös satunnaisotannalla saadulle aineistolle lasketaan oma hälytystaajuus.



Hälytystaajuuksien laskennat suoritetaan kymmeneen kertaan, ja laskujen pohjalta muodostetaan keskiarvot havaittujen ja oletettujen hälytystaajuuksista. Taajuudet lasketaan myös erikseen jokaiselle yksinkertaiselle säännölle.

LIITE 4. Hälytysten priorisointi peliteorian avulla.

Laszka, A., Y. Vorobeychik, D. Fabbri, C. Yan & B. Malin (2017). A game-theoretic approach for alert prioritization. *Proceedings of the AAAI workshop on artificial intelligence for cyber security* [online].

Priorisointipelissä lasketaan budjetin *B* rajoissa havaitsemistodennäköisyyksiä *PD*(***o***,*a*) hälytysten käsittelyyn saaduille priorisointijärjestyksille algoritmilla 1.

Algorithm 1 Computing $PD(\boldsymbol{o}, a)$

Input: prioritization game, prioritization $\boldsymbol{o}$, attack a

1: **for** $b = 0, 1, \ldots, B$ **do**

2: $\quad PD(\boldsymbol{o}, a, |T|, b) \leftarrow R_{a,o_{|T|}} \cdot F^*_{o_{|T|}}\left(\lfloor b/C_{o_{|T|}} \rfloor - 1\right)$

3: **end for**

4: **for** $i = |T| - 1, \ldots, 2, 1$ **do**

5: $\quad$ **for** $b = 0, 1, \ldots, B$ **do**

6: $$PD(\boldsymbol{o}, a, i, b) \leftarrow R_{a,o_i} \cdot F^*_{o_i}(\lfloor b/C_{o_i} \rfloor - 1) + (1 - R_{a,o_i}) \sum_{j=0}^{\lfloor b/C_{o_i} \rfloor} \Big[(F_{o_i}(j) - F_{o_i}(j-1)) \cdot PD(\boldsymbol{o}, a, b - j \cdot C_{o_i}, i+1) \Big]$$

7: $\quad$ **end for**

8: **end for**

9: Return $PD(\boldsymbol{o}, a) := PD(\boldsymbol{o}, a, 1, B)$

Algoritmin syötteenä on peli, kuten Stackelberg, priorisointijärjestys ***o*** ja hyökkäystyyppi *a*. Rivin 2 lauseke on todennäköisyys, millä puolustaja havaitsee hyökkäyksen, jonka todennäköisyys on $R_{a,o|T|}$, ja millä puolustajan budjetti on riittävä tutkimaan sekä väärät hälytykset että itse aidon hyökkäyksen. Rivin 6 lauseke koostuu kahdesta termistä. Ensimmäinen vastaa rivin 2 termiä ja jälkimmäisessä termissä on kerrottu väärien hälytysten määrän *j* todennäköisyys todennäköisyydellä, jolla puolustaja havaitsee hyökkäyksen jäljellä olevaa budjettia ja hälytysluokkia käyttämällä.

LIITE 5. Selitysmallinteen louhinta-algoritmi.

Fabbri, D. & K. LeFevre (2011). Explanation-based auditing. *Proceedings of the very large data bases (VLDB) endowment* [online] 5, 1–12. https://doi.org/10.14778/2047485.2047486

Syötteenä algoritmille annetaan alkuattribuutti (Log.Patient), loppuattribuutti (Log.User), tuki (*S*), max. polun pituus (*M*), rajoitettu määrä tauluja (*T*), janojen joukko (Edges) ja tietokanta (*D*). Ulostulona saadaan joukko hyväksyttyjä selitysmallinteita. Selitysmalline muodostuu algoritmilla 1.

Algorithm 1 One-Way Template Mining Algorithm

Input: Start attribute (Log.Patient), end attribute (Log.User), support (S), max path length (M), restricted number of tables referenced (T), the set of edges from the schema (Edges) and the database instance (D).

Output: Set of supported explanation templates (up to the max length).

```
Length = 1
Paths = {Edges that begin with the start attribute}
Explanations = {}
while Length ≤ M do
    New Paths = {}
    for Path p ∈ Paths do
        for Edge e ∈ Edges do
            if areConnected(p, e) then
                Candidate Path = p.append(e)
                if isARestrictedSimplePath(Candidate Path) then
                    if Support(Candidate Path, D) ≥ S then
                        New Paths.add(Candidate Path)
                        if isAnExplanation(Candidate Path) then
                            Explanations.add(Candidate Path)
    Paths = New Paths
    Length += 1
Return Explanations
```

Algoritmi palauttaa ainoastaan selitysmallinteet, jotka selittävät riittävän määrän lokin tietojen käytöstä. Riittävä määrä eli tuki *S* on annettu algoritmin syötteenä. Algoritmi tarkastaa ensin jokaisen polun (Pituus $\leq$ *M*) ja janan, ja selvittää, ovatko ne toisissaan kiinni. Jos yhteys on olemassa, algoritmi lisää janan polun loppuun (oikeaan reunaan). Seuraavaksi algoritmi tarkastaa, onko polkukandidaatti Candidate Path yksinkertainen polku. Polku on yksinkertainen, jos sen reitti alkaa lokitietueesta, kulkee enintään rajoitetun taulumäärän (*T*) läpi ja päätyy lokitietueeseen, josta aloitettiin.

LIITE 6. Läheisyysmittojen ja poikkeavuuden määrittäminen.

Chen, Y., S. Nyemba, W. Zhang & B. Malin (2012c). Specializing network analysis to detect anomalous insider actions. *Security informatics* [online] 1 (5): 5, 1–24. https://doi.org/10.1186/2190-8532-1-5

Poikkeavien käyttäjien havaitsemisessa keskitytään määrään potilaita, joiden tietoja kukin käyttäjä on käyttänyt. Määrien perusteella saadaan selville, kuinka tärkeä käyttäjä on potilaalle verrattuna muihin potilaisiin. Tärkeyttä kuvataan IDF-mallin avulla, joka muodostetaan kaavalla 1.

$$IDF(u_i) = \log \frac{|S|}{1 + \mathbf{B} \cdot \mathbf{U_i}} \tag{1}$$

IDF-mallin mukaisen matriisin arvojen pohjalta lasketaan kerrallaan yhtä potilasta koskevien käyttäjien välinen samankaltaisuus eli läheisyys. Käyttäjän u_i ja u_j välinen läheisyysmitta määräytyy kosinisen samankaltaisuuden avulla kaavan 2 mukaan.

$$Sim(u_i, u_j) = \frac{\mathbf{IDF\_U_i} \cdot \mathbf{IDF\_U_j}}{\left\| \mathbf{IDF\_U_i} \right\| \times \left\| \mathbf{IDF\_U_j} \right\|} \tag{2}$$

Käyttäjien läheisyysmitoista voidaan muodostaa käyttäjien verkosto, josta koko verkostolle laskea läheisyyssarvo kaikkien käyttäjäparien keskiarvon avulla kaavalla 3.

$$SIM(Net_{s_k}) = \frac{\forall u_i \neq u_j \in U_{s_k} \forall u_j \sum Sim(u_i, u_j)}{\frac{|U_{s_k}| \times (|U_{s_k}| - 1)}{2}} \tag{3}$$

Verkostoista poistetaan yksi käyttäjä kerrallaan ja lasketaan arvot samaan tapaan aliverkostoille. Jos kaavalla 4 saatava poikkeavuusarvo poistetun käyttäjän ja kaikkien käyttäjien keskiarvon välillä on suuri, käyttäjä näyttäytyy sisäisenä uhkana.

$$Score(u_j \rightarrow s_i) = SIM(Net_{s_{ij}}) - SIM(Net_{s_i}) \tag{4}$$

LIITE 7. Varmuuden ja vastavuoroisuuden määrittäminen

Chen, Y., S. Nyemba & B. Malin (2012b). Auditing medical records accesses via healthcare interaction networks. *Proceedings of the American medical informatics association (AMIA) annual symposium 2012,* 93–102.

Varmuus ja vastavuoroisuus saadaan määriteltyä käyttölokien avulla. Käyttölokit muunnetaan matriisimuotoon. Olkoon *P* potilaiden, *U* käyttäjien ja *D* hoito-osastojen joukko. Käyttölokitietueet muunnetaan kahdeksi matriisiksi *A* ja *B*. Matriisi *A* saa koon |*P*| x |*U*| ja matriisi *B* koon |*U*| x |*D*|. Potilaiden ja hoito-osastojen suhde saa muodon $G = AB_N$. Matriisi B_N on matriisin *B* normalisoitu muoto. Osastojen välinen varmuus

$$Cert(d_i \rightarrow d_j) = \begin{cases} \dfrac{Q(i,j)}{Q(i,i)}, & \text{jos } i \neq j \\ \dfrac{\sum_{k=1}^{|P|} \theta(G(k,i))}{\sum_{k=1}^{|P|} \psi(G(k,i))}, & \text{jos } i = j \end{cases} \quad (1)$$

arvioi todennäköisyyttä, jolla osasto d_j käyttää potilaan tietoja, kun saman potilaan tietoja on käyttänyt myös osasto d_i. Yhtälössä $Q(i,j)$ on osastojen d_i ja d_j käyttämien yhteisten potilaiden määrä ja $Q(i,i)$ on osaston d_i käyttämien potilaiden määrä. Tilanteissa joissa osastot ovat samat ($i=j$), osoittajaan on laskettu potilaat, joita on käyttänyt samalta osastolta vähintään kaksi käyttäjää. Nimittäjänä on potilasmäärä, joka saadaan laskemalla yhteen kaikki ne käytöt, joissa osastolta on ainakin yksi käyttäjä. Tämä suhdeluku on arvio tahdista, jolla potilaiden tietoja käyttää useampi käyttäjä samalta osastolta. Verkoston vastavuoroisuus

$$Recip = \frac{\sum_{\forall d_i, d_j \in D,\, i \neq j} \left| (Cert(d_i \rightarrow d_j) - a) \times (Cert(d_j \rightarrow d_i) - a) \right|}{\sum_{\forall d_i, d_j \in D,\, i \neq j} \left| (Cert(d_i \rightarrow d_j) - a)^2 \right|} \quad (2)$$

määräytyy osastojen välisten varmuuksien ja niiden keskiarvojen *a* avulla.

LIITE 8. Potilaan hoitojakson kulun muodostus

Zhang, H., S. Mehotra, D. Liebovitz, C. A. Gunter & B. Malin (2013). Mining deviations from patient care pathways via electronic medical record system audits. *ACM transactions on management information systems* [online] 4 (4), 17. https://doi.org/10.1145/2544102

Potilaiden hoitojaksosta muodostetaan graafi *G* algoritmilla 1. Graafi muodostuu solmupisteistä *V* ja janoista *E*. Pisteet ja niiden väliset janat saavat painoarvot U ja W.

Algorithm 1 Patient flow graph construction

Input: Vectors: An ordered set of attribute R values $[r_{i_1^p}, r_{i_2^p}, \ldots, r_{i_{N_p}^p}]$ for each patient $p \in \mathscr{P}$.

Output: A complete graph $G = V, E$.

Steps:

Let $V = \{v_1, \ldots, v_{K_r}\}$ be the set of attribute R values
Let $G = \{V, E\}$ be a complete graph
Let $U = \{u_1, \ldots, u_{K_r}\}$ be a set of weights for each element of V, initially $u_i = 0$ for each i
Let $W = \{w_{11}, \ldots, w_{K_r K_r}\}$ be a set of weights for each pair of elements in V, initially $w_{ij} = 0$ for each i, j
for each $p \in \mathscr{P}$ **do** ▷ accesses of patient p
    for each $j \in \{1, \ldots, N_p - 1\}$ **do** ▷ each access of the patient p
        $u_{i_j^p} \leftarrow u_{i_j^p} + 1$ ▷ increment the weight for $r_{i_1^p}$
        $w_{i_j^p i_{j+1}^p} \leftarrow w_{i_j^p i_{j+1}^p} + 1$ ▷ increment the weight for transition $r_{i_j^p} \rightarrow r_{i_{j+1}^p}$
    end for
end for
$u_{i_j^p} \leftarrow u_{i_j^p} + 1$ ▷ account for the last vertex of the sequence
return G