LAUREA
AMMATTIKORKEAKOULU
Yhdessä enemmän

# Tiedon varmistaminen ja arkistointi

Enholm, Jesse

2016 Laurea Kerava

**Laurea-ammattikorkeakoulu**
Kerava

# Tiedon varmistaminen ja arkistointi

Jesse Enholm
Tietojenkäsittelyn koulutusohjelma
Opinnäytetyö
Joulukuu, 2016

**Laurea-ammattikorkeakoulu** **Tiivistelmä**
Kerava
Tietojenkäsittelyn koulutusohjelma

Jesse Enholm

**Tiedon varmistaminen ja arkistointi**

Vuosi 2016 Sivumäärä 32

Tämän opinnäytetyön tavoitteena oli perehtyä tiedon varmistamisen, arkistoinnin sekä tietoturvan yleisiin periaatteisiin julkishallinnon virastossa. Tavoitteena oli myös etsiä mahdollisia ongelmakohtia ja riskejä, jotka tulee ottaa huomioon varmistusjärjestelmän käyttöönotossa ja ylläpidossa. Työssä perehdytään tietoturvaan ja arkistointiin pääosin hallinnollisella tasolla. Tiedon varmistamisesta puhuttaessa tutustutaan laajasti erilaisiin käytössä oleviin teknologioihin ja erilaisiin tapoihin varmistaa dataa.

Tämä laadullinen tutkimus toimi osaamisen siirtona järjestelmän tuleville ylläpitäjille. Opinnäytetyöprosessin aikana varmistusjärjestelmän dokumentaatiota päivitettiin sekä viraston toimintatapoja varmistusmenettelyissä kehitettiin. Kehitystyö tapahtui yhteistyössä viraston muiden IT-asiantuntijoiden, sekä uusien ylläpitäjien kanssa.

Prosessin aikana olemassa olevasta varmistusjärjestelmästä löytyi useita parannuskohteita. Parannuskohteiden lisäksi järjestelmän toimintaa oli mahdollista kehittää vikasietoisemmaksi sekä varmistustöiden läpimenoaikoja mahdollista lyhentää jopa puoleen entisestä. Varmistusjärjestelmässä ei ollut yleiseen teoriapohjaan viitaten suuria puutteita. Puutteet liittyivät suureksi osaksi ylläpitäjien käytännön tason prosesseihin. Varmistetun datan palautusta ei ollut testattu riittävän usein, sekä salaamatonta varmistusdataa on saatettu säilyttää lukitsemattomassa tilassa nauhanvaihtoprosessin välivaiheissa.

Järjestelmän jatkokehitysehdotuksia käytiin myös läpi. Dokumentaation kartoituksessa kävi selväksi, että yhdistämällä Linux-varmistukset olemassa olevaan keskitettyyn hallintaan saisi siitä muokattua todella hyvin toimivan kokonaisuuden, jonka toimintatapa olisi looginen ja helppokäyttöinen ylläpitäjien näkökulmasta.

Asiasanat: varmuuskopiointi, tietoturva, arkistointi

**Laurea University of Applied Sciences** **Abstract**
Kerava
Bachelor's Degree Programme in Business Information Technology

Jesse Enholm

**Data backups and archiving**

Year 2016 Pages 32

The purpose of this thesis was to study how the basic principles of data backups, archiving and information security are implemented in a government agency. The aim was also to search for possible problems and risks involved in deploying a backup system for the organization. Archiving and information security were examined mostly on a management level. However, backup systems were given a closer look on a more technical level.

This qualitative research includes the transfer of know-how knowledge to future system administrators of the backup system. During this thesis the documentation of the backup system was updated and management processes were improved.

The research revealed many possible improvements for the backup system in use. In addition, tolerance for technical faults was improved and the time for backup jobs to complete was shortened. Most of the shortcomings in the backup system revolved around administrator processes: The reliability of backups was not tested often enough; and non-encrypted tapes may also have been accessible to other personnel during monthly tape changes.

Future plans for improvement were reviewed. After the documentation of the system was unified, it became clear that more improved synergy benefits could be achieved by adding UNIX backups to the centrally managed backups. The system would be more logical and easier from the management's perspective.

Keywords: backups, information security, archiving

LYHENTEET

| | |
|---|---|
| D2D | Disk-to-Disk |
| D2D2T | Disk-to-Disk-to-Tape |
| DR | Disaster Recovery |
| LTO | Linear Tape Open |
| NAS | Network-attached Storage |
| RAID | Redundant Array of Independent Disks |
| SAN | Storage Area Network |
| SLA | Service Level Agreement |
| SSD | Solid-state Drive |
| SSH | Secure Shell |
| VTL | Virtual Tape Library |

Sisällysluettelo

# 1 Johdanto

Tämä opinnäytetyö käsittelee tiedon varmistamista ja arkistoimista osana organisaation jatkuvuussuunnitelmaa ja tietoturvasuunnitelmaa. Tietoa on mahdollista varmistaa hyvin monilla eri tavoilla, jolloin tärkeäksi aspektiksi nousee myös organisaation oma tietoturvapolitiikka, lait sekä muut poliittiset linjaukset.

Tutkimuksessa käydään läpi kaikki perusaspektit varmistusjärjestelmän ja organisaation tietoturvan suunnittelussa ja kehityksessä niin kattavasti, että perustasoisilla IT-taidoilla tätä opinnäytetyötä on mahdollista hyödyntää organisaation varmistusjärjestelmän suunnittelussa ja kehittämisessä.

Tiedon varmistaminen on osa organisaation tietoturvaa. Tiedon varmistamisella taataan tiedon saatavuus esimerkiksi katastrofitilanteessa, jossa käytössä olevaa tietoa menetetään. Oikein mitoitetuilla varmistusmenettelyillä voidaan taata organisaation toiminnan jatkuvuus erilaisissa virhe- tai katastrofitilanteissa.

Ehkä kattavin ja samalla ajankohtaisin suomeksi varmistusjärjestelmiä yleisesti käsittelevä työ on Tuomas Määtän vuonna 2013 kirjoittama diplomityö ”Palvelinten varmistusjärjestelmät”. Hänen diplomityönsä on kuitenkin teknisempi eikä se käsittele tietoturvaa yhtä yleisellä tasolla kuin tämä tutkimus.

Tutkimuksen alkuun on lisätty lyhenneluettelo, sillä puhuttaessa tiedon varmistamisesta käytetään paljon spesifistä sanastoa sekä lyhenteitä. Lyhenteet on listattu ennen sisällysluettteloa.

# 2 Työn lähtökohdat

Erilaisten organisaatioiden varmistuskäytännöistä on kirjoitettu hyvin vähän suomenkielistä materiaalia. Tällöin tutkimuksen rajaaminen on erityisen tärkeää, jotta aihealue ei ole liian laaja. Kirjoitettavaa aiheesta löytyisi paljon. Toinen rajaava tekijä on aiheen tunnettuus; aiheesta tulee kirjoittaa perusteista lähtien, sillä aihe on monille lähes tuntematon.

Tutkimus toteutettiin Säteilyturvakeskukselle, missä on käytössä HP Data Protector -varmistusratkaisu. Varmistusjärjestelmä koostuu HP MSL4048 -nauhakirjastosta, sekä HP StoreOnce 4430 -varmistuslaitteesta. Opinnäytetyöprosessin alkaessa järjestelmä oli ollut käytössä noin vuoden. Järjestelmän elinkaareksi arvioitiin käyttöönoton aikana 4-6 vuotta. Kehitystyöllä olisi myös mahdollista venyttää järjestelmän elinkaarta pidemmäksi optimoimalla prosesseja sekä sovelluspuolen asetuksia.

Tätä tutkimusta on mahdollista soveltaa sekä julkishallinnon organisaatioissa, että yksityisellä sektorilla. Puhuttaessa yleisesti organisaation tiedon varmistamisesta ja arkistoinnista pätevät pääosin samat nyrkkisäännöt. Mielessä pidettäviä asioita ovat liikesalaisuudet, sekä ministeriöiden laatimat asetukset koskien tiedon säilyttämistä. Yksityisellä sektorilla datan säilyvyys on erittäin tärkeää esimerkiksi big data –keskeisissä yrityksissä. Julkishallinnon näkökulmasta toimintaa määrittelee suurissa osin myös lait.

# 3 Tutkimusmenetelmät

Tutkimusmenetelmänä tässä opinnäytetyössä käytin kvalitatiivista, eli laadullista tutkimusta. Jo työn aiheesta johtuen kvantitatiivinen tutkimusmenetelmä ei ollut pätevä vaihtoehto. Otanta opinnäytetyössä on pieni, mikä mahdollistaa erinomaisen perehtymisen aiheeseen. Opinnäytetyön yksi käyttötarkoitus onkin osaamisen siirto asiantuntijoille, jotka tulevat ylläpitämään järjestelmää tulevaisuudessa.

Kvantitatiivinen tutkimus nojaa vahvasti tilastotieteeseen ja analyyseihin ja siinä käytetään paljon aineistoa, joka on yleensä numeerisessa muodossa. Määrällinen tutkimus on hyvin rajattu ja kurinalainen verrattuna vapaamuotoisempaan kvalitatiiviseen tutkimukseen, ja se sisältää aina numeraalisen matriisin, johon aineisto on tiivistetty. (Tilastokeskus 2016.)

Kvalitatiivisen tutkimuksen heikkoutena voi pitää sen tulkinnanvaraisuutta, sekä mahdollisesti subjektiivista näkemystä aiheeseen. Tätä subjektiivista näkemystä tulee mahdollisuuksien mukaan välttää; tutkijan ei tule sekoittaa omia uskomuksiaan, asenteitaan tai arvostuksiaan tutkimuskohteeseen. (Taloustutkimus 2016.)

Kvalitatiivinen tutkimus perustuu laatuun. Laadulla tarkoitetaan tässä tutkimusstrategiassa sitä, että otanta ei ole suuri, mutta pienellä otannalla päästään syvällisempiin tuloksiin ja pystytään perehtymään paremmin aiheeseen. Kvalitatiivinen tutkimus sopii hyvin erikoisempiin tutkimuskohteisiin, joissa kvantitatiivisen tutkimusstrategian käyttäminen ei ole mahdollista, sillä tarpeeksi suurta otantaa ei ole mahdollista kerätä. (Saaranen-Kauppinen & Puusniekka, 2006.)

Huomattava ero edellä mainituissa tutkimusmenetelmissä on niiden erilainen rakenne ja sen muuttuminen. Määrällisessä tutkimuksessa on vaiheita, joiden jälkeen edeltävään pisteeseen ei ole mahdollista palata. Laadullisessa tutkimuksessa tällaisia kriittisiä pisteitä ei ole, vaan tulkinta jakaantuu koko tutkimusprosessin ajalle. (Tilastokeskus 2016.)

Tutkimuksen ulkopuolelle jätettiin asiakasviraston UNIX-varmistusten kehittäminen, sillä niiden kehitysprojekti tämän opinnäytetyön rinnalla muun kehitystyön ohessa koettiin liian työlääksi. Pelkona oli myös, että opinnäytetyöstä tulisi liian laaja.

## 3.1 Validiteetti ja reliabiliteetti

Tutkimuksen validiteetilla tarkoitetaan tutkimuksen kykyä mitata sitä mihin se on tarkoitettu. Virheellinen tutkimusasetelma voi vaarantaa tutkimuksen validiteetin. Esimerkiksi väärin ajoitettu haastattelu tai haastattelukysymysten väärin muotoilu voivat johtaa harhaan tutkimustulosten tulkinnassa. (KvantiMOTV 2016.)

Reliabiliteetilla tarkoitetaan tutkimuksen luotettavuutta. Luotettavuudella tarkoitetaan sitä, että tutkimusmenetelmät tuottavat ei-sattumanvaraisia tuloksia, eli tulokset ovat toistettavissa. Tutkimuksen reliabiliteetti selviää vertailemalla eri ajankohtana suoritettuja mittauksia. (VirtuaaliAMK 2016.)

## 3.2 Tutkimuskysymykset

Haen opinnäytetyössäni vastausta kahteen tutkimuskysymykseen: minkälaisia ovat hyväksi todetut varmistus- ja arkistointikäytännöt julkishallinnon organisaatiossa? Miten nämä hyväksi todetut käytännöt toteutuvat asiakasvirastossa?

Tutkimuskysymykset valittiin, koska varmistusjärjestelmä oli ollut sopivan aikaa käytössä, mutta sitä ei ollut auditoitu eikä dokumentoitu kunnolla. Järjestelmän yleisestä terveydentilasta ja toimivuudesta ei ollut tarpeeksi näyttöä. Opinnäytetyöprosessin kautta näiden asioiden selvittäminen tuntui toimivalta ratkaisulta.

# 4 Tietoturva

Tietoturva on hyvin laaja käsite. Seuraavassa alaluvussa tarkastellaan lähemmin mihin kolmeen osa-alueeseen tietoturvan eri osa-alueet yleisesti jaotellaan. Voidaan todeta, että ”tiedon käytön mahdollistaminen ja turvaaminen on tietoturvan tärkeimpiä ominaisuuksia.” (Paavilainen 1998, 7).

Tietoturvasta voidaan eritellä omaksi alueekseen myös tietosuoja. Tietosuojasta puhuttaessa tarkoitetaan sellaisen datan käsittelyä, joka sisältää henkilötietoja. Tietosuoja turvaa siis henkilötietojen luottamuksellisuuden ja estää valtuudettomilta tahoilta pääsyn tietoon. (Rousku 2014, 52.)

## 4.1 Tietoturvan osat

Tietoturvan kolme pääosaa ovat tiedon luottamuksellisuus (confidentiality), eheys (integrity), sekä saatavuus (availability). Edellä mainituille hyvä muistisääntö onkin englanninkielinen lyhenne CIA. Joskus saatavuudesta käytetään myös termiä ”käytettävyys”. Käytettävyys sekoitetaan helposti käytettävyyteen (usability) esimerkiksi käyttöliittymäsuunnittelussa, joten käytän opinnäytetyössäni termiä saatavuus. (Chia 2012.)

Luottamuksellisuus, eheys ja saatavuus ovat elinehto tiedolle. Jos edellä mainitut asiat eivät toteudu, ei tietoa yksinkertaisesti voi käyttää. Jos et voi luottaa tietoon tai sen eheyteen tai tieto on saavuttamattomissa, on tieto käyttökelvotonta.

### 4.1.1 Luottamuksellisuus

Luottamus voidaan mieltää jopa tärkeimmäksi edellä mainituista kolmesta osasta, sillä luottamuksellisuudesta on kirjoitettu sekä henkilörekisterilaissa että rikoslaissa, muun muassa yrityssalaisuudet, yksityishenkilön potilastiedot sekä julkisuuslaki. Näin ollen esimerkiksi potilastietojen vuotaminen voi olla rikos joka rikkoo samalla tiedon luottamuksellisuuden. (Paavilainen 1998, 10.)

Julkishallinnossa tietoaineistojen käsittelyyn sovelletaan Valtionhallinnon tieto- ja kyberturvallisuuden ohjausryhmän (VAHTI) ohjeistusta. Julkishallinnon tietoturvasta vastaavien henkilöiden onkin hyvä tunnistaa ja tuntea riskit ja velvollisuutensa koskien asiakirjojen suojaustasoja. (Aho 2010.)

Vahti-ohjeistuksessa kuvataan neljä eri suojaustasoluokitusta. Julkisuuslaki asettaa viranomaisille velvoitteet käytössään olevien tietovarantojen käsittelemisestä hyvän tiedonhallintatavan mukaisesti (Aho 2010). Erilaiset suojaustasot on kuvattu taulukossa: Taulukko 1.

| **SUOJAUSTASO** | **Turvallisuusluokittelun nimi** | **Lyhenne** |
|---|---|---|
| **Suojaustaso I (ST I)** | Erittäin salainen | ERSAL (E) |
| **Suojaustaso II (ST II)** | Salainen | SAL (S) |
| **Suojaustaso III (ST III)** | Luottamuksellinen | LUOT (L) |
| **Suojaustaso IV (ST IV)** | Käyttö rajoitettu | RAJ (R) |

Taulukko 1: Tietoaineistojen luokittelu

### 4.1.2 Eheys

Tiedon eheydellä tarkoitetaan sitä, että tieto on totuudenmukaista ja muuttumatonta. Eheyteen voi vaikuttaa sekä inhimilliset virheet, että erilaiset virheet järjestelmän suunnittelussa tai toiminnassa. Paavilainen on jakanut kirjassaan eheyden vielä kolmeen eri osaan, joilla havainnollistetaan eheyden eri osa-alueita. Osa-alueita ovat:

Alkuperäisyys – onko tiedon alkuperä mikä luulemme sen olevan?
Koskemattomuus – onko tietoa muutettu jossain vaiheessa?
Kiistämättömyys – onko tieto juuri sitä mitä oletamme sen olevan?
(Paavilainen 1998, 10.)

### 4.1.3 Saatavuus

Saatavuus pitää sisällään monia eri näkökulmia. Aihetta onkin helpoin avata esimerkin tavoin. Yrityksellä on omat intranet-sivut, jonka tietoihin pääsee käsiksi vain yrityksen omat todennetut työntekijät. Saatavuus on siis taattu työntekijöille, joiden tulee pystyä tunnistautumaan järjestelmälle. Tämän jälkeen järjestelmä tarkistaa, onko työntekijällä oikeus kyseisen sisällön katsomiseen. Jos henkilöllä ei ole oikeuksia, ei intranet-sivu ole hänen saatavilla. Tämä voi johtua joko järjestelmävirheestä, käyttöoikeuksien puutteesta tai esimerkiksi meneillään olevasta järjestelmän huoltotoimenpiteestä. (Chia 2010.)

Tiedolla on arvoa vain, jos oikeilla henkilöillä on mahdollisuus päästä tietoon käsiksi oikeaaikaisesti. Saatavuuden estämisestä, toisin sanottuna palvelunestohyökkäyksistä on tullut yleinen hyökkäystapa nykypäivänä. (Chia 2010.)

Kaksi nopeasti suosiota kasvattanutta ratkaisua tiedon saatavuuden parantamiseen ovat SAN-tallennuslaitteet (Storage Area Network), sekä NAS-laitteet (Network-Attached Storage). Saatavuutta mitataan usein kahdella eri tavalla. Ensimmäinen tapa on, kuinka usein data on saatavilla. Tässä tapauksessa käytetään yleisesti prosenttimääritettä. Toinen tapa mitata saatavuutta on, kuinka paljon dataa on mahdollista siirtää kerralla. (TechTarget 2005.)

## 4.2 Tietoturvapolitiikka

Tietoturvapolitiikalla (engl. Information Security Policy) tarkoitetaan organisaation yleisiä käytäntöjä sekä oikeanlaista tiedon ja tietovarantojen käyttämistä. Politiikka ei ota kantaa miten viraston sovelluksia tai työvälineitä tulee käyttää. Samalla politiikka turvaa yrityksen työntekijän suorittamien laittomuuksien tai virheiden osalta. (Whitman 2014.)

Tietoturvapolitiikan tulisi ottaa kantaa tietoturvaan liittyviin rooleihin ja ohjeistukseen sekä poikkeusten käsittelyyn. Politiikan tulee myös kertoa säännöt ja ohjeistukset, joita käyttäjien toivotaan noudattavan, sekä mahdolliset sanktiot ohjeistuksen rikkomisesta tai sen noudattamatta jättämisestä. (Bowen, Hash & Wilson 2006.)

### 4.3 Riskienhallinta ja tietoturva

IT-asiantuntijan on tärkeä ymmärtää ero riskienhallinnan ja tietoturvan välillä. Riskienhallinnan voi helposti ymmärtää osana tietoturvaa. Riskienhallinnassa arvioidaan riskien todennäköisyyttä ja taloudellisia menetyksiä niiden tapahtuessa. Näin ei kuitenkaan ole tietoturvassa, jossa tapahtumien todennäköisyyksiä on paljon vaikeampi arvioida.

Kuinka todennäköistä on, että sähköpostisi murretaan ja kolmas osapuoli saa pääsyn kaikkiin sähköposteihisi? Mikä on tämän tapahtuman rahallinen vaikutus? Nämä ovat asioita joita on erittäin vaikea arvioida, sillä tapahtumat ovat hyvin harvinaisia. Riskienhallinnassa taas todennäköisyydet ovat suuremmat, jolloin erilaisten tapahtumien vaikutuksia on helpompi arvioida. (Martin 2013.)

## 5 Tiedon varmistaminen

Tiedon varmistamista ei tule sekoittaa tiedon arkistointiin. Tietoa kuitenkin voidaan varmistaa sekä nauhoille että muille medioille. Magneettinauhat ovat selvästi paras vaihtoehto, jos halutaan arkistoida paljon tietoa pitkäksi aikaa. Varmistamisessa taas varaudutaan lähitulevaisuudessa tapahtuviin tietohävikkeihin. Käyttäjä voi esimerkiksi tuhota vahingossa kansion verkkolevyltä tai palvelin hajoaa. Tällöin data voidaan nopeasti palauttaa vaikkapa D2D-laitteelta suoraan. Jos data on vain arkistoidussa muodossa, vie sen palauttaminen merkittävästi enemmän aikaa.

Magneettinauhoja käytetään myös datan varmistamiseen niiden suuren tallennuskapasiteetin takia. Usein onkin käytössä jonkinlainen hybridimalli, jossa hyödynnetään sekä virtuaalinauhakirjastoja, että perinteisiä fyysisiä nauhakirjastoja. Hybridimallissa voidaan varmistaa tieto alustavasti D2D-laitteelle, josta se kopioidaan tasaisin väliajoin varmistusnauhoille arkistoitavaksi. (Techopedia 2016.)

### 5.1 Varmistusmenetelmät

Yleisimmin käytettyjä varmistusmenetelmiä on kolme. Täysi varmistus (Full backup) sisältää kaiken varmistettavan datan. Differentiaalinen varmistus (Differential backup) sisältää tiedot,

jotka ovat muuttuneet verrattuna edelliseen täyteen varmistukseen. Kolmas tapa on inkrementaalinen backup (Incremental backup), joka sisältää tiedot jotka ovat muuttuneet edellisestä varmistuksesta riippumatta siitä onko edellinen varmistus täysi vai differentiaalinen varmistus. (Types of Backup 2016.)

Täysi varmistus on pakollinen ja tehdään normaalisti yhdestä neljään viikon välein. Tämän lisäksi käytetään vaihtelevasti sekä differentiaalisia varmistuksia että inkrementaalisia varmistuksia. Erilaisten varmistusmenetelmien hyvät ja huonot puolet on kuvattu liitteessä yksi.

Varmistusten suunnitteluvaiheessa tulee kiinnittää huomiota siihen, minkälaista dataa säilötään, paljonko datan määrän arvioidaan kasvavan tulevaisuudessa, kuinka kauan datan tulee olla palautettavissa lyhyellä palautumisajalla ja kuinka arkistointi suunnitellaan toteutettavan.

## 5.2 Varmistusmediat

Tiedon varmistamiseen voidaan käyttää hyvin erilaisia medioita. Varmistusmedioiden säilyvyydessä ja niiden toiminnallisuuksissa on paljon eroja ja ne soveltuvat moniin erilaisiin varmistustapoihin. Tämä opinnäytetyö käsittelee myös lyhyesti varmistusmetodit, joita voidaan käyttää pienen mittakaavan organisaatioissa tai henkilökohtaisiin varmistuksiin.

Suurissa yrityksissä suuri datamäärä aiheuttaa ajallisia haasteita varmistusten suorittamisessa. Suuren datamäärän siirtäminen vaatii paljon organisaation tietoliikenneyhteyksiltä. Nykypäivänä suuret organisaatiot suosivat nauhajärjestelmiä sekä pilvivarmistuspalveluja. Jotkin organisaatiot säilyttävät omat nauhavarmistusjärjestelmänsä arkistointikäyttöä varten vähäisten ylläpitokustannusten takia. (Erickson 2012.)

### 5.2.1 D2D (Disk-to-Disk) varmistus

Levypohjaisten varmistuslaitteiden käyttö on yleistynyt jatkuvasti vuodesta 2005 lähtien. Vaikkakin edelleen suositaan paljon perinteisiä nauhakirjastoja, ovat D2D-ratkaisut valtaamassa tilaa perinteisiltä varmistusratkaisuilta (Types of Backup 2016).

D2D-laitteella tarkoitetaan varmistuksia varten räätälöityä levyjärjestelmää. Kapasiteetti tämmöisessä laitteessa voi olla nykypäivänä muutaman teratavun (TB) laitteista jopa satoihin teratavuihin. Perinteisten nauhojen sijaan laitteissa on joko mekaanisia kovalevyjä tai SSD-levyjä (Solid-state drive). Laitetta voidaan hallita joko käyttökonsolista, verkkosivulta, SSH:n (Secure shell) ylitse tai Telnetillä.

Laitteeseen luodaan virtuaalisia nauhakirjastoja (Virtual Tape Library, VTL). Nämä virtuaaliset nauhakirjastot ovat toiminnallisuuksiltaan täysin yhteneviä oikeiden nauhakirjastojen kanssa.

D2D-varmistuksella tarkoitetaan varmistusta, joka tehdään esimerkiksi tiedostopalvelimen kovalevyltä D2D-laitteen kovalevylle. Teknologia on vallannut lähivuosina markkinoita perinteisiltä nauhavarmistuksilta, kun kovalevyjen koot ovat kasvaneet ja hinnat laskeneet. D2D:n vahvuuksia verrattuna nauhavarmistukseen:

- Mahdollista kirjoittaa ja lukea samanaikaisesti monia eri medioita, sillä nauhurit ovat virtuaalisia. Nauhureita voidaan siis luoda lähes määrättömästi.
- Palautuksien tekeminen on nopeampaa, sillä data voidaan hakea satunnaishakuina suoraan kovalevyltä, eikä laitteen tarvitse lukea koko nauhaa lävitse.
- Data on koko ajan laitteessa. Tällöin ylläpitäjän ei tarvitse etsiä nauhaa arkistosta ja syöttää sitä laitteelle luettavaksi, jos hän haluaa tehdä palautuksen.
- Nykyaikaisissa laitteissa on kehittyneet deduplikointiominaisuudet. (Brewerton, 2010.)

Deduplikointi tapahtuu D2D-laitteessa niin sanottujen siivousikkunoiden (Housekeeping) aikana. Tällöin laite vertailee levyllä olevia blockeja keskenään. Jos samanlaisia blockeja löytyy, poistetaan duplikaatit ja luodaan viittaus ainoastaan yhteen tällaiseen blockiin. Näin ollen sama tieto ei löydy monesta eri paikasta useilta levyiltä. Varmistettava data on usein muuttumatonta. Samanlaisia blockeja voi siis löytyä kymmeniä tai satoja, kun esimerkiksi tiedostopalvelimesta otetaan kahden viikon välein täysi varmistus. Siivoamalla duplikaatit laitteen deduplikointisuhde (deduplication ratio) voi olla jopa 1:15. 10 Teratavun järjestelmään mahtuisi siis 150 teratavua varmistettavaa dataa. Tätä tapaa kutsutaan Target-based deduplikoinniksi. (Keegan 2012.)

Deduplikointi voidaan myös suorittaa jo ennen datan siirtoa varmistuslaitteelle. Tätä menetelmää kutsutaan Source-side deduplikoinniksi. Lähettävän laitteen deduplikoinnissa on monia etuja verrattuna Target-side-deduplikointiin.

- Lähetettävän datan määrä laskee merkittävästi
- Vähemmän rasitusta verkkoyhteyksille
- Nopeammat varmistussessiot pienemmän datamäärän ansiosta
- Varmistettaessa kolmannen osapuolen pilvipalveluun mahdollisuus kustannussäästöihin, jos laskutus tapahtuu siirretyn datan perusteella.

Lähettävässä palvelimessa suoritettu deduplikointi vie kuitenkin paljon resursseja. Koska Target-side deduplikoinnissa työn tekee siihen suunniteltu laite, on itse deduplikointiprosessi paremmin optimoitu ja nopeampi tehdä kuin lähettävässä laitteessa. Palvelimen käytettävyys voi siis laskea, jos deduplikointi suoritetaan jo ennen datan siirtoa D2D-laitteeseen. (Moir 2010.)

Jotkut valmistajat tarjoavat nykyään myös hybridimalleja, mutta tämänlaiset ratkaisut ovat alkaneet yleistyä vasta viimeisen parin vuoden aikana. Tällaisessa toimintatavassa hallinta tapahtuu yhdestä räätälöidystä pääkonsolista, joka hallitsee koko varmistusympäristöä. (Keegan 2013.)

### 5.2.2 Nauhavarmistus

Nauhavarmistusten vahvuutena on nopea kirjoitusnopeus, hyvät arkistointiominaisuudet ja hyvä hinta-kapasiteetti-suhde. LTO7 (Linear Tape-Open) –nauhojen natiivikapasiteetti on 6 teratavua. LTO Ultrium –yhtymän etenemissuunnitelman mukaan yhdeksännen sukupolven LTO-nauhojen tavoitekapasiteetiksi ilmoitetaan 62,5TB ja LTO10-nauhoille jopa 120TB kapasiteettia. (LTO Ultrium 2015.)

Nauhoille varmistamisessa on myös huonot puolensa. Nauhurien teknisestä toteutuksesta johtuen varmistaminen voi olla myös hidasta. Nauhuri vaatii tarpeeksi suuren syöttökaistan, jotta nauha pyörii jatkuvasti ja sille voidaan kirjoittaa. Jos dataa ei syötetä nauhurille tarpeeksi nopeasti, joudutaan nauha pysäyttämään hetkittäin. Tämä kuluttaa sekä nauhaa että hidastaa varmistamista, kun nauha pysähtyy lyhyin väliajoin. Käytännössä tämä vaatii, että laite on suoraan kuituyhteydellä (Fibre channel – FC) yhteydessä varmistuspalvelimeen. Suuremmassa ympäristössä vaaditaan kuitukytkin, jos halutaan yhdistää useampia palvelimia kuituyhteydellä suoraan nauhakirjastoon. Tämä taas nostaa nopeasti varmistusjärjestelmän hintaa drastisesti. (Manes 2012.)

Nauhateknologioiden toinen huono puoli on niiden toimintavarmuus. Sekä nauhakirjastot että itse varmistusnauhat ovat mekaanisia. Nauhakirjasto liikuttaa nauhoja mekaanisesti sisällään. Kun nauhoja liikutetaan satoja ja tuhansia kertoja pois kaseteista nauhuriin ja takaisin, syntyy kulumia ja osia voi hajota. Vikasietoisuus ei siis ole nauhavarmistusten vahvoja puolia. Tästä syystä varmistusnauhoille määritellään yleensä jo varmistussovelluksessa tietty määrä ylikirjoituksia, jonka jälkeen nauha saa merkinnän ”poor”. Tämä tarkoittaa sitä, että nauhalle ei enää kirjoiteta dataa. (Lock 2010.)

Myös nauhojen hallinta on hidasta. Jokainen nauha otetaan aina erilliseen käsittelyyn, ja pelkästään nauhan hakeminen kasetista nauhuriin voi kestää useita minuutteja. Tämän jälkeen

nauhaa käsitellään: skannataan, luetaan tai formatoidaan. Kaikki nämä toiminnot voivat viedä useita minuutteja, D2D-laitteella samat toiminnot suoritetaan sekunneissa. Myös nauhojen fyysinen käsittely on tietoturvariski; käsiteltäessä kryptaamattomia nauhoja tuleekin noudattaa erityistä varovaisuutta. Jos kolmas osapuoli saa nauhan hallintaansa, on tieto suoraan luettavissa nauhalta. (Lock 2010.)

Nauhateknologiat kehittyvät edelleen muun teknologian mukana. Ei ole siis syytä uskoa, että nauhateknologia olisi vanhentumassa vielä lähivuosina, vaan sille löytyy edelleen oma käyttötarkoituksensa tiedon varmistamiseen ja erityisesti arkistointiin. (Lock 2010.)

### 5.2.3 Muut varmistusmediat

Pilvipalvelut ovat käytännössä D2D-varmistuksia palveluntarjoajan ylläpitämässä pilvipalvelussa. Tällöin data säilötään suuriin konesaleihin, joista ostetaan kapasiteettia tarpeen mukaan. Pilvipalvelujen ehdottomia vahvuuksia ovat halpa hinta ja hyvä skaalautuvuus eri käyttötarkoituksiin. Puhuttaessa suurien datamäärien varmistamisesta ulkopuoliselle taholle on kuitenkin ongelmana esimerkiksi organisaation kaistanleveys ulospäin, jos varmistettavaa dataa on paljon.

Toinen suuri ongelma pilvipalvelujen käyttämisessä on tietoturva-aspekti. Valtiovarainministeriö on tehnyt ohjeistuksen organisaatioille koskien pilvipalveluiden tietoturvaa. (Valtiovarainministeriö 2010.)

Hyvin pienissä organisaatioissa tietoa voidaan varmistaa myös DVD tai Blu-Ray-levyille. Näiden vaihtoehtojen hyvänä puolena on halpa hinta ja helppokäyttöisyys, kun erillisiä varmistuslaitteita ei tarvita lukuun ottamatta DVD- tai Blu-Ray-asemaa. Huonoina puolina taas on erittäin rajallinen kapasiteetti, DVD-levyillä noin 4,7GB ja Blu-ray-levyillä 24-54GB. (Types of Backup 2016.)

Ulkoiset kovalevyt voivat toimia myös varmistusmediana. Kovalevyille pystyy hintaan suhteutettuna tallentamaan paljon tietoa. Nykypäivänä pystyy ostamaan jo usean teratavun levyjä kuluttajahinnalla. Ongelmana on kuitenkin kovalevyjen kestävyys pitkällä aikavälillä. Ne rikkoutuvat herkästi, varsinkin väärin varastoituna. Tällöin on mahdollista menettää kaikki kovalevylle tallennettu data. Tästä syystä perinteisiä kovalevyjä ei tulisi missään nimessä käyttää pitempiaikaiseen arkistointiin. (Yurin 2016.)

NAS-verkkotallennus on varteenotettava vaihtoehto pienissä ja keskisuurissa organisaatioissa. NAS-palvelin on tavallisen pöytäkoneen kokoinen laite, jossa on valmistajan asentama käyttöjärjestelmä. Laite voidaan kytkeä yritysverkkoon, jossa sille voidaan tehdä hakemistoja kuin mille tahansa kovalevylle. Kovalevyjä NAS:ssa voi olla kahdesta jopa kahteentoista.

NAS:ssa on ominaisuuksia, jonka takia se soveltuu hyvin varmuuskopiointiin. Tyypillisesti kovalevyjä on useita, jolloin voidaan käyttää erilaisia RAID-kokoonpanoja. Jo kolmen kovalevyn kokoonpanolla päästään RAID5-tasoon, joka on sekä vikasietoinen että nopea lukea. Pakasta voi hajota yksi levy ilman että tietoa menetetään. Kun rikkoutunut kovalevy vaihdetaan uuteen, korjaa laite automaattisesti tiedostojärjestelmän. (Yurin 2016.)

Pitkäaikaiseen arkistointiin NAS ei kuitenkaan sovellu, vaan tulisi käyttää esimerkiksi Blu-ray-levyjä tai paremmin arkistointiin soveltuvaa nauhavarmistusratkaisua. NAS-kovalevyt eivät ole pitkäaikaisessa arkistoinnissa toimiva ratkaisu lyhytikäisyytensä vuoksi. (Types of Backup 2016).

### 5.3 Suojausajat

Varmistuksia ja arkistointia suunnitellessa on tärkeää kiinnittää huomiota datan suojausaikoihin. Suojausaika varmistaa, että mediaa ei ylikirjoiteta niin kauan, kun sen tiedetään sisältävän suojattua tietoa.

Suojausaikaan tulee kiinnittää erityistä huomiota, jos suoritetaan D2D2T –tyyppisiä varmistuksia ja nauhoja tai D2D:llä sijaitsevaa dataa ylikirjoitetaan usein. Suojausaika voi olla tyypillisesti noin kaksi viikkoa D2D-laitteella, jonka jälkeen se kopioidaan nauhalle ja suojausaika muutetaan organisaation varmistuspolitiikan mukaiseksi. Tämän jälkeen media on valmis arkistoitavaksi. (HPE 2012.)

Vaihtoehtoisesti taas suojausaika voi olla hyvinkin lyhyt. Esimerkiksi päivittäisiä varmistuksia voidaan ajaa vuorotellen 7 eri nauhalle, jolloin joka päivä yhden nauhan suojausaika umpeutuu, ja seuraava varmistustyö ajetaan tälle kyseiselle medialle. Näin saadaan jaettua ylikirjoitusprosessi seitsemälle eri nauhalle. Kerran viikossa yhdeltä nauhalta voidaan ajaa kopio arkistoitavalle nauhalle. (HPE 2012.)

Joissain varmistussovelluksissa, muun muassa HP Data Protectorissa varmistustyölle voidaan antaa myös katalogin suojausaika (Catalogue protection time). Katalogin suojausajalla tarkoitetaan tietoa, kuinka kauan nauhan indeksitietoa säilytetään varmistusjärjestelmän omassa sisäisessä tietokannassa (Internal Database). Suuren datamäärän ympäristöissä katalogitietoja

ei voida säilöä loputtomasti, sillä se kasvattaa sisäisen tietokannan kokoa, eritoten jos kantaan syötetään otsikkotiedon lisäksi myös tarkat tiedostonimet. Tässä tapauksessa ongelmaksi muotoutuu sisäisen tietokannan hallitsematon kasvaminen. Esimerkiksi palautustyö tämmöisessä tilanteessa on ongelmallista, kun kansiorakenteen muodostamisessa voi kestää useita minuutteja etsittäessä tiettyä hakemistoa palvelimen hakemistorakenteesta. (HPE 2012.)

Edellä mainitun tapaisissa tilanteissa voidaan joutua siis tilanteeseen, jossa data löytyy nauhoilta, mutta varmistussovellus ei tiedä mitä nauha sisältää. Jokainen nauha sisältää kuitenkin oman katalogitietonsa. Tämän avulla katalogitieto voidaan luoda uudestaan tuomalla (engl. import) nauha takaisin kirjastoon arkistosta. Tämän jälkeen palautustyö on mahdollista suorittaa. (HPE 2012.)

### 5.4 Replikointi

Replikointi ja levypeilaus (disk mirroring) ovat käytännöllisesti katsoen hyvin lähellä toisiaan. Puhuttaessa replikoinnista tarkoitetaan perinteisesti verkon yli tapahtuvaa tiedon kopiointia, kun taas levypeilauksella tarkoitetaan RAID-kokoonpanojen suorittamaa lyhyen matkan kopiointia. (Vangie 2016.)

Replikoinnilla tarkoitetaan toimintoa, jossa data kopioidaan muuttumattomana kahteen tai useampaan paikkaan. Tämä on mahdollista esimerkiksi kahdella maantieteellisesti eri paikoissa sijaitsevilla D2D-laitteilla.

Synkronisella replikoinnilla tarkoitetaan mallia, jossa kirjoitetaan data samanaikaisesti primääriin sekä sekundääriseen sijaintiin. Synkronisen replikoinnin vahvuutena on pieni virhemarginaali. Se onkin suositeltava tapa tilanteessa, jossa dataa ei saada missään nimessä menettää. Tämä toimintatapa vaatii kuitenkin, että replikointikohde on joko samassa tilassa tai hyvin lähellä primääriä sijaintia, mieluiten samassa SAN-verkossa. Hintansa puolesta synkroninen replikointi on myös hyvin kallista. (Elberg 2014.)

Asynkroninen replikointi toimii hyvin samalla tavalla kuin synkroninen replikointi. Data kirjoitetaan ensin primääriin sijaintiin, josta se tietyin väliajoin kirjoitetaan sekundääriseen sijaintiin. Asynkronisen replikoinnin vahvuutena on, että sitä pystytään hyödyntämään maantieteellisesti hyvinkin suurilla etäisyyksillä, missä synkroninen replikointi ei ole enää toimiva vaihtoehto latenssista johtuen. Asynkroninen replikointi on myös selvästi halvempi kuin synkroninen replikointi. (Elberg 2014.)

## 5.5 Offsite-varmistukset

Edellisessä luvussa käsiteltiin automatisoitua replikointia, joka on suositeltava tapa keskisuurissa ja isoissa organisaatioissa. Pienemmissä organisaatioissa, joissa replikointi on liian kallis vaihtoehto, tuleekin suorittaa perinteisiä offsite-varmistuksia. Käytännössä offsite-varmistukset voi suorittaa kahdella tavalla, joko sijoittamalla fyysisesti varmistusnauhoja organisaation ulkopuolisiin tiloihin, tai hyödyntämällä pilvipalveluita. Offsite-varmistus tarkoittaa varmistusta, joka ei sijaitse organisaation palvelintilassa tai konesalissa. (Knight 2014.)

Vaikka offsite-varmistukset voikin käsittää viimeisenä varokeinona, ei niitä tulisi käsitellä semmoisena. Katastrofitilanteen sattuessa, esimerkiksi tulipalotilanteessa offsite-varmistus voi olla ainoa paikka missä organisaation dataa on säilynyt tuhoutumattomana. Jos näitä varmistuksia on viety ulkopuoliseen sijaintiin esimerkiksi kerran kuukaudessa, tarkoittaa se, että dataa voidaan menettää jopa kuukauden ajalta. Viikon sykleillä dataa olisi menetetty enimmillään viikon ajalta. (Knight 2014.)

Offsite-varmistusten nyrkkisäännöt:

- Tee varmistukset tarpeeksi usein
- Testaa varmistusten toimivuus tasaisin väliajoin
- Varmista nopea palautuminen
- Salaa varmistettu data kryptauksella
- Pidä kirjaa ja merkitse nauhat huolellisesti.
  (Knight 2014.)

Pilvipalvelut ovat myös varteenotettava vaihtoehto tietyin esiehdoin. Tässä asiassa suosittelen perehtymään jo aiemmin viitattuun VAHTI-ohjeeseen, jossa kerrotaan pääpiirteittäin mitä tulee ottaa huomioon, kun dataa säilötään pilvessä. Samaa ohjeistusta voi soveltaa myös tiedon varmistamiseen pilvessä.

Todennäköisin pullonkaula pilveen tehtävissä varmistuksissa on organisaation verkkoyhteys. Päiväsaikaan varmistuksia ei voi suositella ajettavaksi pilveen, sillä tällöin kaistanleveys ulospäin voi koitua ongelmaksi. Tiedostot myös pääosin muuttuvat päiväsaikaan, jolloin asynkronista varmistusta ei voi suositella tehtävän. Käytännössä siis varmistukseen käytettävä aika on enimmillään 12 tunnin luokkaa. Huom. datamäärien kasvaessa myös varmistuksiin kuluva aika kasvaa jatkuvasti. Varmistuksia suunnitellessa tuleekin pitää mielessä jatkuvasti kasvava datan määrä, jotta varmistuksia pystytään tekemään samalla tavalla ainakin vuosi eteenpäin. (Manes 2012.)

Myös sisäänpäin tuleva kaista voi olla ongelma. Katastrofitilanteen sattuessa data pitää saada siirrettyä takaisin organisaation tiloihin. Jos dataa on varmistettu vuosia, voi sen määrä olla niin suuri, että katastrofitilanteesta palautuminen voi kestää päiviä tai viikkoja. Tällöin parempi vaihtoehto taas voisi olla fyysiset offsite-nauhat, joilta palautus voitaisiin tehdä päivän aikana. (Manes 2012.)

Kryptattu varmistusdata vaatii myös avaimen, jolla kryptaus on mahdollista purkaa. Kryptausavain tulisikin säilöä omana offsite-varmistuksenaan, kuitenkin jossain muussa sijainnissa kuin missä itse varmistettu data sijaitsee (Manes 2012).

Pilvivarmistuspalveluja käytettäessä tulee olla myös selvillä siitä, mitä varmistusmenetelmiä käytetään. Kuinka kauas historiassa taaksepäin voidaan mennä? Säilytetäänkö dataa päiviä vai viikkoja. Onko mahdollista palata esimerkiksi kuukauden takaiseen täysivarmistukseen?

### 5.6 Varmistetun tiedon salaaminen

Tietoturvan kannalta väheksytty osa-alue tiedon varmistamisessa on tiedon salaaminen. Erilaiset varmistusjärjestelmät mahdollistavat tiedon salaamisen erilaisilla metodeilla, mutta samalla se tarkoittaa lisää ylläpitotyötä. Jos varmistettua dataa ei ole salattu, on sen lukeminen varmistusnauhoilta äärimmäisen helppoa. Skenaarioita joissa tieto voi joutua vääriin käsiin on useita.

Varmistettu data tulee olla ehdottomasti lukitussa tilassa. Samalla tulee huolehtia, tilan lämpötila on viitearvojen sisällä, kuten myös ilmankosteus. Jos nauhoja säilytetään väärin, voivat ne olla käyttökelvottomia silloin kun niitä eniten tarvitsee. Datan säilöminen lukitussa tilassa vaikeuttaa huomattavasti nauhojen varastamista. (Mah 2012.)

Riskit voivat olla myös sisäisiä. Jos tieto ei ole kryptattua, on käytännössä kaikilla säilytystilaan pääsyn omaavilla henkilöillä mahdollisuus ”lainata” nauhaa, lukea data siltä ja viedä nauha takaisin omalle paikalleen. Näin työntekijä voisi varastaa erittäin arkaluonteista tietoa: henkilötietoja, potilastietoja, käyttäjätunnuksia, pankkitietoja ja taloushallinnon tietoa kuten palkkatietoja.

Teollinen vakoilu voi olla riski virastolle tai yksityiselle yritykselle. Tällaista toimintaa voi suorittaa esimerkiksi kilpaileva yritys tai toinen valtio. Myös hakkerit voivat yrittää kaivaa tämmöisiä tietoja joko pelkästä mielenkiinnosta, tai todistaakseen kuinka helppoa on murtautua yrityksen tietojärjestelmään ja varastaa sieltä dataa. Näissä tapauksissa ongelma on edelleen

sama: jos tieto on täysin selkokielisessä tilassa, on sen lukeminen helppoa. Myös tässä tapauksessa varmistetun datan kryptaaminen on halpa keino estää tiedon joutumista vääriin käsiin. (Suojelupoliisi 2016.)

# 6 Tiedon arkistointi

Nykymaailmassa tiedon määrä kasvaa jatkuvasti. Lisääntyvä tiedon määrä tuokin monia haasteita organisaatioille sekä tietoturvanäkökulmasta, että rahallisesta näkökulmasta. Pitkäaikaisessa säilytyksessä arkistointi onkin kustannustehokkain tapa säilöä tietoa. (Yuhanna 2011.)

Kirjanpitolaki velvoittaa säilyttämään kirjanpidot ja tilinpäätökset. Myös kirjanpitoon ja tilinpäätökseen käytettävä materiaali tulee säilyttää, eli arkistoida. Vastuu arkistoinnin järjestämisestä on yrityksen johdolla. Tiedot on mahdollista säilyttää myös sähköisesti koneellisille tietovälineille. Kaikki kirjanpidon materiaali tasekirjaa lukuun ottamatta voidaan arkistoida paperittomana versiona. (Taloushallintoliitto 2016.)

Taloushallintoliitto toteaa julkaisussaan, että ”Tilinpäätös, kirjanpidot sekä käyttöikää koskevien merkinnöin varustettu tililuettelo kirjanpidoista ja aineistoista on säilytettävä vähintään 10 vuotta tilikauden päättymisestä. Aineisto on säilytettävä järjestelmällisellä tavalla.” (Taloushallintoliitto 2016).

Tiedon arkistointi eroaa monin tavoin tiedon varmistamisesta, ja sisältää erilaisia haasteita siihen verrattuna. Lyhyesti varmistamisella tarkoitetaan datan varmistamista niin, että se on palautettavissa nopealla aikataululla takaisin tuotantoon ja käyttöön. Arkistoinnilla tarkoitetaan semmoisen tiedon arkistoimista, mitä ei enää tarvita päivittäin, mutta tulee silti säilyttää historiatietoa ja mahdollisia katastrofitilanteita varten. (Posey 2010.)

Arkistoinnin perusperiaate on sama sekä pienissä että suurissa organisaatioissa, mutta datan määrä sekä sen kriittisyys voi erota toisistaan paljon. Tietoa on kyettävä säilömään jopa kymmeniä vuosia niin, että sitä on mahdollista lukea ja hyödyntää vuosien kuluttua.

Arkistointi on mahdollista suorittaa myös täysin sähköisesti. Puhuttaessa digitaalisesta taloushallinnosta, tarkoitetaan taloushallintoa, missä kaikki kirjanpitotyön osaprosessit ovat sähköisiä. Digitaalista taloushallintoa voidaan kuvata myös automaattisena taloushallintona. (Lahti & Salminen 2008, 19.)

Digitaalisen datan pitkäaikaissäilyttäminen muodostaa monia haasteita teknisestä, hallinnollisesta ja sosiaalisesta näkökulmasta. Tekninen ongelma pitkäaikaisessa digitaalisessa arkistoinnissa on teknologian kehitys, ja miten muuttuvat standardit tallennusmuodoissa, tallennuslaitteissa ja tallennusmedioissa otetaan mahdollisimman hyvin huomioon (Raymond 2001).

# 7 Kehittämistyö

Opinnäytetyön kohteena oleva varmistusjärjestelmä on Hewlett-Packardin toimittama tuotekokonaisuus. Kyseessä on hybridijärjestelmä, johon kuuluu HP MSL 4048 nauhakirjasto sekä HP StoreOnce 4430 Backup 3G D2D-laite. Järjestelmän ylläpitopalvelin on yhdistetty suoraan kuitukaapelilla nauhavarmistuslaitteeseen ja D2D-laite kuitukytkimen välityksellä palvelimiin, joilla sijaitsee suurin osa varmistettavasta datasta. Laitteiston tiedettiin olevan tarkoituksenmukainen, sillä nykyinen järjestelmä on laitteistotasolla lähes identtinen Säteilyturvakeskuksesta vuonna 2014 vanhentuneen ja käytöstä poistuneen laitteiston kanssa.

Järjestelmän perusperiaatteena on, että pitkäaikaiseen säilytykseen tarkoitettu data varmistetaan magneettinauhoille ja lyhyen ajan varmistukset suoritetaan D2D-laitteelle. Toimintaperiaatteena on siis D2D2T. Tämä metodi hyödyntää molempien varmistustapojen hyviä puolia. Ensin data varmistetaan D2D-laitteelle, josta se kopioidaan myöhemmin nauhakirjastossa sijaitseville medioille. Tämän jälkeen mediat siirretään paloturvalliseen kassakaappiin. Palauttaminen D2D-laitteelta on huomattavasti nopeampaa kuin nauhakirjastosta palauttaminen, jolloin D2D-laitteessa on suositeltavaa säilyttää muun muassa käyttäjä- sekä käyttäjäryhmälevyjen varmistuksia.

## 7.1 Dokumentaatio

Järjestelmään liittyvä dokumentaatio koettiin tärkeimmäksi kehityskohteeksi. Pääosin kaikki toimi jo oikein. Koska varmistusjärjestelmän tiedettiin siirtyvän lähitulevaisuudessa uusille ylläpitäjille, koettiin aihe erityisen tärkeäksi. Hyvin dokumentoitu järjestelmä tuo monia hyötyjä, erityisesti jos ylläpitäjiä on jatkossa useita. Ajantasaisella dokumentaatiolla ja oikeilla käyttäjäoikeuksilla varustettu henkilö voi hoitaa työnsä tehokkaasti, kun asiat tehdään kerralla oikein.

Varmistusjärjestelmän kehitystyö alkoi järjestelmän nykytilan kartoituksella. Kartoituksella selvitettiin järjestelmään jo valmiiksi tehty dokumentaatio, kuvatut toimintaprosessit ja toimintatavat. Ylläpitoon liittyviä dokumentteja löytyi organisaation IT-osaston verkkolevyltä ja dokumentinhallintajärjestelmästä. Ylläpitäjille hyödyllisiä dokumentteja olivat ohjeet siitä, kuinka palautustyö tehdään palvelinohjelmiston avulla, sekä kuinka kuukausittaiset nauhat vaihdetaan.

Varmistettavat kohteet oli listattu omassa dokumentissa. Tämä koettiin hyödylliseksi, sillä vain varmistusjärjestelmässä olevat määritykset katoavat helposti. Dokumentissa kerrotaan yksityiskohtaisesti, mihin verkko-osoitteeseen varmistus tehdään, mihin aikaan ja mitkä hakemistot palvelimesta varmistetaan. Dokumentin avulla on helppoa tehdä ristiin tarkistus siitä, vastaavatko varmistusmääritykset sitä mitä palvelimen ylläpitäjä on pyytänyt.

Viraston tietoturvapolitiikka asettaa erilaisia vaatimuksia varmistuskäytännöille. Näihin lukeutuu esimerkiksi varmistettavan datan suojausaika, sekä loppukäyttäjän ja tietohallinnon vastuut palveluun liittyen. Tietoturvapolitiikkaohje oli päivitetty juuri ennen opinnäytetyöprosessin alkua, joten sen muuttamiselle ei nähty tarvetta. Nykyiset käytännöt todettiin tarkoituksenmukaiseksi ja niihin ei tehty muutoksia.

Virastossa oli otettu käyttöön IT-wiki, jonka tarkoituksena oli yhtenäistää eri järjestelmiin ja palveluihin liittyvä tieto yhteen paikkaan niin, että tieto on helppo pitää ajantasaisena ja se on helposti saatavilla. Varmistusjärjestelmästä ei ollut olemassa olevaa kuvausta yleisellä tasolla ja semmoinen koettiin tarpeelliseksi, kun järjestelmän ylläpitäjät vaihtuisivat. Luotiin Wiki-sivu ”Varmistusjärjestelmä”, mihin kirjattiin palvelimen tiedot tarkalla tasolla. Näin dokumentaatio pysyisi ajantasaisena ja koottuna yhteen paikkaan myös jatkossa.

Huomasimme, että wiki-pohjaisilla sivuilla on myös mahdollista tuottaa hyödyllisiä raportteja. Luotiin raporttisivu ”Varmistuslista” joka kerää kaikista palvelinten tiedoista niille kirjatut varmistuksiin liittyvät huomautukset. Kun raportti oli luotu, huomattiin myös, että monilla wikissä olevalla palvelintiedolla on vanhentuneet tiedot varmistuksiin liittyen. Kaikki palvelimille luodut wiki-sivut käytiin raportin perusteella läpi ja päivitettiin ajan tasalle.

## 7.2 Varmistaminen ja palauttaminen

Viraston varmistuskäytännöt oli toteutettu käyttäen osaan varmistuksista pelkästään täysiä varmistuksia, sillä esimerkiksi Exchange-integraation kautta varmistettava data on pääosin niin muuttumatonta ja datan määrä pieni, että varmistukset voidaan ajaa päivittäin täysinä varmistuksina. Käyttäjä ja käyttäjäryhmälevyillä oli käytössä malli, jossa täysi varmistus suoritetaan kahden viikon välein, ja muina kertoina suoritetaan differentiaalivarmistus. Palautustilanteessa differentiaalivarmistukset on koettu käytännössä paremmin toimivaksi ratkaisuksi kuin inkrementaaliset varmistukset. Differentiaalivarmistus on vikasietoisempi vaihtoehto koska se tarvitsee vain edellisen täyden varmistuksen toimiakseen, inkrementaalinen varmistus taas kaikki muut inkrementaalivarmistukset ennen viimeisintä täyttä varmistusta. Tähän asiaan ei koettu tarpeelliseksi tehdä muutoksia.

Jo opinnäytetyöprosessin alussa tiedettiin joistain järjestelmässä olevista ongelmista, mihin kaivattiin kipeästi korjausta. Varmistusten ajastuksista johtuen jotkut varmistustyöt epäonnistuivat, kun määritelty varmistus ei saanut käyttöönsä nauhuria nauhakirjastosta. Ongelmia ilmeni vielä lisää, jos haluttiin tehdä palautus samaan aikaan, kun D2T-varmistus on käynnissä. Kuukausittainen D2T-varmistus vei aina vähintään 30 tuntia, mikä aiheutti ongelmia seuraavan vuorokauden varmistuksissa.

Ongelmana edellä mainitussa tapauksessa oli se, että varmistusprosessien annettiin varata kaksi nauhuria käyttöönsä, vaikka prosessi toimi hyvin yhdellä nauhurilla. Tämä ongelma oli ollut tiedossa järjestelmän määrityksistä lähtien. Ratkaisuna määritykset korjattiin niin, että yksikään varmistustyö ei saa varata kumpaakin nauhuriansa käyttöön. Joissain tilanteissa tämä ratkaisu saattaa hidastaa yksittäisiä varmistustöitä, mutta parantaa huomattavasti vikasietoisuutta.

Loppukäyttäjille tarjotaan User Backup-toimintoa. User Backup käynnistetään jokaisen työaseman työpöydällä olevasta kuvakkeesta. Sen avulla varmistetaan työasemasta tärkeät tiedostot verkkolevylle. Toiminnossa oli kuitenkin havaittu epävakautta, minkä takia varmistustyö ei joskus varmistanut lainkaan dataa, eikä loppukäyttäjä saanut vastaavaa palautetta. Käyttäjä siis saattoi elää käsityksessä, että ajamalla komennon varmistus tapahtui, vaikka todellisuudessa palvelimelle ei koskaan päätynyt minkäänlaista dataa.

Alkuperäistä ongelmaa miksi jotkut varmistustyöt epäonnistuivat ei onnistuttu korjaamaan, mutta User Backup-toimintoa varten tehtiin ohje, jossa kerrotaan, miten varmistus tapahtuu ja miten toimia, jos haluaa varmistua siitä että tiedot ovat varmasti tallessa. Tämä ohje lisättiin Intranetiin ja Varmistusjärjestelmä-wikisivulle.

Varmistuksien onnistumista seurataan hallintasovelluksella avattavan sisäisen tietokannan, sekä järjestelmän itse lähettämien SMTP-viestien perusteella. Varmistuspalvelin on myös verkonvalvonnan piirissä, jotta huomataan esimerkiksi kovalevyjen täyttyminen tai epänormaali määrä verkkoliikennettä. Palvelimen saatavuutta testataan siis jatkuvasti verkonvalvonnan kautta.

Järjestelmän mahdollisuuksia hyödyntää replikointitoiminnallisuuksia käytiin lyhyesti läpi. Kävi selväksi, että HP:n toimittaman ratkaisun replikoinnin kustannukset nousisivat niin korkeaksi että sitä ei pidetty järkevänä vaihtoehtona. Jos konesalissa olevat datat haluttaisiin replikoida toiseen sijaintiin, vaatisi se toisen D2D-laitteen sekä lisenssit replikointia varten.

### 7.3 Tietoturva

Tietoturvan näkökulmasta löydettiin joitain kehityskohteita. Opinnäyteprosessin aikana HP Data Protectorissa havaittiin kriittinen haavoittuvuus vuoden 2016 alussa. Haavoittuvuus havaittiin verkkoon tehdyn kolmannen osapuolen suorittaman verkkoskannauksen kautta. Havaittu tietoturva-aukko mahdollisti haitallisen koodin ajamisen palvelimessa järjestelmäoikeuksilla. Tietoturva-aukko korjattiin heti kun se havaittiin, sekä tietoturva-aukon olemassaolosta tehtiin kirjaus poikkeamaraporttiin. (Zero Day Initiative 2016.)

Mainitusta tietoturva-aukosta tekee huomionarvoisen se, että ongelma on löydetty jo vuonna 2014, mutta oletusarvoilla sovellus ei edelleenkään käytä hallintapalvelimen ja asiakaspalvelimien välisessä yhteydenpidossa kryptattua yhteyttä.

Vastoin teoreettista viitekehystä organisaatiolla ei ollut käytössä varmistusnauhojen kryptaustoimintoa. Kryptaamaton data on suuri tietoturvariski, jos varmistusnauhoja ei käsitellä tarkoin. Kuukausivarmistusnauhojen vaihdon aikana nauhoja säilytettiin lyhyitä aikoja avokonttorin pöydällä, jolloin kuka tahansa IT-osastolle päässyt henkilö olisi voinut napata nauhan mukaansa. Jos nauhoja olisi käsitelty vain palvelinsalissa, niin tilaan olisi pääsy vain luotetuilla henkilöillä.

## 8 Yhteenveto ja johtopäätökset

Opinnäytetyöprosessi kesti kokonaisuudessaan pitkään. Kehitystyötä Säteilyturvakeskukselle tehtiin lähes kahden vuoden ajan. Tänä aikana järjestelmään liittyvä dokumentaatio kartoitettiin sekä kuvattiin siihen oleellisesti liittyvät prosessit. Tänä aikana otettiin myös käyttöön wiki-pohjainen tietämyssivusto. Sivulle tallennettiin kaikki varmistusjärjestelmään liittyvä dokumentaatio ja tieto. Wiki-toteutuksella tuotettu sivu koettiin ylläpitäjien mielestä toimivaksi ratkaisuksi, muun muassa siitä löytyvien raportointi- ja mallipohjien käytännöllisyyden takia.

Kirjoitustyön loppuvaiheessa päävastuu järjestelmästä oli jo siirretty uusille ylläpitäjille. Uudet ylläpitäjät vastaanottivat päivitetyn dokumentaation järjestelmästä, sekä järjestelmää esiteltiin heille kahdessa eri Lync-verkkokokouksessa. Järjestelmää pidettiin selkeänä sekä helposti ymmärrettävänä. Käytetyt ratkaisut olivat lähes poikkeuksetta toimivia ja varmistusjärjestelmän vikasietoisuus oli hyvällä tasolla. Myös dokumentaation paikkansapitävyyttä ja selkeyttä kehuttiin.

Pilvipalveluiden käyttö varmistamiseen sisältää paljon kysymyksiä. Varsinkin julkishallinnosta puhuttaessa tulee kiinnittää erityistä huomiota varmistettavan datan luonteeseen, sekä viraston hallinnonalan vaikutukset varmistuskäytäntöihin.

Riittävä tekninen osaaminen on tärkeää varsinkin varmistusjärjestelmän käyttöönottovaiheessa. Väärin mitoitetut varmistusmenettelyt eivät ole tarkoituksenmukaisia ja voivat tulla todella kalliiksi. Pienen mittakaavan organisaatioissa hyvin pienelläkin varmistusjärjestelmällä on mahdollista päästä tilanteeseen, jossa data on varmistettu turvallisesti ja hyväksi todettujen toimintatapojen mukaisesti.

Kehitysmahdollisuuksia löydettiin useita ja ne välitettiin uusille ylläpitäjille. Tämä työ ei käsitellyt varmistusten UNIX-puolta, johon virastossa kuuluu lähes puolet varmistettavista palvelimista. Tämä johtui käytetyn varmistussovelluksen teknisistä rajoituksista, kun osa UNIX-jakeluista eivät ole tuettu sovelluksessa. Kyseisiä palvelimia ei siis olisi ollut mahdollista liittää keskitetyn hallinnan piiriin. Lähitulevaisuudessa järjestelmä tullaan yhtenäistämään myös UNIX-varmistusten osalta niin, että kaikki varmistustyö koko viraston organisaatiossa tapahtuu saman keskitetyn hallinnan kautta.

Toinen kehitysmahdollisuus olisi tehostaa nauhakiertoa, jotta arkistoitavat nauhat kiertäisivät aktiivisemmin. Näin saataisiin jaettua kirjoitusten määrä eri nauhoille niin, että tietyt nauhat eivät kuluisi muita nopeammin ja säilyvyys sekä käyttöikä pitenisi.

Lähteet

Kirjalliset lähteet

Curtis Preston, W. 2007. Backup & Recovery: Inexpensive Backup Solutions for Open Systems. O'Reilly Media, Inc.

Lahti, S. & Salminen, T. 2008. Kohti digitaalista taloushallintoa – sähköiset talouden prosessit käytännössä. Juva: WS Bookwell Oy.

Paavilainen, J. 1998. Tietoturva. Espoo: Suomen ATK-kustannus.

Rousku, K. 2014. Kyberturvaopas. Tietoturvaa kotona ja työpaikalla. Viro: Print Best.

Whitman, E. 2004. Management of Information Security. Thomson Course Technology.

Sähköiset lähteet

Aho, T. 2010.Tietoaineistojen luokittelu. Valtiovarainministeriö. Viitattu 1.3.2016. https://www.vahtiohje.fi/web/guest/tietoaineistojen-luokittelu

Arkistolaitos. Usein kysytyt kysymykset: Sähköinen säilyttäminen. Viitattu 5.10.2016. http://www.arkisto.fi/fi/palvelut/usein-kysytyt-kysymykset/asiakirjahallinto/saehkoeinen-saeilyttaeminen/

Brewerton, A. 2009. Understanding Disk-to-Disk backup. Viitattu 2.3.2016. http://www.continuitycentral.com/feature0690.html

Bowen, P. & Hash, J & Wilson, M. 2006. NIST Special Publication 800-100 Information Security Handbook: A Guide for Managers. Viitattu 16.10.2016 http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-100.pdf

Chia, T. 2012. Confidentiality, Integrity, Availability: The three components of the CIA triad. Viitattu 20.10.2016. http://security.blogoverflow.com/2012/08/confidentiality-integrity-availability-the-three-components-of-the-cia-triad/

Data Backup and Disaster Recovery Software. Viitattu 1.3.2016. https://kb.acronis.com/content/1536

Elberg, A. 2014. Difference between Synchronous & Asynchronous Replication. Paranet. Viitattu 3.3.2016. http://www.paranet.com/blog/bid/133845/Difference-between-Synchronous-Asynchronous-Replication-Table

Erickson, T. 2012. Protectign Petabytes: Best practices for big data backup. TechTarget. Viitattu 20.10.2016. http://searchdatabackup.techtarget.com/feature/Protecting-petabytes-Best-practices-for-big-data-backup

Hewlett-Packard Enterprise. 2012. HP Data Protector 7.00 Getting Started Guide. Viitattu 5.10.2016. http://community.hpe.com/hpeb/attachments/hpeb/itrc-251/138074/1/GettingStarted.pdf

Housekeeping. 2013. Wikipedia. Viitattu 2.3.2016. https://en.wikipedia.org/wiki/Housekeeping_%28computing%29

Lock, I. 2010. Tape backup vs disk backup. Computer Weekly. Viitattu 2.3.2016. http://www.computerweekly.com/podcast/Tape-backup-vs-disk-backup

LTO Ultrium Press release. 14.9.2015. Hewlett-Packard Enterprise, IBM and Quantum. Viitattu 4.3.2016.
http://www.lto.org/2015/09/the-lto-program-announces-upcoming-generation-7-specifications-for-licensing/

Keegan, C. 2013. Source vs Target Based Deduplication. Viitattu 3.3.2016.
http://www.storage-switzerland.com/Articles/Entries/2013/1/2_Source_vs_Target_Based_Data_Deduplication.html

Knight, L. 2014. Five Best Practices for Offsite Backup. Zetta.net. Viitattu 3.3.2016.
http://www.zetta.net/blog/five-best-practices-offsite-backup/

KvantiMOTV – Menetelmäopetuksen tietovaranto. Mittaaminen. Viitattu 19.10.2016.
http://www.fsd.uta.fi/menetelmaopetus/mittaaminen/luotettavuus.html

Mah, P. 2012. The Importance of Encrypting Small Business Data Backups. Viitattu 3.3.2016.
http://www.smallbusinesscomputing.com/News/Security/the-importance-of-encrypting-small-business-data-backups.html

Manes, C. 2012. What are the risks of backup up your business data in the cloud. Disaster Recovery Journal. Viitattu 3.3.2016.
http://www.drj.com/articles/online-exclusive/what-are-the-risks-of-backing-up-your-business-data-in-the-cloud.html

Moir, A. 2010. Deduplication: The Pros and Cons end users are not always told. Viitattu 10.3.2016.
http://www.continuitycentral.com/feature0796.html

Raymond, A. 2001. Long term preservation of digital information. Viitattu 10.3.2016.
http://dl.acm.org/citation.cfm?doid=379437.379726

Kvalitatiivinen tutkimus. Taloustutkimus Oy. Viitattu 5.10.2016.
http://www.taloustutkimus.fi/tuotteet_ja_palvelut/tiedonkeruuratkaisut_ja_monitila/kvalitatiivinen_tutkimus/

Techopedia. 2016. Disk-to-Disk-to-Tape (D2D2T). Viitattu 17.10.2016.
https://www.techopedia.com/definition/1075/disk-to-disk-to-tape-d2d2t

Types of Backup. Viitattu 2.3.2016.
www.typesofbackup.com

Managing Information Security Risk. National Institute of Standards and Technology. Viitattu 2.3.2016.
http://csrc.nist.gov/publications/nistpubs/800-39/SP800-39-final.pdf

Martin, L. Voltage Security. Viitattu 2.3.2016.
www.bsc.org/content/conWebDoc/15870

Mustonen, M. 2013. Kirjanpitoaineistojen sähköinen arkistointi ja siihen liittyvät vaatimukset.
https://www.theseus.fi/bitstream/handle/10024/57447/Mustonen_Maiju.pdf?sequence=1

Suojelupoliisi. Tietoverkkovakoilu. Viitattu 8.10.2016.
www.supo.fi/vastatiedustelu/tietoverkkovakoilu/

Posey, B. 2010. 10 things you should know about long-term data archiving. Viitattu 3.3.2016.
http://www.techrepublic.com/blog/10-things/10-things-you-should-know-about-long-term-data-archiving/

Taloushallintoliitto. Miten kauan kirjanpitoja tulee säilyttää? Viitattu 8.10.2016. https://taloushallintoliitto.fi/kirjanpidon-abc-mita-jokaisen-tulisi-tietaa-kirjanpidosta/miten-kauan-kirjanpitoja-taytyy-sailyttaa

TechTarget. 2005. data availability. Viitattu 5.10.2016. http://searchstorage.techtarget.com/definition/data-availability

Tilastokeskus. Viitattu 5.10.2016. https://www.stat.fi/virsta/tkeruu/01/07/

Vangie, B. RAID – Redundant array of independent disks. Viitattu 4.10.2016. http://www.webopedia.com/TERM/R/RAID.html

Viestintävirasto. 2014. Pilvipalveluiden tietoturva organisaatioille. Viitattu 8.10.2016. https://www.viestintavirasto.fi/attachments/tietoturva/Pilvipalveluiden_tietoturva_organisaatioille.pdf

VirtuaaliAMK. Tutkimuksen reliabiliteetti. Viitattu 20.10.2016. http://www2.amk.fi/digma.fi/www.amk.fi/opintojaksot/0709019/1193463890749/1193464185783/1194413792643/1194415307356.html

Yuhanna, N. 2011. Your Enterprise Data Archiving Strategy. IBM. Viitattu 3.3.2016. ftp://public.dhe.ibm.com/software/data/sw-library/data-management/optim/papers/your-enterprise-data-archiving-strategy.pdf

Yurin, M. The History of Backup. SoftLogica. Viitattu 2.3.2016. www.backuphistory.com

Zero Day Initiative. Hewlett-Packard Data Protector EXEC_INTEGUTIL Remote Command Execution Vulnerability. Viitattu 26.10.2016. http://www.zerodayinitiative.com/advisories/ZDI-14-344/

Taulukot

Liitteet

Liite 1: Full vs Incremental vs Differential backup

Liite 1: Full vs incremental vs differential backup