Opinnäytetyö (AMK)

Tietojenkäsittelyn koulutusohjelma

2020

Marko Parkkonen

KESKITETYN LOKIENHALLINTAJÄRJESTELMÄN KEHITYS

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tietojenkäsittelyn koulutusohjelma

2020 | 48 sivua

Marko Parkkonen

KESKITETYN LOKIENHALLINTAJÄRJESTELMÄN KEHITYS

Tämän opinnäytetyön toimeksiantajana toimi 2M-IT Oy. Opinnäytetyön tavoitteena on dokumentoida ja toteuttaa yhtiön kyberturvatiimille keskitetty lokienhallintajärjestelmä käyttäen jo olemassa olevaa Graylog-järjestelmää. Tarkoituksena oli kehittää kyberturvatiimille enemmän työkaluja yhtiön tietoturvan parantamiseen hakemalla tiettyjä lokitietoja tietyistä tapahtumista.

Lokitiedolla tarkoitetaan jonkin laitteen tai palvelun keräämää tietoa tietyn ajan tapahtumasta. Lokitiedot sisältävät vaihtelevasti erilaista tietoa, joista yleisimpiä ovat aikaleimat, käyttäjätiedot, autentikointitiedot ja palvelimen tai laitteen vastauskoodi. Lokitietoja keräävät mm. tietokoneet, verkkolaitteet ja palvelimet. Yhtiön tietoturva paranee huomattavasti, kun käytössä on lokienhallintajärjestelmä, josta voidaan tarkastella tapahtumia ja niiden tietoja.

Toteutettavan järjestelmän päätehtävä oli kerätä ja tallettaa yrityksen tietoturvalokeja lyhyen aikaa, jotta niitä kaikkia voitaisiin tarkastella keskitetystä paikasta. Lokeja piti kyetä rajaamaan niin, että vain tietyt käyttäjät pystyivät tekemään hakuja ja näkymiä tietoturvalokeille. Järjestelmän oli myös tuettava määräaikoja jotka oli annettu yhtiön lokikuvauksessa, sekä mahdollistaa lokien kierrättäminen.

Järjestelmän laajentamisessa käytettiin apuna nykyistä järjestelmää ja sen dokumentaatiota uusien lähteiden lisäämisestä. Kyberturvatiimin lähteiden kanssa vaadittiin lisäselvityksiä esimerkiksi lokitiedostojen siirtämisestä Logstash-kerääjän kautta klusteriin.

Lokienhallintajärjestelmän laajennettu osa on kyberturvan käytettävissä, kun sen pitää päästä käsiksi DC- & DHCP-lokitietoihin. Lokitietojen haku on rajattu vain admin-käyttäjille, ettei niitä päästä tarkastelemaan muilta tasoilta. Tietoturvalokien tuominen klusteriin auttaa kyberturvatiimiä löytämään tarvitsemansa tiedon yhdestä sijainnista eri laitteiden lokien selaamisen sijaan.

ASIASANAT:

loki, Graylog, Logstash

BACHELOR´S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Business Information Technology

2020 | 48 pages

Marko Parkkonen

IMPROVEMENT OF A CENTRALIZED LOG MANAGEMENT SYSTEM

This Bachelor's thesis was assigned by 2M-IT. The aim for this thesis is to document and implement a centralized log management system for the company's cybersecurity team by using the existing Graylog system. The aim was to offer the cybersecurity team more tools to improve information security within the company by retrieving certain log information about certain events.

Log information refers to the information collected by some device or service about a certain event. The information within a log varies but the most common values are time stamps, user information, authentication information, and the response code from the device or service. Logs are collected for example by computers, network devices and servers. The information security within the company is greatly improved with a centralized log management system that can be used to examine events and the related information.

The main task for the implemented system was to collect and store the company's security logs for a short time to enable viewing form a centralized location. The logs were to be restricted such that only specific users could access information within the security logs. The system also had to support the time limits that were set in the company's log description and enable log rotation.

Parts of the current system and its documentation concerning adding a new log source were deployed to expand the system. Additional research was required about the cybersecurity team's log sources when for example transferring them to the cluster via a Logstash collector.

The expanded part of the centralized log management system is now available for the cybersecurity team for access to DC & DHCP logs. The search for the logs is restricted only to admin users to prevent examination by other user levels. By bringing the security logs to the cluster the cybersecurity team can access the required information from one place instead of searching the logs on different devices.

KEYWORDS:

log, Graylog, Logstash

SISÄLLYS

LYHENTEET **6**

1 JOHDANTO **7**
1.1 Mitä lokienhallinta on? 7
1.2 Toimeksiantaja 7
1.3 Toimeksianto 8
1.4 Tavoitteet 8

2 LOKIT **10**
2.1 Lokin määritelmä 10
2.2 Lokien luokittelu 10
2.3 Lokienhallinnan osa-alueet 11
2.4 GDPR 13

3 LOKITIETOJEN KERÄÄMISEN TYÖKALUJA **16**
3.1 Graylog 16
3.2 TCP & UDP verkkoprotokollat 20
3.3 Syslog lokiprotokolla 23
3.4 Logstash 25
3.5 Elasticsearch 27

4 KÄYTÄNNÖN OSA **29**
4.1 Nykyinen järjestelmä 29
4.2 Uuden lähteen lisäys Logstashin kautta 30
4.3 Uuden lähteen lisäys suoraan Graylogiin 36

5 LOPPUTULOKSET **40**

6 POHDINTA **44**

LÄHTEET **46**

KUVAT

Kuva 1 Graylog-käyttöliittymän aloitusnäkymä 17
Kuva 2. Graylogin virtojen prosessointi (Graylog docs – Streams, 2020) 18
Kuva 3. Esimerkki GELF-viestistä. (Graylog Docs – GELF, 2020) 19
Kuva 4 TCP toiminta 21
Kuva 5 UDP toiminta 22
Kuva 6 Syslog-tasot & funktiot 25
Kuva 7 Yrityksen NOC-arkkitehtuuri 29
Kuva 8 Uusi Graylog Sisääntulo 31
Kuva 9 DC-lokien indeksijoukon asetukset 32
Kuva 10 Onnistunut testiviesti Graylogissa 33
Kuva 11 Graylogiin saapuvaa lokidataa 36
Kuva 13 DHCP-lokien sisääntulo 40
Kuva 14 DC-lokien sisääntulo 41
Kuva 15 Lokien kierrätyksen määritelmät 42

LYHENTEET

AMQP	Advanced Message Queuing Protocol
CSV	Comma-Separated Values
GELF	Graylog Extended Log Format
GDPR	Global Data Protection Regulation
ICT	Information and Communication Technology
IP	Internet Protocol
SATSHP	Satakunnan sairaanhoitopiiri
SIEM	Security Information and Even Manager
SSH	Secure Shell
SQL	Structured Query Language
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VSHP	Vaasan sairaanhoitopiiri
VSSHP	Varsinais-Suomen sairaanhoitopiiri
XML	Extensible Markup Language

1 JOHDANTO

1.1 Mitä lokienhallinta on?

National Institute of Standards and Technology (NIST) kuvailee SP800-92 julkaisussaan lokienhallinnan olevan prosessi, jossa luodaan, siirretään, säilötään, analysoidaan ja hävitetään tietojärjestelmien lokidataa. Maanläheisemmin lokienhallinnalla tarkoitetaan tietojärjestelmien luomien tapahtumatiedostojen käsittelyä. Mitä tietoa halutaan kerätä? Mitense halutaan tehdä? Kauanko tietoa säilytetään?

Ilman lokienhallintaa IT-alan ongelmat olisivat paljon työläämpiä. Järjestelmien ylläpitäjät tietäisivät jonkin olevan vialla ja sen selvittämiseen kuluu paljon enemmän aikaa hakemalla yksittäiset tiedot jokaiselta laitteelta. Keskitetyllä lokienhallinnalla ongelmatilanteet tulevat nopeasti näkyviin yhteen paikkaan ja tiimi kykenee aloittamaan korjaustyöt ajoissa. Lokienhallinnan käyttäminen on ennakoiva toimenpide, jota jokaisen yrityksen tulisi käyttää hyväkseen. Käyttämällä lokienhallinnan koko työkaluskaalaa yritys pysyy potentiaalisten ongelmien edellä, saa paremman tilannekuvan järjestelmien sisäisestä toiminnasta, tehokkaamman tietoturvan, nopeamman ongelmanratkaisun kriisitilanteissa ja ylivoimaiset toimenpiteet ratkaisemaan yleiset ja ei-toivotut ongelmat. (Carstensen, 2018)

Lokienhallinnan avulla yrityksen kyberturvatiimi voi toimia tehokkaammin esimerkiksi viranomaisen pyytäessä yritykseltä tietoja liittyen vaikkapa tietomurtoihin. Tiimi kykenee keskitetyn lokienhallinnan avulla parsimaan lokitiedoista kaikki normaalilta vaikuttavat tapahtumat ja täten vähentämään viranomaisten työmäärää. Lokeja voidaan myös yhdistellä toisiinsa ja verrata poikkeamia esimerkiksi kirjautumisien ja liikennelokien välillä.

1.2 Toimeksiantaja

Työn toimeksiantajana toimi 2M-IT Oy. Yhtiö on Suomen suurin sosiaali- ja terveydenhuollon tietoteknisiä palveluja tuottava julkisomisteinen yhtiö. Yhtiö on fuusio Medbitin ja Medi-IT:n yhdistyttyä ja yrityksen toimipisteitä on ympäri Suomen. Yhtiön pääasialliset asiakkaat ovat muun muassa Varsinais-Suomen sairaanhoitopiiri, Satakunnan

sairaanhoitopiiri, Vaasan sairaanhoitopiiri ja Porin Perusturva, joille yhtiö tarjoaa kattavat tietotekniset palvelut kaikilla eri tasoilla.

”2M-IT auttaa asiakkaitaan strategisena kumppanina ja alaa perusteellisesti ymmärtävänä toimijana niin sote-alan päivittäisessä tekemisessä kuin pitkäjänteisessä kehitystyössä. Kokonaisratkaisuihin kuuluvat strateginen kehittäminen, asiakkuuden ja palvelu-hallinta sekä asiakkaan toiminnan mukaiset palvelut. Tuotetut ICT-palvelut ovat helposti skaalautuvia, minkä lisäksi turvataan toimintaympäristö ja palvelutuotanto.” (2M-IT, 2020.)

1.3 Toimeksianto

Opinnäytetyön toimeksiantona oli dokumentoida ja toteuttaa yhtiön kyberturvatiimille keskitetty lokienhallintajärjestelmä, jolla kyettäisiin tarkastelemaan tietoturvakomponenttien keräämää lokitietoa eri tapahtumista turvallisuuden lisäämiseksi. Järjestelmä sai vaatimusmäärittelyn, jossa tarkasteltiin, mihin kaikkeen järjestelmän olisi tarkoitus soveltua.

Toteutuksessa käytettiin yrityksen tietoliikennetiimin käytössä jo olevaa Graylog-järjestelmää, järjestelmää venytettiin ja jaettiin, jotta kyberturvatiimi saisi tarvittavat lokikeräykset järjestelmään. Graylog on avoimeen lähdekoodiin pohjautuva lokienhallintajärjestelmä. Järjestelmä on myös modulaarinen, joten sen toiminnallisuutta kyetään muuttamaan ja kasvattamaan erilaisin lisäosin käyttöönoton jälkeenkin.

1.4 Tavoitteet

Päätavoitteena järjestelmälle oli kerätä ja tallettaa laitteiden ja palvelujen lokitietoja käyttäen olemassa olevaa Graylog-järjestelmää. Järjestelmän haluttiin kykenevän vastaanottamaan lokeja uusista lähteistä, venyvän uusien ”nodejen” lisäämiseen suorituskyvyn parantamiseksi, suorittamaan tarkennettuja hakuja, jotka eivät keskeytä lokien keräämistä, ja asettavan katseluoikeudet lokikohtaisesti, jotta vain kyberturvatiimi pääsisi näkemään turvalokit.

Pääasiallisesti kyberturvatiimi halusi Graylogiin näkyviin Dynamic Host Control Protocol-palvelimen (DHCP) lokeja tarkastelemaan eri laitteille annettuja Internet Protocol-osoitteita (IP) ja Windowsin Domain Controller-palvelimen (DC) lokeja, joilla nähtäisiin Windows-laitteiden keräämiä tietoja käyttäjistään. Lokilähteiden lisäämistä varten tuli toteuttaa pohja ja suunnitelma, joilla tietoturvakomponenttien lokeja pystyttäisiin jatkossa lisäämään järjestelmään.

Lokeja varten piti myös toteuttaa kerättyjen lokien oikeaoppinen kierrätys Graylogin säädöksien avulla. Kierrätyksen täytyy täyttää yrityksen lokiperiaatteet. Järjestelmää piti myös tarkastella yleiskuvallisesti ja selvittää suorituskyvyn ja levykapasiteetin nykyinen tilanne laajennuksia varten.

Järjestelmää tulee muokata, jotta lokit voitaisiin toimittaa myös Security Information and Even Manageriin (SIEM) joko Logstashin tai Graylogin kautta. Uusien noodejen lisäämistä, lokikuvauksen yleiskirjaamista, roolioikeuksien muokkaamista lokikohtaisiksi sekä lokien parsimisen optimointia pyydettiin myös tarkastelemaan, jos se sopii aikatauluun

Kyberturvatiimi koostuu useista eri tietoturva-asiantuntijoista, jotka työskentelevät ympäri Suomea ja lokitiedostojen tarkastelun avulla he pystyvät paremmin tutkimaan ongelmatilanteita ja tarjoamaan viranomaisille heidän pyytämiään tietoja tarvittaessa.

2 LOKIT

2.1 Lokin määritelmä

Loki tarkoittaa aikajärjestyksessä kirjattua tallennetta tapahtumista ja niiden aiheuttajista. Tapahtumat ja muutokset tietojärjestelmissä, sovelluksissa, tietoverkoissa ja tietosisällöissä kirjataan lokiin eli lokitetaan. Lokia syntyy koko ajan kaikkialla. Tietokone kerää käyttölokia, langattoman verkon tukiasema ja lankaverkon reititin tallentavat tapahtumalokia, puhelimen operaattori kirjaa viestintälokia, jokapäiväiset ohjelmistot pääsynvalvontalokia, virhelokia ja niin edelleen. (Kyberturvallisuuskeskus, 2020.)

Lokeihin voidaan tallettaa useita eri tietoja ja seuraavassa listauksessa esitellään niistä muutamia:

- aikaleima: lokiin kirjatun tapahtuman aikaleima
- tapahtuma & toimija: kirjaa lokiin, mitä tehtiin tai yritettiin tehdä ja tekijä
- käyttöoikeus: millä valtuuksilla tai oikeuksilla tapahtuma tehtiin
- tapahtuman lähde: mistä tapahtuma tehtiin ja mistä muutostieto on peräisin
- tapahtuman tila: onnistuiko tekijä aikeissaan ja jos teko epäonnistui, kirjaa epäonnistumisen syyn.

(Kyberturvallisuuskeskus, 2020.)

Lokitietojen avulla reagoidaan häiriötilanteisiin ja etsitään vastauksia tapahtuneisiin muutoksiin tai väärinkäytöksiin järjestelmissä, laitteissa ja palveluissa. Lokit talletetaan yleensä ASCII-merkistöä tukevissa tekstitiedostoissa, jotta niitä voidaan tarkastella yleisimmillä tekstieditoreilla. Jotkut palvelut tallentavat lokit erilaisiin tiedostomuotoihin, esimerkiksi jotkin verkkolaitteet käyttävät pcap-tiedostomuotoa lokien tallettamiseen ja ne voidaan avata Wiresharkin tai WinDumpin avulla.

2.2 Lokien luokittelu

Lokeja voidaan luokitella muotonsa ja käyttötarkoituksensa mukaan. Muutamia eri lokityyppejä ovat esimerkiksi ylläpitolokit, käyttölokit eli tapahtumalokit, muutoslokit, virhelokit, viestintälokit ja haltijalokit.

- Ylläpitolokit kertovat järjestelmään tehdyistä muutoksista ja virhetilanteista. Näitä lokeja käytetään esimerkiksi versionhallinnassa ja toimintaympäristön kokonaisarkkitehtuurin seurannassa.
- Käyttölokit ovat tavallisin lokiformaatti. Käyttölokeihin rekisteröidään käyttäjän sisään- ja uloskirjautumiset, sekä tietoa järjestelmän suorittamista prosesseista. Tulostustapahtumat ja tietosisältöjen lukeminen jättävät myös merkinnän käyttölokiin.
- Muutosloki tallettaa tietoa mm. tietojen lisäämisestä, poistosta ja muutoksista.
- Virhelokeja käytetään ongelmatilanteiden ratkaisemisessa. Virheen tarkka kirjaaminen helpottaa korjaamista ja auttaa havaitsemaan sen aiheuttajan.
- Viestintälokit tallettavat tietoa kulkeneesta viestinnästä. Se tallettaa viestin alkuperän, päätepisteen ja tietoja ajankohdasta, määrätä, ylätunnisteen tiedoista ja tilasta. Monet sähköpostipalvelimet on asetettu kirjaamaan viestintälokia.
- Haltijaloki kertoo kenen omistukseen jokin tietty nettiosoite, puhelinnumero tai vuokra-auto on kuulunut. Haltijatieto voidaan yhdistää henkilöön, organisaatioon tai järjestelmään.

(Kyberturvallisuuskeskus, 2020.)

Luokittelemalla lokeja niiden sisältämiä tietoja voidaan löytää helpommin. Hakemalla tiettyyn tapahtumaan kuuluvia lokeja, turhan tiedon saa suljettua pois ja tuotua kriittisen tiedon nopeammin esiin. Tietokantaan tallettaessa tietyille lokityypeille voidaan antaa sopivat ylätunnisteet ja ne voidaan säilöä samaan sijaintiin nopeaa hakua varten.

2.3 Lokienhallinnan osa-alueet

Lokienhallintaan on monia eri syitä: tapahtumien seuranta, lisäturvan tarjoaminen, auditointi ja niin edelleen. Oli syy lokien keräämiseen mikä tahansa, lokienhallinnan prosessissa on useita eri osa-alueita: lokiperiaatteen määrittely, konfigurointi, keräys, normalisointi, indeksointi, säilöntä, korrelointi, vertailukohdat, hälytykset ja raportoinnit. Lokienhallinnan ratkaisujen valinnassa on syytä arvioida järjestelmän ominaisuudet ja kapasiteetti koko prosessi mielessä pitäen. (Grimes, 2010)

Lokiperiaate toimii yrityksen kulmakivenä sille, mitä aiotaan auditoida ja mistä hälytetään. Kattaako suunnitelma vain työasemat ja palvelimet, vai kuuluuko siihen myös

sovellukset ja verkkolaitteet? (Grimes, 2010.) Toimivaan lokiperiaatteeseen kuuluvat teknologia, ihmiset ja prosessit. Teknologia kattaa lokiperiaatteessa käytettävät työkalut. Ihmisillä tarkoitetaan lokien parissa työskenteleviä henkilöitä ja heidän roolejaan, ylläpitäjiä ja lokien käsittelijöitä. Prosessit ovat lokiperiaatteessa määritellyt keräys- ja säilöntätavat sekä lokien analysointi. Prosessien dokumentaation ylläpito on tärkeää jatkuvuuden varmistamiseksi sekä uuden työntekijän perehdyttämisessä. Dokumentoinnilla varmistetaan myös lokienhallinnan toiminta oikeusistuimen sitä vaatiessa. (Chuvakin ym. 2012.) Lokiperiaatteisiin vaikuttavat myös lainsäädännölliset vaatimukset, kuten 2018 eteenpäin EU:n asukkaiden yksityisyyttä suojaava tietosuoja-asetus GDPR.

Konfiguroinnissa päätetään mitä halutaan tallettaa lokiin ja miten. Monet lokienhallinnan palvelut tarjoavat valmiita ratkaisuja konfigurointeihin. Jokaisen käyttäjän täytyy itse tarkastaa mitä kaikkea heidän järjestelmänsä voivat tallettaa lokeihin ja hälyttää. Konfigurointi muuttaa lokiperiaatteessa määritellyt toimet käytännöksi. Konfiguroinnin jälkeen dataa aletaan keräämään. Kerääminen sisältää viestin lähettämisen laitteelta lokienhallinnan palvelimelle. Kerätty data parsitaan ja jaotellaan yksittäisiksi tunnistekentiksi. Parsittu data on helpompi indeksoida ja hakea. Parsimaton raakadata voidaan myös säilöä sellaisenaan, mutta sen indeksointi ja hakeminen on parsittua dataa vaikeampaa. Normalisointi on datan jakamista samantyyppisiin formaatteihin tietokannan sisällä. Indeksoinnilla nopeutetaan datan suodatusta ja hakua tietokannasta erilaisten ylätunnisteiden avulla. Kerätty data on säilöttävä johonkin. Kaikki tuotteet tallettavat tietoa paikallisille levyille, jotkut saattavat tallettaa sen erilaisiin verkkolevysijainteihin. (Grimes, 2010.)

Korreloinnilla tarkoitetaan samantyyppisten tapahtumien tunnistamista ja yhdistämistä toisiinsa. Korreloinnin avulla jotkin lokienhallinnan palvelut kykenevät tunnistamaan vaikkapa verkkohyökkäyksen tavallisesta raportoinnista. Vertailukohdilla järjestelmät tarkkailevat mikä on normaalia kyseisessä verkkoympäristössä. Vertailukohtana voidaan pitää vaikkapa useaa epäonnistunutta kirjautumisyritystä. Vertailukohtaan määritellään, kuinka monta epäonnistunutta kirjautumisyritystä sallitaan, kunnes se raportoidaan epäilyttäväksi toiminnaksi. (Grimes, 2010.)

Hälytyksen tapahtuessa, siitä on pikaisesti toimitettava tieto tiimille korjaustoimenpiteiden aloittamisen nopeuttamiseksi. Viesti voidaan toimittaa tekstiviestinä tai sähköpostina asianomaisille. Jossain tapauksissa hälytyksistä kirjautuu automaattisesti tiketti järjestelmään ja sitä kautta se kulkeutuu oikealle vastuuhenkilölle. Raportoinnin avulla kaikista kerätyistä tapahtumista voidaan muodostaa pitkäaikainen vertailukohta. (Grimes, 2010.)

Lokitietoja käsitellessä on hyvä ottaa huomioon pykälät 137 ja 138 sähköisen viestinnän palvelujen laissa, joissa mainitaan sähköisten viestien ja välitystietojen käsittelyn vain siinä määrin kuin se on tarpeellista. Pykälässä 137 mainitaan tietojen luovutuksen olevan sallittua vain niille, joilla on oikeus käsitellä tietoja asianomaisessa tilanteessa, sekä tietojen jälkikäsittely tilanteen jälkeen. (Laki sähköisen viestinnän palveluista, §137, 2014/917.) Pykälän 137 sisältö nousee esiin esimerkiksi luovuttaessa lokitietoja viranomaisille rikostutkintaa varten. Pykälässä 138 kerrotaan viestinnän välittäjän olevan velvoitettu ilmoittamaan tilaajalle tai käyttäjälle, millaisia tietoja käsitellään ja kuinka kauan. Pykälässä 138 viitataan myös pykälään 272, jossa käydään läpi erilaisia toimia, joihin viestin välittäjä, lisäarvopalvelujen tarjoaja sekä niiden lukuun toimiva voi ryhtyä tietoturvasta huolehtiakseen. (Laki sähköisen viestinnän palveluista, §138, §272, 2014.)

Pykälän 272 toimiin kuuluvat:

1) viestintäverkkojen tai niihin liitettyjen palvelujen sekä tietojärjestelmien tietoturvalle haittaa aiheuttavien häiriöiden havaitsemiseksi, estämiseksi, selvittämiseksi ja esitutkintaan saattamiseksi;
2) viestin lähettäjän tai viestin vastaanottajan viestintämahdollisuuksien turvaamiseksi; tai
3) viestintäpalvelujen kautta laajamittaisesti toteutettavien rikoslain 37 luvun 11 §:ssä tarkoitettujen maksuvälinepetosten valmistelun ehkäisemiseksi.

- (Laki sähköisen viestinnän palveluista, §272, 2014/917.)

Edellä 1 momentissa tarkoitetut toimet voivat käsittää:

1) viestin sisältöä koskevan automaattisen selvittämisen;
2) viestien välittämisen ja vastaanottamisen automaattisen estämisen tai rajoittamisen;
3) tietoturvaa vaarantavien haitallisten tietokoneohjelmien automaattisen poistamisen viesteistä;
4) muut 1–3 kohdassa tarkoitettuihin rinnastettavat teknisluonteiset toimenpiteet.

- (Laki sähköisen viestinnän palveluista, §272, 2014/917.)

2.4 GDPR

Tietosuoja-asetus

Global Data Protection Regulation eli GDPR on Euroopan Unionin yleinen tietosuoja-asetus, jonka tarkoituksena oli tarjota suojaa henkilön perusoikeuksia ja -vapauksia kohtaan. Asetus annettiin ja hyväksyttiin EU:n parlamentissa 27.4.2016 ja yritysten täytyi ottaa se soveltavasti käyttöön viimeistään 25.5.2018. Yleisessä tietosuoja-asetuksessa asetetaan vaatimukset koskien henkilötietojen keräämistä, säilytystä ja hallintaa varten. Vaatimukset koskivat sekä EU:n että sen ulkopuolella toimivia yrityksiä, jos ne käsittelivät henkilötietodataa EU:n asukkaista.

Asetus korvasi vuonna 1997 voimaan astuneen direktiivin 95/46/EY, joka koski henkilötietojen käsittelyä ja tiedon vapaata liikkuvuutta. Uuden tietosuoja-asetuksen tavoitteena oli yksinkertaistaa ja selkeyttää henkilötietoja koskevaa lainsäädäntöä kaikissa Euroopan Unionin maissa. ((EU) 2016/679)

Henkilötietoihin kuuluvat

- nimi
- osoite
- henkilökortin/passin numero
- tulot
- kulttuurinen profiili
- IP-osoite
- terveydenhuollossa käytettävät yksilöivät tiedot.
  (Sinun Eurooppasi, Yleinen tietosuoja-asetus. 2020)

Tietosuoja-asetuksen vaikutus lokien keräämiseen

Tietosuoja-asetuksen alaisena hallintalokien, virhelokien ja tietoturvalokien katsotaan sisältävän henkilötietoja. Yritysten täytyy suojata IP-osoitteet ja selainkeksien tiedot, sillä näiden avulla voidaan tunnistaa henkilöitä. Lisäksi henkilötietoja ei voida kerätä tai säilöä ilman todisteita yksilön hyväksynnästä. Henkilötietoja voidaan kuitenkin kerätä ja säilöä verkkopalvelimien lokeilla väärinkäytöksien ja luvattoman käytön estämiseksi. (Graylog, 2020c. GDPR)

Lokitietoja kerätessä tulisi rajoittaa tietojen kerääminen vain tarpeellisimpiin osiin. Jos tapahtumia seurataan liian yksityiskohtaisesti, se saattaa rikkoa tietosuoja-asetusta ja yksilön tietosuojaa.

Määrätyn rekisterinpitäjän tulee säilyttää rekisteröityjen tietoa vain niin kauan kuin se on tarpeellista ja niistä on säädettävä määräaikainen säilytysaika, jonka jälkeen rekisterinpitäjän on tarkastettava tai poistettava tiedot. Säilytettävien tietojen päätyminen varmuuskopiointeihin ei vaadi tietojen välitöntä poistoa, jos siitä aiheutuisi vaivaa pitäjälle. Rekisterinpitäjä on kuitenkin vastuussa myös varmuuskopioiduista henkilötiedoista ja tietoturvaloukkauksen sattuessa menetetyistä tiedoista. Rekisteröity voi pyytää rekisterinpitäjää poistamaan tietonsa myös varmuuskopioinneista ja jos poistoa ei voida suorittaa välittömästi pyynnön seurauksena, tulee siitä ilmoittaa rekisteröidylle ja asettaa määräaika tietojen poistoa varten. (EU 2016/679)

Läpinäkyvyys, ohjeidenmukaisuus ja vastuullisuus ovat yleisen tietosuoja-asetuksen pääpiirteitä, joihin yrityksien täytyy myös pyrkiä. Yritysten täytyy tarjota selvä kommunikointi liittyen käyttäjien henkilökohtaiseen dataan ja miten sitä aiotaan käyttää, sekä tarjota vaadittu ilmoitus mahdollisesta tietoturvarikkeestä. Yrityksen on kerrottava asiakkailleen, miten dataa käsitellään ja prosessoinnin on vastattava annettua kuvausta. (Graylog 2020e. GDPR)

Jotta yritys voisi toimia täysin tietosuoja-asetuksen mukaisesti, yrityksen on dokumentoitava rekisteröityjen datan käyttö täysin. Yrityksen asiakkailla on myös oikeus tulla unohdetuiksi ja rekisteröidyn tätä pyytäessä on yrityksen tarjottava heille tämä oikeus poistamalla data yrityksen järjestelmistä. Datastruktuurin on oltava kykeneväinen suoriutumaan poistosta missä vaiheessa tahansa ilman, että se aiheuttaa vahinkoa muihin järjestelmiin. Yrityksen on myös oltava valmis tulemaan tarkastetuksi valvovan viranomaisen puolesta. Datan pitää olla tarjottavana tarkastavalle viranomaiselle sellaisessa muodossa, joka voidaan katselmoida turvallisesti. (Graylog 2020e. GDPR)

Seuraavissa kappaleissa tarkastelemme yrityksen lokienhallintajärjestelmässä käytettäviä työkaluja kuten Graylog, Logstash ja Elasticsearch.

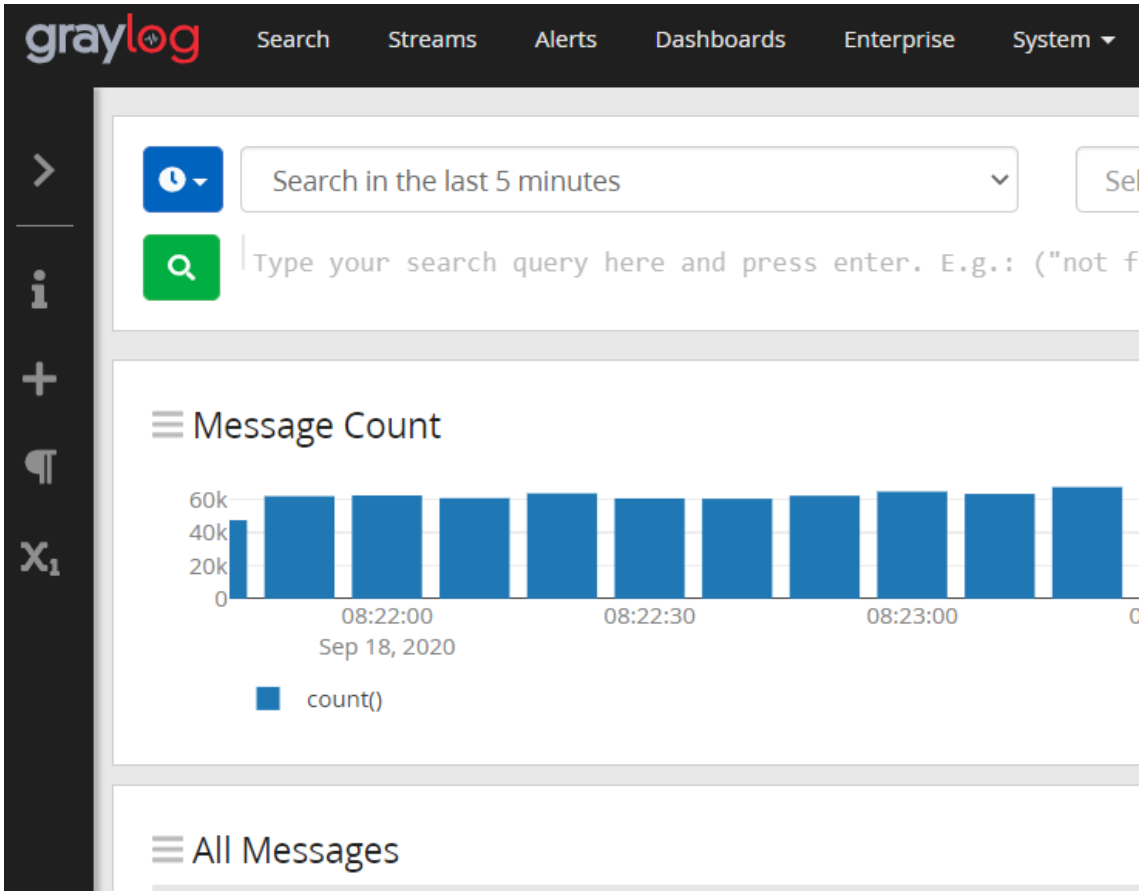
3 LOKITIETOJEN KERÄÄMISEN TYÖKALUJA

3.1 Graylog

Graylog on avoimeen lähdekoodiin perustuva modulaarinen lokienhallintajärjestelmä. Järjestelmän avulla pystytään tarkastelemaan laitteilta ja palveluilta tulevaa lokidataa reaaliaikaisesti ja piirtämään niistä datavirtoja tarkkailemaan esimerkiksi saapuvien lokitiedostojen määrää ja niiden sisältöä. Palvelun taustalla toimii palvelin, joka vastaanottaa verkon yli lähetettyjä TCP-, UDP- tai AMQP-protokollan viestejä, jotka se tallettaa Elasticsearch-tietokantaan, josta ne voidaan hakea käyttöön myöhemmin. Järjestelmän modulaarisuudella tarkoitetaan sitä, että siihen voidaan lisätä käyttöönoton jälkeenkin uusia liitännäisiä ja tukipalveluja suorittamaan työtehtäviä.

Käyttöliittymä

Graylog käyttää verkkopohjaista käyttöliittymää, johon voidaan luoda näkymiä saapuvista lokiviesteistä. Käyttöliittymässä kyetään myös suorittamaan lokikohtaisia hakuja tietokantaan, joiden avulla saadaan haettua esimerkiksi tietyn ajankohdan tapahtumia.



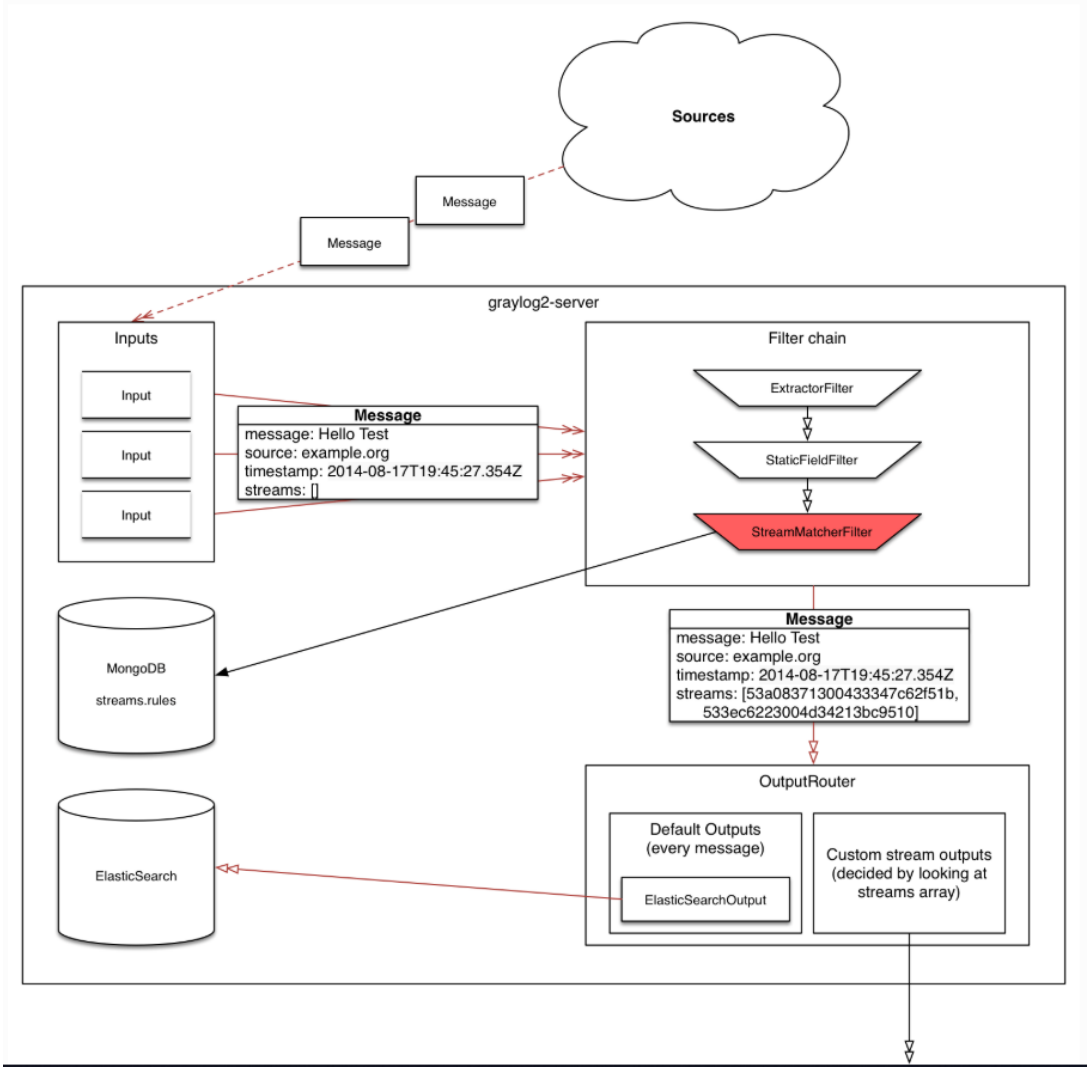
Kuva 1. Graylog-käyttöliittymän aloitusnäkymä.

Verkkokäyttöliittymän aloitusnäkymä näyttää saapuneiden lokiviestien määrän ja niiden aikaleimat, eri välilehdet yläpalkissa, sivupalkin lisäkomennot, sekä selaimen yläoikealla noodien lähettämän ja vastaanottaman viestimäärän summattuna.

Streams-välilehdellä Graylogiin saapuvaa dataa voidaan jakaa virtoihin, jotka toimivat reaaliaikaisina ja jatkuvina tallennettuina hakuina. Graylogiin asetetut noodit vastaanottavat saapuvat lokitiedot ja jakavat ne omiin virtoihinsa näiden konfiguroinnin mukaan, käyttöliittymän Inputs-välilehdellä voidaan tehdä muutoksia konfigurointeihin ja määrittää, millaisessa muodossa noodi vastaanottaa lokiviestit esimerkiksi Raw/Plaintext TCP. Talletettujen hakujen ja virtojen välinen ero on se, että virtoja prosessoidaan reaaliaikaisesti. Tämä mahdollistaa reaaliaikaisen varoituksen ja sen lähettämisen eteenpäin kaikille järjestelmille. (Graylog 2020d. Streams)

Streams-välilehti näyttää Inputs-välilehdellä tehtyjen konfigurointien virtaukset listana, josta käyttäjä voi helposti klikata haluamansa virran auki tarkkailuun. Välilehdellä voi myös muokata virran sääntöjä, ulostuloja ja hälytyksiä. Sääntöjä voi myös testata

pysäyttämättä tuotannossa olevaa virtaa lataamalla muokatun säännön avulla haettu lokiviesti. Kuva 2 on lisätty havainnollistamaan Graylogin virtojen toimintaa.



Kuva 2. Graylogin virtojen prosessointi. (Graylog 2020d. Streams, 2020.)

Käyttöliittymän Alerts-välilehdelle käyttäjä saa näkymän Graylogin kaappaamista häiriötekijöistä. Välilehdellä voi myös suorittaa hakuja hälytyksille tietyltä aikaväliltä tai palvelimelta. Jos yrityksellä on käytössä Enterprise-ominaisuus ja lisenssi siihen, Alerts-välilehdellä voi myös tehdä korrelointeja tapahtumista tarkemman yleiskuvan saamiseksi tapahtumien kulusta. Hälytyksien ilmoituksen voi konfiguroida siten, että käyttäjä saa hälytyksestä esimerkiksi sähköpostiviestin käyttöliittymän dashboardissa näkyvän ilmoituksen lisäksi. (Graylog 2020a. Alerts)

Käyttöjärjestelmässä on myös Dashboards-välilehti, johon käyttäjä voi tehdä ennalta määriteltyjä hakutoimintoja. Käyttäjän ei siis tarvitse suorittaa hakuja erikseen, vaan haut

voidaan tallentaa yhteen välilehteen, josta ne saa helposti valmiiksi näkymiksi. Vain ylläpitäjä voi tehdä muutoksia Dashboard-näkymässä, lukuoikeuden omaavilla on nähtävissä vain ne dashboardit, jotka adminit ovat heille määrittäneet näkyviksi.

GELF-lokiformaatti

Graylog käyttää lokiformaattinaan kehittäjien itsensä luomaa GELF-formaattia. GELF-formaatin alkuperäinen tavoite oli korjata Syslog-protokollassa esiintyvät heikkoudet, joita olivat esimerkiksi rajattu 1024 tavun pituus sekä lokitietojen kompressoinnin puutteellisuus.

Koska GELF tukee suuremman bittimäärän rajaa, protokollalla lähetetyt lokitiedot voivat kasvaa hyvinkin suuriksi. Graylog käyttää ”chunked” GELF-muotoa, jossa data jaetaan osiin, jotta ne olisi helpompi lähettää vastaanottajalle. Viestiä paloiteltaessa niiden otsikoihin sisällytetään erilaisia tunnistekenttiä, kuten viestin GELF ID, järjestysnumero ja muiden osien määrät. Vastaanottaja kokoaa palat yhteen kenttien avulla ja saa itselleen kokonaisen GELF-viestin. GELF on JSON-merkkijono, joka kompressoidaan GZIP- tai ZLIB-pakkausmuotoon. (Graylog 2020b. GELF)

Kuvassa 3 on esimerkki tyypillisestä GELF-viestistä.

```
{
  "version": "1.1",
  "host": "example.org",
  "short_message": "A short message that helps you identify what is going on",
  "full_message": "Backtrace here\n\nmore stuff",
  "timestamp": 1385053862.3072,
  "level": 1,
  "_user_id": 9001,
  "_some_info": "foo",
  "_some_env_var": "bar"
}
```

Kuva 3. Esimerkki GELF-viestistä. (Graylog 2020b. GELF)

3.2 TCP & UDP verkkoprotokollat

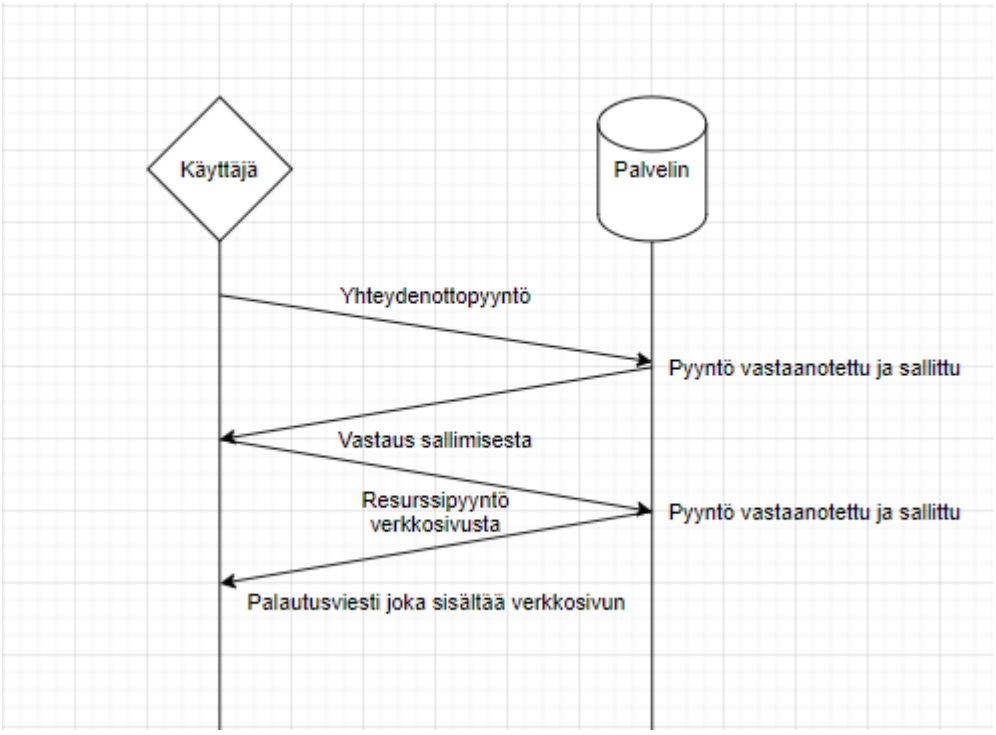
Yleistä

Jokaisella verkkoon kytketyllä laitteella on oma IP-osoite (Internet Protocol), joka muodostuu numeroista ja desimaaleista. IP-osoitteet mahdollistavat laitteiden löytävän toisensa, sekä kommunikoinnin eri laitteiden välillä ja täten muodostavat Internetin. Kommunikoidessaan toistensa kanssa, laitteet lähettävät dataa edestakaisin käyttäen protokollia. Internetprotokollat jakavat datan pieniin lähetettäviin paketteihin. Paketit eivät välttämättä aina saavu samaan aikaan tai samassa järjestyksessä perille, joten niitä varten tarvitaan erillisiä protokollia kasaamaan ne oikeaan järjestykseen vastaanottavalle laitteelle. Näitä protokollia ovat esimerkiksi tämän työn järjestelmissä viestien lähettämiseen käytetyt TCP ja UDP. Protokollat ovat käytössä yrityksen järjestelmässä niiden helppokäyttöisyyden ja yhteensopivuuden vuoksi.

TCP

TCP (Transmission Control Protocol) on yksi yleisimmistä protokollista tiedostonsiirrossa, siitä voidaan myös käyttää nimitystä TCP/IP. IP jakaa paketin pienempiin osiin, paketteihin, ja lähettää ne internetin yli kohdelaitteeseen. TCP kehitettiin kasaamaan saapuvat paketit alkuperäiseen järjestykseen käyttämällä pakettien ylätunnisteissa olevia järjestysnumeroita. TCP vaatii kommunikointia lähettäjältä ja vastaanottajalta, tämän vuoksi viestien toimitus on usein hitaampaa. Jos yksittäinen paketti katoaa tai jää puuttumaan, koko lähetysprosessi pysähtyy kunnes se saadaan lähetettyä uudestaan. Vaikka tämän ongelman ajallinen pituus on vain joitain millisekunteja se voi silti vaikuttaa suuresti järjestelmien toimintaan. TCP varmistaa viestien saapumisen perille mikä tekee siitä todella monimuotoisen protokollan, tämän vuoksi se onkin yleisimmin käytetty protokolla Internetissä. (Internet Engineering Task Force, RFC 793)

TCP Painottaa toiminnassaan luotettavuuteen. Toiminta perustuu vuoropuheluun lähettäjän ja vastaanottajan välillä. Tekemällä pyynnön esimerkiksi verkkoselaimelle, lähettäjä toimittaa pyynnön verkkosivusta TCP-pakettien avulla selaimen palvelimeen. Palvelin vastaa palauttamalla TCP-paketteja, jotka sisältävät verkkosivun. Sama periaate toimii kaikissa käyttäjän tekemissä toiminnoissa verkkosivulla. Linkin klikkaaminen, kirjautuminen, päivityksen jakaminen, kaikista näistä käydään vuoropohjainen keskustelu TCP-pakettien avulla.



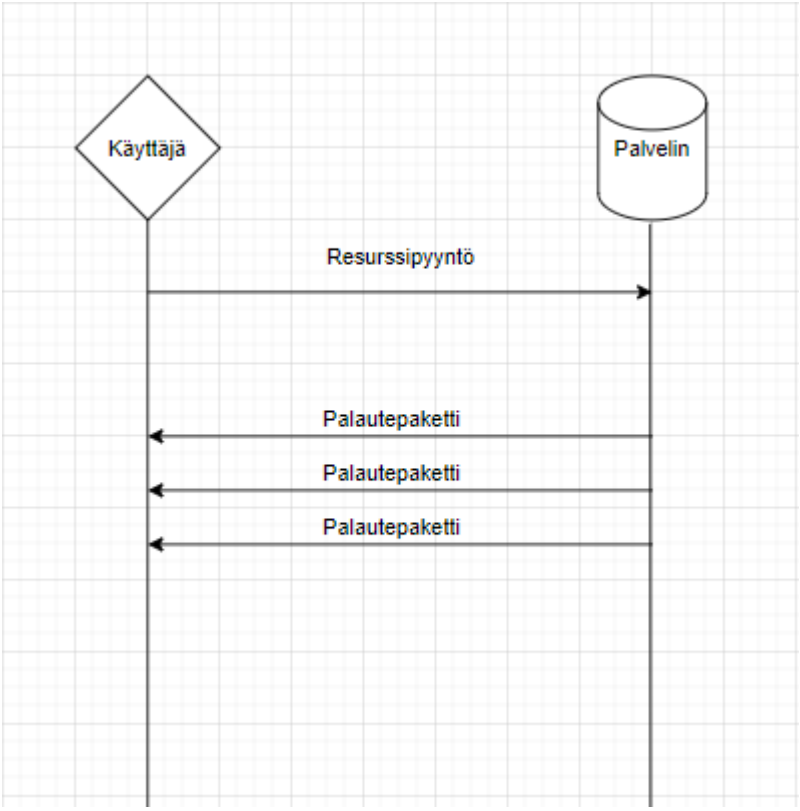
Kuva 4. TCP toiminta.

UDP

UDP (User Datagram Protocol) on TCP:n yksinkertaistettu ja nopeampi versio. UDP ei vaadi kohteelta vastausta pakettien vastaanottamisesta kuten TCP. Lähetyksen aikana kadonneita paketteja ei lähetetä uudestaan kohteelle, vaan ne pudotetaan. Datapaketin pudottaminen esiintyy esimerkiksi pätkivänä ääni- tai videoyhteytenä. UDP:n ansiosta laitteet kykenevät kommunikoimaan nopeammin keskenään, mutta datan eheydestä ei ole varmuutta. UDP-paketeilla ei myöskään ole ylätunnisteessa olevaa järjestysnumeroa, paketit saattavat siis saapua vastaanottajalle väärässä järjestyksessä. Paketit

sisältävät tarkastussumman (checksum) jota käytetään virheiden tunnistamiseen. (Internet Engineering Task Force, RFC 768)

UDP painottaa viestinnässä nopeutta ja sitä käytetään silloin kun pakettien eheydellä ja virheiden korreloinnilla ei ole suurta merkitystä. Yleisimpiä käyttökohteita ovat suoratoistot, ääni- ja videopuhelut verkon yli (Voice over Internet Protocol - VoIP) ja verkossa pelattavat videopelit.



Kuva 5. UDP toiminta.

Seuraavassa listassa vertaillaan lyhyesti TCP ja UDP protokollia

TCP	UDP
Tarkastaa lähetettyjen pakettien määrän vastaanotettujen viestien ylätunnisteesta	Ei tarkasta pakettien määrää.
Tarkistaa virheelliset paketit ja lähettää ne uudestaan kohdelaitteelle.	Tarkastaa ja pudottaa virheelliset paketit lähettämättä niitä uudestaan.
Käyttää kättelyä (handshake) pyyntöjen varmistukseen.	Ei suorita kättelyä prosessissa
Hidas protokolla, tähtää datan saapuvan varmasti perille	Nopeampi protokolla, ohittaa virheiden korjaamisen ja tähtää nopeuteen.
Käytetään pääasiallisesti tilanteissa, joissa datan odotetaan saapuvan ehjänä ja varmasti perille, kuten sähköposteissa	Käytetään tilanteissa, joissa nopeudella on väliä ja varmuudella ei. Esimerkkeinä suoratoistot ja videopuhelut.

3.3 Syslog lokiprotokolla

Syslogin historia

Syslog-protokolla kehitettiin 1980-luvulla Eric Allmanin toimesta sendmail-sovellusta varten ja se on toiminut UNIX-järjestelmien standardina lokituksen käytäntönä tähän päivään asti. Protokolla sallii laitteiden lähettävän tapahtumailmoituksia toisilleen IP-verkkojen yli ilmoituksia keräävälle laitteille, Syslog-palvelimille. Protokolla yksinkertaisesti kuljettaa tapahtumaviestin sen luojalta viestin keräävälle laitteelle. Keräävä osapuoli ei lähetä palauteviestinä ilmoitusta paketin saapumisesta. (Deveriya, 2005)

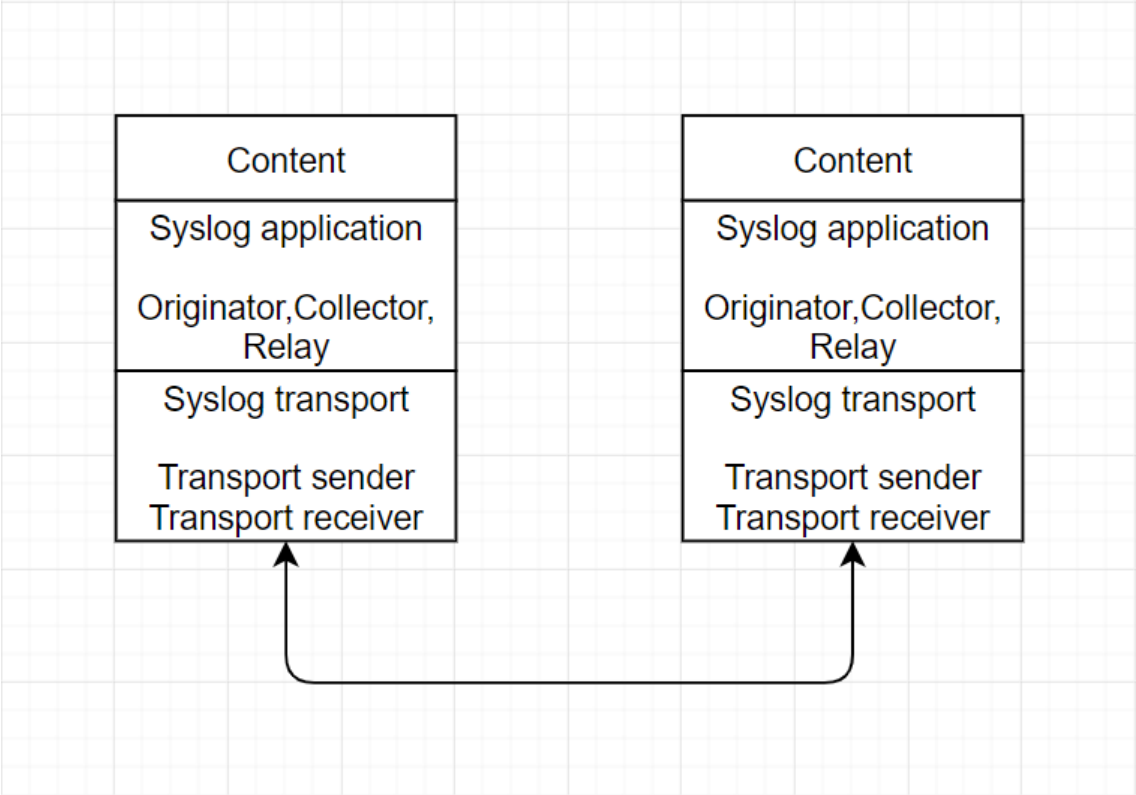
Protokollalla ei ollut pitkään aikaan määriteltyä standardia, joten siitä oli käytössä monia toteutuksia, jotka eivät olleet yhteensopivia. Vasta vuonna 2001 IETF julkaisi protokollalle ensimmäisen standardin, RFC 3164-dokumentin nimeltään ”The BSD Syslog Protocol”. Standardin avulla protokollaa saatiin yhtenäistettyä ja se muuttui yhteensopivammaksi muiden järjestelmien kanssa. Vuonna 2009 julkaistiin toinen standardi ”The

Syslog Protocol”, jonka tarkoituksena oli lisätä selkeyttä ja tuoda parempi tietoturva protokollan käytölle asettamalla lisää määrittelyjä protokollan käytöstä. (Gerhards, 2009)

Rakenne & toiminta

Syslog pohjautuu kerrostettuun arkkitehtuuriin. Arkkitehtuurin tarkoituksena on erotella viestien sisältö sitä kuljettaneesta metodista. Syslog käyttää hyödykseen kolmea kerrosta, jotka ovat viestien sisältö, sovellus ja kuljettaminen. Sisällöllä tarkoitetaan Syslog-viestien sisältämää dataa, sovellustaso käsittelee luonnin, tulkitsemisen, reitityksen ja säilyttämisen. Kuljetustaso hoitaa viestien siirtämisen ja vastaanottamisen. (Gerhards, 2009)

Jokaisella tasolla suoritetaan niille tyypillisiä funktioita. Lähde (originator) luo viestiin sen sisällön. Kerääjä (collector) kerää viestit analysointia varten. Välittäjä (relay) lähettää viestit eteenpäin, vastaanottaa viestit lähteistä tai muista välittäjistä ja toimittaa ne kerääjiin tai välittäjiin. Lähettäjä (transport sender) välittää syslog-viestit tietyille kuljetusprotokollille. Vastaanottaja (transport receiver) vastaanottaa viestit tietyiltä kuljetusprotokollilta. (Gerhards, 2009)



Kuva 6. Syslog-tasot & funktiot.

3.4 Logstash

Yleisesti

Logstash on avoimeen lähdekoodiin perustuva, datan keräämiseen käytettävä työkalu. Sillä voidaan kerätä, parsia, lähettää ja tallentaa dataa useista eri lähteistä. Logstash on modulaarinen työkalu ja tällä hetkellä siihen on saatavissa yli 200 erilaista lisäosaa. Avoimen lähdekoodin ansiosta käyttäjien on mahdollista luoda myös omia lisäosia ja lisätä ne käyttöön. Tässä projektissa Logstashia käytetään keräämään lokeja esimerkiksi DC- ja DHCP-palvelimilta, mutta sillä voidaan kerätä myös syslog-viestejä tai muita verkkoliikenneviestejä. (Logstash Reference [7.8]. 2020)

Logstashin toiminta

Logstash pipeline koostuu kolmesta tasosta, inputs, filters, outputs. Inputs-tasolla luodaan tapahtumat, filters-taso muokkaa niitä ja outputs-taso lähettää ne edelleen muualle. Filters-taso on näistä osuuksista ainoa valinnainen osa. (Logstash Reference [7.8])

Sisääntulot (Inputs) määrittävät vastaanotettavien viestien sijainnit ja tavat, joilla niitä otetaan vastaan. Yleisimmin käytetyt inputs-lisäosat ovat

- file: lukee tiedostosta kohteen sijainnista, toimii kuten UNIX-komento ”*tail -0F*”
- syslog: vastaanottaa RFC 3164:n mukaisia syslog-viestejä TCP- tai UDP-protokollan avulla. Yleisimmin käytetty inputs-lisäosa
- redis: lukee tietoja redis-palvelimelta, käyttäen redis-kanavia ja redis-listoja; redisiä käytetään Logstashin ”brokerina”, joka asettaa Logstashin lähettäjiltä saapuvat tapahtumat jonoihin
- eventlog: kerää Windows-järjestelmien Event Logeja ja lähettää ne keskitettyyn kohteeseen erilliselle palvelimelle

Suodattimet (Filters) parsivat saapuneen datan. Suodattimien avulla datalle voidaan asettaa kriteerejä ja täten jättää pois turhat osat datasta. Suodattimiin kuuluvat

- grok: grok-lisäosa parsii järjestämättömän datan helpommin käsiteltävään muotoon. Jäsennelty data on täten helpompi hakea ja indeksoida. Grok-malleja on tällä hetkellä yli 120 kappaletta, joten jokaiseen käyttöön löytyy varmasti sopiva lisäosa
- mutate: mutate-suodattimella voidaan suorittaa tavallisia muutoksia datan kenttiin, kuten uudelleennimetä, poistaa, korvata ja muokata.
- drop: drop-suodattimen avulla tapahtumia voidaan pudottaa kokonaan.
- clone: clone-suodattimella voidaan luoda kopio tapahtumasta ja lisätä tai poistaa kenttiä.
- geoip: geoip-suodattimella voidaan lisätä tietoa IP-osoitteen maantieteellisestä sijainnista.

Ulostulot (Outputs) ovat Logstashin viimeinen vaihe. Ulostuloilla määritetään tapahtumien tallentaminen ja eteenpäin lähetys. Tapahtuma voi kulkeutua usean ulostulon läpi, mutta kaikkien ulostulojen jälkeen prosessi määritetään valmiiksi ja tapahtuma on suorittanut toimeenpanonsa. Ulostuloihin kuuluvat esimerkiksi nämä:

- Elasticsearch: prosessin data lähetetään Elasticsearch-tietokantaan, Elasticsearch on tarkoitettu erilaisten datalähteiden tallettamiseen ja nopeaan hakuun.
- file: prosessin data kirjoitetaan tiedostoon, joka säilötään levylle.
- graphite: prosessin data lähetetään graphiteen, avoimeen lähdekoodiin perustuvaan työkaluun, jota käytetään tallentamiseen ja graafisen datan luontiin.

Sisään- ja ulostulot tukevat myös erillisiä virtojen suodattimia (Codecs), joiden avulla voidaan helposti erottaa viestien kuljetusprosessi ja sarjoitusprosessi. (Logstash Reference [7.8]. 2020)

3.5 Elasticsearch

Yleisesti

Elasticsearch on avoimeen lähdekoodiin perustuva tietokanta- ja hakupalvelin. Elasticsearch ottaa esimerkiksi Logstashin keräämän datan ja säilöö sen kantaansa, tämän jälkeen Elasticsearch hoitaa datan indeksoinnin, hakemisen ja analysoinnin. Elasticsearch tarjoaa lähes reaaliaikaisen hakemisen ja analysoinnin kaikille datatyypeille. Data voi tulla Elasticsearchiin jäsenneltynä, jäsentämättömänä tai numeraalisena datana ja se voidaan silti säilöä tehokkaasti ja indeksoida nopeaa hakua varten. (Elasticsearch Reference [7.9]. 2020)

Elasticsearch on rakennettu Apache Lucene-nimisen hakukirjaston pohjalta. Lucene on yksi kehittyneimpiä ja korkeatehoisimpia hakukone-kirjastoja maailmassa tällä hetkellä, mutta se on pelkkä kirjasto. Lucene on myös erittäin monimutkainen ja vaatii Javan, jotta siitä saa irti mahdollisimman paljon. (Gormley;ym., 2015)

Elasticsearchin toiminta

Kommunikointi Elasticsearch-tietokannan kanssa riippuu siitä, käytetäänkö siinä Javaa vai ei. Java ohjelmointirajapinnan kanssa Elasticsearchissa on kaksi sisäänrakennettua palvelinta, joita voidaan hyötykäyttää yhteyden rakentamisessa. (Gormley;ym., 2015)

Node-palvelin

Node-palvelin liittyy Elasticin paikalliseen klusteriin, eli saman nimiattribuutin jakavien noodien ryhmään, ”non data” noodina. Se ei siis pidä sisällään mitään dataa, mutta tietää mitä dataa mikäkin klusterin osa sisältää ja se voi lähettää pyyntöjä eteenpäin klusterin noodeille. (Gormley;ym., 2015)

Välittäjäpalvelin

Välittäjäpalvelin toimii kevyemmin ja sillä voidaan lähettää pyyntöjä etäklusterihin asti. Välittäjäpalvelin ei liity klusteriin, vaan toimii reillisenä palvelimena, joka lähettää pyynnöt klusterin noodeille.

Molemmat Java-palvelimet keskustelevat klusteriin portin 9300 kautta ja käyttävät Elasticin omaa kuljetusprotokollaa. Klusterin noodit kommunikoivat myös keskenään portin 9300 kautta, klusteria ei voida muodostaa portin ollessa suljettuna tai käytössä. (Gormley;ym., 2015)

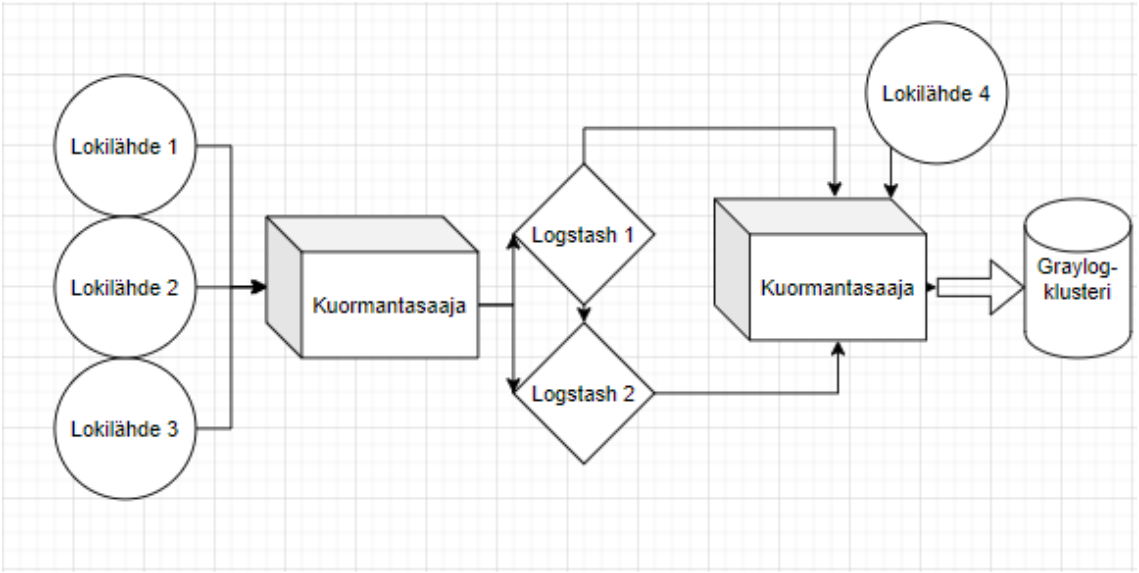
Jos käytössä ei ole Javan ohjelmointirajapintaa, kaikki muu kommunikointi tapahtuu REST-ohjelmointirajapintaan HTTP-protokollan avulla. REST-rajapinta keskustelee portin 9200 kautta ja sitä voidaan käyttää verkkoselaimen tai komentorivin kautta. Hakuja voidaan suorittaa HTTP-protokollan komentojen avuilla, joita ovat GET, POST, PUT, HEAD ja DELETE. (Gormley;ym., 2015)

4 KÄYTÄNNÖN OSA

4.1 Nykyinen järjestelmä

Tämän työn alussa mainittiin, että yrityksellä on jo valmiina käytössä Graylog-järjestelmä.

Seuraava havaintokuva näyttää nykyisen lokijärjestelmän-arkkitehtuurin kuvauksen karkeasti:



Kuva 7. Yrityksen lokijärjestelmän-arkkitehtuuri.

Valtaosa klusteriin saapuvista lokitiedostoista kulkeutuu kuormantasaajan kautta Logstashiin, jossa niille määritetään erilaisia suodatteita riippuen tulevasta lokilähteestä. Lokilähteet toimittavat viestit Syslog TCP/UDP-protokollilla. Lokilähteet ovat yrityksen ja asiakkaiden erilaisia laitteita, jotka lähettävät lokitietoja.

Jotkin erilliset lokilähteet toimittavat lokiviestit Logstashin ohi suoraan toiseen kuormantasaajaan, josta ne toimitetaan Graylog-klusteriin parsimattomana raakadatana. Kuormantasaajien tarkoituksena on toimia työnjakajina laitteiston resursseille, jotta järjestelmä ei ylikuormittuisi työmäärän kasvaessa. Kuormantasaajien avulla vältetään työmäärän kasautuminen vain yhdelle laitteistoresurssille kerrallaan.

Graylog-klusteri koostuu viidestä konesaliin sijoitetusta laitteesta, joihin tulevat lokitiedot päätyvät. Klusterin laitteilla on määriteltyinä Graylogin yhteydet Elasticsearchiin ja Mongo-tietokantaan, jotka esiteltiin Graylogin virtojen kuvauksessa (Kuva 2). Laitteilla on myös erikokoisia kovalevyjä 250Gb ja 2Tb välillä, jotka säilövät saapuvaa lokitietoa laitteille lyhytaikaisesti.

4.2 Uuden lähteen lisäys Logstashin kautta

Uutta lähdettä lisätessä järjestelmään pitää suorittaa konfigurointi sekä lähettävässä, että vastaanottavassa päädyssä. Graylogissa konfiguroinnit kulkeutuvat Inputs-välilehden alle ja lokiviestin lähettävässä päädyssä luodaan polku tarvittavien reittien kautta klusteriin, esimerkiksi jos viestit kulkeutuvat Logstashiin, luodaan polku lodfwdl-koneille ja sieltä jatkoreitti kuormantasaajan kautta klusteriin.

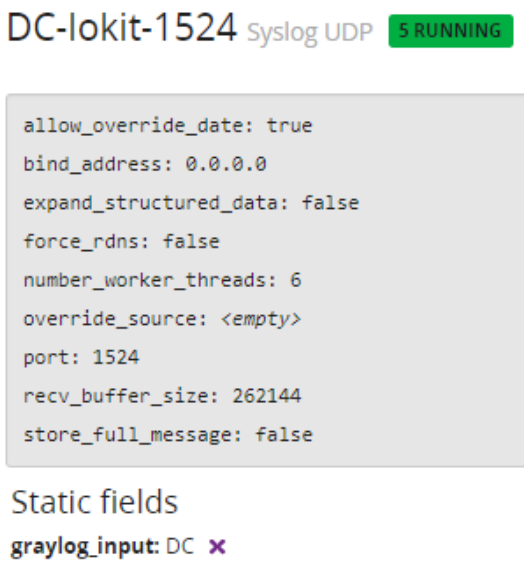
Graylog-sisääntulo

Uutta lähdettä lisätessä täytyy aloittaa Graylog-sisääntulon konfiguroimisesta. Päätepiste on hyvä olla auki ja täten välttää palvelun mahdollinen uudelleenkäynnistys, varsinkin jos palvelu on jo tuotantokäytössä.

Uuden sisääntulon konfigurointi on yksinkertaista, valitaan valikosta ”Inputs” ja sitä kautta saapuvan lokitiedoston haluttu muoto, esimerkiksi tässä tapauksessa ”GELF UDP”, jonka jälkeen valitaan ”Launch new input.” Ruudulle ilmestyvästä listasta valitaan ensin ”Global” painike, joka määrittää sisääntulon kulkeutuvan kaikille klusterin laitteille käyttöön. Tämän jälkeen sisääntulo nimetään ja annetaan kuunteluosoite esimerkiksi ”0.0.0.0”, sekä portti. Graylog tarjoaa oletuksena listasta seuravan vapaan portin automaattisesti, mutta käyttäjä voi halutessaan muuttaa porttia mieleisekseen.

Puskurointikoko ja työsäikeiden määrä määrittävät verkkoyhteyden nopeuden sisääntulolle. Nämä ovat täysin valinnaisia muokkauksia ja yleensä ne jätetään oletusarvoonsa.

Tämän jälkeen uusi sisääntulo asetetaan valmiiksi ja se tulee näkyviin ”Inputs” välilehdelle kuvan 8 mukaisesti.



Kuva 8. Uusi Graylog Sisääntulo.

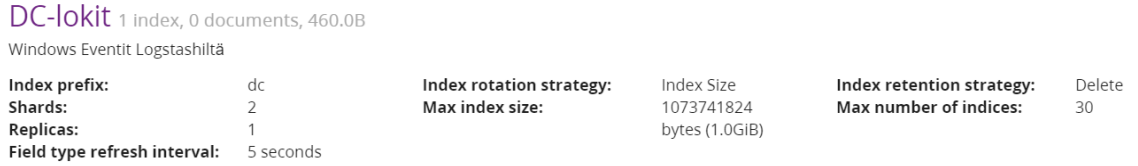
Sisääntulolle lisätään vielä erillinen staattinen kenttä, jonka avulla voidaan rajata sisääntulon ottavan vastaan vain DC-arvon sisältävän graylog_input kentän.

Indeksointi Graylogissa

Lokitiedostojen virtausta varten Graylogissa täytyi konfiguroida indeksointijoukko (Index Set). Indeksointijoukko valitaan välilehden ”System” osion ”Indices” alavalikosta. Tällä välilehdellä on nähtävissä muutkin konfiguroidut indeksointijoukot. Indeksointijoukko luodaan ”Create New Index Set”-painikkeella, jonka jälkeen sille annetaan tarvittavat asetukset. Kuva 9 näyttää valitut asetukset ennen lopullista käyttöönottoa. DC-lokien indeksointiin valikoituivat nämä asetukset ja niistä lyhyet kuvaukset:

- Nimi: DC-lokit
- Kuvaus: Windows Eventit Logstash
- Indeksin etuliite: dc, tätä käytetään Elasticsearch-tietokannan hakemistoissa tunnistautumiseen.

- Indeksisiru: 2 kpl Siruja voidaan pitää tietokannan hakemistojen osina, joiden avulla data kulkeutuu klusterissa. Suurempi sirumäärä tarjoaa paremman nopeuden datalle klusterissa.
- Indeksikopio: 1 kpl Elasticsearch luo indeksistä arvon verran kopioita, jotka säilöytyvät toiseen siruun. Kopioiden avulla voidaan nopeuttaa hakutoimintoja kantaan ja luoda indeksin varmuuskopio toiminnallisuuden jatkamiseksi.
- Päivitysväli: 5 sek kuinka usein indeksi päivittää tietoja Graylogille
- Indeksin kierrätyksen strategia: Indeksin koko, määrätään tietty tiedostokoko, jonka jälkeen indeksin tietojen kierrätys alkaa. Muita vaihtoehtoja ovat indeksien lukumäärä ja indeksien aikaleimat.
- Indeksikoko: 1 Gb aiemmassa asetuksessa määritellyn levytilan optimoitu koko, jonka jälkeen indeksin kierrätys alkaa.
- Indeksien säilytyksen strategia: Poisto, poistaa kierrätetyn indeksin Elasticsearchista resurssien käytön minimoimiseksi. Muita vaihtoehtoja ovat indeksin sulkeminen ja indeksin ohitus.
- Hakemistojen maksimimäärä: 30 kpl kuinka monta hakemistoa säilytetään Elasticsearchiin, kunnes vanhimpia aletaan poistamaan kierrätyksen yhteydessä.



Kuva 9. DC-lokien indeksijoukon asetukset.

Lokilähteen konfigurointi

Tämän esimerkin tarkoituksena oli tuoda DC-laitteiden Windows Event-lokeja Logstashin kautta Graylogiin. Graylog ei itsessään tue Windows-järjestelmien lokiformaattia, joten vaadittiin erillinen lisäys DC-laitteille, NXLog. Koska venytimme jo

käytössä olevaa järjestelmää, NXLog oli jo asennettu DC-laitteille ja se oli konfiguroitu lähettämään Event Logit Logstashin koneille kuormantasaajan kautta.

Yhteyttä testattiin kirjautumalla Logstash-koneelle ja menemällä konfiguraatiokansioihin polussa ”/etc/logstash/conf.d/nxlog_fsecurepm” ja ajamalla komentokehoitteessa seuraava testikomento:

logger -n <kuormantasaajan ip-osoite> -d -P 1524 testingggg

Komento luo lokiviestimäisen merkkijonon testaukseen ja lähettää sen kuormantasaajalle porttiin 1524, jota Graylog-klusteri kuuntelee. Testiviestiksi komento kirjaa aikaleiman, lähettäjän tunnisteen ja portin perässä olevan merkkijonon. Graylogin ”Inputs” alavalikossa määritellyn ”DC-lokit-1524” sisääntulon alla näkyy vastaanotettu testiviesti (Kuva 10), joka on parsittu Logstashissä määriteltyjen suodattimien avulla.

Kuva 10. Onnistunut testiviesti Graylogissa.

Onnistunutta DC-lokien lähetystä varten muokattiin Logstash-koneelta löytyvää NX-Log-konfiguraatiotiedostoa. Tiedostoon lisättiin tarvittavat muuttujat, jotta lokitiedostot lähetetään eteenpäin klusteriin. Koodikatkelmaan on lisätty numeroita eri fontilla, jotta alla oleva selitys olisi helpompi tunnistaa koodikatkelmasta. Kohdekoneiden nimet muutettiin tähän katkelmaan niiden sisältämän arkaluontoisen tiedon vuoksi.

```
## DC (NXLog)

  else if [host] =~ " DC" or [host] =~ "osoite.fi" or [host] =~ "paikallinen.local" or [host] =~ "sairaala.net" or [host] =~ "ER" or [host] =~ "kohdekone001" or [host] =~ "kohde80" or [host] =~ "sijainti001" or [host] =~ "paikka03" {
```

```
(2)mutate {
  add_field => {
    "logsource" => "windows"
  }
 }
(1) if [host] =~ " DC1" or [host] =~ " DC2" or [host] =~ " DC3" or [host] =~ " DC4"
 else {
  mutate {
   add_field => {
   (3)"[@metadata][do_not_send_noc]" => true
   }
  }
 }
}
```

jos-lauseke kaappaa MN-DC-koneilta saapuvat lokitiedostot muiden joukosta (1) ja toimittaa ne yrityksen Network Operations Centerin (NOC) sijaan suoraan Graylog-klusteriin, täten klusteriin ei päädy ei-haluttujen lähteiden lokeja.

”Mutate”-ominaisuudella lähteisiin voitiin lisätä tunnistekenttiä, joden avulla lokitietoja pystyttiin luokittelemaan muista erilleen. Tässä tapauksessa lisättiin lokilähdekentäksi tunniste ”windows” lokilähteen käyttöjärjestelmän tunnistamiseksi (2) ja metadatakenttä ”do not send noc”, jolla tunnistetaan lokitiedon ohittavan NOC-sijainnin (3).

Tiedostossa määritellään myös ulostulot Graylog-klusteriin seuravasti. Koodikatkelmaan on lisätty numeroita eri fontilla, jotta alla oleva selitys olisi helpompi tunnistaa koodikatkelmasta.

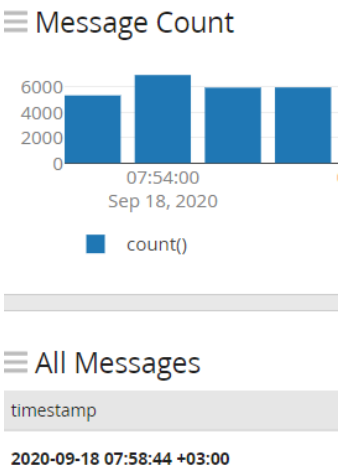
```
##### Graylog outputs #####
(1) if ![@metadata][do_not_send_noc]){
 if [logsource] == "windows" {
```

```
(2)udp {
    host => <klusterin-ip-osoite>
    port => 1524
    (3)codec => line {
      format => "%{message}"
    }
    id => "output-graylog-windows"
  }
 }
 }
```

End of Graylog outputs

Graylogin ulostulossa määritellään aiemman if-lausekkeen lokilähteiden kohdeosoite ja muutokset. Lauseke tarkistaa metadatakentän sisällön ja varmistaa lokilähteeksi arvon ”windows”(1), tämän jälkeen lauseke määrittelee käytettävän kuljetusprotokollan lokiviestille ”udp”, jonka alle määritellään klusterin ip-osoite, sekä portti (2). Koodekki-osuudessa viestin muoto määritellään ja annetaan sille formatointityyppi tarvittaessa, ”%{message}” muuttujalla vastaanotettu lokiviesti lähetetään klusteriin raakadatana (3). Lopuksi viestiin lisätään id ”output-graylog-windows”, jonka avulla klusteri osaa toimittaa viestin oikeaan sijaintiin kannassa.

Muutoksien jälkeen prosessi tapettiin ja käynnistettiin uudelleen. Kahdennetun järjestelmän vuoksi uudelleenkäynnistys ei aiheuttanut katkoa lokiviestien kulussa klusteriin ja SIEMiin. Tämän jälkeen uutta lokidataa alkoi saapumaan Graylogiin, kuten kuvassa 11 nähdään.



Kuva 11. Graylogiin saapuvaa lokidataa.

Itse raakadatan sisältöä ei voida näyttää sen sisältämien arkaluontoisten tietojen vuoksi.

4.3 Uuden lähteen lisäys suoraan Graylogiin

Graylog-järjestelmään voidaan tuoda myös suoraa lokidataa laitteelta. NOC-arkkitehtuurin kuvauksessa näkyy esimerkki, miten lokilähde voi kulkeutua myös pelkän yksittäisen kuormantasaajan kautta klusteriin. Esimerkkinä käytettiin Cisco-kytkimiä lähettämään lokitietojaan suoraan klusteriin. Vaikka esimerkki olikin tietoliikennetiimin vuoksi tehty venytys, samaa tapaa käytetään myös kyberturvatiimin tarpeisiin, kun tuodaan lokilähde samankaltaisesta palvelimesta tai laitteesta klusteriin. (kuva 7)

Graylogin asetukset uudelle lokilähteelle

Cisco-kytkimiä varten luotiin samankaltainen sisääntulo kuin Logstash-koneilta tuleville DC-lokeille. Määritellyksi portiksi annettiin vain uusi vapaa portti ja lokiviestin kuljetusprotokollaa muutettiin tarpeen mukaiseksi. Uuden sisääntulon staattiselle kentälle pakotettiin arvo ”cisco”, joten vain Cisco-tunnisteella olevat lokiviestit päätyvät tähän sisääntuloon. Sirumäärää laskettiin pienemmäksi, koska kytkinlokin vaatima kapasiteetti on DC-lokien vaatimaa kevyempi.

Kytkinlokien indeksointi muodostui myös lähes samanlaiseksi DC-lokeihin verrattuna, kapasiteettimäärät vain laskivat, kun viestit eivät kulje Logstashin kautta, vaan saapuvat suoraan kuormantasaajan kautta klusteriin.

Asetukset suoran lokilähteen päädyssä

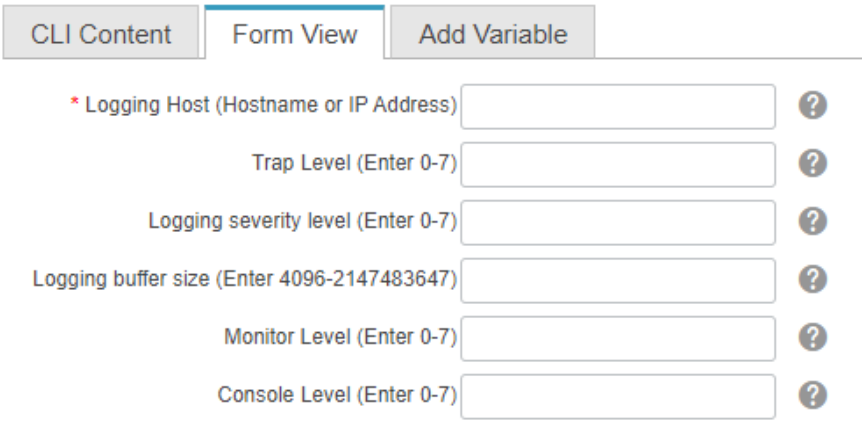
Lokilähteen päätyyn tehtiin muutoksia, jotta lokit tallentuvat hetkellisesti laitteelle, mutta lähtevät myös pidempiaikaiseen säilöön Graylogiin. Täten lokeja voidaan tarkastella sekä laitteelta itseltään, että keskitetystä lokienhallintajärjestelmästä.

Cisco-kytkimien tapauksessa käytössä on Cisco Prime-hallintatyöpöytä, johon pystyi luomaan mallipohjan kytkimien lokiasetuksia varten. Mallipohjassa määritellään, mitä lokitetaan, mihin kohteeseen ja lokien vakavuuden taso.

```
logging on
logging host $hostname transport udp port 1525 session-id hostname
#if($severity_level != "")
   logging buffered $severity_level
#end
#if($buffer_size != "")
   logging buffered $buffer_size
#end
#if($trap_level  != "")
  logging trap $trap_level
#end
#if($monitor  != "")
  logging monitor $monitor
#end
#if($console  != "")
logging console $console
```

#end

Tämän avulla saatiin luotua kuvan 12 mukainen pohja lokitusta varten.



Kuva 12. Valmis lokituspohja Cisco-kytkimille.

Kyseisen pohjan avulla määritellään lokituksen kohdeosoite ja muut tiedot lokitarpeita varten, kuten tietojen vakavuustaso ja kaappaustaso.

Esimerkiksi antamalla Graylogin klusterin IP-osoite ja asettamalla kaikkiin valintoihin valvontatasoksi 6, kytkimen konfiguraatiotiedostoon luodaan seuraavat lokitusrivit:

logging origin-id hostname

logging facility syslog

logging snmp-trap emergencies

logging snmp-trap alerts

logging snmp-trap critical

logging snmp-trap errors

logging snmp-trap warnings

logging snmp-trap informational

logging host <klusterin ip-osoite> transport udp port 1525 session-id hostname

Tällä konfiguraatiomuutoksella kytkin kaappaa lokitietoihin määritellyt kohteet ja lähettää tallentamansa lokit myös kohdeosoitteeseen samassa muodossa kuin ne on laitteelle tallennettuna, käyttäen protokollana UDP:tä ja kohdeporttia 1525, joka määriteltiin Graylogin klusterissa (kuva 8).

Konfiguraatiotiedoston asetuksia voidaan soveltaa myös muiden kytkinvalmistajien laitteisiin, kirjoitusasu saattaa vaihdella eri valmistajien välillä, mutta ydinrakenne säilyy samanlaisena. Täten voidaan lisätä myös muita kytkinlokeja suoraan Graylogiin.

Lokituksen vakavuustasojen kanssa täytyy olla tarkkana, tasolla 0 kytkin ei käytännössä tallenna lokeihin yhtään tietoa, kun taas tasolla 7 annetaan kytkimelle ”debug all” lokitustaso. Tason 7 asetuksilla kytkin yrittää tallettaa kaiken laitteella kulkevan tiedon lokitiedostoihin, jonka seurauksena laitteet hidastuvat huomattavasti ja useimmissa tapauksissa lakkaavat kokonaan vastaamasta, eli laitteet ns. ”kaatuvat”.

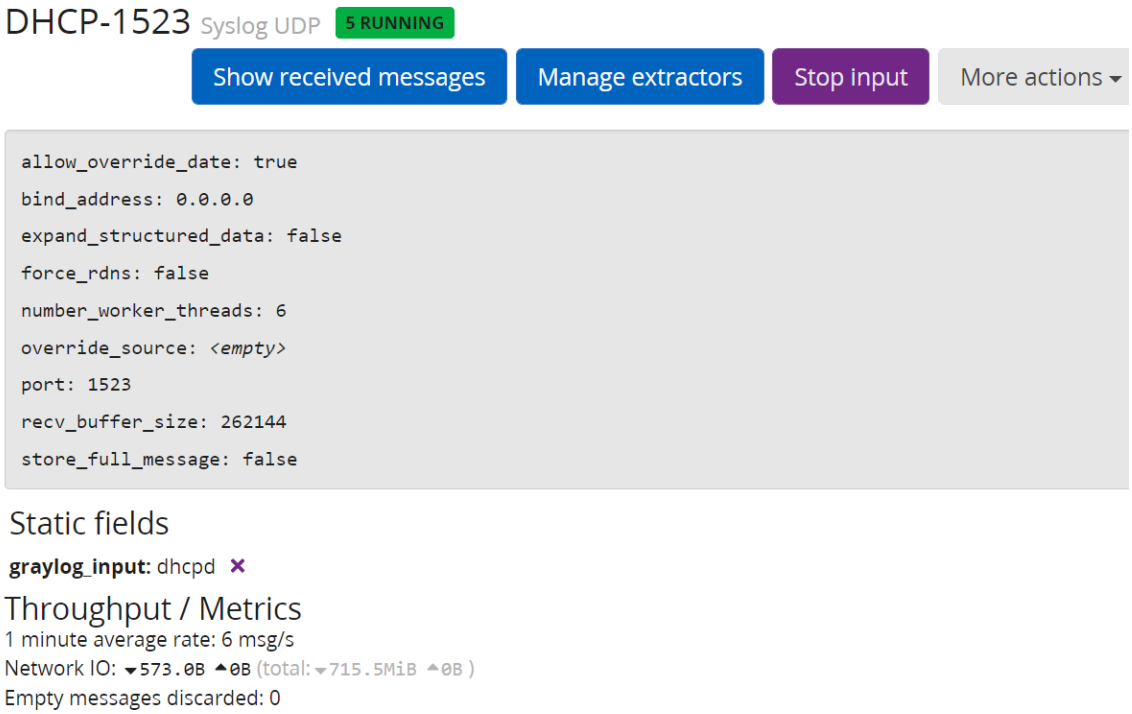
5 LOPPUTULOKSET

Uuden lokilähteen lisäys

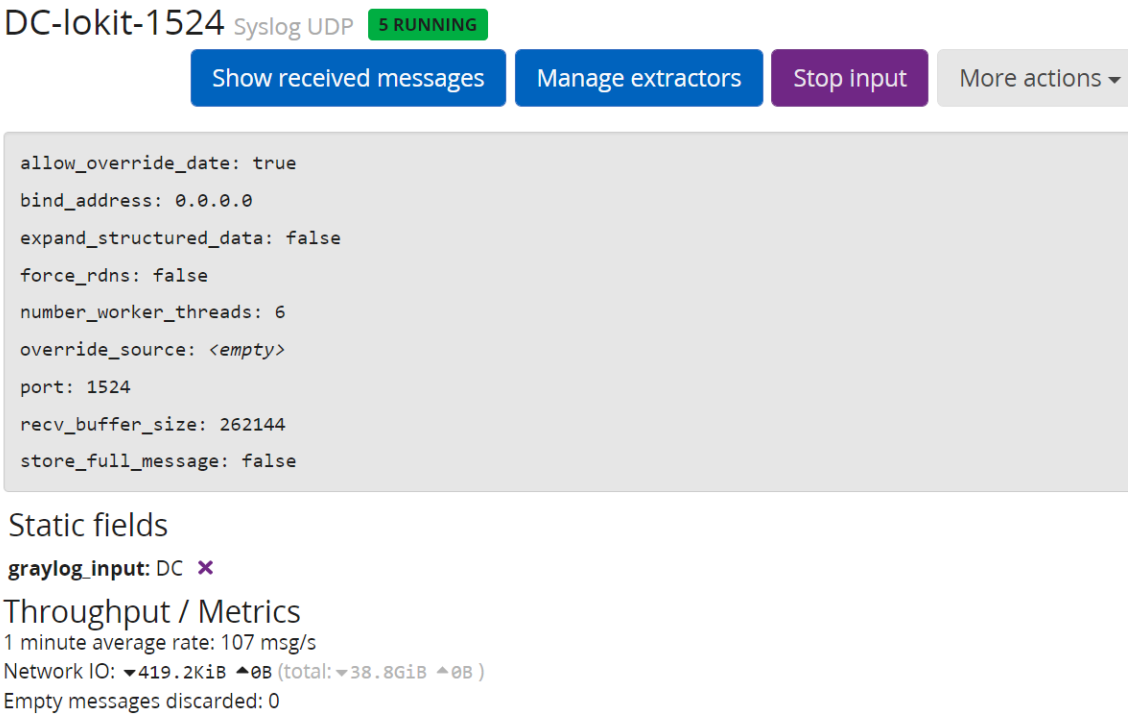
Päävaatimuksena oli selvittää uusien lokilähteiden tuonti Graylogiin kyberturvatiimin käyttöön.

DC- ja DHCP-lokien tuonti Graylogiin Logstashin kautta osoitti järjestelmän ottavan helposti vastaan uusia lokilähteitä. DC-lokien esimerkin avulla Graylogiin tuotiin lokitiedostoja useista eri lähteistä ja samaa tapaa käytettiin myös DHCP-lokien lisäämiseksi. Jatkokehitystä varten konfigurointimallia voidaan käyttää roskapostisuodattimien ja muiden tietoturvakomponenttien lähteiden lisäämiseen Graylog-klusteriin.

Kuvissa 13 & 14 nähdään, miten DC- ja DHCP-lokeja varten luodut sisääntulot ovat toiminnassa ja ovat toimittaneet lokeja Graylogiin tarkasteltavaksi.



Kuva 13. DHCP-lokien sisääntulo.



Kuva 14. DC-lokien sisääntulo.

Skaalautuvuus

Vaatimusmäärittelyissä pyydettiin tutkimaan nykyisen järjestelmän kapasiteettia ja laajentamismahdollisuutta. Uutta lokilähdettä lisätessä tämä nousi pakolliseksi selvittää ja asiasta keskusteltiin yrityksen konesalitiimin kanssa. Konesalitiimi vastaa yrityksen laitteista sekä palvelimista ja niiden sijoituspaikoista. Useiden kokouksien ja keskustelujen jälkeen saimme selville nykyisen kapasiteetin olevan riittävä ainakin kyberturvatiimin tietoturvalokien lisäystä varten.

Keskustelimme myös useista vaihtoehdoista laajennustapauksia varten ja päädyimme sekä fyysisten levyjen lisähankintaan, että mahdollisten pilvipalvelujen käyttöönottoon pitkäaikaissäilöntää varten. Pilvipalvelusta esimerkkinä olisi konesalitiimin käytössä oleva Cloudian-palvelu, joka toimii s3-rajapinnan avulla. Selvitystyö jatkuu Cloudianin osalta siten, että tukeeko käytössä oleva Graylog-järjestelmä s3-rajapintaa täysin vai vain osittain. Fyysisien levyjen hankinnassa tarkasteltavat kohteet olivat niiden halpa hinta per teratavu ja helppokäyttöisyys asennettaessa fyysiseen laitteeseen.

Uusien noodejen lisääminen järjestelmään todettiin mahdolliseksi virtuaaliympäristön vuoksi. Klusteriin kyetään lisäämään uusi virtuaalikone ja kopioimaan sille tarvittavat asetukset aiemmista laitteista. Uuden noodin levykapasiteettia voidaan muokata virtuaalikoneen asetuksista ja se jakaa fyysisen levytilan muiden klusterin laitteiden kanssa.

Lokien kierrätys & poisto

Vaatimusmäärittelyssä haluttiin lokien kierrätyksen ja poiston vastaavan yrityksen lokiperiaatteita. Eri lokitiedoille on määrätty omat säilytysajat, tämän työn kohdalla ne vaihtelivat kolmen viikon ja kolmen kuukauden välillä riippuen lokilähteestä. Säilytysajan täyttyessä lokitiedot tulee joko poistaa tai anonymisoida GDPR:n mukaan.

Lokilähteitä lisätessä Graylogiin annettiin kierrätykseen ja poistoon datakatto ja indeksien määrät, jonka jälkeen kierrätys alkaa (Kuva 15). Kierrättäminen esimerkkitapausten lokitiedostoissa oli suoraan lokitiedon poistaminen.

Index rotation strategy:	Index Size	**Index retention strategy:**	Delete
Max index size:	1073741824 bytes (1.0GiB)	**Max number of indices:**	30

Kuva 15 Lokien kierrätyksen määritelmät.

Turvallisuus

Palvelimiin ja laitteisiin pääsevät kirjautumaan vain ne henkilöt, joille on myönnettynä oikeanlaiset oikeudet. Kirjautumiseen käytetään yrityksen työntekijän henkilökohtaisia AD-tunnuksia, täten lokeihin jää ylös merkintä kirjautumisista. Palvelimien root-oikeuksia pääsee käyttämään, jos AD-tunnuksille on sallittuna siirtyminen root-käyttäjäksi. Palvelimet sijaitsevat yrityksen sisäisessä verkossa ja ulkopuolelta niihin pääsee vain käyttämällä VPN-tunnelia ja erikseen määrättyä hallintapöytää erillisin tunnuksin.

Klusteriin kirjautuminen tapahtuu myös verkkokäyttöliittymän kautta, jonka osoite täytyy tietää. Kirjautumiseen käytetään myös AD-tunnuksia ja käyttäjä on lisättävä Graylogiin ennen tätä. Käyttäjää lisätessä asetetaan myös hänelle oikeudet eri näkymiin.

Lokikohtaisia näkymiä voidaan muokata ”Authentication” välilehdellä. Tämän avulla yksittäisille käyttäjille voidaan sallia tai evätä pääsy tiettyjen lokien näkymiin ja hakutoimintoihin. Täten vaatimusmäärittelyn pyyntö salata tietoturvalokit muilta kuin Graylogin admin-käyttäjiltä ja kyberturvatiimiltä olisi mahdollista toteuttaa.

6 POHDINTA

Avoimeen lähdekoodiin pohjautuvien järjestelmien kanssa on syytä muistaa kattava dokumentaatio. Alkuvaiheessa työn suurin ongelma oli dokumentaation hakeminen ja nykyisen järjestelmän toiminnan selvittely. Jouduin niputtamaan yhteen Graylogin omaa dokumentaatiota ja yrityksen sisäistä dokumentaatiota saadakseni selville kokonaiskuvan. Rakentamalla Graylogin alusta saakka itse olisin varmaan paremmin selvillä kaikista sen toiminnallisuuksista. Tämä ei olisi ollut tuotannon kannalta järkevä ratkaisu. Onneksi tietoliikenne- ja kyberturvatiimissä vastattiin kaikkiin kysymyksiini järjestelmästä ja sen toiminnasta. Näin sain suurimmat katvealueet dokumentoinnista selvitettyä, ja omaksi asiakseni jäi paljon pienempi työmäärä. Klusteriin saapuva data tuotiin alkuun raakadatana, viestit lähetettiin suoraan eteenpäin samassa muodossa kuin niiden saapuessa. Tietoja pyritään parsimaan jatkossa hyödyntämällä esimerkiksi GELF-formaattia NXLogin tuomiin viesteihin Logstash-koneille.

Ongelmakohtia oli muutamia, esiin nostettavana esimerkiksi S3-ohjelmointirajapinnan selvitystyön osuus. Rajapinnan testaukseen tarvittava virtuaaliympäristö on vielä toteutuksessa. Virtuaaliynpäristön valmistuttua voimme turvallisesti kokeilla rajapinnan vaikutusta klusterin toimintaan. Selvitystyö rajapinnan sopivuudesta järjestelmään jatkuu tämän työn palautuksen jälkeen lokien pidempiaikaista säilöntää varten. Pienempiä ongelmia oli esimerkiksi dokumentaation puutteellisuus ja päivitykset. Yrityksen dokumentaatio Graylogista oli jäänyt vajaavaiseksi ja sitä täydennetään jatkossa enemmän.

Mielestäni vaatimusmäärittelyssä pyydettyihin kohtiin saatiin vastattua, vaikkakaan DHCP-sisääntulo ei rakentunut minun toimestani, vaan se toteutettiin samalla kaavalla kuin DC-lokien tuonti. Vaatimuksiin vastaaminen oli jatkuvasti mielessäni ja sen vuoksi listaa karsittiin ensimmäisistä versioista pienemmäksi. Aiemmassa osassa vaatimusmäärittelyt katsottiin yhdessä ja kaikkiin niihin saatiin vastauksia. Aikaa jäi olettamaani runsaammin, joten lisää vaatimuksia saatetaan lisätä tähän työhön vielä loppupuolellakin. Jatkokehityksen osalta jää myös paljon tehtävää, mutta tämän työn avulla on saatu luotua pohja ja tutkimustyö, joiden avulla laajennusta päästään jatkamaan.

Olen tyytyväinen tuloksiin, sain tehdä opinnäytetyötä työpaikalla ja tämän vuoksi se eteni verkkaisesti. Muut työprojektit eivät päässeet pahasti katkomaan kirjoitus- ja selvitystyötä, koska olin määritellyt tietyt päivät eri työtehtäville. Työtä tehdessäni pääsin opettelemaan täysin uutta järjestelmää ja sen käyttöä.

Lokienhallinta kiinnostaa itseäni nyt paljon enemmän tietoturvan- ja tietoliikenteen näkökulmista.Olisin voinut omasta mielestäni olla vielä enemmän aktiivinen työn suhteen. Pienensimme vaatimusmäärittelyjä alkuperäisestä, koska en luottanut omaan osaamiseeni. Nyt jälkeenpäin huomaan, että vaatimuksia olisi voinut ottaa mukaan enemmänkin.DC-lokien esimerkin avulla voidaan tuoda muidenkin palvelimien lokeja järjestelmään, mutta niitä ei esitellä tässä työssä, koska ne toistavat lähes samaa kaavaa.

LÄHTEET

2M-IT 2020. Ratkaisut sosiaali- ja terveyspalvelujen kehittämiseksi. (viitattu 20.7.2020)

https://2m-it.fi/

Carstensen, Nick. 2018 Why Is Log Management Important (luettu 31.7.2020)

https://www.graylog.org/post/why-is-log-management-important

Chuvakin, A., Schmidt, K. & Crishtopher, P. 2012. Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management. (uusintapainos , Newnes, 2012.) (luettu & viitattu 27.7.2020)

Clinton Gormley & Zachary Tong, 2015. Elasticsearch: The Definitive Guide: A Distributed Real-Time Search And Analytics Engine (verkkokirja, luettu & viitattu 24.8.2020)

Elasticsearch Reference [7.8] >> What is Elasticsearch? (luettu ja viitattu 14.8.2020)

https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html

Euroopan parlamentin ja neuvoston asetus (EU) 2016/679 luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuoja-asetus) Annettu 27.4.2016. (Luettu 3.8.2020)

https://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:32016R0679&from=FI

Gerhards, R. The Syslog Protocol – 2009 (luettu 7.8.2020)

https://www.hjp.at/doc/rfc/rfc5424.html

Graylog 2020a. Alerts (luettu 22.7.2020)

https://docs.graylog.org/en/3.3/pages/alerts.html

Graylog 2020b. GELF (luettu 27.7.2020)

https://docs.graylog.org/en/4.0/pages/gelf.html

Graylog 2020c. Get one step closer to GDPR compliance (luettu 3.8.2020)

https://www.graylog.org/webinars/get-one-step-closer-to-gdpr-compliance

Graylog Documents 2020d. Streams. (luettu 22.7.2020)

https://docs.graylog.org/en/3.3/pages/streams.html

Graylog 2020e. What GDPR means for Log Management? (luettu 27.7.2020)

https://www.graylog.org/resources/what-gdpr-means-for-log-management

Grimes, R. 2010. Living the log management lifecycle. InfoWorld 4.8.2010. (luettu 27.7.2020)

https://www.infoworld.com/article/2625977/living-the-log-management-lifecycle.html

Internet Engineering Task Force – Request for Comments UDP 768, elokuu 1980 (luettu 12.8.2020)

https://tools.ietf.org/html/rfc768

Internet Engineering Task Force – Request for Comments 793 TCP, syyskuu 1981 (luettu 12.8.2020)

https://tools.ietf.org/html/rfc793

Laki sähköisen viestinnän palveluista, Finlex. 7.11.2014/917 (luettu ja viitattu 12.12.2020)

https://www.finlex.fi/fi/laki/ajantasa/2014/20140917

La rosa Alexander 2018. Log Monitoring: not the ugly sister (luettu 22.7.2020)

https://web.archive.org/web/20180214153657/https://blog.pandorafms.org/log-monitoring/

Logstash introduction [versio 7.8] (luettu ja viitattu 14.8.2020)

https://www.elastic.co/guide/en/logstash/current/introduction.html

Logstash Pipeline [versio 7.8] (luettu ja viitattu 14.8.2020)

https://www.elastic.co/guide/en/logstash/current/pipeline.html

Lonvick, C. The BSD Syslog Protocol – 2001 (luettu 7.8.2020)

https://www.hjp.at/doc/rfc/rfc3164.html

National Institute of Standards and Technology – Special Publication 800-92, syyskuu 2006 (luettu ja viitattu 11.8.2020)

https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf

”Näin keräät ja käytät lokitietoja – Kyberturvallisuuskeskus 30.4.2020” (viitattu 20.7.2020)

https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/ohjeet-ja-oppaat/nain-keraat-ja-kaytat-lokitietoja

Sinun Eurooppasi, Yleinen tietosuoja-asetus (luettu 27.7.2020)

https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm

Tämä sivu jätetty tyhjäksi tarkoituksella