# Tietomurrot henkilötietoja sisältävissä terveydenhuollon big data -ympäristöissä

Turun yliopisto
Tietotekniikan laitos
LuK-tutkielma
Tietojenkäsittelytiede
Maaliskuu 2026
Veera Salonen

TURUN YLIOPISTO
Tietotekniikan laitos

Veera Salonen: Tietomurrot henkilötietoja sisältävissä terveydenhuollon big data -ympäristöissä

LuK-tutkielma, 26 s.
Tietojenkäsittelytiede
Maaliskuu 2026

---

Terveydenhuollossa käsitellään koko ajan kasvavia määriä potilaiden dataa. Valtavat määrät dataa pitää kerätä, säilöä ja käsitellä uusilla menetelmillä ja järjestelmillä, jotka muodostavat yhdessä terveydenhuollon big data -ympäristöt. Nämä ympäristöt ovat tuoneet mukanaan paljon positiivista edistystä, mutta myös uusia tietoturvariskejä. Tämä voidaan huomata tietomurtojen määrän valtavana kasvuna järjestelmien digitalisoitumisen myötä. Terveydenhuollon tuottama big data sisältää myös paljon arkaluonteista dataa ja henkilötietoja, mikä tekee tietomurroista erityisen vakavan ongelman.
Tässä kirjallisuuskatsauksessa keskitytään terveydenhuollon big data -ympäristöihin kohdistuviin tietomurtoihin. Erityisesti keskitytään henkilötietoja sisältäviin järjestelmiin. Tarkoituksena on selvittää tietomurtoihin johtavia syitä sekä miten tietomurtoja voidaan ehkäistä. Lisäksi perehdytään kahteen tapausesimerkkiin, Anthem ja Vastaamo, joiden avulla mietitään konkreettisia toimenpideideoita tietomurtojen ehkäisyyn.
Tutkielmassa selviää, että tietomurtojen taustalla on usein monien eri tekijöiden yhdistelmä. Tämän vuoksi myös niiden ehkäisemiseen ei riitä yksittäiset toimenpiteet, vaan organisaatioilta vaaditaan kokonaisvaltaista lähestymistä tietoturvaan. Yksi tutkielmassa toistuva tekijä on huolimattomuuden ja ihmisten virheiden suuri rooli tietomurroissa. Lisäksi voidaan todeta, että suurin osa tietomurroista on vältettävissä tarpeeksi huolellisella tietoturvasuunnittelulla ja valvonnalla.

Asiasanat: big data, tietomurrot, henkilötiedot, terveystiedot, terveydenhuolto, tietosuoja, tietoturva, Anthem, Vastaamo

# Sisällys

# 1 Johdanto

Nopeasti edistyvän digitalisaation ansiosta nykyisin organisaatiot voivat hyödyntää big dataa monipuolisesti esimerkiksi edistämään innovaatiota ja tehostamaan päätöksentekoa [1]. Terveydenhuollossa potilastietojärjestelmät ovat muuttuneet sähköisiksi ja sairaalaan laiteet ovat yhdistetty nettiin, mikä on tehnyt terveydenhuollosta turvallisempaa, tehokkaampaa ja helpommin saavutettavaa. Samalla kuitenkin potilaan yksityisyys vaarantuu, sillä digitalisaatio tekee järjestelmät haavoittuvaisemmiksi tietomurroille. [2]

Terveydenhuoltoala on yksi eniten tietomurroista kärsineistä big dataa hyödyntävistä toimialoista. Yhdysvaltalaisen the Privacy Rights Clearinghouse -järjestön esittämän datan mukaan vuosina 2005–2019 rekisteröidyistä 6355 big datan tietomurrosta 3912 kohdistui terveydenhuoltoalaan. [2] Terveydenhuollossa tietomurrot ovat myös erityisen haitallisia, koska varastettu tieto sisältää usein henkilötietoja ja muuta arkaluontoista dataa, kuten potilaiden terveys- ja maksutietoja [3]. On siis erityisen tärkeää, että terveydenhuollon organisaatiot pysyvät digitalisaation perässä ja osaavat suojautua tietomurroilta.

Tietomurroilta voidaan suojautua useilla erilaisilla toimenpiteillä. Organisaatioilla on usein omat tietoturvamallit ja -mekanismit. Näihin sisältyy erilaisia teknisiä toimenpiteitä, kuten datan salaus tai varmuuskopiointi. Organisaatioilla on myös omat ohjeet ja käytänteet, joiden avulla pidetään yllä tietoturvaa. Esimerkiksi tietoturvakoulutus ja salasanojen säännöllinen vaihtaminen ovat tällaisia käytänteitä.

Lisäksi on valtakunnallisia ja kansanvälisiä säädöksiä, joita organisaatioiden tulee noudattaa käsitellessään henkilötietoja. [3]

Tässä tutkielmassa tarkastellaan terveydenhuollon big data -ympäristöihin kohdistuvia tietomurtoja ja keskitytään erityisesti järjestelmiin, jotka sisältävät henkilötietoja. Tarkoituksena on selvittää tietomurtoihin johtavia tekijöitä ja niissä käytettyjä menetelmiä. Selvitetään myös, millaisilla toimenpiteillä terveydenhuollon organisaatiot voivat vähentää tietomurtoja. Lisäksi tarkastellaan kahta tapausesimerkkiä terveydenhuoltoon kohdistuneista tietomurroista ja pohditaan, miten nämä tapaukset olisi voitu mahdollisesti ehkäistä.

Tutkielmassa vastataan seuraaviin tutkimuskysymyksiin:

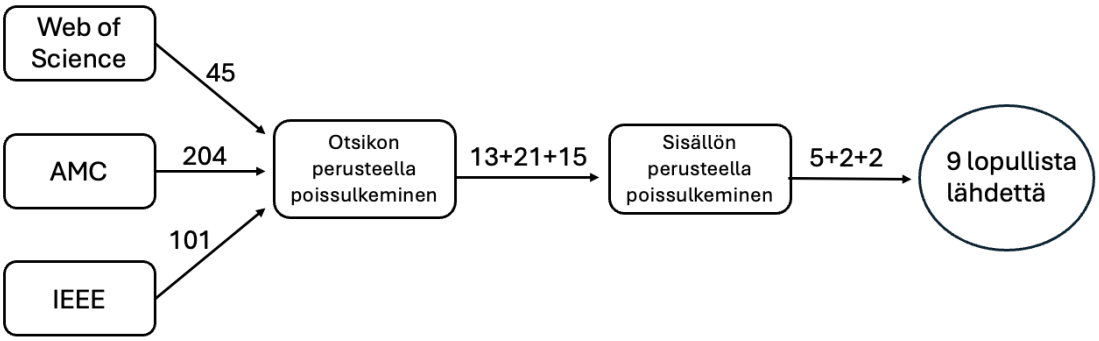
1. Millaisia tietomurtoja terveydenhuollon big data -ympäristöissä tapahtuu ja mitkä tekijät altistavat niille?

2. Miten terveydenhuollon organisaatiot voivat ehkäistä tietomurtoja big data -ympäristöissä?

Tutkielman alussa taustoitetaan aihetta kahden luvun verran. Luvussa 2 määritellään big data käsitteenä ja kerrotaan terveydenhuollon big datasta sekä sen haasteista. Luvussa 3 taas kerrotaan tietoturvan ja -suojan perusperiaatteista sekä määritellään tietomurrot ja kerrotaan tietomurroissa käytetyistä menetelmistä. Luvussa 4 kerrotaan tietomurroista terveydenhuollossa ja miten niitä voidaan ehkäistä. Lisäksi käsitellään tapausesimerkit Anthemin ja Vastaamon tietomurroista. Lopuksi tutkielman pääkohdat ja päätelmät kootaan yhteen luvussa 5.

Tutkielma on toteutettu kirjallisuuskatsauksena. Tärkeimmät lähteet on haettu pääosin Web of Science- , AMC- ja IEEE-tietokannoista. Lisäksi lähteitä taustalukuihin on haettu erityisesti Google Scholar -hakukoneen avulla. Hakulauseena on käytetty *”big data” AND ”data breach” AND "healthcare"* ja haku on kohdistettu julkaisuvuosiin 2021–2025. Kuvassa 1.1 on havainnollistettu, miten tutkielman tärkeimmät lähteet on haettu. Hakulauseella ja julkaisuvuosirajauksella saadut tulokset

on käyty ensin otsikoiden ja tiivistelmien perusteella läpi ja karsittu pois epäsopivat. Sen jälkeen aineistojen koko sisällön perusteella on valittu parhaiten aihetta tukevat tekstit.




Kuva 1.1: Tutkielman tärkeimpien lähteiden hakuprosessi.

# 2 Big data ja terveydenhuolto

## 2.1 Big data

Termille big data eli massadata on olemassa monenlaisia määritelmiä, jotka tarkentuvat ja vakiintuvat teknologian kehityksen mukana. Yksinkertaisesti sanottuna big data kuitenkin viittaa todella suuriin datamääriin, joita on mahdotonta hallita perinteisillä datan hallintakeinoilla ja työkaluilla. Määrä ylittää perinteisesti käytössä olleen tallennus-, käsittely- ja analysointikapasiteetin. [4]

Yleinen tapa big datan määrittelyyn on kuvata sitä Doug Laneyn ns. kolmen V:n (engl. 3 Vs) avulla. Kolmella V:llä viitataan kolmeen big datalle tyypilliseen ominaisuuteen: volyymiin (engl. volume), nopeuteen (engl. velocity) ja vaihteluun (engl. variety). Volyymi kuvaa datan suurta määrää, nopeus datan määrän erittäin nopeaa kasvua, ja vaihtelu datan lähteiden, tyypin ja luonteen moninaisuutta. Tähän määritelmään on myöhemmin lisätty myös muita V-kirjaimia, joista tunnetuin on luotettavuus (engl. veracity). [4]

On olemassa kolmea erityyppistä big dataa: Strukturoitu big data (engl. structured big data), rakenteeton big data (engl. unstructured big data) ja puolistrukturoitu big data (engl. semi-structured big data). Strukturoitu data on hyvin jäsenneltyä dataa, jota voidaan hakea ja tallentaa helposti. Rakenteeton data taas on tietoa, jolla ei ole selkeää rakennetta tai muotoa. Sen käsittely ja analysointi on hankalaa. Puolistrukturoitu data sisältää sekä strukturoitua että rakenteetonta dataa. Sitä voi-

daan usein hankkia jäsennellyssä muodossa, mutta sitä ei voida säilöä perinteisillä tietorakenteilla, kuten taulukoilla. [5]

Big data -ratkaisut mahdollistavat organisaatioille suurten tietomassojen keräämisen, käsittelyn ja analysoinnin. Ne tukevat parempaa päätöksentekoa, tehostaa toimintaa ja edistää innovaatioita. Samaan aikaan datan määrän, käsittelynopeuden ja monimuotoisuuden kasvu on tuonut mukanaan uusia tietoturvariskejä, kuten tietomurrot, sisäiset uhat ja datan manipulointi. [3]

## 2.2 Terveydenhuollon big data

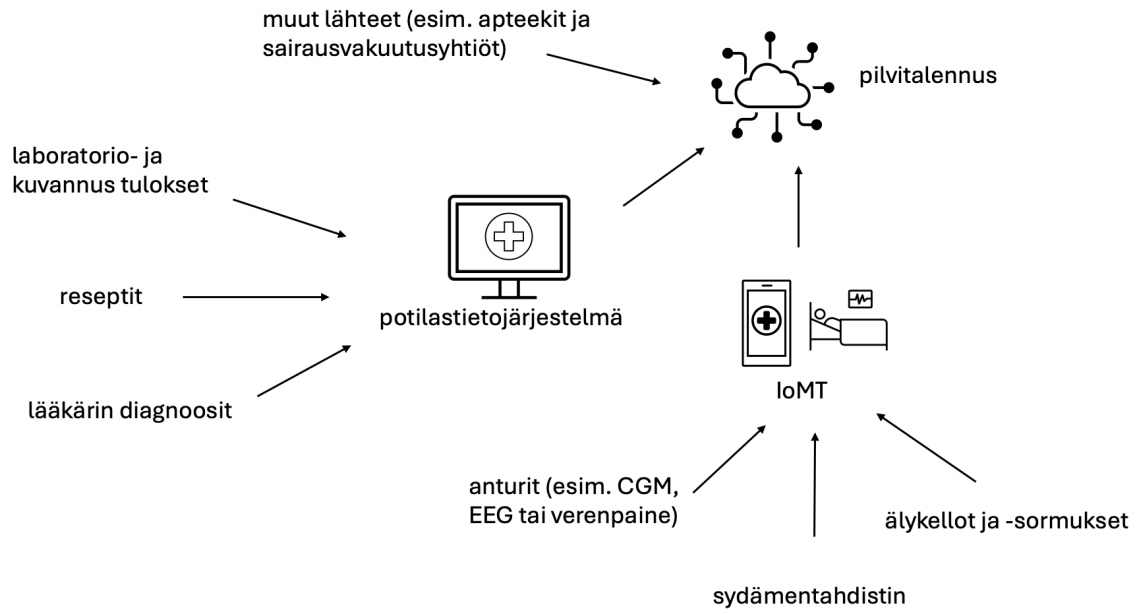
Terveydenhuolto tuottaa isoja määriä monentyyppistä dataa. Eri datatyyppejä ovat esimerkiksi potilaan terveyteen ja hoitoon liittyviä tietoja sisältävä kliininen data ja erilaisten laitteiden tuottamat mittaustulokset eli biometrinen data. Lisäksi terveydenhuollon tuottama data sisältää tutkimusdataa, taloudellista dataa ja perinteisiä henkilötietoja. Tämä data kertyy monista eri lähteistä, kuten sairaalan järjestelmistä, laboratorioista ja apteekeista. [6] Kaksi suurta terveydenhuollon big datan lähdettä ovat lääketieteellisten esineiden internet ja potilastietojärjestelmät [4].

**Lääketieteellisten esineiden internetillä** (engl. Internet of Medical Things, lyh. IoMT) tarkoitetaan terveydenhuollon järjestelmiä, joissa laitteet pystyvät keräämään dataa sekä jatkuvasti analysoimaan, tuottamaan ja siirtämään sitä internetyhteyden avulla. Tällaisia laitteita ovat esimerkiksi elintoimintojen seurantaan käytetyt kliiniset laitteet, kuten älykkäät glukoosimittarit, sekä puettava terveysteknologia, kuten älykellot. IoMT-järjestelmät mahdollistavat potilaan terveyden tarkkailun paremmin myös etänä. Niiden avulla potilaasta saadaan runsaasti dataa, jota voidaan hyödyntää hoidossa yhdessä muun terveysdatan kanssa. [4][7]

**Potilastietojärjestelmä** (engl. electronic health record, lyh. EHR) on järjestelmä, jonka avulla terveydenhuollon henkilöstöllä on pääsy kaikkiin potilaan terveystietoihin. Järjestelmä sisältää potilailta kerättyä lääketieteellistä ja kliinistä dataa,

kuten diagnooseja, testituloksia, reseptejä, allergiatietoja ja muuta hoidon kannalta oleellista tietoa. EHR-järjestelmät helpottavat koordinaatiota eri terveyspalveluntarjoajien välillä sekä vähentävät logistisia virheitä ja tarpeettomia tutkimuksia. [4]

Uudet digitalisaation mukanaan tuomat dataa tuottavat järjestelmät vaativat myös datan säilömiseen käytettyjä järjestelmiä mukautumaan datan suureen määrään ja uusiin muotoihin. Terveydenhuollolle parhaaksi säilöntäjärjestelmäksi on muodostunut pilvitallennus (engl. cloud storage). Se on skaalautuva, joustava ja turvallinen tapa säilöä suuria määriä dataa. Pilvitallennus on korvannut tai tullut avuksi aiemmin käytetylle ratkaisulle eli tietojen säilömiselle paikallisesti. Pilven sisällä data on säilötty tiedostoissa, lohkoissa tai objekteina. [8]




Kuva 2.1: Terveydenhuollon bigdata -ympäristö.

Kuvassa 2.1 havainnollistetaan esimerkkiä terveydenhuollon big data -ympäristöstä. Kuvassa on erilaisia esimerkkejä terveydenhuollon tuottamasta datasta, kuten diagnoosit, reseptit ja anturien mittaustulokset. Tämä data on peräisin potilastietojärjestelmästä, IoMT-laitteista sekä muista lähteistä, kuten apteekeista ja vakuutusyh-

tiöistä. Näistä tiedoista muodostuu terveydenhuollon big data, jota voidaan säilöä pilvessä.

Terveydenhuollon dataa voidaan säilöä, analysoida ja hyödyntää useisiin eri tarkoituksiin. Sitä voidaan käyttää auttamaan diagnosoinnissa ja sopivan hoitomenetelmän valitsemisessa. Sen avulla voidaan ennakoida ja ehkäistä sairauksia. Lisäksi dataa voidaan käyttää lääketieteeseen ja terveyteen liittyvässä tutkimuksessa sekä tarkkailemaan ja kuvamaan erilaisten ihmisryhmien terveydentilaa. [6]

Suuret määrät dataa ja sähköiset järjestelmät voivat olla erittäin hyödyllisiä terveydenhuollolle. Datan keräämiseen ja säilyttämiseen liittyy kuitenkin suuri vastuu, koska data sisältää henkilötietoja (engl. personal data). Se tekee datasta erityisen arkaluonteista, minkä seurauksena myös tietoturvariskit ovat suurempia. [3]

## 2.3 Haasteet ja riskit

Big dataan liittyy myös monia haasteita ja riskejä, koska sitä ei pystytä käsittelemään perinteisillä ratkaisuilla. Suurimmat haasteet liittyvät tietoturvaan sekä datan rakenteeseen, prosessointiin, standardointiin ja säilömiseen. Lisäksi nopeasta kehityksestä johtuva yleinen osaamattomuus ja tietämättömyys aiheuttaa haasteita esimerkiksi datanhallinnassa. [6] Tässä tutkielmassa keskitytään erityisesti terveydenhuollon big datan haasteisiin.

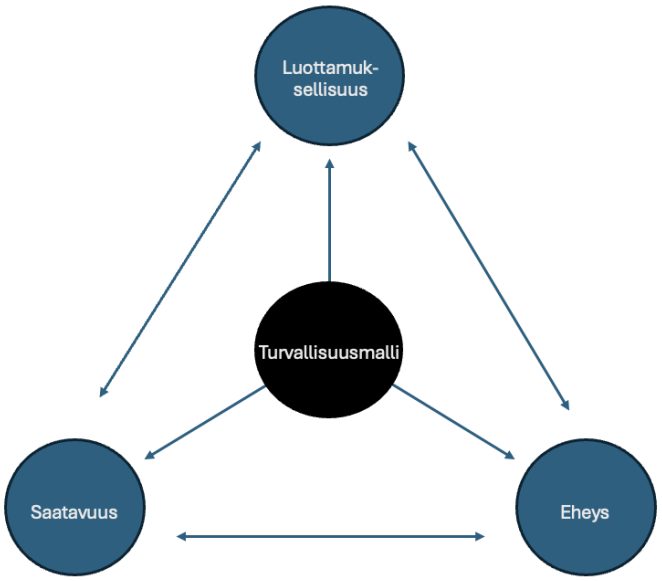
Datan valtava määrä, rakenteen puute, kasvun nopeus ja monimuotoisuus on yksi suurimmista haasteista [9]. Terveydenhuollossa dataa kerääntyy päivittäin suuria määriä ja perinteiset tietokannat eivät pysty säilömään nopeasti kasvavaa datan määrää. Data on myös usein rakenteetonta ja monimuotoista, kuten lääkärin muistiinpanoja, testituloksia ja kuvannustuloksia. Se tekee datan turvallisesta ja käytännöllisestä analysoimisesta haastavaa. Pilvipalvelut ovat tuoneet tähän ongelmaan helpotusta, sillä ne mahdollistavat big datan säilömisen ja ylläpidon perinteisiä tapoja paremmin. [10]

Toinen big datan suurimmista haasteista on sen tietoturva. Erityisesti terveydenhuollossa tähän täytyy kiinnittää runsaasti huomiota, sillä terveydenhuollon data on usein arkaluonteista. Eräs datan tietoturvaongelma on se, että terveydenhuollon henkilöstöllä pitää olla pääsy potilaan dataan useissa eri sijainneissa. Pitää siis löytää ratkaisu, jossa potilastietoihin on tarpeeksi laaja pääsy tietoturvasta liikaa tinkimättä. [10]

# 3 Big datan tietoturva ja tietosuoja

## 3.1 Tietoturvan perusperiaatteet

Tietoturva perustuu perinteisesti kolmeen perusvaatimukseen eli luottamuksellisuuteen, eheyteen ja saatavuuteen [11]. Näistä muodostuu CIA-malli, joka on yksi tunnetuimmista datan suojamiseen käytetyistä malleista. Kuvassa 3.1 mallinnetaan, miten CIA-mallin avulla voidaan suojata dataa. Nuolilla kuvataan eri tekijöiden suhdetta toisiinsa ja ne painottavat erityisesti tasapainon tärkeyttä. Jos halutaan vahvistaa esimerkiksi eheyttä ja luottamuksellisuutta, voidaan mahdollisesti joutua tinkimään saatavuudesta. [12]




Kuva 3.1: CIA-malli. Mukaillen lähdettä [12].

**Luottamuksellisuus** (engl. confidentiality) viittaa tietojen suojaamiseen luvattomalta käytöltä [11]. Organisaatoilla on dataa, johon vain tietyillä tahoilla saa olla pääsy. Luottamuksellisuus varmistaa, että vain nämä tahot pääsevät käsiksi kyseiseen dataan. Jos arkaluontoista dataa päätyy väärille henkilöille, se saattaa vaarantaa organisaation tai sen asiakkaat. [12] Esimerkki luottamuksellisuuden rikkoutumisesta on tilanne, jossa ei-toivottu taho pääse tarkastelemaan potilaan tietoja.

**Eheys** (engl. integrity) viittaa tietojen suojaamiseen luvattomilta muutoksilta [11]. Tällä tarkoitetaan datan suojaamista kaikilta tahoilta riippumatta siitä, onko niillä luvallinen pääsy dataan tai ei. Lisäksi pitää varautua datan korruptoitumiseen. Eheyden varmistaminen sisältää datan suojaamista ei-toivotuilta muutoksilta ja valmiutta toipua näistä muutoksista niiden tapahtuessa. [12] Esimerkki eheyden rikkoutumisesta on tilanne, jossa potilastiedot muuttuvat tai katoavat tahattomasti niiden siirron aikana.

**Saatavuus** (engl. availability) viittaa siihen, että valtuutetuilla käyttäjillä on pääsy tietoihin tarvittaessa [11]. Tasapainon löytäminen tarvittavan turvallisuuden ja riittävän saatavuuden välillä voi olla haastavaa. Datan suojaaminen on todella tärkeää, mutta se vähentää saatavuutta. Jos data ei ole saatavissa tarvittaessa, sillä voi olla todella vakaviakin seurauksia. [12] Saatavuus voi vaarantua esimerkiksi potilastietojärjestelmän palvelimen kaatuessa, jolloin hoitohenkilökunnalla ei välttämättä ole pääsyä potilastietoihin.

Tavat säilöä, siirtää ja suojata dataa ovat muuttuneet ja kehittyneet paljon viime vuosina. Tämän vuoksi myös tietoturvaa pitää tarkastella eri näkökulmista riittävän kattavasti. Yksi tapa parantaa tietoturvaa on Parkerian Hexad -mallin avulla. Se lisää CIA-malliin kolme uutta tekijää: todennuksen (engl. authenticity), käyttökelpoisuuden (engl. utility) ja hallinnan (engl. possession). Todennuksella tarkoitetaan, että välitetty tieto on varmasti sen väittämästä lähteestä. Käyttökelpoisuus viittaa siihen, että data on käytettävässä ja ymmärrettävässä muodossa. Hallinnalla

viitataan datan suojaamiseen tilanteissa, jossa data päätyy väärän tahon hallintaa, mutta tällä taholla ei ole pääsyä siihen. [12]

## 3.2 Tietosuojan perusperiaatteet

Tietosuoja on jokaisella oleva oikeus, joka turvaa yksityisyyden säilymisen ja henkilötietojen lakiin perustuvan käsittelyn. Henkilötiedot ovat tietoja, joista henkilö voidaan tunnistaa suoraan tai välillisesti [13]. Suoraan henkilön yksilöiviä henkilötietoja ovat esimerkiksi nimi, henkilötunnus ja puhelinnumero. Välillisesti tunnistettavia henkilötietoja ovat esimerkiksi IP-osoite, paikannustiedot ja osa potilastiedoista. Näitä ei voi ilman lisätietoja yhdistää tiettyyn henkilöön. Yksi tärkeimmistä tietosuojaa turvaavista laeista on yleinen tietosuoja-asetus (engl. General Data Protection Regulation, lyh. GDPR) [14].

GDPR on vuonna 2018 voimaan tullut laki, joka antaa henkilöille lisää oikeuksia päättää heidän henkilötietojensa käsittelystä. GDPR on Euroopan Unionissa säädetty laki, joka koskee kaikkia EU-alueella toimivia henkilöitä ja organisaatioita. Sen lisäksi se koskee kaikkia organisaatioita, jotka käsittelevät EU-kansalaisten dataa. GDPR:n perusperiaatteet ovat laillisuus, oikeudenmukaisuus, läpinäkyvyys, tarkkuus, datan minimointi, vastuullisuus, säilytyksen rajoittaminen ja käyttötarkoitussidonnaisuus. [14]

Vastuu tietosuojalain noudattamisesta on lähtökohtaisesti rekisterinpitäjällä eli taholla, joka määrittelee henkilötietojen käsittelyn tarkoitukset ja menetelmät. Rekisterinpitäjän pitää varmistaa myös, että tietosuoja on huomioitu henkilötietojen käsittelijän toiminnassa asianmukaisilla teknisillä ja hallinnollisilla toimenpiteillä. Henkilötietojen käsittelijä on kolmannen osapuolen taho, esimerkiksi yritys tai viranomainen, joka käsittelee henkilötietoja rekisterinpitäjän alaisuudessa. [15][16]

GDPR takaa monia oikeuksia rekisteröidylle eli henkilölle, jonka henkilötietoja käsitellään. GDPR:n mukaan rekisteröidyllä on oikeus tietää, mitä hänen tietojansa käsitellään ja mihin tarkoitukseen. Hänellä on myös oikeus saada pääsy häntä koskeviin käsiteltäviin tietoihin sekä oikeus vaatia oikaisua niiden ollessa virheellisiä puutteellisia tai epätarkkoja. Lisäksi rekisteröidyllä on oikeus tietyissä tilanteissa pyytää tietojen poistamista tai rajoittaa niiden käsittelyä, kuten silloin, jos rekisteröity kokee, että hänen tietojansa on käsitelty lainvastaisesti. Näiden oikeuksien avulla halutaan parantaa yksilön asemaa sekä luoda läpinäkyvyyttä ja luottoa henkilötietojen käsittelyyn. [17]

Yhdysvalloissa vastaavanlainen laki on Health Insurance Portability and Accountability Act eli HIPAA, joka koskee erityisesti terveydenhuollon dataa. Suurin ero GDPR:n ja HIPAA:n välillä on se, että GDPR suojaa kaikkea dataa samalla tavalla sen muodosta, lähteestä ja sijainnista huolimatta, kun HIPAA taas soveltaa erilaisia sääntöjä dataan esimerkiksi riippuen siitä, mistä se on peräisin ja kenen hallinnassa se on. On myös olemassa paljon dataa, jota HIPAA ei koske ollenkaan, kuten älylaitteiden keräämää terveysdataa. [18]

## 3.3 Tietomurrot ja tietomurtomenetelmät

Tietomurrot ovat yksi suurimmista big datan tietoturvauhkista. Viiden viime vuoden aikana noin 157 miljoonaa ihmistä on jollain tavalla joutunut tietomurron kohteeksi. [10] Big data -järjestelmissä on haavoittuvuuksia, koska niissä säilötään valtavia määriä monimutkaista dataa, johon useilla tahoilla on pääsy. Nämä haavoittuvuudet voivat altista tietomurroille, joissa hakkeri saa haltuunsa arkaluontoista dataa ja voi paljastaa sitä tai myydä sitä rikolliseen käyttöön. [9] Myös standardien ja sääntöjen noudattamatta jättäminen sekä huolimattomuus aiheuttavat haavoittuvuutta ja lisäävät tietomurtojen riskiä [3].

Tietomurtoja tehdään monenlaisilla eri menetelmillä. Uhkat voivat olla ulkoisia tai sisäisiä. Taulukossa 3.1 on lueteltu ja kuvattu yleisimpiä menetelmiä. Yleisiä ulkoisia uhkia on tietojenkalastelu (engl. phishing), haittaohjelmat (engl. malware), näppäilytallennin (engl. keylogger) ja erilaiset salasanahyökkäykset. Myös vanhentuneet ja vialliset ohjelmistot tekevät järjestelmän haavoittuvaiseksi ulkoisille uhkille. Sisäisiä uhkia taas on ihmisten inhimilliset virheet, vahingot sekä pahantahtoiset sisäpiiriläiset. [10]

Taulukko 3.1: Tietomurtomenetelmiä. Käännetty ja muokattu lähteestä [10].

| Tietomurtomenetelmä | Uhan tyyppi | Selitys |
|---|---|---|
| Ohjelmistohaavoittuvuus | ulkoinen | Virhe tai puute ohjelmistossa, jota hyökkääjä voi hyödyntää. |
| Tietojenkalastelu | ulkoinen | Hyökkääjä yrittää huijata yksilöä luovuttamaan arkaluontoisia tietoja, kuten salasanan tai pankkitiedot. |
| Haittaohjelma | ulkoinen | Ohjelma tai sovellus, jonka tarkoitus on häiritä tai vahingoittaa järjestelmää. Esimerkiksi madot ja virukset. |
| Näppäilytallennin | ulkoinen | Ohjelma, joka tallentaa kaikki näppäinpainallukset ja mahdollisesti lähettää ne ulkopuoliselle taholle. Voidaan käyttää esimerkiksi varastamaan salasanoja. |
| Salasanahyökkäys | ulkoinen | Hyökkääjä yrittää arvata, varastaa tai murtaa salasanan. |
| Inhimillinen virhe | sisäinen | Ihmisen tekemät vahingolliset teot, kuten päivityksen tekemättä jättäminen ja heikot salasanat. |
| Tahaton häviäminen | sisäinen | Tilanne, jossa data tai laite on vahingossa kadotettu tai tehty saavuttamattomaksi. |
| Pahantahtoinen sisäpiiriläinen | sisäinen | Henkilö, joka käyttää valtuuksiaan väärin. Hän voi esimerkiksi varastaa dataa tai sabotoida järjestelmiä. |

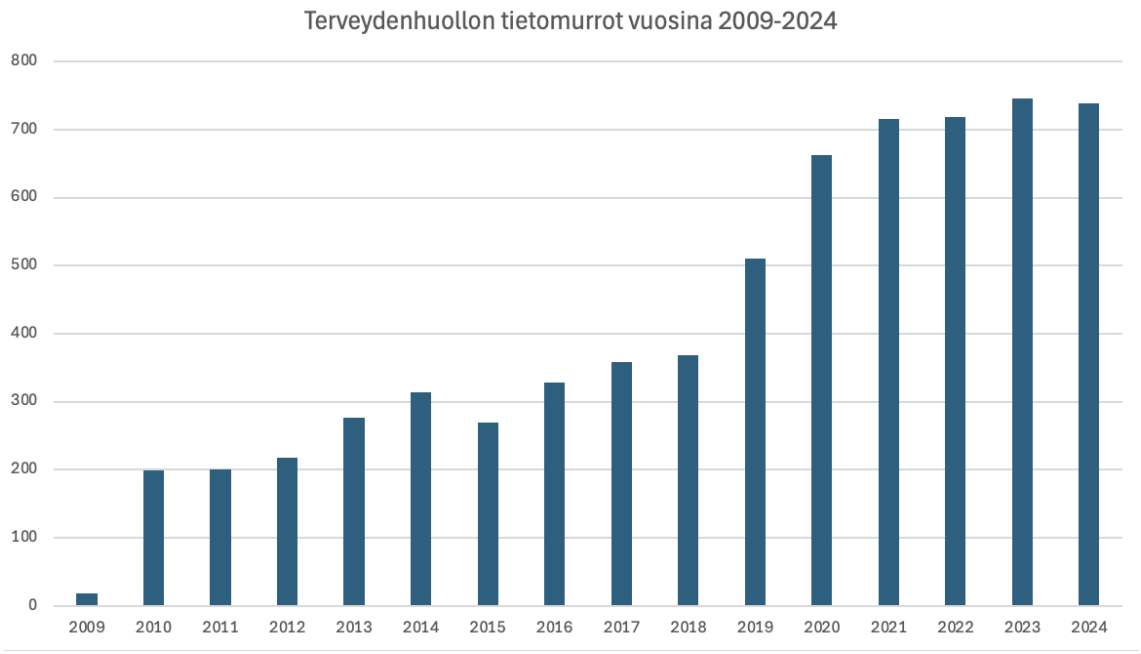
Suurinta osaa tietomurtomenetelmistä yhdistää se, että ne vaativat joko käyttäjän tai kehittäjän virheitä [10]. Ihminen onkin tietoturvan suurin riskitekijä, koska koulutus ja osaaminen eivät pysy nopeasti kehittyvän teknologian perässä [19]. Ihmisellä on myös tunteita, joita hakkerin on helppo käyttää hyväksi eri tavalla kuin teknologian heikkouksia [10].

# 4 Tietomurrot terveydenhuollon järjestelmissä ja niiden ehkäisy

## 4.1 Tietomurrot terveydenhuollossa

Terveydenhuollossa tietomurrot ovat erityisen suuri ongelma, sillä esimerkiksi vuonna 2018 maailmanlaajuisesti raportoidusta 2216 tietomurrosta terveydenhuoltoalaan kohdistui 536. Terveydenhuollon data on myös usein todella arkaluonteista ja sisältää henkilötietoja, minkä takia tietomurtojen seuraukset ovat vakavia. Tietomurrot voivat johtaa suuriin taloudellisiin ja maineellisin haittoihin terveydenhuollon toimijoille sekä loukata vakavasti potilaan yksityisyyttä. [10] Henkilötietojen joutuminen vääriin käsiin voi johtaa esimerkiksi petokseen, identiteettivarkauteen tai jopa kuolemaan.

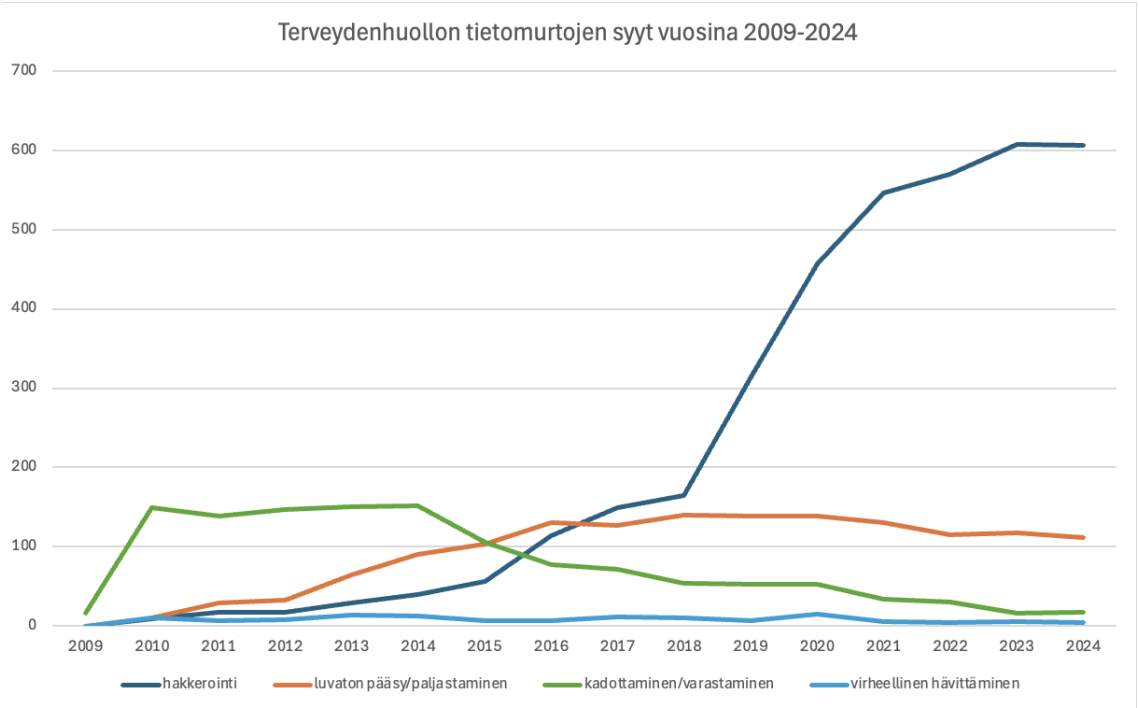
Kuvassa 4.1 on Yhdysvalloissa tapahtuneiden terveydenhuollon tietomurtojen määriä vuosina 2009–2024. Kuten pylväsdiagrammista näkyy, tietomurtojen määrä terveydenhuollossa on kasvanut valtavasti 15 viime vuoden aikana. Tietomurtojen lisääntymiseen on monia syitä, mutta erityisesti palvelujen digitalisoituminen ja laitteiden yhdistäminen internetiin ovat varmasti vaikuttaneet asiaan. Digitalisaatio on myös edennyt todella nopeasti, joten organisaatioiden tietoturvamallit ja koulutus eivät ole välttämättä pysyneet mukana nopeassa muutoksessa.




Kuva 4.1: Terveydenhuollon tietomurrot vuosina 2009–2024 Yhdysvalloissa. Mukaillen lähdettä [20].

Kuva 4.2 taas kuvastaa terveydenhuollon tietomurtojen syitä Yhdysvalloissa vuosina 2009–2024. Tällä hetkellä nopeimmin yleistyvä tietomurtotyyppi on hakkerointi. Tähän on todennäköisesti vaikuttanut sama syy kuin tietomurtojen yleistymiseen ylipäätään eli digitalisaation nopea eteneminen. Luvaton pääsy ja paljastaminen ovat myös kasvaneet tarkastelujakson alkupuolella, mutta saatu viime vuosina laskuun. Näihin tapauksiin lukeutuu esimerkiksi henkilökunnan virheitä, laiminlyöntejä ja pahantahtoisten sisäpiiriläisten toimintaa [20]. Taustalla lienee digitalisaation nopean etenemisen ja koulutuksen puutteen yhdistelmä.

Kuvasta myös nähdään, että tarkastelujakson alkupuolella yleisin tietomurtoihin johtava syy oli datan kadottaminen tai varastaminen. Näissä tapauksissa usein on ollut kyseessä paperiset potilastiedot [20], mikä on varmasti syy niiden vähentymiseen digitalisaation myötä. Lisäksi kuvassa näkyy, että ajoittain datan virheellinen hävittäminen voi myös johtaa tietomurtoon. Näissä tapauksissa on usein kyse paperisista potilastiedoista, joita ei ole silputtu oikeaoppisesti [20].




Kuva 4.2: Terveydenhuollon tietomurtojen syyt vuosina 2009–2024 Yhdysvalloissa. Mukaillen lähdettä [20].

Kuvissa 4.1 ja 4.2 on vain Yhdysvaltojen raportoidut tietomurrot ja vain tietomurrot, jotka ovat vaikuttaneet vähintään 500 henkilöön. Kuvia voi myös hieman vääristää se, että ennen tietomurtoja ei huomattu ja raportoitu yhtä hyvin kuin nykyään [20]. Kuvat eivät siis anna täysin tarkkaa tai maailmanlaajuista informaatiota, mutta ne antavat hyvää suuntaa sille, miten tietomurtojen määrä ja syyt ovat muuttuneet viime vuosina.

## 4.2 Tietomurtojen ennaltaehkäisy

Tietomurtojen ennaltaehkäisyllä viitataan toimenpiteisiin, joilla vähennetään tai poistetaan tietomurroille altistavia tietoturvariskejä. Niihin sisältyy riskien ja haavoittuvuuksien tunnistaminen sekä näiden uhkien ja niiden vaikutusten arvioiminen. Lisäksi näihin uhkiin varautuminen erilaisten tietoturvamallien ja -menetelmien sekä

lakien, standardien ja käytäntöjen avulla. Tavoitteena on erityisesti varmistaa datan luottamuksellisuus, eheys ja saatavuus. [9][3] Seuraavaksi käydään läpi yleisiä tapoja suojata big dataa tietomurroilta ja keskitytään niihin erityisesti terveydenhuollon näkökulmasta.

**Jatkuva valvonta ja uhkien havaitseminen** varmistaa, että tietomurto havaitaan ajoissa ja siihen pystytään vastaamaan nopeasti. Tämä toteutetaan järjestelmätoimintojen jatkuvan valvonnan ja tietoturvalokien perusteellisen analyysin avulla. [10] Analyysissä ja valvonnassa voidaan käyttää apuna esimerkiksi algoritmeja ja analyyttisia työkaluja, jotka tunnistavat poikkeamia lokitiedoissa ja käyttäjien toiminnassa [9]. Lisäksi on hyvä tehdä jatkuvia tarkastuksia järjestelmien heikkouksien tunnistamiseksi, jotta niihin voidaan puuttua ajoissa [1].

**Datan luokittelu ja salaus** (engl. data classification and encryption) ovat keskeisiä vahvan tietoturvamallin osia. [10]. Datan luokitellulla tarkoitetaan datan erittelyä eri luokkiin sen arkaluonteisuuden ja tietosuojaa koskevien riskien perusteella [1]. Datan salaus taas muuntaa tiedot sellaiseen muotoon, että vain valtuutetut henkilöt voivat lukea ja käyttää sitä. Vaikka hyökkääjä onnistuisi murtautumaan järjestelmään, salaus estää häntä pääsemästä tietoihin käsiksi ilman tarvittavia käyttäjätunnuksia. [9] Luokittelua ja salausta voidaan usein myös hyödyntää muiden menetelmien yhteydessä ja apuna.

**Käyttöoikeuksien hallinta** (engl. access control) on tapa suojata erityisesti datan luottamuksellisuutta ja eheyttä. Tarkoituksena on varmistaa, että arkaan dataan pääsy vain tarpeellisilla tahoilla. Käyttöoikeuksien hallinta voi sisältää esimerkiksi datan luokittelun ja roolipohjaisen pääsynvalvonnan (engl. role-based access control). [9][10] Esimerkiksi sairaalassa hoitohenkilökunnalla voi olla pääsy vain heidän hoitovastuullansa olevien potilaiden tietoihin. Toimistotyöntekijöillä taas voi olla pääsy potilastietoihin rajatulla näkymällä ja sairaala-apulaisilla ei ollenkaan.

**Todennus** (engl. authentication) on keino suojata erityisesti luottamuksellisuutta ja eheyttä. Tarkoituksena on varmistaa, että dataa pääsevät tarkastelemaan ja käsittelemään vain ne henkilöt, joilla on siihen oikeus. Todennusta voidaan toteuttaa esimerkiksi monivaiheisen tunnistautumisen (engl. multi-factor authentication, lyh. MFA) avulla. [10] Esimerkiksi potilastietoihin käsiksi pääseminen ja niihin tehtävät muutokset voivat vaatia salasanan lisäksi toisen todisteen identiteetistä, kuten varmennuskoodin tai tunnistautumissovelluksen.

**Tietojen peittäminen** (engl. data masking) on menetelmä, joka suojaa erityisesti datan luottamuksellisuutta. Tarkoituksena on tehdä arkaluontoisesta datasta lukukelvotonta luvattomille henkilöille. Peittämällä dataa voidaan käyttää esimerkiksi tutkimustoiminnassa ilman riskiä tietoturvaloukkauksista. Tietojen peittäminen onnistuu esimerkiksi korvaamalla arkaluontoinen data tokenilla tai satunnaisesti arvotulla arvolla. [9] Esimerkiksi sairaalan luovuttaessa potilastietoja tutkimusryhmälle, arkaluontoiset tiedot, kuten nimi ja henkilötunnus, piilotetaan. Tutkijat saavat käyttöönsä diagnoosit, hoitotiedot ja laboratoriotulokset, mutta eivät voi yhdistää niitä yksittäiseen potilaaseen.

**Varmuuskopiointi ja hajautus** (engl. data backup and distribution) ovat menetelmiä, jolla suojataan erityisesti datan saatavuutta. Luomalla säännöllisiä varmuuskopioita voidaan varmistua siitä, että tiedot voidaan palauttaa niiden korruptoituessa tai järjestelmän kaatuessa. Parhaan mahdollisen suojan varmistamiseksi varmuuskopiot voidaan säilyttää myös fyysisesti eri sijainnissa suojamaan esimerkiksi luonnonkatastrofien varalta. Hajauttamalla dataa ja järjestelmiä voidaan varmistaa, että kaikki tiedot eivät korruptoidu kerralla ongelmatilanteessa. [9][10] Esimerkiksi potilastietojärjestelmän kaatuessa on tärkeää, että hoitohenkilökunnalla on pääsy potilastietoihin muilla keinoilla, kuten varmuuskopion tai tulostettujen potilastietojen avulla.

**Tietoturvakoulutus- ja suunnittelu** on yksi tärkeimmistä tavoista suojautua tietomurroilta. Kuten aikaisemmin todettiin, ihminen on tietoturvan suurin heikkous ja tietomurrot vaativat usein inhimillisen virheen. Siksi eräs tehokas tapa vähentää tietomurtoja on hyvä tietoturvasuunnittelu ja henkilökunnan kattava koulutus. On myös tärkeää, että organisaation sisällä kommunikaatio toimii sujuvasti ja tietoturvaongelmille ja -kysymyksille on oma toimiva kanava [1].

Tietoturvasuunnitelmaa tehdessä voidaan ensin arvioida organisaatiolle ominaisia tietoturvariskejä ja sen jälkeen niiden pohjalta suunnitella sopiva tietoturvamalli ja -käytänteet [1]. On tärkeää, että suunnittelussa noudatetaan organisaatiota koskevia säädöksiä ja suosituksia mahdollisimman tarkasti [3]. Lisäksi on hyvä, että organisaation sisäisessä kulttuurissa pyritään painottamaan tietoturvaa ja yksityisyyttä [1]. Tietoturva voidaan ottaa huomioon myös rekrytoinnissa kysymällä hakijalta tietoturvaan liittyviä kysymyksiä, jos työtehtävään liittyy henkilötietojen tai muiden arkaluontoisten tietojen käsittely [19].

Tietoturvakoulutuksissa henkilökunnalle kerrotaan tietoturvauhista, tietosuojasta ja siihen liittyvistä säädöksistä sekä organisaation omista tietoturvakäytänteistä. Voidaan myös järjestää roolipohjaista koulutusta, jotta jokainen henkilökunnan jäsen tietää oman roolinsa tietoturvan varmistamisessa. [1] Lisäksi on tärkeää, että uudet järjestelmät ja menettelytavat koulutetaan aina henkilökunnalle huolellisesti ja ajoissa [3]. Kun esimerkiksi sairaalassa käyttöön on tulossa uusi järjestelmä, tulee se ja siihen liittyvät tietoturvakäytänteet kouluttavat sitä käyttävälle henkilökunnalle ennen sen virallista käyttöönottoa.

## 4.3 Tapausesimerkit

Tietomurrot terveydenhuollossa aiheuttavat suuria riskejä potilas- ja henkilötietojen yksityisyydelle, sekä organisaation maineelle ja taloudelle [10]. Niihin varautuminen ja oikeaoppiset tietoturvatoimet ovat siis erityisen tärkeitä terveydenhuoltoalan or-

ganisaatioille. Yksi tapa selvittää parhaita käytäntöjä tietomurtojen ehkäisemiseen on tarkastella aikaisempia tietomurtoja ja millaiset virheet ja puutteet tietoturvassa ovat mahdollistaneet murrot. Seuraavaksi käsitellään kaksi esimerkkiä terveydenhuollon organisaatioihin kohdistuneista tietomurroista ja pohditaan, miten nämä tietomurrot olisi mahdollisesti voitu ehkäistä.

### 4.3.1 Anthemin tietomurto

Anthemin tietomurto on yksi kaikkien aikojen suurimmista terveydenhuoltoalan tietomurroista. Anthem on suuri yhdysvaltalainen sairausvakuutuksia tarjoava yritys. Vuonna 2015 se joutui valtavan tietomurron kohteeksi, jossa noin 78,8 miljoonan henkilön tiedot vaarantuivat. [21] Varastettuihin tietoihin sisältyi henkilötunnuksia, syntymäaikoja, osoitteita sekä tulo- ja työllisyystietoja. Tietomurto tehtiin Anthemin tietovarastointiin käyttämän kumppaniyrityksen järjestelmään. Aluksi Anthem kuvaili hyökkäystä “kehittyneeksi”, mutta myöhemmin ilmeni, että perinteiset suojauskeinot, kuten vahvempi tunnistautuminen, olisivat mahdollisesti voineet estää sen. [22]

Selvitysten mukaan hyökkäys toteutettiin tietojenkalastelun ja haittaohjelman avulla. Hyökkääjä lähetti Anthemin työntekijöille kalastelusähköposteja, jotka mahdollisesti sisälsivät haittaohjelman. Näiden sähköpostien avulla hyökkääjä sai erään työntekijän kirjautumistiedot käyttöönsä. Niiden avulla hyökkääjä sai pääsyn järjestelmään ja pystyi liikkumaan Anthemin tietokannoissa ja varastamaan dataa. Lisäksi on selvinnyt, että järjestelmän tunnistautumisessa oli heikkouksia, kuten monivaiheisen tunnistautumisen puute ja salasanojen säännöllisen vaihtamisen unohtaminen. Hyökkääjän uskotaan olleen kiinalainen valtion rahoittama ryhmä. [21]

Tietomurto aiheutti vakavia seurauksia sekä Anthemille että sen asiakkaille. Anthem kiisti syyllisyytensä tietomurtoon, mutta organisaatio määrättiin lisäämään tietoturvan rahoitusta ja tekemään muutoksia tietoturvajärjestelmiinsä. Määrättiin

myös, että tietoturvaan tehtävät muutokset julkaistaan kaikkien nähtäville. Anthemin tekemiin muutoksiin kuului esimerkiksi kaikkien yhteistyökumppanien salasanojen vaihtaminen ja kolmevaiheisen todennuksen käyttöönotto laajojen käyttöoikeuksien yhteydessä. Lisäksi Anthem joutui maksamaan noin 115 miljoonaa dollaria vahingonkorvauksia asiakkaille siviilioikeudenkäynneissä. [22]

Anthemin tapauksessa tietojenkalastelu oli suuressa osassa tietomurrossa. Kalastelusähköposti saattoi sisältää myös haittaohjelman, mutta työntekijän inhimillinen virhe sekä heikko tunnistautuminen edesauttoivat silti murron onnistumista. [21] Inhimillisten virheiden ehkäisemiseen paras tapa on henkilöstön jatkuva ja perusteellinen kouluttaminen. Jos työntekijä olisi pystynyt tunnistamaan kalastelusähköpostien tyypilliset piirteet, hän ei todennäköisesti olisi päätynyt avaamaan sitä. Yksin se olisi mahdollisesti voinut ehkäistä tietomurron tai pienentää sen laajuutta. Myös paremman salasanapolitiikan kouluttaminen ja harjoittaminen olisi ollut tärkeää.

### 4.3.2 Vastaamon tietomurto

Vastaamon tietomurto tehtiin vuosina 2018–2019 ja tuotiin julkiseksi vuonna 2020. Vastaamo on suomalainen yksityinen mielenterveyspalveluja tarjoava yritys. Vuoden 2020 lokakuussa mediassa julkaistiin tieto, että noin 30 000 asiakkaan potilas ja -henkilötiedot oli varastettu Vastaamon tietokannasta. Varas vaati ensin Vastaamolta lunnaita uhaten julkaista tiedot kaikkien nähtäville. Kun Vastaamo ei suostunut lunnaisiin, lähetti varas tietonsa menettäneille henkilöille kiristysviestejä. Koska kiristys ei tuottanut tuloksia, tekijä alkoi julkaisemaan potilas- ja henkilötietoja anonyymisti Tor-verkossa. [23]

Selvitysten perusteella tietomurron mahdollistivat useat eri tekijät. Hallinnollisesti Vastaamo ei ollut suojannut potilastietoja riittävän vahvasti ja henkilöstön tietoturvakäytännöt olivat heikkoja. Yritys käytti itse rakennettua EHR-järjestelmää,

joka ei täyttänyt terveydenhuollon tietoturvastandardeja. Lisäksi Vastaamo oli saanut tietoturvaan liittyviä varoituksia ja huomautuksia tarkastuksissa havaituista puutteista, mutta ei ollut tehnyt riittäviä korjaavia toimia. Tietomurtoon reagointi ei myöskään ollut riittävällä tasolla, sillä tietomurtoja oli todennäköisesti tapahtunut jo aiemmin, mutta niistä ei ollut raportoitu. [24]

Teknisestä näkökulmasta tietomurron mahdollistivat potilastietojen salauksen puute, monivaiheisen tunnistautumisen puuttuminen, puutteelliset palomuuriratkaisut sekä heikko salasanapolitiikka. Vastaamon käyttämä EHR-järjestelmä ei salannut potilastietoja levossa tai siirrossa, minkä vuoksi tietoturva perustui käytännössä vain palomuureihin ja salasanoihin. Palomuurit oli selvitysten mukaan kuitenkin mahdollisesti poistettu käytöstä. Tietokantaa suojasi käytännössä siis vain yksi kirjautumistunnus, minkä vuoksi murron onnistuessa hyökkääjä sai pääsyn koko tietokantaan. [24]

Vastaamon tietomurrolla oli vakavat seuraukset yritykselle ja sen asiakkaille. Noin 30 000 potilaan tietojen vuotaminen aiheutti suurta henkistä kärsimystä ja vakavimmillaan ajoi henkilöitä jopa itsemurhaan. Vastaamo ajautui lopulta konkurssiin tietomurron aiheuttamien mainehaittojen seurauksena. [24][23] Vastaamon ex-toimitusjohtaja tuomittiin alun perin tietosuojarikoksesta vuonna 2023, mutta tuomio kuitenkin kumottiin myöhemmin vuonna 2025 [25]. Murron tekijä Aleksanteri Kivimäki taas tuomittiin vuonna 2024 kuuden vuoden ja kolmen kuukauden vankeusrangaistukseen tietomurtoon liittyvistä rikoksista [26].

Tietomurto ei johtunut mistään yksittäisestä virheestä vaan sen mahdollisti pitkään jatkunut tietoturvan laiminlyönti. Näin ollen sen estäminenkään ei olisi onnistunut yksittäisillä toimenpiteillä. Kokonaisvaltaisemmilla tietoturvamuutoksilla tietomurron olisi voitu kuitenkin mahdollisesti ainakin rajata pienemmäksi. Tällaisia muutoksia olisi voinut olla henkilöstön parempi kouluttaminen, laajempi tietoturvasuunnittelu, tiukempi salasanapolitiikka sekä tietoturvasuositusten huolellisempi

noudattaminen. Lisäksi valtion puolesta olisi voitu suorittaa tarkempaa valvontaa, että tietoturvaan liittyviä lakeja ja säädöksiä noudatetaan ja niistä tehtyihin huomautuksiin reagoidaan.

Merkittävin ehkäisevä toimenpide olisi ollut terveydenhuollon tietoturvastandardit täyttävä potilastietojärjestelmä, joka salaa tiedot myös levossa ja siirrossa. Toinen tapa estää tietomurto tai ainakin rajata sitä olisi ollut datan luokittelu ja hajauttaminen sekä käyttöoikeuksien parempi hallinta. Tällöin hyökkääjän päästessä tietokantaan kaikki data ei olisi ollut saatavilla kerrallaan. Lisäksi jatkuvalla valvonnalla ja uhkien havaitsemisella olisi voitu havaita tietomurto aiemmassa vaiheessa, jolloin sen aiheuttamat haitat eivät olisi olleet välttämättä niin laajat.

# 5 Johtopäätökset

Terveydenhuollon big data -järjestelmiin kohdistuvat tietomurrot ovat merkittävä ja kasvava ongelma nopeasti digitalisoituvassa yhteiskunnassa. Henkilötietoja ja muuta arkaluonteista tietoa sisältävät järjestelmät ovat erityisen haavoittuvia hyökkäyksille, koska niihin kohdistuvilla tietomurroilla on usein vakavia seurauksia sekä potilaille että organisaatioille. Tässä tutkielmassa tarkasteltiin terveydenhuollon big data -ympäristöihin kohdistuvia tietomurtoja, niihin johtavia tekijöitä sekä keinoja niiden ehkäisemiseen.

Tutkielman perusteella voidaan todeta, että tietomurrot terveydenhuollossa johtuvat harvoin yhdestä yksittäisestä syystä. Useimmiten taustalla on useiden teknisten, hallinnollisten ja inhimillisten tekijöiden yhdistelmä. Järjestelmien monimutkaisuus, datan määrä ja rakenteettomuus sekä nopea teknologinen kehitys lisäävät haavoittuvuutta. Keskeisimmiksi tietomurtoja mahdollistaviksi tekijöiksi tutkielmassa nousi erityisesti inhimilliset ja hallinnolliset tekijät, kuten työntekijöiden virheet, heikko salasanapolitiikka sekä puutteellinen tietoturvan koulutus, suunnittelu ja valvonta.

Tietomurtojen ehkäisy terveydenhuollon big data -ympäristöissä edellyttää kokonaisvaltaista lähestymistapaa. Tekniset ratkaisut, kuten salaus, monivaiheinen tunnistautuminen ja varmuuskopiointi ovat tärkeä osa tietoturvaa, mutta ne eivät yksin riitä. Niiden tukena täytyy muistaa perusteellinen tietoturvasuunnittelu, ajantasaiset tietoturvakäytännöt sekä henkilöstön jatkuva kouluttaminen ja tietoisuuden

lisääminen. Myös organisaation tietoturvakulttuurin merkitys on suuri, koska tietoturva on kaikkien järjestelmiä käyttävien vastuulla.

Tutkielmassa käsitellyt tapausesimerkit Anthemin ja Vastaamon tietomurroista havainnollistivat konkreettisesti, millaisia seurauksia puutteellisella tietoturvalla voi olla. Molemmissa tapauksissa olisi ollut mahdollista joko estää tietomurto kokonaan tai ainakin lieventää sen vaikutuksia, jos tietoturvaan olisi käytetty enemmän resursseja ja sitä olisi valvottu tarkemmin. Tapausesimerkit osoittavat, että tietomurtojen ehkäisemiseen ei välttämättä tarvita uutta teknologiaa vaan melko perinteisetkin tietoturvakäytännöt voivat olla ratkaisevia suurten tietomurtojen estämisessä.

# Lähdeluettelo


[1] W. A. Daden et al., "Cybersecurity Enabled Improved BigData Privacy Management Measures to Preserve Information With Privacy Concerns", teoksessa *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, IEEE, 2024, s. 1–7.

[2] A. H. Almulihi, F. Alassery, A. I. Khan, S. Shukla, B. K. Gupta ja R. Kumar, "Analyzing the Implications of Healthcare Data Breaches through Computational Technique.", *Intelligent Automation & Soft Computing*, vol. 32, nro 3, 2022.

[3] M. Aslam et al., "Getting smarter about smart cities: Improving data security and privacy through compliance", *Sensors*, vol. 22, nro 23, s. 9338, 2022.

[4] S. Dash, S. K. Shakyawar, M. Sharma ja S. Kaushik, "Big data in healthcare: management, analysis and future prospects", *Journal of big data*, vol. 6, nro 1, s. 1–5, 2019.

[5] A. Attaallah, H. Alsuhabi, S. Shukla, R. Kumar, B. K. Gupta ja R. A. Khan, "Analyzing the Big Data Security Through a Unified Decision-Making Approach.", *Intelligent Automation & Soft Computing*, vol. 32, nro 2, 2022.

[6] K. Batko ja A. Ślęzak, "The use of Big Data Analytics in healthcare", *Journal of big Data*, vol. 9, nro 1, s. 3, 2022.

[7] F. A. Alzahrani, M. Ahmad ja M. T. J. Ansari, "Towards design and development of security assessment framework for internet of medical things", *Applied Sciences*, vol. 12, nro 16, s. 8148, 2022.

[8] S. Armoogum ja P. Khonje, "Healthcare data storage options using cloud", teoksessa *The Fusion of internet of things, artificial intelligence, and cloud computing in health care*, Springer, 2021, s. 25–46.

[9] N. Thapliyal ja M. S. Gaur, "Security threats in healthcare big data: a comparative study", teoksessa *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, IEEE, 2023, s. 32–37.

[10] M. J. Samonte, L. M. Achacoso, A. C. Amper ja A. L. Eco, "An In-Depth Analysis on Mitigating Data Breaches in Healthcare Systems through Big Data Structure", teoksessa *Proceedings of the 2024 12th International Conference on Computer and Communications Management*, 2024, s. 139–147.

[11] E. Bertino, "Data security and privacy: Concepts, approaches, and research directions", teoksessa *2016 IEEE 40th annual computer software and applications conference (COMPSAC)*, IEEE, vol. 1, 2016, s. 400–407.

[12] G. Pender-Bey, "The parkerian hexad", *Information Security Program at Lewis University*, 2019.

[13] *Tietosuoja*, Luettu 10.1.2026. url: `https://tietosuoja.fi/tietosuoja`.

[14] R. N. Zaeem ja K. S. Barber, "The effect of the GDPR on privacy policies: Recent progress and future promise", *ACM Transactions on Management Information Systems (TMIS)*, vol. 12, nro 1, s. 1–20, 2020.

[15] *Henkilötietojen käsittelijät*, Luettu 10.1.2026. url: `https://tietosuoja.fi/henkilotietojen-kasittelijat`.

[16] *What is a data controller or a data processor?*, Luettu 10.1.2026. url: https://commission.europa.eu/law/law-topic/data-protection/rules-business-and-organisations/obligations/controllerprocessor/what-data-controller-or-data-processor_en.

[17] *Rekisteröidyn oikeudet tieteellisessä tutkimuksessa*, Luettu 10.1.2026. url: https://tietosuoja.fi/rekisteroidyn-oikeudet-tieteellisessa-tutkimuksessa.

[18] W. N. Price ja I. G. Cohen, "Privacy in the age of medical big data", *Nature medicine*, vol. 25, nro 1, s. 37–43, 2019.

[19] S. Alrobaian, S. Alshahrani ja A. Almaleh, "Cybersecurity awareness assessment among trainees of the technical and vocational training corporation", *Big Data and Cognitive Computing*, vol. 7, nro 2, s. 73, 2023.

[20] S. Alder, *Healthcare Data Breach Statistics*, lokakuu 2025. url: https://www.hipaajournal.com/healthcare-data-breach-statistics/.

[21] P. R. Vaka et al., "Anthem Health Insurance Breach or Ransomware Attacks", *International Scientific Journal of Contemporary Research in Engineering Science and Management*, vol. 2, nro 1, s. 41–49, 2017.

[22] N. Shankar ja Z. Mohammed, "Surviving data breaches: A multiple case study analysis", *Journal of Comparative International Management*, vol. 23, nro 1, s. 35–54, 2020.

[23] M. Kortesoja, "Tapaus Vastaamo: Symptomaattinen luenta potilastietosuojan murtumisen yhteiskunnallisista syistä ja seurauksista", *Tutkimus & kritiikki*, vol. 2, nro 1, s. 9–32, 2022.

[24] J. C. Looi et al., "Cybersecurity lessons from the Vastaamo psychotherapy data breach for psychiatrists and other mental healthcare providers", *Australasian Psychiatry*, vol. 33, nro 1, s. 106–110, 2025.

[25] J. Mäntysalo. ”Vastaamon ex-toimitusjohtajan Ville Tapion syyte kaatui”. Luettu 31.1.2026. url: https://yle.fi/a/74-20200364.

[26] M. Rautio. ”Aleksanteri Kivimäelle yli kuusi vuotta vankeutta Vastaamon tietomurrosta – Kivimäki pettynyt”. Luettu 31.1.2026. url: https://yle.fi/a/74-20084812.