Opinnäytetyö (AMK)

Tieto- ja viestintätekniikka

2023

Pekka Näyskä

Tietokeskuksen lokienhallintapolitiikan suunnittelu

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

Opinnäytetyö (AMK) | Tiivistelmä

Turun ammattikorkeakoulu

Tieto- ja viestintätekniikka

2023 | 35 sivua

Pekka Näyskä

Tietokeskuksen lokienhallintapolitiikan suunnittelu

Tietoverkkojen turvaamiseksi ja verkkoliikenteen eheyden, saatavuuden ja todenmukaisuuden varmistamiseksi liikenteestä ja verkkojärjestelmistä täytyy kerätä tietoa eli lokia. Lokitus ja lokienhallinta ovat avainasemassa tietoturvapoikkeamien ennaltaehkäisyssä ja selvittämisessä jälkikäteen. Ilman lokitusta monien tietoturvapoikkeaminen havaitseminen ja niihin reagoiminen on hankalaa ja jopa mahdotonta. Lokienhallintapolitiikka edesauttaa, että lokitusta tehdään yhteisten sääntöjen mukaisesti.

Vuonna 2022 Tietokeskukselle teetetyssä tietoturva-auditoinnissa todettiin sen sisäisessä lokituksessa ja lokienhallinnassa olevan kehitettävää. Opinnäytetyö toteutettiin osana 2023 perustettua sisäisen lokienhallinnankehittämisprojektia. Kehittämisprojektin tavoitteena oli saattaa toimeksiantaja lähemmäksi ISO 27001 -standardisointia, käytännössä parantaa tietoturvatoimintaa ja tarjota parempia työkaluja sisäisten järjestelmien vianselvitykseen. Ennen kehitysprojektin aloitusta toimeksiantajalla ei ollut kirjoitettua lokienhallintapolitiikkaa.

Opinnäytetyön tavoitteena oli tutkia lokitusta, siihen liittyviä standardeja ja lainsäädäntöä, lokienhallintaa ja käyttöön otettavaa keskitetyn lokienhallinnan Elastic Stack -ympäristöä. Tärkeänä tavoitteena oli selvittää ja suunnitella keskeisiä lokienhallintapolitiikan osa-alueita. Tutkimus toteutettiin kartoituksena ja lähteinä käytettiin alan kirjallisuutta, asiantuntijaorganisaatioiden ohjeistuksia sekä Suomen ja EU:n lainsäädäntöä. Tietokeskuksen vaatimuksia ja tarpeita selvitettiin lokienhallintaprojektiryhmän palavereissa sekä keskusteluissa Tietokeskuksen tietoturvapäällikön kanssa. Työn tuloksena luodaan Tietokeskukselle sisäinen lokienhallinnanpolitiikka, joka perustuu sen tarpeisiin, lainsäädäntöön, standardeihin ja yleisiin hyviin käytänteisiin alalla.

Asiasanat:

lokienhallintapolitiikka, lokienhallinta, lokitus

Bachelor’s Thesis | Abstract

TURKU UNIVERSITY OF APPLIED SCIENCES

Information and Communications Technology

2023 | 35 pages

Pekka Näyskä

Designing a log management policy for Tietokeskus

In order to secure information networks and ensure the integrity, availability, and confidentiality of network traffic, logs of this traffic and network systems must be collected. Logging and log management are a key part of preventing and investigating information security breaches. Without logging, detecting, and responding to information security incident, management is difficult or even impossible. A log management policy promotes the use of common rules for logging.

In a security audit conducted for Tietokeskus, an IT company, in 2022, it was found that there was room for improvement in their internal logging and log management. This thesis was conducted as part of a 2023 internal log management development project. The goal of the development project was to bring the client closer to ISO 27001 standardization, improve their information security operations in practice, and provide better tools for troubleshooting internal systems. Before starting the development project, the client did not have a written log management policy.

The aim of this thesis was to study logging, its associated standards and legislation, log management, and the use of the Elastic Stack environment for centralized log management. An important goal was to identify and plan key areas of log management policy. The research was conducted as an information mapping and used literature from the field, guidelines from expert organizations, and Finnish and EU legislation as sources. The requirements and needs of Tietokeskus were determined in project group meetings and in discussions with the information security manager of Tietokeskus. As a result of the thesis, an internal log management policy will be created for Tietokeskus based on their needs, legislation, standards, and best practices in the industry.

Keywords:

log management policy, log management, log files

Sisältö

Kuvat

Taulukot

Sanasto

AD-autentikointi	Active Directory on Microsoft Windows -toimialueen käyttäjätietokanta ja hakemistopalvelu. AD-autentikointi tarkoittaa käyttäjien todentamista Active Directoryllä. (Wikipedia 2021.)
ASCII	American Standard Code for Information Interchange eli amerikkalainen standardisointi tiedonsiirtoon. Yleisesti käytössä oleva 128 merkkipaikan tietokoneiden merkistö. (Wikipedia 2022.)
ICT	Information and communications technology eli tieto- ja viestintätekniikka. (Ite wiki 2019.)
IP-osoite	Internet Protocol-osoite eli internetin protokollaosoite. Kaikille verkkoon kytketyille laitteille yksilöllinen osoite, jolla laite voidaan tunnistaa verkoissa. (Jensen, P. 2018.)
SIEM	Security Information and Event Management eli suojaustietojen ja tapahtumien hallinta. SIEM:in avulla voidaan tunnistaa, analysoida ja reagoida tietoturvauhkiin. (Microsoft n.d.)
Avoin lähdekoodi	Jonkin ohjelman lähdekoodi, jota kuka tahansa voi hyödyntää, jakaa ja muokata, koska se on julkaistu julkiseksi. (OpenSource 2019.)
HTTP-metodi	Hypertext Transfer Protocol eli hypertekstin siirtoprotolla. Selaimet ja www-palvelimet siirtävät tietoa HTTP:n avulla. (Vanhatapio, J. 2020.)
MD5-algoritmi	Messade Digest -algoritmi on kryptograafinen tiivistefunktio. (Wikipedia 2023b.)
SHA-algoritmi	Secure Hash Algorithm on kokoelma tarkistussummia, joiden avulla voidaan todentaa tiedonsiirrossa tiedon eheyttä. (LinuxWiki 2016.)

JSON	JavaScript Object Notation on yksinkertainen ja kevyt avoimen standardin tiedostomuoto tiedonvälitykseen ja tallennukseen. (Wikipedia 2023a.)
API	Application Programming Interface on sovellusliittymä. Se on ohjelmointiliitäntä, joka sallii useiden sovellusten kommunikoida toistensa kanssa. (Visma n.d.)

1 Johdanto

Jotta tiedon aitoutta ja saatavuutta sekä tietoverkkojen turvallisuutta voidaan mitata, täytyy verkkoliikenteestä ja verkkojärjestelmistä kerätä tietoa. Yritysten kasvaessa myös niiden verkkoliikenne ja verkkoinfrastruktuuri kasvaa. Nykyisin maailmassa lähes kaikki verkkoon kytkettävät laitteet reitittimistä autoihin keräävät lokitietoja. Verkkoliikennettä ja järjestelmien toimintaa voidaan havainnoida valvomalla lokeja joko reaaliajassa tai tallentamalla ne pitkäaikaisesti. Ilman lokien keräämistä vikatilanteiden ja virheiden korjaaminen on haastavaa tai jopa mahdotonta. Keskitettyllä lokienhallinnalla parannetaan lokien saatavuutta, eheyttä ja suojausta (AWS Whitepaper n.d.).

Tässä työssä tutkitaan ja määritellään, mitä ovat lokikirjaukset, mitä on lokienhallinta ja mitä lokienhallintapolitiikan tulee sisältää. Tarkoituksena on myös syventää omaa osaamista lokeista ja lokienhallinnasta. Työn pohjalta luodaan Tietokeskus Finland Oy:lle lokienhallintapolitiikka, joka perustuu yrityksen tarpeisiin, vaatimuksiin sekä yleisiin standardeihin ja lakeihin. Lokienhallintapolitiikan avulla Tietokeskus pystyy säilyttämään, analysoimaan, valvomaan, indeksoimaan ja keräämään lokeja hallitusti ja yhteisillä säännöillä. Tietokeskukselle tulee käyttöön Elastic Stack -lokienhallinnan ympäristö, josta työssä käydään läpi perusteet sekä reflektoidaan politiikan suunnittelua ja toteutumista Elastic Stackin kanssa.

Opinnäytetyön aluksi luvuissa 2 ja 3 tutkitaan, mitä lokit ja lokienhallinta ovat, ja miksi niitä kerätään, sekä määritellään lokien ja lokienhallinnan termistö ja peruskäsitteet. Luvussa 4 käydään läpi Elastic Stack -ympäristön toimintaa ja osia, jotta lokienhallintapolitiikassa voidaan viitata sen käyttöön ja käytön sääntöihin. Luvussa luku 5 annetaan raamit sille, mitä politiikan tulee sisältää ja miten lokipolitiikka kirjoitetaan pohjautumaan lainsäädäntöön ja standardeihin.

Tutkimus toteutettiin kartoituksena ja lähteinä käytettiin alan kirjallisuutta, asiantuntijaorganisaatioiden ohjeistuksia sekä Suomen ja EU:n lainsäädäntöä.

Tietokeskuksen vaatimuksia ja tarpeita selvitettiin lokienhallintaprojektiryhmän palavereissa sekä keskusteluissa Tietokeskuksen tietoturvapäällikön kanssa. Lokienhallintaprojektiryhmään kuuluivat tietoturvapäällikön ja itseni lisäksi kaksi tietoverkkoasiantuntijaa. Työssä suunnitellaan pohjaa politiikalle vain Tietokeskuksen sisäisen toiminnan lokienhallinnalle, jolloin esimerkiksi luottokorttitietojen tallentamista käsittelevä PCI-DSS (Payment Card Industry Data Security Standard) -standardi jätetään täysin huomiotta, koska kerättävät lokit eivät tule sisältämään luottokorttitietoja.

Theseuksesta löytyi opinnäytetöitä samankaltaisista aiheista. Niissä käytiin läpi lokienhallintapolitiikan suunnittelemista sekä tyypillisesti jotain lokienhallinnan ympäristöä. Näiden töiden valmistumisen jälkeen on kuitenkin Suomen ja EU:n lokeja koskevaa lainsäädäntöä päivitetty. Esimerkkitöissä ei myöskään käydä läpi Elastic Stackin toimintaa eikä politiikan perustamista sille. Johannes Matilaisen (2019) työssä ”Tietoturvapoikkeaminen- ja lokienhallinta” kerrotaan lokienhallinnasta AlienVault USM Appliance SIEM -ratkaisulla sekä esiteltiin lokienhallinnan standardeja, mutta ei käydä läpi lokienhallinnan lainsäädäntöä tai Elastick Stack -ympäristöä. Jesse Kumlanderin (2019) tekemässä ”Saas-palvelun tietoturvan kannalta olennaiset lokit ja niiden hallinta” -opinnäytetyössä kerrotaan lainsäädännön vaatimuksista ja lokienhallinnasta LogDNA-palvelulla. Opinnäytetyön valmistumisen jälkeen lainsäädäntöä on kuitenkin päivitetty ja LogDNA-palvelu on hyvin erilainen Elastic Stack -ympäristön kanssa.

2 Lokit

2.1 Lokikirjaus

Lokikirjaus on jonkin laitteen esimerkiksi reitittimen, tulostimen tai tietokoneen ennalta määrätyssä muodossa luoma dokumentti jostakin tapahtumasta tiettynä hetkenä. Tapahtuma voi liittyä esimerkiksi käyttäjän autentikointiin, järjestelmässä tapahtuneeseen virheeseen tai kovalevyn täyttymiseen. Lokikirjauksen sisältämä tieto auttaa järjestelmän ylläpitäjiä ja käyttäjiä oikeusturvan toteuttamisessa, järjestelmien ja verkon toiminnan optimoinnissa ja erilaisten tietoturva- ja muiden poikkeamien havaitsemisessa ja selvittämisessä. (Lokiohje 2009. 13.)

Lokien tutkimiseksi määritellään ensin, mitä yleisimmät lokitermit tarkoittavat. Tapahtuma on yksittäinen esiintymä ympäristössä, josta kirjataan tietoja. Tätä yksittäistä tietoa kutsutaan tapahtumakentäksi. Tapahtumakenttä voi olla esimerkiksi esiintymän aika, mitä on tapahtunut tai missä laitteessa. Tapahtumakenttien indeksoinnilla tarkoitetaan lokien luokittelua jonkin tapahtuman kentän mukaan, esimerkiksi kronologisesti ajan suhteen. Lokikirjaus tarkoittaa dokumenttia yhdestä tapahtumasta, ja se rakentuu tapahtumakentistä. Loki taas on kokoelma lokikirjauksia. Lokitus tarkoittaa lokien keräämistä. Esimerkissä havainnoidaan lokitermistöä suhteessa toisiinsa, lähtien esiintymän ajasta vasemmalta (Kuva 1). (MITRE 2010.)

Kuva 1. Lokitermistö. (MITRE 2010.)

Lokikirjaukset voidaan jakaa esimerkiksi seuraaviin yleiskategorioihin niiden tyypin mukaisesti (Chuvakin ym. 2012, 3.):

- Informatiivinen kertoo esimerkiksi laitteen käynnistyneen uudelleen tai käyttäjän kirjautuneen sisään. Jos tapahtuma on normaalia kyseiselle lokin lähteelle, ei informatiivisista lokeista tarvitse huolestua.
- Debug-loki on tietokoneohjelmiston tuottama viesti, jonka tarkoitus on auttaa ohjelmakehittäjiä löytämään ongelmia ja virheitä koodista.
- Varoitus kertoo, että ajetussa ohjelmassa saattaa olla puute, mikä ei estä ohjelman ajamista, mutta tieto saattaa olla hyödyllinen ohjelman käyttäjälle tai kehittäjälle.
- Virhe tarkoittaa, että ohjelma tai laite on ajanut virheeseen yrittäessään tehdä jotain. Virheloki saatta syntyä esimerkiksi ohjelman yrittäessä avata tiedostoa, jota ei ole.
- Hälytyslokit liittyvät yleensä tietoturvalaitteisiin tai laitteiden tietoturvaan. Kirjaus saattaa syntyä esimerkiksi siitä, että järjestelmä on kaatunut odottamattomasti.

Tarkastelemalla kokoelmaa lokeja saadaan luotua kokonaiskuva. Lokikirjaukset tehdään kronologisessa järjestyksessä, jolloin niitä tarkastelemalla voidaan luoda aikajana tapahtumien kulusta. (Chuvakin ym. 2012, 1–3.)

2.2 Lokitus

Lokien keräämiselle on kolme pääsyytä: tietoturva, toiminnallisuus sekä ohjelmakehitys. Tietoturvan lokitusta tehdään tietoturvauhkien huomaamiseksi ja niihin vastaamiseksi. Tietoturvalokikirjaus voi olla esimerkiksi käyttäjän sisäänkirjautumisilmoitus. Toiminnallisuuslokeja kerätään, jotta voidaan varmistaa kaikkien laitteiden ja järjestelmien toimivan oikein ja optimoidusti. Toiminnallisuusloki on esimerkiksi järjestelmän ilmoitus hupenevasta tallennustilasta. Ohjelmavirhe- eli debug-lokeja kerätään ohjelmistojen kehitysvaiheessa tutkimaan ohjelmien ja sovellusten toimintaa. Monet ohjelmointisovellukset pystyvät tunnistamaan tarkalleen, millä rivillä koodia ongelma ilmeni ja miksi. (Graylog Team 2022.)

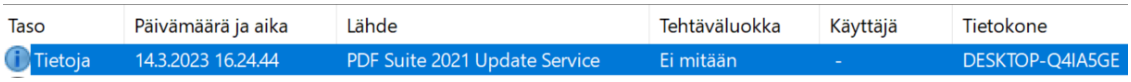
2.3 Lokien syntaksi

Lokeja voidaan kerätä joko tekstitiedostoina esimerkiksi ASCII-formaatissa tai vain binääritiedostoina. Laitteiden ja ohjelmistojen on helpompaa lukea vain binäärimuotoista lokia. Binäärimuotoisia lokeja ei voi pakata pienemmiksi tiedostoiksi samassa suhteessa kuin tekstimuotoisia, mutta ne ovat lähtökohtaisesti tiiviimmin tallennettuja ja vievät levyllä vähemmän tilaa. Tekstinä tallennetut lokit ovat tiedostokooltaan suurempia ja vaikeampi lukuisia tietokoneille. Tekstilokeja ei tarvitse välttämättä muuntaa binäärimuotoon tietokoneiden käsiteltäväksi, mutta tiekoneet saattavat käsitellä vain niille tarkoitettua binäärimuotoista lokia nopeammin. (Chuvakin ym. 2012, 38–39.) Lähes kaikkia tekstinä tallennettuja lokeja voidaan tarkastella millä tahansa tekstieditorilla, esimerkiksi Windowsin sisäänrakennetulla Muistiolla (Sematext n.d.).

Binäärimuotoinen loki on joidenkin mielestä tietoturvallisempi, koska sitä ei pysty ilman oikeaa tulkkaavaa ohjelmaa lukemaan. Tämä viittaa kuitenkin turvallisuuskäytäntöön, jossa pyritään turvaamaan järjestelmän tai tiedon turvallisuus salaamalla se tai pitämällä se piilossa sen sijaan, että turvallisuus perustuisi vahvoihin salaustekniikoihin ja hyviin tietoturvakäytäntöihin. Tämä käytäntö on yleisesti ottaen heikko tapa suojata järjestelmiä, sillä salailun tai salaisuuksien ylläpitäminen ei yksinään riitä suojaamaan järjestelmää tai tietoa.

Kerättävien lokien tarkoitus kuitenkin viime kädessä on olla ihmisen tulkittavissa. Jos lokeja tallennetaan tekstitiedostoina, ne kannattaa jo lähtökohtaisesti tallentaa mahdollisimman ihmisystävällisellä tavalla. Tämä tarkoittaa esimerkiksi tapahtumakenttien erottelemista selvillä merkeillä, eikä kenttien kirjaamista yhdeksi putkeksi ilman selvää erottelua. Vaikka lokit tallennettaisiin binäärimuodossa, helpottaa yhtenäinen syntaksi lokikirjausten välillä lokien luokittelussa myöhemmin. (Hombergs 2018.)

Tarkastellaan tyypillistä lokikirjausta, joka on otettu Windows 10 -tietokoneen tapahtumienvalvonnasta.



Taso	Päivämäärä ja aika	Lähde	Tehtäväluokka	Käyttäjä	Tietokone
Tietoja	14.3.2023 16.24.44	PDF Suite 2021 Update Service	Ei mitään	-	DESKTOP-Q4IA5GE

Kuva 2. Windowsin Tapahtumienvalvonta lokimerkintä.

Esimerkissä Windowsin Tapahtumienvalvonta erottaa selvästi lokin eri osat ihmiselle helppolukuiseksi (Kuva 2). Nopealla vilkaisulla nähdään tapahtuman vakavuus, mitä on tapahtunut ja milloin sekä lokin lähde eli mikä laite on luonut lokin.

Jotta lokia voidaan pitää hyödyllisenä, sen pitäisi sisältää ainakin seuraavat kentät (Kyberturvallisuuskeskus 2019):

- kellonaika, päivämäärä, aikavyöhyke
- lokin tyyppi (esim. virhe vai hälytys)
- tärkeysaste
- lokin lähde
- oliko tapahtuma onnistunut vai epäonnistunut
- käyttäjä joka tapahtuman aiheutti.

3 Lokienhallinta

3.1 Lokienhallinta ja sen keskitys

Lokienhallinta tarkoittaa yksinkertaisesti lokikirjauksien käsittelyä niiden koko elinkaaren ajan. Se kattaa lokin keräyksen, kuljettamisen, säilyttämisen, analysoinnin, seuraamisen ja siitä raportoinnin. Oikein suunnitellulla ja toteutetulla lokienhallinnalla vähennetään ylimääräistä verkkoliikennettä sekä säästetään aikaa, rahaa ja tallennustilaa. Hyvällä lokienhallinnalla saavutetaan lokien hyvä saatavuus, todenmukaisuus ja luottamuksellisuus. (Sharif 2022a.)

Keskitetty lokienhallinta tarkoittaa lokitiedon keräämistä verkoista, verkkoinfrastruktuurista ja ohjelmista yhteen paikkaan sen analysointia ja tallentamista varten. Keskitetyllä lokienhallinnalla mahdollistetaan lokitiedon säilyttäminen yhdessä paikassa. Jos verkossa oleva palvelin hakkeroidaan ja sen sisältämä loki pyyhitään jälkien peittämiseksi, keskitetyn lokinhallinnan omaavassa järjestelmässä lokit on jo lähetetty muualle turvaan. Yleisesti käytössä olevat keskitetyn lokinhallinnan ohjelmistot pystyvät lukemaan ja tulkitsemaan monenlaisia lokeja tarvitsematta eritellä niitä tiedostotyypin mukaan. Lokitiedon keskittäminen helpottaa sen indeksointia ja tätä kautta myös analysointia ja seuraamista. (Sharif 2022b.)

Lokienhallinnan keskityksellä on myös haittoja. Lokitiedon kuljettaminen kaikista laitteista keskitettyyn palveluun kuormittaa verkkoa huomattavasti. Jos lokin kuljetusta lähteestä kerääjälle ei ole suojattu, kulkee se myös vapaasti luettavana ja kaapattavana verkon läpi. Kaiken lokitiedon kerääminen yhteen paikkaan vaatii myös huomattavan määrän tallennustilaa. (Chuvakin ym. 2012, 78–79.)

3.2 Lokien lähteet ja keräys

Lokilähde tarkoittaa lokia tuottavaa järjestelmää, laitetta tai ohjelmaa, esimerkiksi palvelinta. Loki voidaan joko hakea tai vastaanottaa keskitettyyn

lokienhallintaan. Tässä opinnäytetyössä käsitellään lokilähteinä myös protokollia, jotka eivät aiemman määritelmän mukaan ole lokilähteitä, mutta joita voidaan sellaisina pitää keskitetyn lokienhallinnan kannalta. Näistä yleisimmät vastaanotettavat lokilähteet verkoissa ovat Syslog-protokolla, SNMP-paketit ja Windows event log. Lokienhallinnan kannalta esimerkiksi kaikki SNMP-lokit ovat keskenään yhteneväisiä. Syslog-protokolla on yleisimmin Linux-käyttäjärjestelmien käytössä oleva lokien keräys-, tallennus- ja siirtomenetelmä. Sitä voidaan pitää lokilähteenä keskitetyn lokienhallinnan kannalta. (Chuvakin ym. 2012, 53.)

SNMP on standardiprotokolla TCP/IP-verkoissa olevien laitteiden, erityisesti verkkoinfrastruktuurilaitteiden, hallintaan. Muun muassa useimmat reitittimet, kytkimet, palvelimet, työasemat ja tulostimet tukevat sitä. Vaikka SNMP-protokolla ei kokonaisuudessaan ole lokitusjärjestelmä, SNMP-ansat ja -ilmoitukset voidaan katsoa tyypillisiksi lokilähteiksi. (Mauro & Schmidt 2001, 2) (Chuvakin ym. 2012, 59.)

Windows event log on eli Windows-tapahtumaloki on Windows-käyttöjärjestelmien tuottama ja keräämä järjestelmä-, tietoturva- ja ohjelmistoloki. Teknisesti Windows-tapahtumaloki käsite sisältää protokollan, siirtomenetelmän, tallennuksen ja haun. Windows-tapahtumaloki tallennetaan binäärinä, eikä sitä voi lukea tavallisella tekstieditorilla. Tarkastelu onnistuu Windows Tapahtumienvalvonnalla. (Solarwinds 2023.)

3.3 Lokien suodatus ja normalisointi

Lokien suodatus tarkoittaa tarvittavan lokidatan pitämistä ja ylimääräisen hylkäämistä. Oikein toteutetulla suodatuksella on suuri vaikutus kerätyn lokin käytettävyyteen ja hyödyllisyyteen. Ylimääräisen hyödyttömän lokin pitäminen kuormittaa lokienhallintajärjestelmää turhaan.

Normalisointi on erilaisen lokin yhtenäistämistä. Lokilähteitä ja täten erilaisia lokityyppejä on lukemattomia. Lokien suodatuksen yhteydessä lokikirjaukset puretaan yksittäisiin tapahtumakenttiin ja sen jälkeen kootaan yhteen yleiseen

muotoon. Suodatuksen ja normalisoinnin jälkeen lokidata voidaan tallentaa analysointia ja raportointia varten (Chuvakin ym. 2012, 145–149.)

3.4 Tietokeskus yrityksenä ja lokienhallinan kehitys

Opinnäytetyön toimeksiantajana toimii Tietokeskus Finland Oy. Yritykselle on vuonna 2022 tehdyssä tietoturva-auditoinnissa todettu, että tiettyjä tietoturvan osa-alueita tulee kehittää, että se pääsee ISO 27001 -tietoturvassertifiointiin. Tietoturvasertifiointiin pääsemisen lisäksi tarkoitus on käytännössä parantaa tietoturvatoimintaa ja tarjota parempia työkaluja sisäisten järjestelmien vianselvitykseen. Tietokeskus Finland Oy on suomalainen vuonna 1989 perustettu ICT-palveluyhtiö, joka tuottaa tietotekniikan palveluita laitetoimituksista kokonaisvaltaisiin yritysten IT-infrastruktuurin ja tietoturvan auditointeihin. Tietokeskuksessa työskentelee tällä hetkellä yli 300 työntekijää ympäri Suomea. (Tietokeskus 2023.)

3.5 Lokienhallintajärjestelmä Tietokeskukselle

Lokienhallinnankehitysprojektin alkuvaiheessa käytiin läpi eri vaihtoehtoja lokienhallintajärjestelmälle. Tutkimusta tehtiin lokienhallintaprojektiryhmän toimesta palaverien ja eri vaihtoehtojen läpi käymisen muodossa. Tällä hetkellä käytössä olevasta Graylogin ilmaisversiosta on poistettu käytöstä AD-autentikointi, joka on ehdoton vaatimus toimeksiantajalle. Tarkoituksena ei aluksi ole kehittää kokonaisvaltaista SIEM järjestelmää, vaan kerätä lokit keskitetysti. Tutkittuja vaihtoehtoja lokienhallinta-palvelulle olivat Splunk, Sematext, SolarWinds, Graylog ja Elastic Stack.

Vaihtoehdoista lopullinen valinta tehtiin Elastic Stackin ja jo ilmaisversiona käytössä olevan Graylogin maksullisen version väliltä. Ratkaisevat tekijät olivat mm. järjestelmän lisenssin hinta, skaalautuvuus, käytettävyys ja yhteensopivuus jo käytössä olevien järjestelmien kanssa. Suurin osa muista palveluista karsiutui korkean hinnan ja monimutkaisen käyttöliittymän vuoksi.

Muun muassa Elastic Stackin hyviä puolia olivat sen skaalautuvuus myöhemmin SIEM-järjestelmäksi, joustavuus sekä projektiryhmän aiempi kokemus sen käyttäjäystävällisestä käyttöliittymästä ja konfiguroinnista.

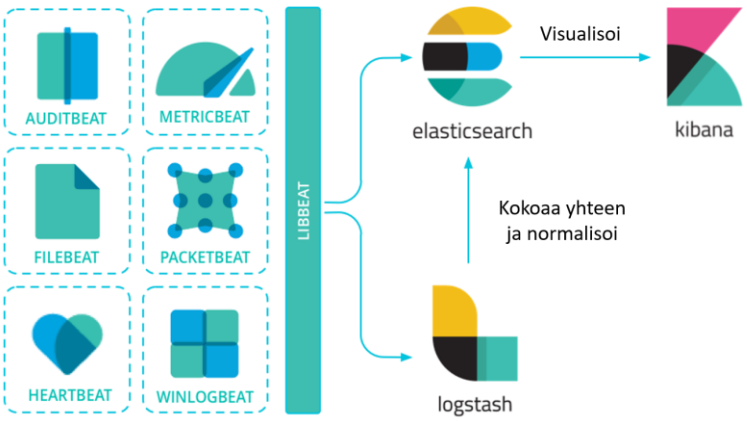
Tuotteiden vertailun jälkeen esitettiin lokienhallintakehitysprojektin ohjausryhmälle projektihenkilöiden havainnot ja tulokset, kuten suuntaa antavat kuukausittaiset lisenssimaksut ohjelmistoille. Ohjausryhmä hyväksyi projektiryhmän ehdotuksen Elastic Stackin käyttöönotosta.

4 Elastic Stack -ympäristö

Elastic Stack on kokoelma Elastic N.V:n kehittämiä ja ylläpitämiä avoimen lähdekoodin tuotteita, joiden tarkoitus on helpottaa lokien käsittelyä, keräämistä, tallentamista ja analysointia. Elastic Stack lyhennetään usein sanaksi ELK. ELK koostuu Elasticsearchistä, Logbeatistä, Beateistä ja Kibanasta. Esimerkissä näytetään Elastick Stack -ympäristön työnjakoa ja lokin liikkumista niiden välillä (Kuva 3). Toisin kuin esimerkki antaa ymmärtää Beateistä voidaan siirtää lokitietoa suoraan Elasticsearchiin. (Yasar 2022.)

Kuva 3. Elastic Stackin osat ja vastuualueet. (Wutzke 2019.)

4.1 Beat-agentit



Kuva 4. Lokitiedon kulku Beateistä visualisoitavaksi. (Elastic 2023a.)

Beatit ovat Elastic-ympäristössä kevyitä avoimen lähdekoodin datasensoreita, joita käytetään lokin keräämiseen ja siirtämiseen lähteeltä, esimerkiksi palvelimelta joko Logstashiin ja sieltä Elasticsearchiin tai suoraan Elasticsearchiin. Esimerkissä näytetään, millä eri tavoilla lokitieto voi kulkea Beateistä Elasticsearchiin (Kuva 4). Beatit asennetaan agentteina laitteille, joista lokia halutaan kerätä. Beatit keskustelevat keskitetyn lokienhallinnan kanssa lokidatan lähetyksen yhteydessä, jotta ne pysyvät kartalla siitä, mitä ne ovat jo lähettäneet. Beat-agentit voidaan määritellä valitsemaan jo lähdelaitteessa, mitä lokeja kerätään ja lähetetään, mikä vähentää huomattavasti keskitetyn lokienhallinnan prosessointikuormaa ja verkkoliikennettä. Elastic N.V tarjoaa tällä hetkellä seitsemän virallista Beatiä, jotka on lueteltu alla olevassa taulukossa (Taulukko 1). Lisäksi tarjolla on yhteisön ylläpitämiä, tarkempiin käyttötarkoituksiin olevia Beatejä. Libbeat on API, jota kaikki Beatit käyttävät tiedonsiirrossa. (Elastic 2023a.) Työssä esitellään muutama toimeksiantajalle käyttöön tuleva Beat tarkemmin luvuissa 4.2 ja 4.3.

Taulukko 1. Elastic Stackin tarjoamat Beatit.

Beatit	**Funktio**
Auditbeat	Linux-auditointidata
Filebeat	Lokit
Functionbeat	Pilvidata
Heartbeat	Järjestelmien saatavuus
Metricbeat	Järjestelmämittarit
Packetbeat	Verkkoliikenne
Winlogbeat	Windows lokit

4.2 Winlogbeat-agentti

Winlogbeat on Windows-käyttöjärjestelmäpohjaisille laitteille Elasticin oma virallinen Beati, joka lähettää Windows-tapahtumalokit joko Logstashiin tai Elasticsearchiin. Winlogbeat toimii Windowsin taustaprosessina ja seuraa Windows-tapahtumalokeja Windows-pohjaisilla palvelimilla ja työasemilla. Se kerää tapahtumalokeista käyttäjän ennalta määrittämät kriteerit täyttävät lokit ja lähettää ne eteenpäin keskitetylle lokienhallinnalle. Winlogbeat pystyy myös säilyttämään lokeja hetkellisesti, esimerkiksi verkkokatkoksen ajan ja jatkamaan lähettämistä, kun verkon tilanne palautuu normaaliksi. (Elastic 2023b.)

4.3 Filebeat-agentti

Filebeat on beat-agentti, joka kerää ja lähettää lokeja muun muassa eri Linux-pohjaisista lähteistä niiden keskittämiseksi. Se asennetaan agenttina palvelimille monitoroimaan ja keräämään käyttäjän määrittelemiä lokitiedostoja ja lähteitä. Tietyin ennalta määritellyin välein se lähettää lokitiedostot joko Logstashiin tai suoraan Elasticsearchiin.

Kun Filebeat käynnistetään se alkaa seurata ennalta määriteltyjä lähteitä eli laitteiden lokitiedostoja. Filebeat-kerääjä lukee löydettyjä lokitiedostoja rivi riviltä ja lähettää eteenpäin vain uudet rivit lokitiedostosta. Se myös pitää kirjaa lukemistaan lokikirjauksista. Näin se pitää huolen siitä, ettei lokikirjauksia kerätä ja lähetetä moneen kertaan, vaikkei keskitetty lokienhallinta vastaisikaan ensimmäiseen lähetykseen. (Elastic 2023c.)

4.4 Logstash-työkalu

Alun perin Logstash oli avoimen lähdekoodin työkalu, jonka tarkoitus oli vain ottaa vastaan lokeja monista eri lähteistä ja lähettää ne eteenpäin yhtenä datastriiminä. Myöhemmin Elastic N.V sisällytti Logstashin Elastic Stackkiin, kun sitä kehitettiin hoitamaan suurempaa osaa lokienhallinnasta. (Horowits 2021.)

Nykyään Logstash on tiedonkeräysmoottori. Sen tarkoitus on edelleen ottaa vastaan lokeja Beateistä ja muista lähteistä, kuten Syslog-lähteistä ja normalisoida ne. Logstash pystyy reaaliaikaisesti yhtenäistämään ja puhdistamaan sisään tulevat lokitiedostot, ennen niiden lähettämistä eteenpäin Elastisearchiin. (Elastic 2023d.)

Osa Logstashin monipuolisuutta on sen kyky tukea lisäosia. Käyttäjät voivat valita vain ne osat, joita tarvitsevat, eikä Logstash kuormitu turhasta. Lisäosia (engl. plug-in) löytyy muun muassa tulevan lokidatan vastaanottamiseen eri lähteistä esimerkiksi dokumenteista, Beateistä, Syslogista ja HTTP:sta. Eri lokilähteistä tulevat lokit tarvitsevat erilaisia määrityksiä, kuten vastaanottoportti ja protokolla. Myös lokien normalisointi, suodatus ja lähetys eteenpäin tapahtuu lisäosien avulla. (Horowits 2021.)

4.5 Apache Lucene -projekti

Apache Lucene -projekti, joka on jatkuva, kehittää mm. avoimen lähdekoodin hakumoottoria nimeltä Lucene Core. Se on alun perin kirjoitettu Javalla, mutta sittemmin käännetty monelle ohjelmointikielelle. Lucene Core pystyy suorittamaan tehokasta datan indeksointia sekä hakemista ja on laajalti käytössä erilaisissa hakukoneissa, kuten Elasticsearch ja MongoDB Atlas search. Sen hyviä puolia ovat oikeinkirjoituksen tarkistus, hakujen korostaminen ja datan analysointi. (The Apache Software Foundation 2023.)(Bridgeman 2021.)

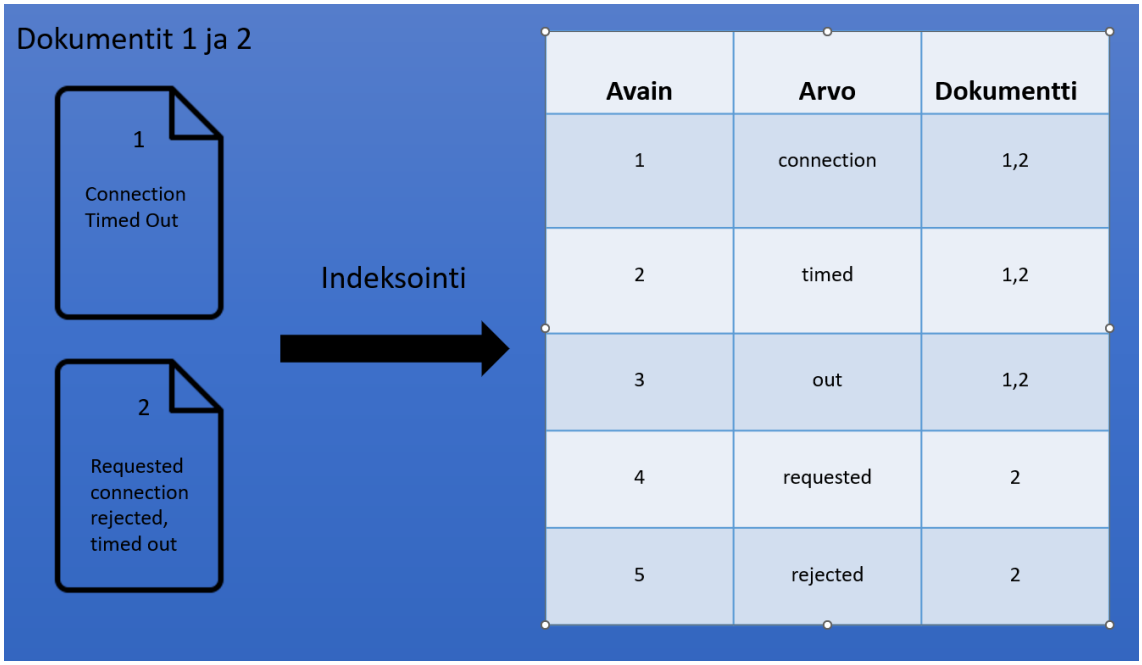
4.6 Elasticsearch-tietokanta

Elasticseach on Elastic Stackin sydän. Se julkaistiin alun perin avoimella lähdekoodilla vuonna 2010, modernina haku- ja analysointityökaluna, joka perustuu Apache Luceneen. Nykyään ohjelmisto on lisensoitu ja sen käyttö edellyttää kuukausimaksuja (Elastic 2023e). Sen tehtävä Elastic Stackissa on

lokidatan indeksointi ja säilytys. Elasticsearchia voidaankin kuvata hajautetuksi dokumenttivarastoksi. (Reback & Berman 2023.)

4.6.1 Indeksointi

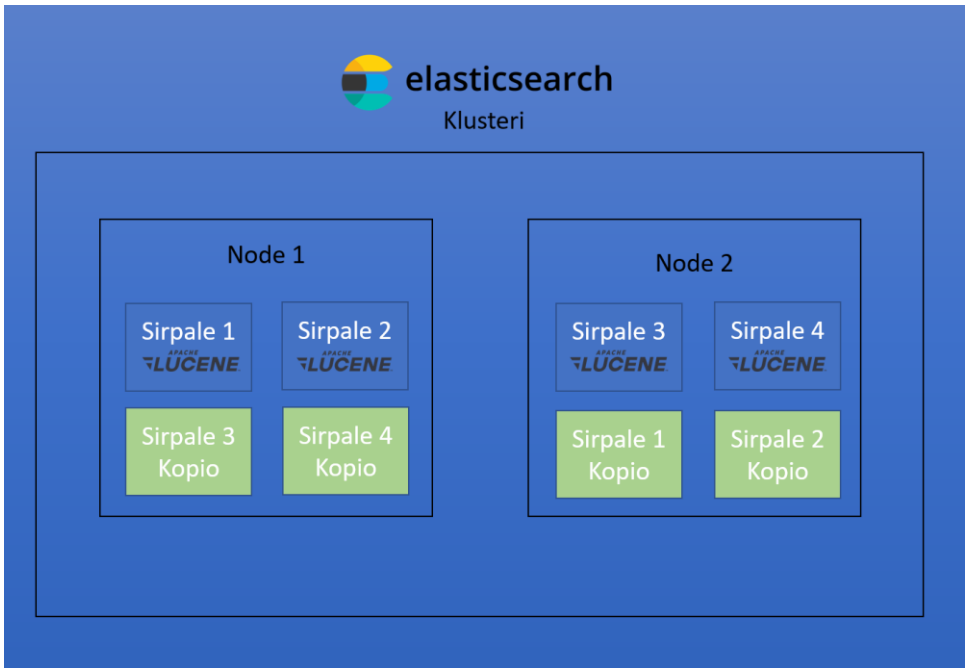
Elasticsearch tallentaa monimutkaisen datan jäsenneltyihin JSON-dokumentteihin. Dokumentit ovat kokoelmia kenttiä eli yksittäisiä lokikirjauksia. Ne rakentuvat avain-arvo pareina, eli jokaiselle arvolle on olemassa oma avaimensa, jolla sen löytää. Kun lokidataa tallennetaan dokumentteihin, jokainen tämän datan sisältämä esimerkiksi sana indeksoidaan, jolloin jokainen niistä eritellään hakukelpoiseksi ja kaikkia dokumentteja voidaan etsiä niiden sisältämillä sanoilla. Hakemalla tiettyä indeksiä, esimerkiksi varoitustyyppistä lokikirjausta, Elasticsearch hakee kaikki dokumentit, jotka sisältävät varoitustyyppisiä lokikirjauksia ja antaa vastaukseksi jokaisen yksittäisen varoitus-lokikirjauksen. (Elastic 2023f.) Esimerkissä kuvataan dokumenttien indeksointia hakukelpoisiksi avain-arvopareiksi (Kuva 5).



Kuva 5. Dokumenttien indeksonti avain-arvopareiksi. (Abueg 2020.)

4.6.2 Nodet ja sirpaleet

Elasticsearch-klusteri jakautuu nodeihin eli yksittäisiin palvelimiin. Nodeihin tallennetaan indeksejä, jotka voivat kasvaa todella suuriksi, jolloin tiedon hakeminen niistä vaikeutuu. Elasticsearch jakaakin indeksit pienempiin osiin eli sirpaleisiin (engl. Shard). Jokainen sirpale sisältää oman Apache Lucene -hakumoottorinsa. Yksittäinen node sisältää monta sirpaletta ja näin myös monta Apache Lucene -hakumoottoria. Esimerkissä kuvataan, miten Elasticsearch klusteri rakentuu nodeista ja sirpaleista (Kuva 6). (Bridgeman 2021.)



Kuva 6. Elasticsearch klusterin rakenne.

Elasticsearch luo sirpaleista kopioita toisiin nodeihin, joita kutsutaan kopioiksi (engl. Replica). Esimerkissä näkyy miten sirpaleet ja niiden kopiot jakautuvat nodeihin (Kuva 6). Tämä mahdollistaa sen, että kaikki data on tallennettu monena kappaleena ja vaikka yksittäinen node ei olisikaan saatavilla, tieto voidaan hakea sirpaleen kopiosta. Datan kopiointi toisiin nodeihin varmistaa sen saatavuuden, mutta myös vaatii tallennustilaa moninkertaisesti. (Elastic 2023f.)

4.7 Kibana-käyttöliittymä

Kibana on Elastic Stackin käyttöliittymä. Sen avulla pystytään hakemaan, seuraamaan ja analysoimaan dataa. Sillä myös hallitaan Elastic Stackin asetuksia ja vaikutetaan esimerkiksi siihen, kenellä käyttäjistä on oikeus mihinkin dataan. Kibana, kuten Logstash, rakentuu lisäosien päälle, jolloin käyttäjä voi itse valita mitä ominaisuuksia tarvitsee. Datan visualisointiin ja hakemiseen löytyy myös kattavasti erilaisia sisäänrakennettuja vaihtoehtoja. Etenkin isommille yrityksille, kuten Tietokeskus, on erityisen hyödyllistä, että lokidataa voi etänä tarkastella monta henkilöä samanaikaisesti. (Yigal n.d.)

5 Lokienhallintapolitiikka

Lokienhallintaa uudistettaessa tai kehitettäessä, kannattaa luoda selkeä lokienhallintapolitiikka. Politiikan tulisi sisältää määritelmät siitä, mitä lokeja kerätään ja miten, kuinka pitkäksi aikaa, miten lokeja säilytetään ja hävitetään, sekä kenellä on oikeus päästä lokeihin käsiksi. Politiikka tulee toteuttaa noudattaen lakeja, standardeja ja yleisiä käytänteitä. Valmis, kaikille selkeä politiikka varmistaa johdonmukaisen lähestymistavan lokienhallintaan koko organisaatiossa. Politiikalla voidaan myös varmistaa, että päätökset ja toimintatavat täyttävät lait ja säädökset lokienhallinnasta. Säännöllisellä valvonnalla voidaan vahvistaa, että kirjoitettu politiikka toteutuu ja ohjeita seurataan. (Kent & Souppaya, 2006.)

Seuraavissa alaluvuissa on käsitelty keskeisiä osa-alueita, jotka tulee sisällyttää lokienhallintapolitiikkaan. Vähintään tulee määritellä keskeiset käsitteet ja asiat sekä esittää niitä koskevat yleiset periaatteet. Osa-alueita ja niihin kohdistuvia vaatimuksia on selvitetty osittain yleisestä osittain toimeksiantajan näkökulmasta.

5.1 Keskeiset käsitteet

Politiikan alussa on suositeltavaa käydä läpi käytössä olevat lokienhallinnan käsitteet ja termit. Yhteiset ja selkeät käsitteet, jotka on määritelty politiikassa, auttavat eri tahojen analysointia, raportointia ja kommunikointia lokienhallinnasta. Toimeksiantajan käyttöön tulevat käsitteet kannattaa pohjata johonkin alan yleisistä standardeista, kuten MITRE:n yhteinen tapahtumailmaisu CEE (Common Event Expression). (MITRE 2010.)

5.2 Lokien suojaus

Jotta lokeja voidaan käyttää luotettavana tietolähteenä, niiden eheydestä täytyy varmistua. Lokit ovat yleensä todiste jostakin tapahtumasta, ja ne sisältävät monesti tietoturvaan ja järjestelmiin liittyviä kirjauksia. On tärkeää varmistaa, ettei lokikirjauksien sisältöä voi muuttaa, vaan ne voi ainoastaan poistaa

määritellyn säilytysajan päätyttyä. Tämä tarkoittaa sitä, että toimeksiantajan on suojattava lokia tuottavat järjestelmät ja ohjelmistot mahdolliselta manipuloinnilta. (Lokiohje 2009 61–63.)

Käyttäjien, mukaan lukien niiden, joilla on järjestelmänvalvojan oikeudet, ei tule pystyä muuttamaan, poistamaan tai kytkemään pois omista toimistaan kirjattavaa lokia. Toimeksiantajan lokienhallinnan tulee pystyä varmistamaan, ettei lokityyppejä tai -kirjauksia voi muuttaa tai poistaa. On myös varmistettava, että kaikki tarvittava loki tallennetaan onnistuneesti, eikä esimerkiksi tallennustilan loppuminen aiheuta lokien ylikirjoittamista. (SFS-EN ISO/IEC 27002 2022, 114–115.)

Eheyden varmistamiseksi on myös suojattava lokien siirto lokilähteestä keskitettyyn lokienhallintaan. Yksi keino tämän toteuttamiseen ovat tarkistussummat. Tarkistussummalla pyritään siihen, että alkuperäinen kirjattu loki pysyy muuttumattomana sen siirron ajan. Tarkistussumma on tiedoston digitaalinen allekirjoitus, jonka avulla voidaan varmistua siitä, ettei lokikirjaukseen ole tehty pientäkään muutosta. Yleisimmin käytettyjä suojaustekniikoita ovat MD5 ja SHA algoritmit. (Lokiohje 2009 63.)

Toimeksiantajan käyttöön tuleva Elastic Stack käyttää oletusarvoisesti lokitiedostojen siirtämiseen TLS-protokollaa (Transport Layer Security). TLS-protokolla käyttää sertifikaatti-avain pareja, jolloin data salataan ennen sen lähettämistä. Lähetetyn datan katselu vaatii oikeaa avainta salauksen purkamiseksi vastaanotto päässä. Elastic Stack tukee montaa TLS-protokollaversiota, jolloin on mahdollista vaikuttaa käytössä olevaan tiivistefunktioon. Muun muassa TLS 1.2 tukee turvallisena pidettyä SHA-384-hajautusta. (Elastic 2023g.)(Cloudflare 2021.)

5.3 Lokien säilytys

Lokien säilytyksellä tarkoitetaan lokitapahtumien, erityisesti tietoturvaan liittyvien, arkistointia. Lokien säilytys voidaan jakaa karkeasti kolmeen osaan. Ensimmäisenä on lokit, jotka täytyy lain mukaan säilyttää määrätyn ajan tai joita

saatetaan tarvita mahdollisissa oikeudellisissa ongelmissa. Esimerkiksi tietoturvalokien säilytysaika on hyvä olla vähintään kaksi vuotta, mutta suositus on vähintään viisi vuotta. Toisena ovat lokit, joita yritys haluaa säilyttää, mutta joita lainsäädäntö ei vaadi säilyttämään, esimerkiksi järjestelmien ylläpitolokit. Näille tyypillinen säilytysaika on kolmesta kuukaudesta vuoteen. Kolmantena ovat lokit, jotka halutaan poistaa, esimerkiksi henkilötietoja sisältävät lokit, joita ei ole perusteltua pitää. (LogicMonitor 2022.)

Lokia säilytetään yleensä kuuden ja 24 kuukauden välillä. Tietosuoja-asetus ei määrittele tarkkoja aikoja, kuinka kauan henkilötietoja pitää säilyttää, mutta niitä saa säilyttää vain niin kauan, kun niiden voidaan katsoa olevan tarpeellisia käyttötarkoituksen kannalta. Toimeksiantajan on pystyttävä perustelemaan säilytysajat. Lokiensäilytysajat sekä menetelmät säilytysaikojen umpeutuessa on määriteltävä politiikkaan eri lokia tuottavien järjestelmien ja lokityyppien osalta (Turun kaupunki Konsernihallinto Tietosuojavastaava 2019.). Toimeksiantaja voi määritellä politiikkaan samaan taulukkoon lokityypit, jotka se tulee säilyttämään ja niiden säilytysajat. (Kyberturvallisuuskeskus 2023.)

5.4 Lokien käyttöoikeudet ja vastuut

Lokien käsittely edellyttää tarkoin harkittua pääsynvalvontaa, joka tulee dokumentoida lokienhallintapolitiikkaan. Lokienhallinnan ylläpidon ohjausryhmän on annettava riittävät oikeudet ylläpitäjille päättää, toteuttaa ja valvoa lokienhallintaa. Lokienhallintaa ylläpitävien henkilöiden tulee vastata siitä, kenellä on oikeudet lokien tarkasteluun ja järjestelmän valvontaan; lokien käyttöoikeudet tulee luovuttaa vain sellaisille henkilöille, jotka niitä työtehtävissään tarvitsevat. Lokien tarkasteluun on annettava oikeudet vain valikoiduille käyttäjille (Lokiohje 2009 61–63). On varmistuttava siitä, että oikeudet omaavat henkilöt ovat riittävästi koulutettuja käsittelemään lokitietoja, sekä kenen vastuulla koulutus on. Toimeksiantajan on määriteltävä, kuka henkilö tai mikä ryhmä on viime kädessä vastuussa, jos ennalta määriteltyä politiikkaa rikotaan. (Turun kaupunki Konsernihallinto Tietosuojavastaava 2019.)

Elastic Stackiin on mahdollista helposti luoda käyttäjiä eri oikeuksilla. Käyttäjäoikeuksia on eritasoisia, ja niiden pääsy ja oikeudet on dokumentoitava. Toimeksiantajan tulee määritellä, mihin lokikirjauksiin kukin käyttäjä pääsee käsiksi ja millä tarkkuudella, esimerkiksi yksittäisiin lokikirjauksiin vai vain kirjausten tilastointiin. Kibanalla on mahdollisuus auditoida käyttäjien sisäänkirjautumisia ja yhteydenottoja, joista on kieltäydytty. Nämä lokikirjaukset sisältävät muun muassa aikaleiman, yhteyden IP-osoitteen ja laitteen nimen sekä tapahtuman.

5.5 Toimintaperiaatteet tietoturva- ja tietosuojarikkomuksissa

Lokien tietoturvaloukkauksella tarkoitetaan tapahtumaa jossa, tässä tapauksessa toimeksiantajan, vastuulla olevien lokien salassapito, saatavuus tai eheys vaarantuvat. Euroopan komission tietoturvaloukkausta käsittelevällä sivulla todetaan seuraavasti: ”Tietoturvaloukkaus asettaa yksilön oikeudet ja vapaudet vaaraan, joten kyseisessä tilanteessa yrityksesi/organisaatiosi on ilmoitettava asiasta valvontaviranomaiselle ilman aiheetonta viivytystä ja viimeistään 72 tunnin määräajassa tietoturvaloukkauksen havaitsemisesta.” (Euroopan komissio n.d.). Jos tietoturvaloukkauksen kohteeksi joutunut toimeksiantajan lokienhallinta on henkilörekisteri, on loukkauksesta ilmoitettava valvontaviranomaiselle. Kaikki tietoturvaloukkaukset eivät kuitenkaan aseta yksilön oikeuksia tai vapauksia vaaraan. Esimerkiksi lyhyt palvelunestohyökkäys ei aseta henkilötietoja vaaraan. On tärkeää varmistua, että toimeksiantajalla on riittävät tekniset ja hallinnalliset toimenpiteet, että tietoturvaloukkauksilta vältytään. (Euroopan komissio n.d.)

Lokienhallinnan ylläpidolla on oltava riittävä ohjeistus ja valtuutukset tietoturva- ja tietosuojarikkomusten selvittämiseksi. Politiikkaan on määriteltävä menettelyt kyseisissä rikkomuksissa. Menettelyssä tulee todeta, kuka tekee ja mitä tehdään, jos rikkomuksia epäillään. Tietoturvapoikkeamiin on vahvasti sidoksissa lokienhallinnan käyttöoikeudet ja vastuut. Lokienhallinnan ylläpidolle tulee määritellä henkilö tai ryhmä, jolta saa apua tietosuojarikkomuksissa, esimerkiksi toimeksiantajan juristi(t). Politiikkoihin voidaan sisällyttää

mahdollisia rangaistuksia sisäisistä tahallisista ja tahattomista tietoturva- ja tietosuojapoikkeamista. (Chuvakin ym. 2012, 267–303.)(Turun kaupunki Konsernihallinto Tietosuojavastaava 2019.)

5.6 Lokituksen lainsäädäntö

Lainsäädäntö asettaa vaatimuksia ja rajoituksia lokin sisällölle, säilytysajoille, käyttötarkoitukselle ja eheyden varmistamiselle. Lokin keräämisen ja käsittelyn on perustuttava lainsäädäntöön, niiden käsittelytapa ja -oikeus riippuvatkin siitä, millaista tietoa ne sisältävät ja mihin tarkoitukseen ne on alun perin kerätty. (Kyberturvallisuuskeskus 2023.)

Henkilötietoja ovat kaikki tiedot, joiden avulla voidaan tunnistaa henkilö suorasti tai epäsuorasti. Epäsuoralla tunnistamisella viitataan kahden tiedon yhdistämistä henkilön tunnistamiseksi. Yleisimpiä henkilötietoja tietoverkoissa ja lokeissa ovat nimi, sähköpostiosoite, henkilötunnus, sijaintitieto ja IP-osoite. Yrityksen yleinen sähköpostiosoite myynti@yritys.fi ei ole henkilötieto, toisin kuin matti.meikalainen@yritys.fi. (Tietosuojavaltuutetun toimisto 2019a.)

Jos lokitiedot sisältävät henkilötietoja, muodostuu lokista henkilörekisteri. Oikeusministeriön laatimassa tietosuojalaissa 5.12.2018/1050 todetaan, että mikäli henkilötietojen käsittely tapahtuu Euroopan unionin alueella rekisterinpitäjän toimesta, jonka toimipaikka on Suomessa, sovelletaan Suomen lakia. Tämä tarkoittaa sitä, että toimeksiantajan on noudatettava Suomen tietosuojalakia, joka täsmentää ja täydentää EU:n yleistä tietosuoja-asetusta. (Kyberturvallisuuskeskus 2023.) (Finlex 2018.)

EU:n yleinen tietosuoja-asetus GDRP, eli General Data Protection Regulation on henkilötietojen käsittelyä sääntelevää lainsäädäntöä. Asetusta sovelletaan kaikissa EU-maissa vuodesta 2018 lähtien. GDPR:sin tarkoituksena on parantaa ja yhtenäistää henkilötietojensuojaa ja tietosuojaoikeutta kaikissa EU-maissa. GDPR määrittelee, että organisaatio saa käsitellä henkilötietoja, mikäli sille on laissa määritelty peruste. Toimeksiantajalle näitä perusteita ovat mm.

rekisterinpitäjän tai kolmannen osapuolen etu, sopimus tai lakisääteinen velvoite. (Tietosuojavaltuutetun toimisto 2019b.)

Toimeksiantaja tulee kerämään lokia vain sisäisestä toiminnasta ja sen tarkoituksena on valvoa verkkoliikennettä ja käyttäjien toimintaa sisäisissä palveluissa, sisäisen kyberturvan ylläpitämiseksi. Tietosuoja-asetuksen johdanto-osassa perustelukappaleessa 49 todetaan, että mikäli tarkoituksena on kerätä esimerkiksi luvattoman sähköisiin viestintäverkkoihin pääsyn, palvelunestohyökkäyksien ja sähköisille viestintäjärjestelmille koituvien vahinkojen estäminen, on henkilötietojen käsittely oikeutettua (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, johdanto-osan perustelukappale 49).

Lokituksen kannalta EU:n tietosuoja-asetus mainitsee ja määrittelee muutamia tärkeitä kohtia (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5):

- Henkilötietoja saa käsitellä vain ja ainoastaan sitä tarkoitusta varten, jota varten ne on alun perin kerätty.
- Henkilötietojen keräys täytyy rajoittaa sellaisiin, jotka ovat asianmukaisia, olennaisia ja tarpeellisia siihen tarkoitukseen, mihin niitä kerätään.
- Henkilötietoja tulee säilyttää vain niin kauan, kuin se on tarpeellista niiden keräystarkoitukseen.
- Henkilöille joiden tietoja kerätään, on annettava selkeällä ja yksinkertaisella kielellä tietoa siitä, mitä, miten ja miksi heidän tietojaan kerätään ja käsitellään. Tämän voi toteuttaa esimerkiksi tietosuojakäytännöllä.

6 Pohdinta

Lokienhallinnankehittämisprojekti tavoitteena oli saattaa toimeksiantaja lähemmäksi ISO 27001 -standardisointia, käytännössä parantaa tietoturvatoimintaa ja tarjota parempia työkaluja sisäisten järjestelmien vianselvitykseen. Lokienhallintapolitiikan keskeisen sisällön tutkiminen ja suunnittelu määritettiin tämän opinnäytetyön tavoitteeksi.

Opinnäytetyön tavoitteena oli myös tutkia ja määritellä, mitä ovat lokikirjaukset, mitä on lokienhallinta sekä syventää omaa osaamista lokeista ja lokienhallinnasta. Työ käynnistyi tutustumalla alan kirjallisuuteen, asiantuntijaorganisaatioiden ohjeistuksiin sekä Elastic Stack -ympäristöön.

Työssä selvitetään kattavasti Elastic Stackin osia ja toiminnallisuutta. Järjestelmän toiminnan ymmärtäminen auttaa myös lokienhallintapolitiikan suunnittelussa. Reflektoitaessa lokienhallintapolitiikan suunnittelua ja toteutumista Elastic Stackin kanssa, todettiin sen oleva hyvä työväline politiikan tavoitteiden toteuttamiseen käytännössä.

Lopuksi selvitettiin lokienhallintapolitiikan sisältöä kirjallisuuden ja muun materiaalin pohjalta. Työssä saatiin määriteltyä ja kuvattua keskeiset käsitteet ja asiat, jotka tulee sisällyttää lokienhallintapolitiikkaan vähintään yleisten periaatteiden tasoisina. Aikarajoitteista johtuen jouduttiin opinnäytetyössä jättämään käsittelemättä muun muassa NIS2-direktiivi, joka on direktiivi toimenpiteistä yhteisen korkeatasoisen kyberturvallisuuden varmistamiseksi koko Euroopan unionissa. Aihetta tullaan tutkimaan ja siitä tullaan tekemään lisäyksiä Tietokeskuksen politiikkaan.

Opinnäytetyössä esitetyn pohjalta voidaan luoda yhteisesti hyväksytty, ajantasainen, selkeä lokienhallintapolitiikka, joka luo edellytykset yhtenäiselle lokienhallinnalle koko organisaatiossa. Opinnäytetyön kartoitusta voidaan hyödyntää myös muiden lokienhallintajärjestelmien toteutuksessa.

Lähteet

Abueg, R. (2020). What is Elasticsearch and what is it used for? Viitattu 11.5.2023 https://www.knowi.com/blog/what-is-elastic-search/.

AWS Whitepaper (n.d.). Benefits of centralized logs - Establishing Your Cloud Foundation on AWS. Viitattu 11.5.2023 https://docs.aws.amazon.com/whitepapers/latest/establishing-your-cloud-foundation-on-aws/benefits-of-centralized-logs.html.

Bridgeman, G. 2021. What Are Elasticsearch Shards? Why Do They Matter? Elasticsearch Cluster Architecture Explained. Viitattu 6.4.2023 www.youtube.com/watch?v=NxpZyQVO0K4&t=399s.

Chuvakin, A.; Schmidt, K. & Phillips, C. 2012. Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management. Waltham: Syngress.

Cloudflare 2021. What Is Transport Layer Security? Viitattu 9.4.2023 www.cloudflare.com/learning/ssl/transport-layer-security-tls/.

Elastic 2023a. “What Are Beats? Viitattu 1.4.2023 www.elastic.co/guide/en/beats/libbeat/current/beats-reference.html.

Elastic 2023b. Winlogbeat Overview. Viitattu 1.4.2023 www.elastic.co/guide/en/beats/libbeat/current/beats-reference.html.

Elastic 2023c. Filebeat Overview. Viitattu 1.4.2023 https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-overview.html.

Elastic 2023d. Logstash Introduction. Viitattu 1.4.2023 https://www.elastic.co/guide/en/logstash/current/introduction.html.

Elastic 2023e. Elastic Subscriptions. Viitattu 3.4.2023 https://www.elastic.co/subscriptions.

Elastic 2023f. Data in: documents and indices. Viitattu 3.4.2023 https://www.elastic.co/guide/en/elasticsearch/reference/current/documents-indices.html.

Elastic 2023g. Start the Elastic Stack with security enabled automatically. Viitattu 9.4.2023 https://www.elastic.co/guide/en/elasticsearch/reference/current/configuring-stack-security.html.

Euroopan komissio (n.d.). Mikä on tietoturvaloukkaus ja miten sellaisen sattuessa pitää toimia? Viitattu 22.4.2023 https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-data-breach-and-what-do-we-have-do-case-data-breach_fi.

Euroopan parlamentin ja neuvoston asetus (EU) 2016/679. Annettu 27.4.2016. https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex:32016R0679.

Finlex. 2018. Tietosuojalaki 5.12.2018/1050. Viitattu 9.4.2023 https://www.finlex.fi/fi/laki/ajantasa/2018/20181050.

Graylog Team, 2022. Why is Log Management Important? Viitattu 16.3.2023 https://www.graylog.org/post/why-is-log-management-important/.

Hombergs, T. 2018. Tip: Use a Human-Readable Logging Format. Viitattu 14.3.2023 https://reflectoring.io/logging-format.

Horowits, D. 2021. The Complete Guide to the ELK Stack. Viitattu 1.4.2023 https://logz.io/learn/complete-guide-elk-stack/#logstash.

Ite wiki 2019. ICT-palvelut. Viitattu 22.4.2023 https://www.itewiki.fi/opas/ict-palvelut/.

Jensen, P. 2018. Mikä on IP-osoite? Kotimikro. Viitattu 22.4.2023 https://kotimikro.fi/internet/verkko/mika-on-ip-osoite.

Kent, K. & Souppaya, M. 2006. Guide to Computer Security Log Management Recommendations of the National Institute of Standards and Technology. Viitattu 12.4.2023 https://doi.org/10.6028%2Fnist.sp.800-92.

Kyberturvallisuuskeskus 2023. Näin keräät ja käytät lokitietoja. Viitattu 6.3.2023 https://kyberturvallisuuskeskus.fi/fi/nain-keraat-ja-kaytat-lokitietoja.

LinuxWiki 2016. SHA. Viitattu 22.4.2023 https://www.linux.fi/wiki/SHA.

LogicMonitor 2022. What Is Log Retention? Viitattu 16.4.2023 https://www.logicmonitor.com/blog/what-is-log-retention.

Lokiohje 2009. Verkkodokumentti. Valtiovarainministeriö. Helsinki: Edita Prima Oy. Viitattu 11.3.2023 https://www.suomidigi.fi/sites/default/files/2020-06/pdf_3_2009.pdf.

Mauro, D. & Schmidt, K. 2001. Essentials SNMP. Viitattu 16.3.2023 https://doc.lagout.org/network/Essential%20SNMP%202001.pdf.

Microsoft (n.d.). Mikä on SIEM? Viitattu 22.4.2023 https://www.microsoft.com/fi-fi/security/business/security-101/what-is-siem.

MITRE 2010. Common Event Expression. Viitattu 16.3.2023 https://cee.mitre.org/docs/CEE_Architecture_Overview-v0.5.pdf.

OpenSource 2019. What is open source? Viitattu 22.4.2023 https://opensource.com/resources/what-open-source.

Reback, G. & Berman, D. 2023. An Elasticsearch Tutorial: Getting Started. Viitattu 6.4.2023 www.logz.io/blog/elasticsearch-tutorial/.

Sematext n.d. Log-file Viitattu 11.3.2023 https://sematext.com/glossary/log-file/.

SFS-EN ISO/IEC 27002 2022. Information security, cybersecurity and privacy protection. Information security controls. Helsinki: Suomen Standardisoimisliitto SFS ry.

Sharif, A. 2022a. What is Log Management? Importance & Best Practices. Viitattu 16.3.2023 https://www.crowdstrike.com/cybersecurity-101/observability/log-management/.

Sharif, A. 2022b. What is Centralized Logging? Viitattu 16.3.2023 https://www.crowdstrike.com/cybersecurity-101/observability/centralized-logging/.

Solarwinds 2023. What Is a Windows Event Log? Viitattu 1.4.2023 www.solarwinds.com/resources/it-glossary/windows-event-log.

The Apache Software Foundation 2023. Welcome to Apache Lucene. Viitattu 6.4.2023 https://lucene.apache.org/core/.

Tietokeskus 2023. Tietokeskus yrityksenä, Tietokeskuksen www-sivuilla. Viitattu 11.3.2023 https://www.tietokeskus.fi/tietokeskus/.

Tietosuojavaltuutetun toimisto 2019a. Mikä on Henkilötieto? Viitattu 12.4.2023 www.tietosuoja.fi/mika-on-henkilotieto.

Tietosuojavaltuutetun toimisto 2019b. Mikä on GDPR? Viitattu 13.4.2023 https://tietosuoja.fi/gdpr.

Turun kaupunki Konsernihallinto Tietosuojavastaava 2019. LOKIENHALLINTAPOLITIIKKA. Viitattu 16.4.2023 https://ah.turku.fi/ssnfu/2019/0212001x/Images/1685015.pdf.

Vanhatapio, J. 2020. Mikä on HTTP? Viitattu 22.4.2023 https://www.zoner.fi/tietoturva/http/.

Visma (n.d.). API- Mikä on API. Viitattu 27.4.2023 https://www.visma.fi/epasseli/kirjanpidon-sanakirja/a/api/.

Wikipedia 2021. Active Directory. Viirattu 22.4.2023 https://fi.wikipedia.org/wiki/Active_Directory.

Wikipedia 2022. ASCII. Viitattu 22.4.2023 https://fi.wikipedia.org/wiki/ASCII.

Wikipedia 2023a. JSON Viitattu 27.4.2023 https://fi.wikipedia.org/wiki/JSON.

Wikipedia 2023b. MD5. Viitattu 22.4.2023 https://fi.wikipedia.org/wiki/MD5.

Wutzke, M. 2019. Elastic Stack and Elasticsearch. Viitattu 22.4.2023 https://www.michael-wutzke.com/elastic-stack-and-elasticsearch-basics-and-tips/.

Yasar, K, 2022. Elastic Stack(ELK Stack). Viitattu 1.4.2023 https://www.techtarget.com/searchitoperations/definition/Elastic-Stack.

Yigal, A. n.d. Grafana vs. Kibana: The Key Differences to Know. Viitattu 6.4.2023 www.logz.io/blog/grafana-vs-kibana/.