Opinnäytetyö (AMK)

Tietojenkäsittely

2019

Jesse Kumlander

SAAS-PALVELUN TIETOTURVAN KANNALTA OLENNAISET LOKIT JA NIIDEN HALLINTA

TURKU AMK
TURKU UNIVERSITY OF APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Tietojenkäsittely

2019 | 55 sivua, 1 liitesivu

Jesse Kumlander

SAAS-PALVELUN TIETOTURVAN KANNALTA OLENNAISET LOKIT JA NIIDEN HALLINTA

Lokitiedot ja lokienhallinta on olennainen osa tietoturvallisuuden hallintaa, sillä lokeja tarkastelemalla on mahdollista havaita esimerkiksi tietomurtojen yrityksiä ja väärinkäyttötapauksia. Loki viittaa tallenteeseen, joka sisältää tietoa esimerkiksi tietojärjestelmien ja sovellusten tapahtumista, ajankohdista ja tapahtumien aiheuttajista. Tänä päivänä keskitetty lokienhallintajärjestelmä on lähes välttämätön tietoturvapoikkeuksien havaitsemiseen ja selvittämiseen, sillä suurin osa sovelluksista ja järjestelmistä on siirtynyt pilviarkkitehtuuriin, jonka myötä erinäisten lokien ja niiden lähteiden määrä on kasvanut räjähdysmäisesti.

Opinnäytetyön tavoitteena oli selvittää, minkälaisista tapahtumista toimeksiantajan SaaS-verkkopalvelussa tulee kerätä lokitietoja ja mitä lokien tulee sisältää, jotta niiden avulla voidaan havaita ja selvittää tietoturvapoikkeuksia. Lisäksi työssä selvitettiin, mikä pilvipohjaisista lokienhallintapalveluista on toimeksiantajan käyttötapaukseen soveltuvin. Teoriaosuudessa käsiteltiin lokien ja niiden käsittelyn teoriaa sekä lokien osuutta tietoturvallisuuden hallinnassa. Työssä tarkasteltiin myös lainsäädännön asettamia vaatimuksia ja rajoitteita lokien keräämiselle ja säilyttämiselle.

Opinnäytetyön tuloksena luotiin ohjeistus toimeksiantajan verkkopalvelusta ja sen infrastruktuurista kerättävistä lokeista ja niiden sisällöstä. Tutkimuksessa käytettiin konstruktiivista tutkimusotetta, ja ohjeistuksen lähteinä käytettiin alan kirjallisuutta sekä asiantuntijaorganisaatioiden ohjeistuksia. Lisäksi työn tuloksena syntyi soveltuvuusselvitys, jossa vertailtiin kahta lokienhallintapalvelua, LogDNA:ta ja Datadogia. Selvityksen on tarkoitus esitellä palveluiden ominaisuuksia sen sijaan, että se toimisi yksityiskohtaisena asennusohjeena.

Opinnäytetyön tuloksena syntynyttä ohjeistusta ja selvitystä on tarkoitus hyödyntää Vuosikello-verkkopalvelun lokimerkintöjen sisällön ja lokienhallinnan suunnittelussa. Tuloksia voivat käyttää toimeksiantajan ohella myös muut tahot, sillä ohjeistus on pyritty kirjoittamaan mahdollisimman yleisellä tasolla.

ASIASANAT:

lokienhallinta, tietoturva, lokitieto

BACHELOR´S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Business Information Technology

2019 | 55 pages, 1 appendix

Jesse Kumlander

SAAS APPLICATIONS SECURITY-RELEVANT LOGS AND THEIR MANAGEMENT

Log files and log management are integral part of information security management. By reviewing logs, organizations can detect the data breach attempts and improper use of resources for instance. Log file refers to the recording which contains information about systems and applications events, and their timings and causes. Centralized log management systems are nowadays almost necessary for detecting and investigating security incidents. The main reason for this is the applications and systems movement to distributed cloud architecture which has led to an exponential growth of different logs and log sources.

The purpose of this thesis commission was to find out what events should be logged by the commissioner's SaaS application and what kind of information logs should contain in order to be effective in detecting and investigating security incidents. In addition, the thesis commission was to find out which cloud-based logging service would be the most suitable for commissioner's use case. The theoretical part focuses on logs and log management in general and describes why logs are important in information security management. Theoretical part also describes what kind of requirements and constraints the legislation imposes on collecting and storing logs.

Logging instructions were created as a result of this thesis. Instructions covered what logs should be collected from the web service and its infrastructure, and what information logs should contain. Constructive approach was used for the research and the sources of the guidance consist of relevant literature and the guidelines of several expert organizations. Additionally, a comparison between two log management services, LogDNA and Datadog, was conducted as a part of thesis. The purpose of this comparison was to demonstrate features of the services instead of acting as detailed installation instructions.

Commissioner of this thesis will utilize the results when planning the log content and log management of their web service, Vuosikello. The results of this thesis can also be used by other persons since the instructions are written in as a general manner as possible.

KEYWORDS:

log management, information security, log files

SISÄLTÖ

LIITTEET

KUVAT

SANASTO

Apache	Suosittu avoimen lähdekoodin verkkopalvelin (Apache HTTP Server Project 2019).
HTTP-metodi	Hypertext Transfer Protocol eli hypertekstin siirtoprotokolla. Protokollaa käytetään yleisesti verkkopalvelimen ja selaimen väliseen kommunikaatioon. (W3Schools 2019.)
IP-osoite	Internet Protocol-osoite eli Internetin protokollaosoite. Kaikille verkkoon kytketyille laitteille määritelty osoite, jonka avulla laite kyetään tunnistamaan verkossa. (Jensen 2018.)
Isäntänimi	Järjestelmän tai laitteen nimi, jonka avulla se tunnistetaan esimerkiksi paikallisessa verkossa tai internetissä (Fisher 2019).
Käyttäjäagentti	Merkkijono, jonka avulla voidaan tunnistaa esimerkiksi käyttäjän selain, selaimen versio ja käyttöjärjestelmä (Mozilla 2019).
MAC-osoite	Media Access Control -osoite. Verkkosovittimen uniikki osoite, minkä valmistaja tyypillisesti kirjoittaa laitteeseen. (Wikipedia 2019b.)
MFA	Multi-Factor Authentication eli monivaiheinen tunnistautuminen. Käyttäjä joutuu todentamaan identiteettinsä käyttäjänimen ja salasanan lisäksi jollakin muulla tavalla, kuten älypuhelimeen lähetettävällä koodilla. (Onelogin 2019a.)
Need-to-know basis	Periaate, jonka mukaan henkilöille tulee kertoa tai antaa pääsy ainoastaan tietoon, jota he tarvitsevat työtehtäviensä tekemiseen (SecurityMetrics 2014).
OWASP	The Open Web Application Security Project. Voittoa tavoittelematon järjestö, joka tarjoaa muun muassa ilmaisia dokumentteja ja ohjelmistoja, tavoitteenaan parantaa erityisesti web-sovellusten tietoturvaa. (OWASP 2019a.)
Palvelunestohyökkäys	Hyökkäysmenetelmä, jonka avulla kaapatuilla laitteilla välitetään kohteeseen niin paljon liikennettä, että palvelun käyttö hankaloituu tai estyy (Kyberturvallisuuskeskus 2016).
Peukalointi	Hyökkäysmenetelmä, jossa tarkoituksella muokataan tai tuhotaan esimerkiksi dataa (Study.com 2019).
Privacy Shield	Järjestely, jolla taataan, että tarvittavia tietosuojavaatimuksia noudatetaan, kun henkilötietoja siirretään EU-alueelta Yhdysvaltoihin (Privacy Shield Framework 2019).
Pseudonymisointi	Prosessi, jossa henkilötietoja käsitellään siten, että niiden avulla ei enää kyetä tunnistamaan tiettyä henkilöä ilman tarvittavia lisätietoja (Tietosuojavaltuutetun toimisto 2019c).

S3	Simple Storage Service. Amazonin palvelu, joka tarjoaa tallennustilaa esimerkiksi verkkosivujen ja mobiilisovellusten datalle. (AWS 2019.)
Sanitointi	Menetelmä, jossa syöte muutetaan hyväksyttävään muotoon. Esimerkiksi kentästä, johon syötetään vuosiluku, poistetaan kaikki merkit, jotka eivät ole numeroita. (OWASP 2013).
SIEM	Security Information and Event Management. Järjestelmä, joka kerää lokeja ja dataa eri järjestelmistä ja analysoi niitä erinäisten uhkien varalta. (Petters 2019.)
SSO	Single Sign-On eli kertakirjautuminen. Menetelmä, jonka avulla käyttäjät voivat kirjautua useisiin palveluihin yksillä käyttäjätunnuksilla (Onelogin 2019b).
Säännöllinen lauseke	Metodi, jonka avulla voidaan etsiä merkkijonoja tekstistä (Techopedia 2019a).
TCP	Transmission Control Protocol. Tietoliikenneprotokolla, joka takaa, että lähetetty data saavuttaa määränpäänsä. (Doyle 2018.)
UDP	User Datagram Protocol. Yhteydetön protokolla, joka ei takaa datan onnistunutta toimitusta. Protokollan etu on sen nopeus. (Doyle 2018.)
URL	Uniform Resource Locator, joka tunnetaan myös verkko-osoitteena. Sillä tarkoitetaan internetissä sijaitsevien resurssien osoitteita. (Techopedia 2019b.)
VAHTI	Valtionhallinnon tietoturvallisuuden johtoryhmä. VAHTI vastaa valtionhallinnon tietoturvallisuuteen liittyvistä asioista. (Lokiohje 2009, 5.)
Väsytyshyökkäys	Hyökkäysmenetelmä, jossa hyökkääjä yrittää toistuvasti arvata esimerkiksi käyttäjätunnusta ja salasanaa. Tämä hyökkäys toteutetaan usein automatisoiduilla työkaluilla. (IBM 2019a.)

1 JOHDANTO

Tänä päivänä yhä useammat sovellukset siirtyvät perinteisestä asiakas–palvelinarkkitehtuurista pilvipohjaisiin hajautettuihin järjestelmiin. Samalla myös datan määrä on kasvanut räjähdysmäisesti ja organisaatiot keräävät, käyttävät ja tuottavat dataa jatkuvasti enenevissä määrin. Koska organisaatioiden keräämällä datalla on taloudellista arvoa, se houkuttelee myös hyökkääjiä.

IBM:n ja Ponemon Instituten tekemän tutkimuksen mukaan organisaatioilla meni vuonna 2018 keskimäärin 206 päivää tietomurron havaitsemiseen (IBM 2019b). Syitä tähän voidaan vain arvailla, mutta etenkin verkkosovellusten kohdalla useimmissa tapauksissa perimmäisenä syynä on riittämätön lokienhallinta ja monitorointi (OWASP 2017, 16). Erityisesti modernien pilvipohjaisten sovellusten ja niitä tukevan infrastruktuurin lokienhallinta on tärkeää, jotta kyetään havaitsemaan ja selvittämään tietoturvaloukkauksia.

Lokienhallinta on olennainen osa tietoturvallisuuden hallintaa. Lokitietoja tarkastelemalla voidaan havaita esimerkiksi tietomurtojen yrityksiä, väärinkäyttötapauksia ja järjestelmävirheitä. (Lokiohje 2009, 14.) Lokienhallinta on kuitenkin tänä päivänä huomattavasti monimutkaisempaa kuin aiemmin, sillä organisaatioilla saattaa olla kymmeniä, jopa satoja lokeja tuottavia järjestelmiä, sovelluksia ja laitteita (LogDNA 2019a). Nykyään keskitetty lokienhallintajärjestelmä on välttämätön pilvipohjaisten sovellusten tietoturvatapausten havaitsemiseen ja selvittämiseen (Graylog 2018).

Opinnäytetyön tavoitteena on selvittää, mitä lokitietoja SaaS-verkkopalvelusta tulee kerätä, jotta tietoturvatapahtumia voidaan havaita ja selvittää. Lisäksi opinnäytetyössä kartoitetaan pilvipohjaisia lokienhallintapalveluita toimeksiantajan määrittämien kriteerien perusteella. Määritettyjä kriteereitä ovat muun muassa palveluiden tietoturvallisuus, helppokäyttöisyys sekä mahdollisuus luoda hälytyksiä. Näistä palveluista valitaan kaksi potentiaalisinta palvelua tarkempaan vertailuun.

Opinnäytetyön tuloksena syntyy ohjeistus, jossa käsitellään, mitä lokien tulee sisältää ja mistä tietoturvan kannalta olennaisista tapahtumista niitä tulee kerätä. Ohjeistus käsittelee myös lainsäädännön asettamia vaatimuksia ja rajoitteita lokien keräämiselle ja säilyttämiselle. Ohjeistuksessa käytetään konstruktiivista tutkimusotetta ja ohjeistus perustuu olemassa olevaan tietoon. Lisäksi opinnäytetyön tuloksena syntyy soveltuvuusselvitys kahden lokienhallintapalvelun ominaisuuksista ja käytettävyydestä. Opinnäytetyön

tuloksia on tarkoitus hyödyntää toimeksiantajan Vuosikello-verkkopalvelun lokien sisällön suunnittelussa, kun taas soveltuvuusselvityksen tuloksia voidaan hyödyntää lokienhallintapalvelun valitsemisessa.

Opinnäytetyöni toimeksiantajana toimii Perjantai Markkinointiviestintä Oy, joka valmistautuu lähitulevaisuudessa lanseeraamaan Vuosikello-sovelluksestaan SaaS-verkkopalvelun. Perjantai Markkinointiviestintä Oy on vuonna 2004 perustettu mainostoimisto, joka tarjoaa mainonnan, markkinoinnin ja viestinnän palveluiden ohella monipuolisia digitaalisia ratkaisuja (Finder 2019).

2 LOKITIEDOT

Tässä luvussa käsitellään lokitietoihin liittyvää teoriaa ja lokien käsittelyn eri vaiheita. Luvussa käsitellään myös lyhyesti, mikä merkitys lokeilla on tietoturvallisuuden suhteen.

2.1 Lokien määritelmä ja luokittelu

Loki viittaa tallenteeseen, joka sisältää tietoa esimerkiksi tietojärjestelmien, tietoverkkojen ja sovellusten tapahtumista ja niiden aiheuttajista (Kyberturvallisuuskeskus 2019). Lokit koostuvat lokimerkinnöistä, joista jokainen sisältää tietoa tapahtumasta sekä sen tarkasta ajankohdasta. Alun perin lokeja hyödynnettiin lähinnä tietojärjestelmien ja tietoverkkojen ongelmien kartoittamisessa, kun taas tänä päivänä niiden merkitys korostuu esimerkiksi järjestelmien ja verkon suorituskyvyn optimoimisessa ja tietoturvapoikkeamien tutkimisessa. (Kent & Souppaya 2006, 2-1.)

Erinäiset käyttöjärjestelmät, sovellukset, verkkolaitteet ja tietokannat luovat jatkuvasti lokeja (Lokiohje 2009, 29). Lokien sisältämä data voi poiketa suuresti, mutta yleisesti lokit sisältävät vähintäänkin tapahtuman ajankohdan, tason ja viestin. Lisäksi useat lokit sisältävät tietoa muun muassa isäntänimestä (engl. hostname), lokityypistä, sovelluksesta ja IP-osoitteesta. (LogDNA 2019a.)

Tyypillisimmät lokien luokittelut ovat ylläpitoloki, muutosloki, virheloki sekä käyttöloki. Usein lokeja ei kuitenkaan voi sijoittaa ainoastaan yhteen luokkaan, sillä monet lokit täyttävät useamman luokan kriteerit. (Lokiohje 2009, 29.) Ensimmäinen käsiteltävä lokityyppi on ylläpitolokit, jotka sisältävät tietoa muun muassa järjestelmiin tehdyistä muutoksista sekä virhetilanteista. Lisäksi käyttöoikeuksiin tehdyt muutokset, kuten muokkaukset, lisäykset ja poistot, tallennetaan ylläpitolokiin. (Lokiohje 2009, 30.) Ylläpitolokit ovat hyödyllisiä esimerkiksi toimintaympäristön kokonaisarkkitehtuurin seurannassa sekä versionhallinnassa (Kyberturvallisuuskeskus 2019).

Seuraava lokityyppi on muutoslokit. Ne sisältävät merkintöjä tiedon lisäys-, muutos- ja poisto-operaatioista. Muutoslokeista voidaan tarkistaa muuttuneen tiedon alkuperä ja näin ollen sen oikeellisuus voidaan jäljittää ja varmistaa tarvittaessa. (Kyberturvallisuuskeskus 2019.) Muutosloki sisältää myös tietoa järjestelmäparametrejä ja asetustiedostoja koskevista muutoksista (Lokiohje 2009, 30). Virhelokit sen sijaan sisältävät nimensä

mukaisesti tietoa järjestelmien tai tapahtumien virhetilanteista (Lokiohje 2009, 30) ja niitä kannattaakin hyödyntää ongelmatilanteiden selvittämisessä (Kyberturvallisuuskeskus 2019).

Seuraavana vuorossa on useimmiten yleisin ja tarpeellisin lokityyppi, käyttöloki, josta on suuresti hyötyä esimerkiksi tietoturvapoikkeamien selvittämisessä. Käyttölokia kutsutaan myös tapahtumalokiksi. (Kyberturvallisuuskeskus 2019.) Tyypillisesti käyttöloki sisältää tietoa järjestelmien ja sovellusten käyttäjien sisään- ja uloskirjautumisista, epäonnistuneista kirjautumisista sekä käyttövaltuuksien vaihdoista (Lokiohje 2009, 30). Käyttölokit sisältävät usein myös tietoa järjestelmän muista normaaleista prosesseista sekä tulostustapahtumista (Kyberturvallisuuskeskus 2019). Myös tietokantoihin liittyvät tapahtumat, kuten lukutapahtumat ja kyselytiedot hakuparametreineen, tallennetaan käyttölokeihin (Lokiohje 2009, 30).

Yleisiin lokityyppeihin kuuluvat myös viestintäloki ja haltijaloki. Näistä ensimmäinen sisältää nimensä mukaisesti tietoa viestintätapahtumista, ja sen tarkoitus on ensisijaisesti helpottaa viestintäjärjestelmien häiriötilanteiden selvittämistä. Usein viestintälokeja hyödynnetään myös tietoturvapoikkeamien hallinnassa sekä viestintätapahtumien toteutumisen todentamisessa. (Lokiohje 2009, 30.) Viestintälokit sisältävät tyypillisesti tietoa viestien alkuperästä, päätepisteistä, ajankohdista sekä määristä (Kyberturvallisuuskeskus 2019).

Jälkimmäisenä mainittu haltijaloki puolestaan sisältää tietoa siitä, kenelle tietty puhelinnumero, verkkodomain tai verkko-osoite, kuten IP- tai MAC-osoite, on kuulunut tiettynä ajankohtana (Kyberturvallisuuskeskus 2019). Haltijalokin avulla voidaan muun muassa selvittää tapahtumien kulku ja osapuolet, esimerkiksi tarkistamalla kenellä tietty IP-osoite on ollut käytössä minäkin hetkenä (Lokiohje 2009, 31).

2.2 Lokien käsittely

Lokien käsittely viittaa lokien elinkaaren erinäisiin toimenpiteisiin, joita ovat lokien kerääminen, analysointi, säilyttäminen, poistaminen sekä arkistointi (Lokiohje 2009, 14). Organisaatioiden tulee varata lokien käsittelylle riittävästi laitteisto- ja henkilöstöresursseja, mistä luonnollisesti aiheutuu kustannuksia, jotka tulee ottaa huomioon. Lokien käsittelyssä on myös huomioitava käyttäjien oikeusturva ja tietosuoja. (Lokiohje 2009, 19.)

2.2.1 Lokien kerääminen

Lokien keräämisellä tarkoitetaan lokitietojen lähettämistä keskitettyyn sijaintiin, johon kerätään useiden eri lähteiden lokitiedot. Keskitetyn lokienhallintajärjestelmän käytössä on useita etuja, joista ilmeisin on se, että kaikki lokit sijaitsevat yhdessä paikassa. (Chuvakin ym. 2012, 34.) Organisaatioilla voi olla tänä päivänä kymmeniä, jopa satoja järjestelmiä, sovelluksia ja laitteita, jotka tuottavat jatkuvasti lokeja, ja mikäli lokeja ei siirretä keskitettyyn lokienhallintajärjestelmään, niiden tarkastelu ja analysointi käy mahdottomaksi (LogDNA 2019a).

Keskitetty lokienhallintajärjestelmä myös helpottaa lokien etsimistä ja analysointia, ja monet lokienhallintajärjestelmät mahdollistavat eri järjestelmien tuottamien lokien tarkastelun kronologisessa järjestyksessä. (LogDNA 2019a.) Lokeja on myös huomattavasti helpompi varmuuskopioida, kun ne sijaitsevat yhdessä paikassa (Chuvakin ym. 2012, 35).

Lokitietojen siirtämiseen on tänä päivänä useita eri tapoja. Ennen tyypillisin tapa siirtää lokeja oli Syslog-protokolla (Chuvakin ym. 2012, 35). Syslog-protokollaa käytetään useissa Unix-pohjaisissa järjestelmissä, kuten palvelimissa ja verkkolaitteissa, lokitietojen kuljettamiseen. Syslog-protokollan avulla lokit välitetään tyypillisesti UDP-portin 514 tai TCP-portin 6514 kautta keskitettyyn järjestelmään. TCP-portti tukee myös TLS-protokollaa (Transport Layer Security), joka mahdollistaa lokitietojen salaamisen kuljetuksen aikana. (Rouse 2019a.)

Usein helpoin tapa siirtää lokitietoja keskitettyyn lokienhallintapalveluun on niin sanottu agenttiohjelma. Agenttiohjelma on kevyt ohjelma, jonka tehtävä on kerätä järjestelmästä lokitietoja ja lähettää niitä keskitettyyn lokienhallintapalveluun. Palveluntarjoajasta riippuen agenttiohjelmat kykenevät esimerkiksi käyttämään salattuja yhteyksiä ja piilottamaan arkaluontoisia tietoja. (LogDNA 2017; Sematext 2019.) Useilla palveluntarjoajilla on integraatioita tunnettujen palveluiden kanssa, ja osa palveluntarjoajista tarjoaa myös koodikirjastoja, jotka voidaan integroida suoraan organisaatioiden omiin sovelluksiin. Koodikirjastojen avulla ohjelmistokehittäjät voivat määritellä tarkasti, milloin ja miten lokitietoja lähetetään. (LogDNA 2017.)

2.2.2 Lokien analysointi

Lokien analysointi on yksi haastavimpia vaiheita lokienhallinnassa, mutta sitäkin tärkeämpi. Lokien manuaalinen analysointi on usein työlästä, jonka vuoksi on kannattavaa ottaa käyttöön lokienhallintajärjestelmä, joka automatisoi analysointia mahdollisimman paljon. (Kent & Souppaya 2006, 5-5.)

Ennen kuin lokien analysointi aloitetaan, on tärkeää varmistaa, että lokit sisältävät kaikki tarvittavat tiedot ja että niitä tulkitaan asiayhteyden mukaan (Zhang 2018). Jotta asiayhteys voidaan saada selville, saman järjestelmän eri lokimerkintöjä tulee vertailla keskenään. Kun lokeja tuottavien järjestelmien ja laitteiden normaali toiminta tunnetaan, poikkeavat lokimerkinnät tunnistetaan paremmin. (Lokiohje 2009, 48–49.) Lokitiedot tulee myös normalisoida, jotta ne ovat yhdenmukaisia keskenään (Zhang 2018).

Alkuun on suositeltavaa, että lokeja analysoidaan säännöllisesti, esimerkiksi päivittäin. Näin saadaan tarkka käsitys siitä, mitkä ovat normaaleja lokimerkintöjä ja mitkä epätavallisia. Usein suurin osa järjestelmien lokeista sisältää vain muutamia erityyppisiä lokimerkintöjä, ja niitä säännöllisesti läpi käymällä, pystytään havaitsemaan tehokkaammin mahdolliset poikkeamat. Kun oleelliset ja epäoleelliset lokimerkinnät on tunnistettu, organisaatiot voivat rakentaa erinäisiä suodattimia, jotka kykenevät automaattisesti tunnistamaan epätavallisia tai haitallisia toimia. Tarpeen tullen hälytyksiä ja muita tietoturvakontrolleja voidaan automatisoida siten, että ne aktivoituvat, kun epätavallisia tai haitallisia toimia havaitaan. Lisäksi suodattimet helpottavat mahdollista manuaalista lokitietojen analysointia, sillä niiden avulla voidaan suodattaa epäolennaiset lokimerkinnät. (Lokiohje 2009, 47–48.)

Eri järjestelmien tuottamat lokimerkinnät poikkeavat usein toisistaan huomattavasti. Tämä vaikeuttaa muun muassa lokien analysoinnin priorisointia, sillä lokimerkinnät saattavat käyttää eri priorisointiluokitteluja. Valtiovarainministeriön Lokiohjeessa (2009, 48) määritellään, että organisaatioiden tulee luoda omat priorisointiluokittelut lokimerkinnöille perustuen muun muassa seuraaviin kysymyksiin:

- Mikä on lokimerkinnän tyyppi? (Esimerkiksi kriittinen);
- Onko lokimerkintä harvinainen tai ennennäkemätön?;
- Mikä on lokimerkinnän lähde? (Esimerkiksi kriittinen järjestelmä);
- Mikä on lähteen tai kohteen IP-osoite? (Kuuluuko IP-osoite esimerkiksi kriittiselle järjestelmälle?);

- Mihin aikaan ja minä viikonpäivänä lokimerkintä luotiin? (Esimerkkinä tietyn tyyppiset lokimerkinnät työpäivän jälkeen tai viikonloppuna);
- Kuinka tiheään tahtiin lokimerkintöjä ilmaantuu? (Lokiohje 2009, 48.)

Lokimerkintöjä analysoitaessa lokimerkinnät tulee myös korreloida muihin lokimerkintöihin ja järjestelmiin (Lokiohje 2009, 48). Korreloimisella tarkoitetaan useiden samantyyppisten tai erilaisten tapahtumien yhdistämistä yhdeksi tapahtumaksi, jotta saadaan yleiskuva jostakin suuremmasta tapahtumasta (Chuvakin ym. 2012, 185). Hyvä esimerkki korreloinnin hyödyistä on tilanne, jossa käyttäjä on yrittänyt kirjautua verkkosovellukseen 50 kertaa siinä kuitenkaan onnistumatta. Koska verkkosovellus kerää lokeja sisäänkirjautumisista, muun muassa käyttäjän IP-osoite tallentuu lokimerkintöihin. Tämän jälkeen lokimerkinnöistä näkyy, että samasta IP-osoitteesta on sisäänkirjauduttu onnistuneesti. Jos tapahtumia olisi tarkasteltu yksitellen, tämä väsytyshyökkäys (engl. brute-force attack) olisi saattanut jäädä huomaamatta, sillä yksikään yksittäisistä tapahtumista ei olisi ollut hälytyksen arvoinen. (Graylog 2019.)

2.2.3 Lokien säilyttäminen

Lokien säilytysaika ja arkistointi

Mikäli lokitiedot sisältävät henkilötietoja, niihin sovelletaan EU:n yleistä tietosuoja-asetusta. Koska tietosuoja-asetus määrittelee, että henkilötietoja saa ainoastaan säilyttää niin kauan, kun on tarpeellista niiden käyttötarkoituksen kannalta, organisaatioiden tulee arvioida, kuinka kauan lokitietoja tulisi säilyttää. Säilytysajan pituus on tarvittaessa pystyttävä myös perustelemaan, ja organisaatioiden on poistettava tai vaihtoehtoisesti anonymisoitava lokitiedot, kun niille ei ole enää tarvetta. Tietosuoja-asetus ei kuitenkaan määrittele tarkkaa säilytysaikaa henkilötiedoille, mutta tyypillisesti lokeja on suositeltavaa säilyttää 6-24 kuukautta. (Kyberturvallisuuskeskus 2019.)

Lokitietojen arkistointi tarkoittaa lokien säilyttämistä vielä senkin jälkeen, kun niille ei ole enää aktiivista tarvetta. Tämä mahdollistaa vanhempien lokimerkintöjen tarkastelun vielä myöhemminkin. (Kyberturvallisuuskeskus 2019.) Tyypillisesti lokit arkistoidaan irrotettaville tallennusvälineille, erityisesti siihen tarkoitukseen varatuille laitteille tai keskitettyyn lokienhallintajärjestelmään (Lokiohje 2009, 58). Useat lokienhallintapalvelut, kuten Datadog, Loggly ja Papertrail, mahdollistavat lokien arkistoinnin myös Amazonin S3-palveluun (Datadog 2019a; Loggly 2019; Papertrail 2019a).

Lokien tiivistäminen

Lokien tiivistäminen tarkoittaa usein arkistoinnin yhteydessä tapahtuvaa prosessia, jossa loki pyritään pakkaamaan niin tiiviiseen muotoon kuin mahdollista ilman, että sen sisältö muuttuu. Erityisesti tekstimuotoisten lokien pakkaus on tehokasta (Lokiohje 2009, 58.)

Lokien supistaminen

Lokien supistaminen on prosessi, jossa lokeista poistetaan tarpeettomia kenttiä, jotta niiden koko saadaan minimoitua. Lokien supistamiseen on eri käytäntöjä ja se voidaan toteuttaa esimerkiksi poistamalla kokonaisia lokimerkintöjä tai yksittäisiä kenttiä. Kuten lokien tiivistäminenkin, supistaminen tapahtuu usein lokeja arkistoidessa (Lokiohje 2009, 58–59.)

Lokien normalisointi

Usein eri lähteistä tulevien lokitiedostojen kentät ovat eri muodoissa, joka vaikeuttaa niiden tulkintaa ja analysointia. Lokien normalisointi tarkoittaa prosessia, jossa lokitiedoston kaikki tietokentät muunnetaan yhtenäiseen muotoon ja luokitellaan yhteneväisellä tavalla. Lokien normalisointi tehostaa ja helpottaa erityisesti lokien analysointia, sillä eri lähteistä kerättyjen lokien lokimerkinnät ovat samassa yhteneväisessä muodossa. Yksi yleisimpiä normalisointiin liittyviä toimenpiteitä on lokien sisältämän aikatiedon muuttaminen tiettyyn muotoon. Sovelluksesta, laitteesta tai järjestelmästä riippuen lokien aikatieto saatetaan esittää esimerkiksi 12 tunnin tai 24 tunnin muodossa. Myös aikavyöhykkeisiin liittyvät tiedot ovat usein eri formaateissa. (Lokiohje 2009, 59.) Näiden tietojen muuttaminen yhteneväiseen muotoon helpottaa lokien analysointia ja suodattamista huomattavasti.

2.3 Lokien merkitys tietoturvallisuuden kannalta

Lokien vaikutusta tietoturvallisuuteen ei voi väheksyä. Lokit ovat korvaamattomia tietomurtojen ja muiden tietoturvatapahtumien havaitsemisessa ja selvittämisessä, sillä niiden avulla voidaan saada selville tapahtuman laajuus, syyt sekä osapuolet. Ilman lokeja on haastavaa tunnistaa poikkeuksellisia tai ei-sallittuja tapahtumia ja täten reagoida niihin. Erityisesti huomiota tulee kiinnittää tapahtumiin, jotka tapahtuvat poikkeaviin kellonaikoihin, kuten työajan ulkopuolella. Muita esimerkkejä tapahtumista, joita voidaan saada lokien avulla selville, ovat muun muassa resurssien väärinkäyttötapaukset sekä

valtuuttamaton käyttö. Mikäli lokien luominen ja kerääminen on toteutettu oikeaoppisesti, lokien avulla voidaan luoda luotettava tapahtumaketju (engl. audit trail), joka on korvaamatonta todistusaineistoa rikosprosessissa. Lokitiedot takaavat myös järjestelmien ja sovellusten käyttäjien oikeusturvan, sillä niiden avulla voidaan osoittaa, onko käyttäjä ollut osallisena jossakin tapahtumassa vai ei. (Lokiohje 2009, 13–15.)

Valtiovarainministeriön Lokiohje (2009, 15) määrittelee, että lokitietojen käsittelyn yksi päätavoitteista on selvittää tapahtumien osapuolet ja kulku. Lokitietojen tulisi selkeästi määritellä, kuka tai ketkä osapuolet olivat osallisena tapahtumassa. Mikäli lokeja kerätään useista eri järjestelmistä, sovelluksista ja laitteista, on äärimmäisen tärkeää, että niiden kaikkien kellot ovat oikeassa ajassa ja synkronoitu keskenään. Mikäli lokien lähteiden kellot ovat eri ajoissa, tapahtumien kulkua ja tapahtumaketjua ei voida selvittää luotettavasti. (Lokiohje 2009, 15.)

3 CASE: VUOSIKELLO – LOKIENHALLINTA

Tässä luvussa esitellään lyhyesti toimeksiantajan verkkopalvelu sekä käsitellään opinnäytetyön toimeksiantoa. Lisäksi tässä luvussa kuvaillaan tarkemmin toimeksiannon eri osuuksien sisältöä.

3.1 Vuosikello-sovellus

Vuosikello on Perjantai Markkinointiviestintä Oy:n vuonna 2017 lanseeraama verkkosovellus, jonka tarkoitus on helpottaa organisaatioiden vuosisuunnittelua. Sovelluksen avulla asiakkaat voivat muun muassa hallita markkinoinnin, myynnin ja muiden toimintojen suunnittelua (Vuosikello 2019), jakaa tietoa sekä helpottaa suunnitelmien jalkauttamista ja toimeenpanoa.

Toimeksiantaja aikoo julkaista kyseisestä sovelluksesta SaaS-verkkopalvelun, jotta useammat organisaatiot voivat ottaa sovelluksen käyttöönsä. SaaS eli Software as a Service tarkoittaa ohjelmistojen jakelumallia, jossa SaaS-palveluntarjoajan pilvi-infrastruktuurissa isännöitävää sovellusta tarjotaan loppukäyttäjille Internetin välityksellä (Rouse 2019b). Tyypillisesti palvelun asiakkaat käyttävät sovellusta verkkoselaimen kautta (Mell & Grance 2011, 2), joka mahdollistaa sen, että sovellusta voidaan käyttää useilla eri laitteilla, kuten mobiililaitteilla ja tietokoneilla.

3.2 Toimeksiannon kuvaus

Opinnäytetyön toimeksianto koostuu kahdesta eri osiosta. Ensimmäisessä osiossa selvitetään, minkälaisista tapahtumista Vuosikello-verkkopalvelussa tulee kerätä lokitietoja ja mitä lokimerkintöjen tulee sisältää, jotta niiden avulla voidaan selvittää tietoturvaan liittyviä tapahtumia. Selvityksen alussa kuvaillaan mitä rajoitteita ja velvollisuuksia lainsäädäntö asettaa lokien keräämiselle ja säilyttämiselle. Tämän jälkeen käsitellään, mitä tietokenttiä lokimerkintöjen tulee sisältää, jotta niiden avulla voidaan selvittää tietoturvatapahtumia mahdollisimman tehokkaasti.

Palvelun tietoturvan kannalta olennaisia lokeja koskevassa osuudessa käsitellään ensisijaisesti verkkosovelluksen tietoturvatapahtumia ja lokeja. Kyseinen alaluku on jaettu

selkeyden vuoksi pienempiin osioihin, joista jokainen käsittelee tietyn aihepiirin tapahtumia, joita tulee tallentaa lokitietoihin. Kukin osio sisältää myös mahdollisia perusteluja ja esimerkkejä tietoturvatapahtumista. Selvityksessä käytetään lähteinä alan kirjallisuutta sekä erinäisten asiantuntijaorganisaatioiden, kuten OWASPin, NISTin ja VAHTIn, ohjeistuksia. Selvityksessä käsitellään myös lyhyesti lokitietojen suojaamistoimenpiteitä.

Toimeksiannon toisessa osiossa keskitytään pilvipohjaisiin lokienhallintajärjestelmiin. Jo ennen opinnäytetyöprosessin alkua, toimeksiantaja oli kartoittanut, että pilvipohjainen lokienhallintapalvelu olisi soveltuvin vaihtoehto heille. Syy tälle on se, että pilvipohjainen lokienhallintapalvelu kykenee hoitamaan lokien keräämisen, analysoinnin, säilytyksen sekä arkistoinnin. Lisäksi kyseisten palveluiden käyttöönotto on usein yksinkertaista eikä vaadi paljon resursseja.

Toimeksiantona on kartoittaa erinäisiä pilvipohjaisia lokienhallintapalveluita toimeksiantajan asettamien vaatimusten mukaisesti. Kartoitukseen valitaan viisi tunnettua lokienhallintapalvelua, joista valitaan ominaisuuksien ja toimeksiantajan vaatimusten perusteella kaksi tarkempaan testaukseen. Kartoituksen tarkoituksena on kokeilla kummankin palvelun ominaisuuksia, ja arvioida palveluiden käytettävyyttä ja soveltuvuutta toimeksiantajan käyttötapaukseen. Palveluiden soveltuvuutta arvioidaan luvussa neljä. Tarkoituksena on, että toimeksiantaja hyödyntää opinnäytetyön tuloksia Vuosikello-palvelun lokien sisällön suunnittelussa ja lokienhallintapalvelun valinnassa.

Toimeksiantoa koskevat kappaleet on pyritty kirjoittamaan mahdollisimman yleisellä tasolla toimeksiantajan pyynnöstä siten, että niistä ei paljastu yksityiskohtia sovelluksen järjestelmäarkkitehtuurista tai tietoturvakontrolleista. Neutraalin sävyn vuoksi opinnäytetyön sisältämä tieto toivon mukaan helpottaa toimeksiantajan ohella myös muita asiasta kiinnostuneita henkilöitä ja organisaatioita lokien sisällön ja lokienhallinnan suunnittelussa.

3.3 Lainsäädännön asettamat vaatimukset ja rajoitukset

Lokien käsittelyyn liittyy tiettyjä lainsäädännön asettamia vaatimuksia ja rajoituksia. Vaatimukset koskevat esimerkiksi lokien sisältöä, niiden säilytysaikaa, niiden sisältämien tietojen eheyden varmistamista sekä niiden käyttötarkoitusta. Lokien keräämisen ja käsittelyn tulee aina perustua lainsäädäntöön, ja lokien käsittelytapa ja -oikeus riippuu lokien

sisältämästä tiedosta sekä alkuperäisestä käyttötarkoituksesta. (Kyberturvallisuuskeskus 2019.)

Toimeksiantajan on otettava huomioon minkälaista tietoa lokitiedot sisältävät. Mikäli lokitietoihin on tallennettu henkilötietoja, lokitiedosto muodostaa henkilörekisterin, ja tällöin on otettava huomioon EU:n yleinen tietosuoja-asetus. Henkilötietojen käsittelytoimista on myös luotava seloste, mikäli Vuosikello-palvelun lokeihin tallennetaan henkilötietoja. (Kyberturvallisuuskeskus 2019.)

Henkilötiedoiksi luokitellaan kaikki tiedot, joiden avulla henkilö kyetään tunnistamaan suoraan tai epäsuorasti. Epäsuoralla tunnistamisella tarkoitetaan tilannetta, jossa yksittäisiä tietoja yhdistellään muihin tietoihin, jotka mahdollistavat henkilön tunnistamisen. Henkilötiedoiksi luokitellaan muun muassa nimi, henkilötunnus, sähköpostiosoite, puhelinnumero, kotiosoite, sijaintitiedot sekä IP-osoite. (Tietosuojavaltuutetun toimisto 2019a.)

EU:n yleinen tietosuoja-asetus eli GDPR (General Data Protection Regulation) on kaikissa EU-maissa sovellettava laki, joka säätelee henkilötietojen käsittelyä. Sen tavoitteena on muun muassa parantaa henkilöiden tietosuojaoikeuksia ja henkilötietojen suojaa yleisellä tasolla. GDPR:n mukaan henkilötietoja saa käsitellä ainoastaan laissa määritellyn perusteen mukaan. (Tietosuojavaltuutetun toimisto 2019b.)

Käytännössä tämä tarkoittaa lokitietojen suhteen sitä, että toimeksiantajan tulee pohtia, mihin tarkoitukseen lokitusta käytetään, ja ovatko kaikki lokien sisältämät tiedot tarpeellisia lokien käyttötarkoituksen kannalta (Kyberturvallisuuskeskus 2019). Mahdollisia perusteita henkilötietoja sisältävien lokien keräämiselle ovat sopimus, rekisteröidyn suostumus, lakisääteinen velvoite, elintärkeiden etujen suojaaminen, yleinen etu ja julkinen valta sekä rekisterinpitäjän tai kolmannen osapuolen oikeutettu etu. (Tietosuojavaltuutetun toimisto 2019b.)

GDPR:n mukaan oikeutettu etu voi esimerkiksi olla tietoturvallisuuden varmistaminen (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, johdanto-osan perustelukappale 49). Tietosuoja-asetuksen johdanto-osan perustelukappaleessa 49 todetaan, että välttämätön henkilötietojen käsittely on oikeutettua, mikäli tavoitteena on ”esimerkiksi luvattoman sähköisiin viestintäverkkoihin pääsyn ja vahingollisen koodin jakamisen ehkäiseminen sekä palvelunestohyökkäysten ja tietokoneille ja sähköisille viestintäjärjestelmille koituvien vahinkojen estäminen” (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, johdanto-osan perustelukappale 49). Näin ollen esimerkiksi IP-

osoitteiden kerääminen tietoturvallisuuden varmistamiseksi voi olla perusteltua (Bateman 2018). Organisaatioiden tulee kuitenkin aina arvioida tasapainotestin avulla, onko oikeutettu etu peruste henkilötietojen käsittelylle. Testistä tulee myös laatia kirjallinen kuvaus, jonka avulla voidaan osoittaa, että henkilötietoja käsitellään tietosuoja-asetuksen mukaisesti. (Tietosuojavaltuutetun toimisto 2019d.)

GDPR:ssä määritellään, että henkilötietoja on ”käsiteltävä lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvästi” (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5). Läpinäkyvyydellä tarkoitetaan tässä yhteydessä sitä, että käyttäjille tulee viestiä selkeällä ja yksinkertaisella kielellä, miten heidän henkilötietojansa kerätään ja käsitellään, ja missä määrin niitä tullaan käsittelemään (Tietosuojatyöryhmä 2018, 6). Hyvä tapa demonstroida läpinäkyvyyttä on luoda organisaatiolle tietosuojakäytäntö (engl. privacy policy). Tietosuojakäytännön tulee sisältää muun muassa seuraavat tiedot:

- Organisaation ja tietosuojavastaavan yhteystiedot;
- Perustelut sille, miksi lokitietoja ja muita henkilötietoja täytyy käsitellä;
- Oikeusperusta henkilötietojen käsittelylle;
- Tasapainotestin yksityiskohdat, mikäli oikeusperustana toimii oikeutettu etu;
- Kenelle lokitietoja ja muita henkilötietoja jaetaan ja kuinka kauan niitä säilytetään;
- Tietoa käyttäjien oikeuksista. (Bateman 2018.)

GDPR määrittelee, että henkilötietoja saa käsitellä ainoastaan sitä tarkoitusta varten, jota varten ne alun perin kerättiin. Toisin sanoen tämä tarkoittaa sitä, että esimerkiksi tietoturvallisuuden takaamiseksi kerättyjä tietoja ei voi käyttää markkinointiin. (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5.) Asetuksessa mainitaan myös tietojen minimointiperiaate, jonka mukaan ”henkilötietojen on oltava asianmukaisia ja olennaisia ja rajoitettuja siihen, mikä on tarpeellista suhteessa niihin tarkoituksiin, joita varten niitä käsitellään” (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5). Käytännössä tämä tarkoittaa sitä, että mitään henkilötietoja, joille ei ole asianmukaista tarvetta, ei tule kerätä (Bateman 2018). Organisaatioiden tulee myös varmistaa, että henkilötiedot ovat ajan tasalla, ja että epätodenmukaiset tiedot oikaistaan tai poistetaan välittömästi (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5).

Yksi GDPR:n henkilötietojen käsittelyä koskevista periaatteista liittyy säilytyksen rajoittamiseen. Tällä tarkoitetaan sitä, että organisaatiot saavat säilyttää henkilötietoja

ainoastaan niin kauan kun on tarpeellista niiden käyttötarkoituksen puolesta. Viimeinen periaate liittyy henkilötietojen eheyden ja luottamuksellisuuden takaamiseen. Asetuksen mukaan organisaatioiden tulee käyttää teknisiä ja organisatorisia tietoturvakontrolleja henkilötietojen suojaukseen. (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 5). Vaikka EU:n yleisessä tietosuoja-asetuksessa ei suoranaisesti mainitakaan lokitietojen keräämistä teknisenä kontrollina, lokien avulla voidaan demonstroida (LogSentinel 2018), että organisaatiolla on tarvittavat keinot havaita esimerkiksi tietomurrot. Muita teknisiä toimia, joita GDPR mainitsee, ovat henkilötietojen salaus ja pseudonymisointi (Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, artikla 32), joita kumpaakin tulisi soveltaa myös henkilötietoja sisältäviin lokeihin. Lokitietojen käsittelyoikeudet on aina rajattava tietotarpeen (engl. need to know basis) mukaan (Kyberturvallisuuskeskus 2019).

Muita lokien käsittelyyn vaikuttavia säädäntöjä ovat muun muassa tietosuojalaki (5.12.2018/1050), laki yksityisyyden suojasta työelämässä (13.8.2004/759) ja laki sähköisen viestinnän palveluista (7.11.2014/917). Viimeksi mainitun lain lukua 17 sovelletaan, mikäli lokitiedot sisältävät sähköisen viestinnän välitystietoja, kuten sähköpostiviestien lähettäjien ja vastaanottajien sähköpostiosoitteita tai verkko-osoitteita. Saman lain lukua 18 sovelletaan, mikäli organisaatio aikoo käyttää lokeja tai niitä tuottavia järjestelmiä, organisaation henkilöstön toimien monitorointiin, esimerkiksi tunnistaakseen väärinkäytöstapauksia.Tällöin on huomioitava, että valvonnasta on tiedotettava henkilöstölle ja lisäksi organisaation tulee järjestää yhteistoimintamenettely. (Kyberturvallisuuskeskus 2019.)

3.4 Lokien sisältö

Lokien sisällölle on vaikea määritellä yhtä kriteeristöä erityyppisten lokien suuren määrän vuoksi. Usein myös eri valmistajilla ja järjestelmillä on erilaiset käsitykset siitä, mitä lokien tulisi sisältää. (Chuvakin ym. 2012, 77.) Pääperiaatteena kuitenkin voidaan pitää, että lokien tulisi vastata aina jokaisen tapahtuman kohdalla seuraaviin kysymyksiin:

- Mitä tapahtui ja missä se tapahtui?
- Milloin ja miksi näin tapahtui?
- Miten se tapahtui?
- Kuka tai ketkä olivat osallisena tapahtumassa? (Chuvakin ym. 2012, 363.)

Jotta lokimerkintä olisi hyödyllinen tietoturvatapausten selvittämisessä, sen tulisi vastata kaikkiin edellä mainittuihin kysymyksiin. Tässä alaluvussa käydään läpi, millä tarkkuudella tapahtumien tietoja tulisi tallentaa lokimerkintöihin, jotta ne olisivat mahdollisimman hyödyllisiä.

Ensimmäinen tieto, joka lokimerkinnässä tulisi näkyä, on ajankohta ja aikavyöhyke. Koska verkkopalvelu käyttää hajautettuja järjestelmiä, aikavyöhykkeen tallentaminen on erittäin tärkeää, sillä kaikki järjestelmät eivät sijaitse samalla aikavyöhykkeellä. (Chuvakin ym. 2012, 369.) Myös itse tapahtuman ajankohdan sisällyttämistä lokeihin kannattaa harkita, sillä se saattaa erota lokimerkinnän ajankohdasta. OWASP suosittelee myös sisällyttämään niin sanotun ”interaction identifier”-kentän ajankohdan yhteyteen, jolla voidaan yhdistää kaikki käyttäjän vuorovaikutuksen aiheuttamat relevantit tapahtumat. Koska sovellus tietää jo ennestään, että kaikki nämä tapahtumat ovat seurausta käyttäjän tietystä toimesta, on huomattavasti käytännöllisempää ”yhdistää” nämä tapahtumat jo alussa sen sijaan, että niitä alettaisiin myöhemmin korreloimaan keskenään. (OWASP 2019b.)

Seuraavana lokimerkinnässä tulee näkyä lokimerkinnän prioriteetti, joka kertoo tapahtuman tärkeydestä ja vakavuudesta. Tapahtuman tärkeyteen on vaikeaa antaa yleisiä ohjeistuksia, sillä organisaatioilla on usein eri käsityksiä tapahtumien prioriteeteista ja vakavuuksista. (Chuvakin ym. 2012, 369). Yksi yleinen tapa määritellä tapahtuman vakavuus on syslog-viesteissä käytetty asteikko, joka on seuraavanlainen: ”0 = Emergency, 1 = Alert, 2 = Critical, 3 = Error, 4 = Warning, 5 = Notice, 6 = Informational ja 7 = Debug” (Wikipedia 2019a). Toinen OWASPin suosittelema vakavuusasteikko sisältää seuraavat arvot vakavimmasta vähiten vakavimpaan: ”fatal, error, warning, info, debug, trace” (OWASP 2019b).

Lokimerkinnän tulee sisältää myös tietoa siitä, missä tapahtuma sattui. Tällä tarkoitetaan yleensä järjestelmää, sovellusta tai jotain niiden komponenttia. (Chuvakin ym. 2012, 369). Lokimerkintään tulee sisällyttää vähintään sovelluksen tai järjestelmän nimi ja versio tai jokin muu tieto, josta se voidaan tunnistaa. Lisäksi lokimerkintään tulee sisällyttää sovelluksen tai järjestelmän osoite, kuten isäntänimi, palvelimen IP-osoite sekä portin numero. Nämä tiedot auttavat vastaamaan kysymykseen ”missä?”. OWASP suosittelee sisällyttämään lokeihin myös palveluiden nimet ja protokollat sekä verkkosovelluksen sisääntulo-URL:in ja käytetyn HTTP-metodin. (OWASP 2019b.)

Seuraavaksi lokimerkinnän tulee kertoa, kuka tai ketkä olivat osallisena tapahtumassa. Tekijä voi olla ihminen, järjestelmä tai laite. Lokimerkintään tulee merkitä käyttäjän identiteetti, mikäli se on tiedossa. Tämä on usein käyttäjä-ID tai käyttäjänimi. (OWASP 2019b.) Käyttäjä-ID:n tallentaminen on usein suositeltavin vaihtoehto todennettujen käyttäjien kohdalla, sillä ID ei yleensä muutu toisin kuin käyttäjänimet tai sähköpostiosoitteet. Kuitenkin sisäänkirjautumistapahtumissa on suositeltavaa tallentaa käyttäjänimi, sillä kaikki hyökkäykset eivät välttämättä ole kohdistettu tiettyyn käyttäjätunnukseen, vaan hyökkääjät saattavat kokeilla sisäänkirjautumista satunnaisilla käyttäjänimillä ja salasanoilla. Lisäksi lähdeosoite eli useimmissa tapauksissa käyttäjän IP-osoite, tulee merkitä lokeihin (OWASP 2019b). Myös käyttäjän käyttövaltuustaso kannattaa kirjata lokeihin (Kyberturvallisuuskeskus 2019). Tämän jälkeen lokimerkintään on suositeltavaa kirjata käyttäjän tai järjestelmän suorittama toimi (Chuvakin ym. 2012, 369), kuten sisään- tai uloskirjautuminen.

Seuraavana vuorossa on kohde, joka viittaa tapahtuman kohteena olevaan komponenttiin tai objektiin. Tapahtuman kohteena voivat esimerkiksi olla käyttäjätilit, erinäiset tiedostot tai muut resurssit. Seuraava kenttä sisältää statuksen, johon merkitään, onnistuiko vai epäonnistuiko kohdetta koskeva tapahtuma. Viimeisenä kirjataan tapahtuman syy, joka auttaa vastaamaan kysymykseen ”miksi?”. (Chuvakin ym. 2012, 369). Syy voi olla esimerkiksi riittämättömät käyttöoikeudet.

Näiden tietojen perusteella verkkopalvelu voi luoda esimerkiksi seuraavanlaisen lokimerkinnän:

2019/10/15 10:11:27AM GMT+2 priority: 3, system: testapplication, module: authentication, source: 127.0.0.1, user: testuser, accesslevel: admin, action: login, object: adminpage, status: failed, reason: ”incorrect password”. (Chuvakin ym. 2012, 370.)

Kuten yllä olevasta esimerkistä näkyy, lokimerkintä vastaa tämän alaluvun alussa esitettyihin kysymyksiin ”mitä, missä, milloin, miksi, miten ja kenen toimesta?”.

Viimeinen lokien sisältöä koskeva huomioitava asia on erotinmerkin käyttö. Kuten aiemmin esitetystä lokimerkinnästä näkyy, nimi/arvo-parit ovat eroteltu toisistaan pilkuilla. Vaikka pilkkua käytetäänkin usein erotinmerkkinä, se saattaa aiheuttaa joissain tilanteissa ongelmia. Mikäli lokimerkinnän jokin kenttä sisältää useampia arvoja, erotinmerkiksi kannattaa valita jokin muu merkki, jotta lokienhallintajärjestelmä osaa ryhmitellä lokimerkinnän kentät oikein. (Chuvakin ym. 2012, 370.) Muita hyviä vaihtoehtoja voivat olla esimerkiksi puolipiste tai pystyviiva eli putkimerkki.

3.5 Tietoturvan kannalta olennaiset tapahtumat

Suurin osa käyttöjärjestelmistä, verkkopalvelimista ja tietokantapalvelimista keräävät automaattisesti lokitietoja. Useimmissa tapauksissa pelkät infrastruktuurin luomat lokitiedot eivät kuitenkaan riitä tietoturvapoikkeusten selvittämiseen. Valitettavan usein verkkosovellukset eivät kerää minkäänlaisia räätälöityjä lokitietoja tapahtumista tai lokien keräämisprosessi on huolimattomasti konfiguroitu. (OWASP 2019b.)

OWASPin mukaan verkkosovellusten lähdekoodin tulee aina olla pääasiallinen tapahtumalokien lähde, sillä sovellukset sisältävät usein eniten tietoa käyttäjistä, kuten esimerkiksi heidän rooleistaan ja käyttöoikeuksistaan, sekä tapahtumien kontekstista. Myös infrastruktuurin eri komponentit sisältävät sovelluksen käyttöön liittyvää tietoa, ja niistä tulisi kerätä lokitietoja soveltuvin osin. Näitä ovat esimerkiksi uptime-tasoa mittaavat palvelut, verkkopalvelimien ohjelmiston suodattimet, tietokantasovellukset, verkkosovellusten palomuurit sekä käyttäjien lataamia tiedostoja tarkistavat virustorjuntaohjelmistot. (OWASP 2019b.)

Todennukseen ja käyttövaltuuksiin liittyvät lokitiedot

Tietoturvallisuuden kannalta yksiä tärkeimpiä tapahtumia ovat todennukseen (engl. authentication) ja valtuutukseen (engl. authorization) liittyvät tapahtumat (Chuvakin ym. 2012, 239). Todennus viittaa prosessiin, jossa käyttäjän tai järjestelmän identiteetti varmistetaan ennen kuin hänelle tai sille annetaan pääsy sovellukseen, rajapintaan tai muihin tietolähteisiin (Gebel 2018). Todennus tapahtuu tyypillisesti käyttäjätunnuksella ja salasanalla, mutta muitakin todennustapoja on (Techopedia 2019c). Kun todennus on onnistunut, käyttäjän tai järjestelmän valtuudet eli käyttöoikeudet tarkistetaan. Käyttövaltuudet määrittelevät, mitä toimia käyttäjä tai järjestelmä voi suorittaa millekin resurssille. (Gebel 2018.)

Tietoturvallisuuden varmistamiseksi on tärkeää tallentaa lokitietoihin kaikki todennukseen ja valtuuksiin liittyvät tapahtumat, sillä niitä tarkastelemalla voidaan havaita mahdolliset tietoturvapoikkeamat, kuten tapaukset, joissa käyttäjä on yrittänyt ylittää omat käyttövaltuutensa (Lokiohje 2009, 32). Lokeihin tulee tallentaa kaikki onnistuneet ja epäonnistuneet sisäänkirjautumiset ja niiden yritykset, niin sovellustasolta kuin infrastruktuurin eri komponenteistakin, kuten palvelimilta. Tämä koskee myös sisäänkirjautumisyrityksiä käytöstä poistettuihin ja olemattomiin tileihin (Chuvakin ym. 2012, 239.), sillä niiden yritykset saattavat viitata siihen, että hyökkääjä yrittää päästä sisään satunnaisiin

tileihin. Myös oletustilejä koskevat sisäänkirjautumisyritykset tulee kirjata lokeihin (Chuvakin ym. 2012, 239), sillä hyökkääjät usein pyrkivät kirjautumaan useista järjestelmistä tyypillisesti löytyviin tileihin, kuten ylläpitäjätileihin. Erityistä huomiota tulee kiinnittää epätyypillisiin aikoihin, kuten yöllä tai työajan jälkeen tapahtuviin kirjautumisyrityksiin (Chuvakin ym. 2012, 239), sillä hyökkääjät iskevät usein silloin pienemmän kiinnijäämismahdollisuuden vuoksi.

Erityisesti etuoikeutettuihin tileihin, kuten ylläpitäjien (engl. administrator) ja superkäyttäjien (engl. root user), tileihin liittyviä todennustapahtumia tulee tallentaa ja tarkastella säännöllisesti. Syy tähän on se, että tällaisilla tilityypeillä on usein huomattavasti enemmän käyttöoikeuksia kuin normaaleilla tileillä, ja näin ollen niillä pystytään tekemään huomattavasti enemmän tuhoa. (Chuvakin ym. 2012, 239.) Lisäksi käyttövaltuuksiin tehdyistä muutoksista on hyvä kirjata lokia (OWASP 2019b), sillä hyökkääjä saattaa muokata muiden käyttäjien valtuuksia. Esimerkki tällaisesta tilanteesta on se, että hyökkääjä on saanut haltuunsa pääkäyttäjän tilin. Kiinnijäämisen mahdollisuuden minimoimiseksi, hyökkääjä antaa ylläpito-oikeudet jollekin toiselle käyttäjälle, ja käyttää kyseistä tiliä alkuperäisen pääkäyttäjän tilin sijaan. Mikäli käyttövaltuuksiin tehtyjä muutoksia ei seurata, hyökkääjä saattaa kyetä käyttämään tätä tiliä kenenkään huomaamatta.

Muutoksen hallintaan liittyvät lokitiedot

Järjestelmissä ja sovelluksessa tapahtuvat kriittiset muutokset on suositeltavaa tallentaa lokeihin, sillä muutoksilla on usein suora vaikutus tietoturvallisuuden tasoon. Esimerkiksi luvattomat muutokset määritystiedostoihin, arkaluontoiseen dataan tai käyttäjätileihin saattavat aiheuttaa järjestelmien ja sovelluksen kaatumisen, datan menetyksen tai muita tietoturvatapahtumia. (Chuvakin ym. 2012, 241.)

Lokitiedostoihin tulee tallentaa monipuolisesti tietoa erinäisistä muutoksista. Käyttäjiä ja käyttäjäryhmiä koskevia muutoksia ovat esimerkiksi niiden lisäykset, muokkaukset ja poistot. Hyökkääjät saattavat esimerkiksi lisätä uusia käyttäjiä murrettuihin järjestelmiin ja sovelluksiin, ja poistaa kyseiset käyttäjät myöhemmin, jotta he eivät jäisi kiinni. Ilman lokitietoja kyseisiä tapahtumia on hankala havaita. Erityisesti pääkäyttäjien ja muiden etuoikeutettujen käyttäjien tileihin tehtyjä muutoksia on tärkeää tallentaa ja seurata, sillä kyseisillä tilityypeillä on mahdollista saada aikaan suurempaa tuhoa. (Chuvakin ym. 2012, 241.)

Kaikki salasanojen muutoksia ja nollaamisia koskevat tapahtumat on suositeltavaa tallentaa lokeihin. Mikäli sovellus mahdollistaa sen, että pääkäyttäjät voivat nollata

käyttäjien salasanoja, tulee tällaiset tapahtumat myös tallentaa lokeihin. (Chuvakin ym. 2012, 241.) Näiden lokien avulla voidaan esimerkiksi havaita asiattomia salasanojen muutoksia. Erityisesti ylläpitäjien ja muiden etuoikeutettujen käyttäjien salasanojen muutoksia ja nollaamisia tulee tarkkailla.

Erityisen tärkeää on tallentaa palvelimien, sovellusten ja muiden infrastruktuurin komponenttien järjestelmä- ja määritystiedostoihin tehtyjä muutoksia. Näiden avulla voidaan havaita haitallisia ja esimerkiksi vahingossa tehtyjä toimia. Lisäksi muihin kriittisiin tiedostoihin, sekä niiden käyttöoikeuksiin, tehtyjä muutoksia tulee tallentaa lokitietoihin. Erityisesti tiedostojen käyttövaltuuksiin liittyviä muutoksia on vaikeaa havaita ilman lokeja tai säännöllistä tarkastelua. Myös palvelimiin, sovelluksiin ja muihin komponentteihin tehtyjä päivityksiä ja asennuksia on hyvä tarkkailla. (Chuvakin ym. 2012, 241–242.)

Resurssien käsittelyyn liittyvät lokitiedot

Resurssien käsittelyyn liittyviä tapahtumia on suositeltavaa tallentaa lokeihin, sillä niiden avulla voidaan helposti selvittää, mitä resursseja hyökkääjä on tarkastellut ja mahdollisesti muokannut. Kyseisten lokien avulla voidaan saada selville myös organisaation sisällä tapahtuvia väärinkäyttötapauksia tai petoksia. Resurssien käsittelyyn liittyviä lokeja tulee tallentaa itse sovelluksesta sekä sen erinäisistä komponenteista, kuten palvelimista ja tietokannoista. Erityisesti huomiota tulee kiinnittää ylläpitäjien tietokantoja koskeviin toimiin, sekä työajan ulkopuolella tapahtuvaan kriittisten järjestelmien resurssien käsittelyyn. (Chuvakin ym. 2012, 244–245.)

Erinäisistä tietokantoihin liittyvistä tapahtumista on syytä kirjata lokia, sillä tietokannat usein sisältävät henkilötietoja ja muita arkaluontoisia tietoja (Chuvakin ym. 2012, 245). Koska tietokannoissa säilytetään taloudellisesti arvokkaita tietoja, ne ovat usein tietomurtojen perimmäinen kohde (Lokiohje 2009, 37). Lokeihin tulee kirjata esimerkiksi tietokantojen lukutapahtumat sekä käyttäjien tekemät kyselyt hakuparametreineen (Lokiohje 2009, 30). Mikäli mahdollista, näistä tapahtumista tulisi jättää pois tiedossa olevat sovelluskyselyt, jotta poikkeukselliset kyselyt voidaan havaita helpommin. Erityisesti huomiota tulee kiinnittää lisäys- ja poistokyselyihin, jotta voidaan varmistua, että tietokannan eheys ja saatavuus säilyy. Lisäksi tietokantoja koskevat muokkauskyselyt on suositeltavaa tallentaa lokeihin. (Chuvakin ym. 2012, 244–245.) Muita tietokantakyselyitä, joita tulee seurata, ovat tietokantataulujen luomiseen käytettävä CREATE-kysely sekä GRANT-kysely (Chuvakin ym. 2012, 245), jonka avulla voidaan myöntää

käyttövaltuuksia muun muassa objektien poistamiseen, lisäämiseen ja päivittämiseen (MySQL 2019).

Näitä edellä mainittuja tietokantatapahtumia kirjataan yleensä automaattisesti tietokantojen luomaan transaktiolokiin, mutta siitä huolimatta on suositeltavaa varmistaa, että toiminto on käytössä. Monet tietokannat kykenevät kirjaamaan lokitietoja sisään- ja uloskirjautumisista, ylläpitäjien suorittamista toiminnoista, virhetilanteista, rakenteeseen ja käyttöoikeuksiin tehdyistä muutoksista sekä tietokannan käynnistymisestä ja sammuttamisesta. (Lokiohje 2009, 37–38.) Mikäli tietokanta kykenee kirjaamaan edellä mainittuja tapahtumia lokeihin, asetus kannattaa kytkeä päälle.

Tietokantojen varmuuskopioista on suositeltavaa kirjata merkintä lokitietoihin (Chuvakin ym. 2012, 244–245), jotta voidaan varmistua, että varmuuskopiointi on suoritettu onnistuneesti. Tällöin lokitiedoista voidaan myös havaita luvattomat varmuuskopiot. Tietokantojen varmuuskopiot ovat houkutteleva kohde hyökkääjille, sillä niiden avulla on mahdollista ottaa suuriakin määriä dataa talteen kerralla. (Chuvakin ym. 2012, 244–245.)

Resurssien käsittelyä koskevien lokien keräämisessä kannattaa myös hyödyntää verkkopalvelimen automaattisesti keräämiä käyttölokeja. Verkkopalvelimen käyttölokeihin tallennetaan, millä sivustoilla käyttäjä on milläkin hetkellä vieraillut, ja tästä syystä ne ovat korvaamattomia tietomurtotapauksia selvitettäessä. Verkkopalvelimen käyttölokien avulla on myös mahdollista havaita mahdollisia hyökkäyksiä, kuten palvelunestohyökkäyksiä (engl. Denial of Service, DoS). (LogDNA 2018.)

Kriittisiä virheitä ja vikatilanteita koskevat lokitiedot

Kriittiset virheet ja järjestelmien kaatumiset saattavat viitata mahdolliseen haittaohjelmatartuntaan tai muuhun tietoturvauhkaan. Usein lokeista näkyykin ensimmäiset merkit haitallisesta toiminnasta. Vaikka monet järjestelmien ja sovellusten kriittiset virheet eivät suoranaisesti liitykään tietoturvallisuuteen, ne saattavat ilmaista, että järjestelmässä on haitallista toimintaa. (Chuvakin ym. 2012, 247–248.)

Lokimerkintöjen tulee sisältää tietoa sovellusten ja järjestelmien kaatumisista. Vaikka kaatuminen voikin johtua tietoturvaan liittymättömistä syistä, tapahtumat pitäisi aina tutkia, sillä ne vaikuttavat negatiivisesti sovelluksen saatavuuteen. (Chuvakin ym. 2012, 247–248.) Sovelluksen kaatuminen voi myös johtua erinäisistä hyökkäyksistä, kuten palvelunestohyökkäyksestä. Lisäksi sovelluksen ja sen infrastruktuurin komponenttien sekä

itse lokituksen käynnistymis- ja sammuttamistapahtumat tulee tallentaa lokeihin (OWASP 2019b).

Yksi tärkeimmistä virhelokien lähteistä on verkkopalvelin. Verkkopalvelimien virhelokit sisältävät tietoa palvelimien ja sen komponenttien virhetiloista ja niiden syistä. Lisäksi ne sisältävät usein tietoa esimerkiksi palvelimen kokoonpanoasetuksista. Virhelokien sisältö vaihtelee tapahtuman mukaan, mutta yleisesti ottaen ne sisältävät tietoa tapahtuman ajankohdasta, vakavuudesta ja tapahtuman tuottaneesta komponentistä. (LogDNA 2018.)

Myös palvelimien kapasiteetin ja järjestelmäresurssien, kuten muistin ja prosessorin, käyttötietojen monitorointia tulee harkita, sillä korkea resurssien käyttöaste saattaa johtua esimerkiksi palvelunesto- tai väsytyshyökkäyksestä. Varmuuskopioista ja eritoten niiden epäonnistumisista on myös suositeltavaa tallentaa merkintöjä lokeihin, sillä ne saattavat viitata tietomurron yritykseen. (Chuvakin ym. 2012, 248.)

Muita tietoturvallisuuden kannalta olennaisia lokitietoja

OWASPin mukaan muita tietoturvallisuuden kannalta olennaisia tapahtumia, joista tulisi kirjata lokeja, ovat muun muassa syötteen ja tulosteen tarkastamisen epäonnistumiset. Näihin sisältyy esimerkiksi protokollan rikkomukset, virheelliset parametrien nimet ja arvot sekä tietokantatietueiden yhteensopimattomuus. Myös istunnonhallintaa koskevia epäonnistumisia on suositeltavaa kirjata lokeihin, kuten evästeiden tunnistusarvoja (engl. cookie session identification value) koskevia muutoksia. (OWASP 2019b.)

Verkkopalveluissa on ensisijaisen tärkeää tarkistaa kaikki ladatut tiedostot haittaohjelmien varalta, jotta niiden avulla ei voida saastuttaa järjestelmiä. Tämän vuoksi tiedostoja tarkistavan virustorjuntaohjelmiston havaintoja koskevat lokit tulee lähettää lokienhallintajärjestelmään (OWASP 2019b). Virustorjuntaohjelmiston toiminnallisuutta koskevista tapahtumista, kuten virhetiloista, päivitysten epäonnistumisista ja ohjelmiston kaatumisista, kerättyjä lokeja on suositeltavaa lähettää lokienhallintajärjestelmään (Chuvakin ym. 2012, 246 – 247), jotta kyseiset tapahtumat havaitaan ja niihin voidaan reagoida. Käyttäjien lisäämästä sisällöstä, etenkin ladatuista tiedostoista, on tärkeää kirjata lokimerkintöjä (OWASP 2019b), jotta voidaan tarpeen tullen selvittää, kuka on ladannut minkäkin tiedoston.

Tietoturvapoikkeamia havainnoivien järjestelmien tuottamia lokeja on myös suositeltavaa kerätä lokienhallintajärjestelmään. Näihin järjestelmiin kuuluvat muun muassa

verkkosovelluksen palomuurit (engl. web application firewall, WAF) ja tunkeutumisen havainnointijärjestelmät (engl. intrusion detection system, IDS). (OWASP 2019b.)

Vaikka käyttäjien antamat suostumukset eivät suoranaisesti liitykään tietoturvaan, niitä on suositeltavaa kirjata lokeihin lainsäädännöllisistä ja muista syistä. Tällaisia ovat esimerkiksi henkilötietojen käsittelyyn ja suoramarkkinointiin annettu suostumus sekä vahvistus siitä, että käyttäjä on lukenut ja hyväksynyt käyttöehdot ja palvelun tietosuojakäytännön. (OWASP 2019b.)

Mitä tietoja lokeihin ei tule tallentaa?

Tiettyjä tietoja ei tule ikinä tallentaa lokitietoihin lainsäädännöllisistä tai muista syistä. Lähtökohtaisesti mitään tietoja ei tule tallentaa, jotka eivät ole tarpeellisia käyttötarkoituksen kannalta, etenkään henkilötietoja. (Kyberturvallisuuskeskus 2019.) Seuraavia tietoja ei tule tallentaa lokeihin selkokielisenä:

- Henkilötiedoiksi luokiteltavia tietoja tai muuta arkaluontoista tietoa, kuten henkilötunnuksia, nimiä, syntymäaikoja, puhelinumeroita tai salasanoja (Chuvakin, ym. 2012, 371).
- Terveyteen liittyviä tietoja.
- Sovellukseen liittyviä tietoja, kuten sen lähdekoodia, istunnon tunnisteita (engl. session identification values), käyttöoikeustietueita (engl. access tokens), tietokantojen yhteysjonoja, salausavaimia tai muita salaisuuksia.
- Taloudellisia tietoja, kuten pankkitileihin tai luottokortteihin liittyviä tietoja.
- Kaupallisesti arkaluontoisia tietoja.
- Tietoja, joiden keräämiseen käyttäjä ei ole antanut lupaa tai jossa käyttäjän antama suostumus on vanhentunut. (OWASP 2019b.)
- Henkilöiden lähettämien viestien sisältöä (Kyberturvallisuuskeskus 2019).

Suurinta osaa yllä mainituista tiedoista ei tulisi tallentaa lokitietoihin alkuunkaan. Mikäli lokeihin kuitenkin tallentuu kyseisiä tietoja, ne tulee poistaa, hajauttaa (engl. hash), salata tai pseudonymisoida (OWASP 2019b). Esimerkiksi salasanoja saattaa tallentua selkokielisenä lokeihin, mikäli käyttäjä syöttää vahingossa salasanan käyttäjätunnuksen tilalle (Kyberturvallisuuskeskus 2019). Useat pilvipohjaiset lokienhallintapalvelut sisältävät työkaluja, joiden avulla salasanat ja muut arkaluontoiset tiedot voidaan naamioida tai piilottaa lokeista.

3.6 Lokien suojaus

Koska lokit sisältävät arkaluontoista tietoa, niitä tulee suojata asianmukaisesti. Lokitiedot ja niitä tuottavat järjestelmät tulee suojata peukaloinnilta (engl. tampering), ja toimeksiantajan tulee varmistaa, että niihin ei pääse käsiksi luvattomat henkilöt tai järjestelmät. Toimeksiantajan on tärkeää varmistaa, että lokien tallentamiseen tarkoitetussa järjestelmässä on tarpeeksi tallennustilaa, jotta lokien tallentaminen ei lakkaa tai vanhojen lokien päälle ei tallenneta uusia lokeja. (SFS-EN ISO/IEC 27002 2017, 51.) Useat pilvipohjaiset lokienhallintajärjestelmät ratkaisevat edellä mainitun ongelman, sillä lokien keräämistä jatketaan, vaikka organisaatio ylittäisikin tallennuskapasiteettinsa. Tästä aiheutuu kuitenkin ylimääräisiä kuluja.

Yhdysvaltain kauppaministeriön alainen National Institute of Standards and Technology eli NIST (National Institute of Standards and Technology 2017) määrittelee lokien tietoturvaa koskevassa ohjeistuksessaan useita lokien suojaamista koskevia näkökohtia, joita organisaatioiden tulisi harkita. Ohjeistuksen mukaan organisaatioiden tulee suojella lokitietojen eheyttä rajoittamalla käyttäjien pääsyä ja käyttövaltuuksia lokitietoihin. Tavallisten käyttäjien ei tulisi päästä tarkastelemaan lokitietoja, eikä yhdenkään käyttäjän pitäisi pystyä muokkaamaan, poistamaan tai uudelleennimeämään niitä. (Kent & Souppaya 2006, 5-4.) Mikäli mahdollista, lokitietojen katselusta ja tehdyistä hauista tulisi jäädä merkintä lokeihin (Lokiohje 2009, 62). Mikäli lokitietoja ei ole asianmukaisesti suojattu sekä liikkeessä että levossa, tämä saattaa mahdollistaa lokitietojen muokkaamisen ja tuhoamisen, joka pahimmassa tapauksessa johtaa siihen, että epäilyttäviä tapahtumia ei havaita (Kent & Souppaya 2006, 2-9).

Lokitiedostojen eheyden varmistaminen on tärkeä toimenpide lokienhallinnassa (Lokiohje 2009, 63), sillä jos lokeja voidaan muokata, niiden sisältämään tietoon ei voida enää luottaa. Eheyden varmistaminen on olennainen osa tietoturvallisuuden varmistamista ja se tarkoittaa sitä, että tietoa ei ole muokattu sen kulun aikana tai levossa (Infosec 2018). Lokitietojen eheys voidaan varmistaa esimerkiksi laskemalla lokitiedostoille tarkistussumma, jonka avulla kyetään havaitsemaan, mikäli lokitietoja on muutettu. Jos lokitiedostoon tehdään muutoksia, sen tarkistussumma muuttuu, eikä enää täsmää alkuperäisen tarkistussumman kanssa. Näin ollen voidaan päätellä, että lokitiedostoa on muutettu. Mikäli mahdollista, toimeksiantajan tulee käyttää mahdollisten tarkistussummien laskemiseen vahvoja tiivistefunktioita, kuten SHA-512 (Secure Hash Algorithm) algoritmia. (Lokiohje 2009, 63.)

Lokien suojaamiseksi on tärkeää, että järjestelmien ylläpitäjien tekemiä toimintoja tallennetaan ja monitoroidaan, sillä heillä on usein tarvittavat oikeudet muokata lokitietoja. Minimoidakseen tämän mahdollisuuden, lokitietoja tulee suojata ja katselmoida säännöllisesti. (SFS-EN ISO/IEC 27002 2017, 51.) Lokitietojen muokkausyritysten tulisi myös laukaista hälytys tarvittaessa (Kyberturvallisuuskeskus 2019). Pilvipohjaiset lokienhallintajärjestelmät saattavat ratkaista edellä mainitun ongelman, sillä useimmat niistä eivät mahdollista lokien poistamista tai muokkaamista.

Kaikki kerätty lokidata tulee myös sanitoida (engl. sanitize), jotta vältytään loki-injektioilta (engl. log injection) (OWASP 2019b). Loki-injektio tarkoittaa hyökkäysmenetelmää, jossa hyökkääjä lisää väärennettyjä lokimerkintöjä lokitiedostoihin tai sisällyttää lokeihin haitallista sisältöä. Hyökkäys voidaan esimerkiksi toteuttaa siten, että käyttäjä syöttää sovelluksen lomakkeeseen tiettyjä merkkejä, jolloin sovellus virheellisesti tulkitsee ne lokimerkinnäksi. Kyseisen hyökkäyksen avulla hyökkääjä voi esimerkiksi turmella lokitiedoston tai vääristellä lokimerkintöjä välttyäkseen kiinnijäämiseltä. (OWASP 2019c.)

Lopuksi toimeksiantajan tulee myös varmistaa, että lokitietojen lähettäminen järjestelmistä kohteeseen tapahtuu salatun yhteyden yli (Kent & Souppaya 2006, 5-5). Lisäksi lokitiedoista on tärkeää tallentaa säännöllisiä varmuuskopioita (Kyberturvallisuuskeskus 2019).

4 LOKIENHALLINTAPALVELUIDEN SOVELTUVUUSSELVITYS

Tässä luvussa esitellään ja testataan kahden pilvipohjaisen lokienhallintapalvelun ominaisuuksia. Pilvipohjaisella lokienhallintapalvelulla tarkoitetaan keskitettyä lokienhallintajärjestelmää, joka muun muassa kerää, analysoi ja säilyttää lokitietoja pilvessä. Pilvipohjaisen palvelun tuomia etuja ovat muun muassa niiden käyttöönoton helppous ja skaalautuvuus. (Papertrail 2019b.)

4.1 Palveluiden kriteerit

Esiselvitystä varten valitaan viisi pilvipohjaista lokienhallintapalvelua, joiden ominaisuuksia vertaillaan toimeksiantajan asettamiin kriteereihin. Esiselvitystä varten luodaan taulukko, josta näkyy kunkin valitun palvelun tietoja ja ominaisuuksia. Taulukon tarkoitus on kuvailla lyhyesti, mitä toimeksiantajan asettamia kriteereitä kukin palvelu täyttää. Toimeksiantaja valitsee esiselvityksen perusteella kaksi potentiaalisimmalta vaikuttavaa palvelua, joiden ominaisuuksia ja käytettävyyttä testataan ja arvioidaan. Tarkempaan testaukseen otettavien palveluiden määrä päätettiin rajata kahteen jo opinnäytetyöprosessin alussa aikataulullisista syistä. Esiselvityksessä painotetaan seuraavia seikkoja:

Hinnoittelumalli

Selvityksessä on huomioitava, mitkä seikat vaikuttavat palveluiden hinnoitteluun. Esimerkiksi lokitietojen säilytysaika, käyttäjien määrä ja ominaisuudet.

Tietoturvallisuus

Selvityksessä otetaan myös luonnollisesti palveluiden tietoturva huomioon. Suotavaa olisi, että palvelun datakeskukset sijaitsisivat EU-alueella, mutta tästä voidaan tehdä poikkeuksia, mikäli palvelu osoittaa noudattavansa GDPR:ää ja on sitoutunut noudattamaan Privacy Shield -järjestelyä. Tietoturvallisuuden kannalta on äärimmäisen tärkeää, että palveluun lähetettävät lokit ovat asianmukaisesti salattu sekä liikkeessä että levossa, jotta luvattoman käsittelyn mahdollisuus on minimoitu. Vaatimuksena on myös, että palvelussa voidaan ottaa käyttöön monivaiheinen tunnistautuminen (engl. multi-factor authentication, MFA) tai kertakirjautuminen (engl. single sign-on, SSO). Palvelussa

tulee olla roolipohjainen pääsynhallinta, jotta käyttöoikeuksia voidaan jakaa eri käyttäjien kesken käyttötarpeen mukaan. Hyvänä lisänä pidetään sitä, jos palvelu tarjoaa valmiin työkalun arkaluontoisen datan piilottamiseksi tai poistamiseksi.

Käytettävyys

Palvelun tulee olla helppokäyttöinen ja sen konfiguroinnin tulee olla mahdollisimman yksinkertaista. Palvelun tulee myös mahdollistaa useiden erityyppisten lokien lähetys ja analysointi, ja sen pitää tukea useita eri siirtämismetodeja lokiagenteista koodikirjastoihin. Palvelulla on oltava integraatioita tunnettujen pilvipalveluntarjoajien kanssa, ja lokitietoja tulee pystyä visualisoimaan eri kuvaajien avulla.

Arkistointi

Palvelussa tulee olla mahdollisuus arkistoida lokitietoja automaattisesti myöhempää tarkastelua varten. Erityisesti Amazonin S3-palveluun arkistointi lasketaan eduksi.

Hälytykset ja hälytystavat

Palvelussa tulee olla mahdollisuus lähettää hälytyksiä esimerkiksi poikkeustilanteissa sähköpostin tai Slackin kautta. Muut hälytystavat lasketaan eduiksi.

4.2 Esiselvitykseen valikoidut palvelut

Esiselvitykseen valikoitui seuraavat lokienhallintapalvelut: Datadog, LogDNA, Papertrail, Scalyr ja Sematext. Lähes kaikissa palveluissa on joitain puutteita, ja näistä palveluista ainoastaan Datadog ja LogDNA täyttävät kaikki toimeksiantajan asettamat pakolliset vaatimukset. Tästä syystä nämä palvelut valikoituivat tarkempaan testaukseen. Kaikki esiselvitykseen valitut palvelut sekä niiden kuvaukset löytyvät liitteestä 1.

4.3 Testausympäristö

Oikeassa tuotantoympäristössä lokeja syntyy automaattisesti sitä mukaa, kun käyttäjät suorittavat erinäisiä toimia. Koska soveltuvuusselvitystä varten ei ole mahdollista suorittaa testausta oikeassa tuotantoympäristössä, selvitystä varten luodaan kolme Python-komentosarjaa, joita suoritetaan PyCharm-ohjelmointiympäristön kautta. Ensimmäisen komentosarjan luomat lokimerkinnät pyrkivät jäljittelemään verkkosovelluksen luomia

lokeja. Komentosarja valitsee satunnaisesti listasta erinäisiä objekteja ja kirjaa niistä lokimerkintöjä niin kauan, kunnes komentosarja pysäytetään. Lokimerkinnät sisältävät samat kentät, jotka määriteltiin tarpeellisiksi opinnäytetyön edellisen luvun ohjeistuksessa.

Toinen komentosarja jäljittelee Apache-verkkopalvelimen luomia käyttölokeja, ja se toimii samalla periaatteella kuin ensimmäinen komentosarja. Viimeinen komentosarja sen sijaan pyrkii simuloimaan väsytyshyökkäystä kirjaamalla lokiin 99 epäonnistunutta sisäänkirjautumista samasta käyttäjätunnuksesta ja IP-osoitteesta, jonka jälkeen se kirjaa lokiin yhden onnistuneen sisäänkirjautumisen. Tällä komentosarjalla on tarkoitus kokeilla palveluiden hälytysominaisuutta.

4.4 Palveluiden testaus

Tässä luvussa käsitellään LogDNA:n ja Datadogin ominaisuuksia ja käyttökokemuksia. Kummankin palvelun kohdalla käydään läpi niiden käyttöönotto sekä testataan niiden eri ominaisuuksia. Lopuksi kummankin palvelun ominaisuuksia ja käytettävyyttä arvioidaan lyhyesti.

4.4.1 LogDNA-palvelu

LogDNA on yhdysvaltalainen SaaS-lokienhallintapalvelu, johon asiakkaat voivat lähettää lokeja eri sovelluksista, palvelimista, alustoista ja järjestelmistä. Lokit tallennetaan keskitettyyn järjestelmään, joka sijaitsee joko palveluntarjoajan tai asiakkaan pilvi-infrastruktuurissa. Palvelun toiminnot suoritetaan verkkosovelluksen kautta, joten palvelua on teoriassa mahdollista käyttää useilla eri laitteilla, kuten tietokoneilla ja mobiililaitteilla. (LogDNA 2019b.)

Käyttöönotto

LogDNA tarjoaa ilmaisen kahden viikon kokeilujakson rajattomalla tallennustilalla, jonka aikana palvelun kaikkia ominaisuuksia voi kokeilla (LogDNA 2019f). Palveluun rekisteröityminen tapahtuu LogDNA:n verkkosivujen kautta. Palveluun on mahdollista kirjautua Googlen, GitHubin tai Herokun käyttäjätunnuksilla, mutta kokeilujaksoa varten päätettiin luoda kokonaan uudet tunnukset.

Kun rekisteröinti on suoritettu ja organisaation nimi lisätty, palvelu ehdottaa eri vaihtoehtoja lokien lähettämiselle. Valikoimasta löytyy muun muassa agenttiohjelmia eri palvelimille, integraatioita eri alustojen kanssa ja koodikirjastoja. Palvelu tarjoaa myös alkuun mahdollisuuden kokeilla sen ominaisuuksia näytedatalla, mikä on hyödyllistä, mikäli käyttäjä ei halua lähettää oikeaa lokidataa heti aluksi.

Alkuperäinen tarkoitus oli käyttää koodikirjastoa lokien lähettämiseen, mutta tämä osoittautui haastavaksi, sillä koodikirjasto ei huomioinut lokimerkinnöille asetettua formaattia eikä komentosarjojen satunnaisgeneraattori toiminut. Lopulta ratkaisuksi valittiin lokiagentti, sillä sen avulla on mahdollista kerätä räätälöityjä lokeja suoraan lokitiedostosta. Koska testauksessa käytetään Mac-tietokonetta, lokiagentti asennetaan siihen Homebrew-paketinhallintajärjestelmän avulla. Ennen lokiagentin asentamista Homebrew päivitetään ajan tasalle komennolla ”brew update”. Kun Homebrew on päivitetty, LogDNA:n lokiagentti asennetaan komennolla ”brew cask install logdna-agent”.

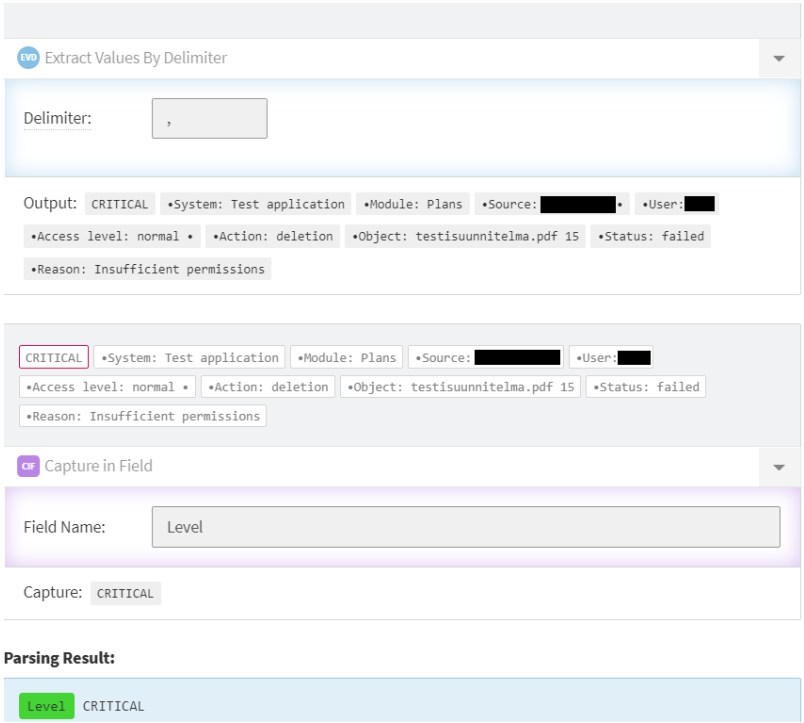
Lokiagentin asentamisen jälkeen lokiagentti tulee aktivoida komennolla ”sudo logdna-agent -k <avain>”, jossa avaimella tarkoitetaan organisaatiolle yksilöllistä avainta. Tämän niin sanotun ”ingestion keyn” avulla määritellään, mille tilille lokeja lähetetään, ja se on tarkoituksella jätetty pois tästä opinnäytetyöstä. Kun avain on lisätty, lokiagentti aktivoituu. Seuraavaksi määritellään, mitä hakemistoja lokiagentin tulee monitoroida. Tämä tapahtuu LogDNA:n komennolla ”sudo logdna-agent -d”, jonka perään lisätään tiedostopolku. Kun tarvittavat hakemistot on määritelty, lokiagentti määritellään toimimaan jatkuvasti taustalla komennolla ”sudo launchctl load -w /Library/LaunchDaemons/com.logdna.logdna-agent.plist”. Tämä komento mahdollistaa sen, että lokiagentti kykenee lähettämään lokitietoja palveluun lähes reaaliajassa.

Räätälöityjen lokien jäsentely

Ennen kuin palvelun ominaisuuksien kokeileminen voidaan aloittaa, LogDNA:han tulee luoda niin sanottu Custom Log Parser -jäsennin. Vaikka LogDNA osaakin tulkita automaattisesti useita eri lokiformaatteja Apache-verkkopalvelimien lokeista JSON-muotoisiin lokeihin, se ei osaa automaattisesti jäsennellä eli parsia komentosarjojen luomia lokeja. Koska komentosarjojen luomat lokit noudattavat räätälöityä formaattia ja sisältävät joitakin sovelluskohtaisia kenttiä, LogDNA:n asetuksista tulee määritellä erikseen, miten tämänkaltaisia lokeja tulee jäsennellä, jotta kenttiä voidaan käyttää kuvaajissa ja suodattimissa. Kyseinen toiminto löytyy LogDNA:n asetusten kohdasta ”Parsing”. Aluksi toiminto pyytää lisäämään manuaalisesti lokimerkinnän tai vaihtoehtoisesti suorittamaan

haun. Koska lokiagentti konfiguroitiin lähettämään lokeja palveluun jo aiemmin, suoritetaan haku, joka noutaa kaikki komentosarjojen luomat lokit.

Seuraavaksi määritellään erotinmerkki. Erotinmerkiksi valitaan pilkku, sillä komentosarjojen luomissa lokimerkinnöissä nimi/arvo-parit erotetaan toisistaan sen avulla. Tämän jälkeen LogDNA erottaa nimi/arvo-parit automaattisesti omiksi kentikseen. Seuraavaksi käydään yksitellen läpi kaikki kentät ja suoritetaan kunkin kohdalla toiminto ”Capture in Field”, joka ”kaappaa” kentän arvon ja mahdollistaa kentän nimeämisen (Kuva 1).



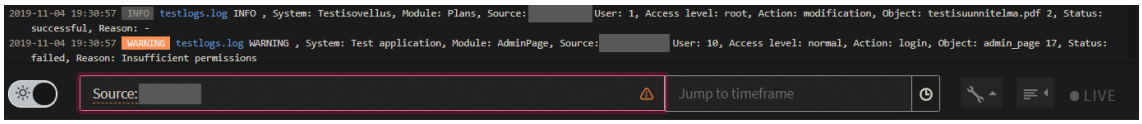
Kuva 1. LogDNA:n Custom Log Parser.

Kun kaikille kentille on suoritettu edellä mainittu toimenpide, LogDNA esittää vastaavanlaisia lokimerkintöjä, joista jokainen pitää validoida manuaalisesti. Tällä varmistetaan, että jäsennin tulkitsee lokimerkintöjä oikein. Todellisessa tuotantoympäristössä tätä osuutta ei tarvitse suorittaa muuta kuin verkkosovelluksen lokien kohdalla, sillä LogDNA osaa automaattisesti jäsennellä esimerkiksi verkkopalvelimien lokeja. Kun tämä toimenpide on suoritettu, aloitetaan palvelun ominaisuuksien testaaminen.

Käyttö

Ensivaikutelma LogDNA:n käyttöliittymästä on, että se näyttää selkeältä ja modernilta. Palvelun aloitusnäkymä on niin sanottu Everything-näkymä, jossa näkyy kaikki

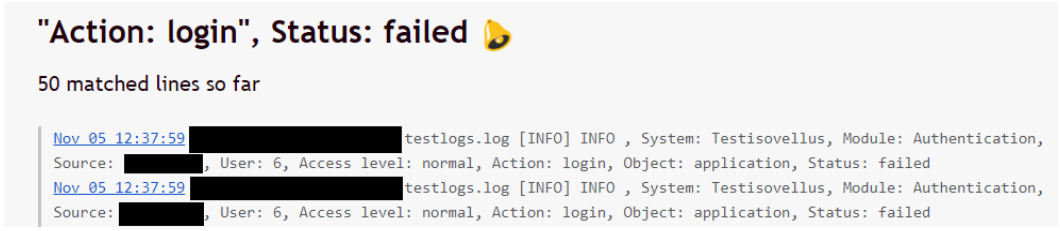
lokimerkinnät kronologisessa järjestyksessä. Lokeja on mahdollista suodattaa lokilähteen, ajankohdan, sovelluksen ja lokimerkinnän tason mukaan, ja suodattaminen on tehty yksinkertaiseksi yläpalkissa sijaitsevien suodattimien avulla. Näkymän alaosasta löytyy myös hakupalkki, jonka avulla voidaan hakea tiettyjä lokimerkintöjä kenttien ja vapaamuotoisten hakusanojen avulla (Kuva 2). Lokeja on myös mahdollista suodattaa ajankohdan mukaan esimerkiksi hakusanoilla ”Last Friday at 15” tai ”today at 10am”.



Kuva 2. LogDNA:n hakutoiminto.

LogDNA:n hakutoiminto toimii hyvin, ja hauissa on mahdollista käyttää Boolen perusoperaattoreita eli AND-, OR- ja NOT-operaattoreita. Esimerkiksi hakukyselyllä ("Access level: root") OR ("Access level: admin") AND (Status: failed) voidaan tarkastella kaikkia epäonnistuneita tapahtumia, joissa on ollut osallisena joko root- tai admin-käyttäjä.

LogDNA:n yksi erityispiirre on niin sanotut näkymät (engl. view), joiden avulla on mahdollista tallentaa tiettyjä suodattimia ja hakukyselyjä. Näkymän tallentaminen tapahtuu hakemalla tai suodattamalla lokidataa, jolloin vasempaan yläkulmaan ilmestyy Unsaved View-painike. Painiketta klikkaamalla on mahdollista luoda uusi näkymä, jonka jälkeen sille annetaan nimi ja vaihtoehtoisesti myös kategoria, jonka avulla on mahdollista ryhmitellä näkymiä (LogDNA 2019c). Samalla on myös mahdollista lisätä hälytyksiä painamalla ”Attach an alert”-painiketta, jonka jälkeen määritellään hälytyksen lähetystapa ja ehdot, joiden tulee täyttyä, jotta hälytys lähetetään. Testausta varten hälytys asetetaan lähtemään sekä sähköpostiin että Slackiin, kun palveluun saapuu minuutin sisällä 50 lokimerkintää epäonnistuneista kirjautumisista (Kuva 3).

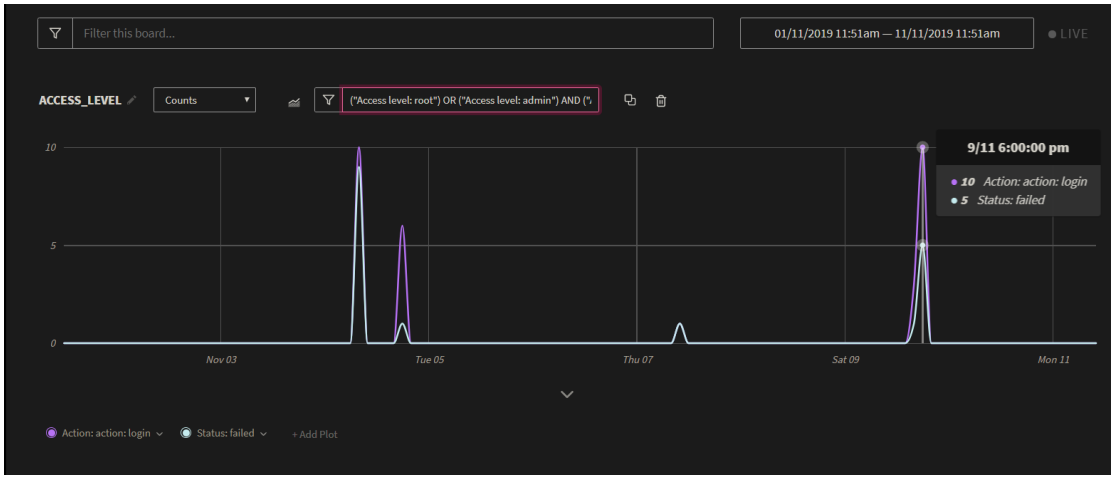


Kuva 3. LogDNA:n hälytys sähköpostin kautta.

LogDNA:han on mahdollista kutsua muita käyttäjiä asetusten Team-välilehden Add Member-painikkeella, jonka jälkeen palvelu pyytää lisäämään kutsuttavan henkilön sähköpostiosoitteen. Team-välilehdellä on myös mahdollista lisätä käyttäjiä eri ryhmiin, ja käyttäjien rooleja voidaan vaihtaa alasvetovalikosta. Itse Owner-roolin lisäksi palvelussa on kolme muuta vaihtoehtoa: Admin, Member ja Read. Admin-tason käyttäjät pystyvät tarkastelemaan kaikkia lokeja, kun taas Member-tason käyttäjillä on pääsy ainoastaan ennalta määrättyihin lokeihin. Read-tason käyttäjät puolestaan voivat ainoastaan tarkastella määrättyjä lokeja ja kuvaajia, suorittaa hakuja sekä exportata lokeja, kun taas muut roolit voivat myös muokata niitä. (LogDNA 2019e.) Read-tason rooli on hyödyllinen esimerkiksi tilanteissa, joissa lokeja halutaan jakaa organisaation ulkopuolisille henkilöille, sillä sen avulla voidaan estää kuvaajien ja näkymien muokkaaminen. Lisäksi sen avulla on mahdollista rajoittaa mitä lokeja käyttäjä voi tarkastella.

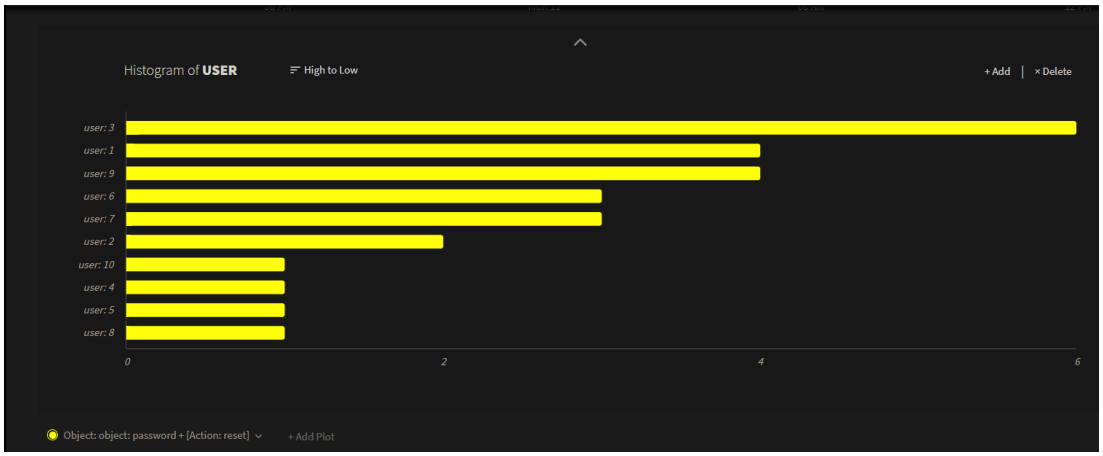
Käyttäjien ja roolien lisäämistä kokeillaan lisäämällä uusi käyttäjä Read-tason käyttöoikeuksilla. Kun uuden käyttäjän rekisteröinti on suoritettu, uudella käyttäjällä yritetään lisätä ja muokata kuvaajia. Kuten oletettua, palvelu ei tarjoa mahdollisuutta suorittaa näitä toimia. Seuraavaksi kokeillaan palvelun Group-ominaisuutta lisäämällä uusi ryhmä, jonka nimeksi annetaan ”Asiakaspalvelu”. Ryhmän jäseneksi määritellään aiemmin luotu uusi käyttäjä. Seuraavaksi määritellään, että ryhmän jäsenillä on pääsy ainoastaan Apache-lokeja tuottavan komentosarjan lokeihin. Ominaisuus toimii halutulla tavalla, eikä kyseisellä käyttäjällä ole pääsyä muiden komentosarjojen luomiin lokeihin.

LogDNA mahdollistaa myös lokidatan visualisoinnin erilaisten kuvaajien avulla, mutta palvelu ei valitettavasti sisällä valmiiksi konfiguroituja kuvaajia. LogDNA:ssa on mahdollista lisätä tauluja, joihin voidaan lisätä useampia kuvaajia. Esimerkkinä luodaan pari erilaista taulua, joista ensimmäinen sisältää admin- ja root-käyttäjien epäonnistuneet sisäänkirjautumiset (Kuva 4). Taulun avulla saadaan nopea yleiskuva kyseisten käyttäjien sisäänkirjautumisyrityksistä.



Kuva 4. LogDNA:n kuvaaja admin- ja root-käyttäjien sisäänkirjautumisista.

Palvelussa on mahdollista luoda histogrammeja ja ympyrädiagrammeja. Esimerkkiä varten luodaan histogrammi käyttäjien salasanojen nollaamisista (Kuva 5). Tämän histogrammin avulla voidaan selvittää helposti, kuinka monta kertaa kukin käyttäjä on yrittänyt nollata salasanansa.



Kuva 5. LogDNA:n histogrammi.

Viimeinen LogDNA:n käsiteltävä ominaisuus on lokien arkistointi. LogDNA tarjoaa mahdollisuuden arkistoida lokitietoja päivittäin muun muassa Amazonin S3-säiliöön. Palvelu tallentaa lokimerkinnät Gzip-tiedostoon JSON-formaatissa, ja salaa kaiken liikkeessä olevan datan. Tämän jälkeen LogDNA pyytää esimerkiksi S3-säiliöön arkistoidessa palvelua käyttämään palvelinpuolen salausta. Arkistointia ei kokeilla käytännössä, sillä palvelun dokumentaatiota lukiessa huomattiin, että palvelu ei mahdollista lokien siirtämistä

arkistosta takaisin palveluun. (LogDNA 2019d.) Tämä ominaisuus olisi ollut toimeksiantajan näkökulmasta huomattavasti mielenkiintoisempi testauskohde kuin itse arkistointi.

Arvio

LogDNA:n yleisilme on selkeä ja sovellus on helppokäyttöinen. Palvelun hakutoiminto on erinomainen, sillä lokeja on mahdollista suodattaa hakusanojen, kenttien ja ajanjakson mukaan. Käyttäjien ja ryhmien hallinta on tehty yksinkertaiseksi, ja LogDNA:ssa on mahdollista määritellä yksityiskohtaisesti, minkälaisia lokeja käyttäjät voivat tarkastella. Tämä ominaisuus on erityisen hyödyllinen, mikäli eri sovellusten ja järjestelmien lokeja halutaan rajata ainoastaan niille käyttäjille, joilla tulee olla niihin pääsy työtehtäviensä puolesta. Palvelun tietoturvallisuus on huomioitu hyvin, sillä palveluun lähetettävä data on salattu sekä liikkeessä että levossa. LogDNA on myös suorittanut Privacy Shield -itsesertifioinnin, joten palvelu noudattaa tiettyjä tietosuojavaatimuksia EU-alueen henkilötietoja käsitellessä.

Palvelu tukee useita eri lokien siirtämismetodeja, kuten lokiagentteja, syslogia, APIa ja koodikirjastoja. Näistä kokeiltiin ainoastaan lokiagenttia ja koodikirjastoa, joista jälkimmäinen ei toiminut halutulla tavalla. Tämä johtuu kuitenkin yksinomaan testausta varten luoduista komentosarjoista, sillä koodikirjastot on tarkoitettu sovellusten lokitukseen. Lokien jäsentely on tehty helpoksi palvelun Custom Log Parser -jäsentimen avulla, ja se ansaitseekin erityismaininnan monipuolisuudestaan. LogDNA sisältää myös joitain integraatioita pilvipalveluntarjoajien kanssa, joka onkin yksi toimeksiantajan asettamista vaatimuksista.

Palvelun visualisointiominaisuudet ovat hienoinen pettymys ja erityyppisiä kuvaajia voisi olla useampia. Valmiiksi konfiguroidut kuvaajat ja kuvaajien räätälöinti olisivat myös hyvä lisä. Palvelu kuitenkin mahdollistaa yksinkertaisten kuvaajien luomisen, joista on varmasti hyötyä toimeksiantajan käyttötapauksessa.

Hälytysten luominen on tehty yksinkertaiseksi, ja palvelun sähköpostin ja Slackin kautta lähetettävät hälytykset toimivat hyvin. Palvelu ei valitettavasti mahdollista arkistoitujen lokien siirtämistä takaisin palveluun, vaan suosittelee käyttämään kolmannen osapuolen työkaluja näiden lokien analysointiin. Kokonaisuudessaan palvelun ominaisuudet ovat hyvät, ja niin käyttöönotto kuin palvelun käyttäminenkin on yksinkertaista. Palvelu täyttää myös toimeksiantajan asettamat kriteerit.

4.4.2 Datadog-palvelu

Datadog on yhdysvaltalainen SaaS-monitorointipalvelu, jonka avulla voidaan monitoroida pilvipohjaisten sovellusten eri komponentteja (Wikipedia 2019c). Palvelu tarjoaa lisäksi muun muassa lokienhallintaan, käyttökokemusten monitorointiin ja sovellusten suorityskyvyn mittaamiseen tarvittavia ominaisuuksia (Datadog 2019c).

Käyttöönotto

Datadog tarjoaa 14 päivän ilmaisen kokeilujakson, jonka aikana kaikkia palvelun ominaisuuksia voi kokeilla (Datadog 2019e). Datadogiin ei suoriteta tässä opinnäytetyössä rekisteröitymistä, sillä toimeksiantaja on ottanut palvelun kokeiluun erään toisen projektin metriikoiden monitorointia varten. Datadog valikoitui soveltuvuusselvitykseen, sillä se vastaa kaikkia toimeksiantajan asettamia vaatimuksia. Hyvänä lisänä pidetään myös palvelun muita ominaisuuksia, kuten infrastruktuurin metriikoiden monitorointia. Tässä soveltuvuusselvityksessä ei kuitenkaan huomioida tai käsitellä muita kuin lokienhallintaan liittyviä ominaisuuksia.

LogDNA:n kokeilusta opittiin, että lokiagentin käyttäminen on suositeltavampaa Python-komentosarjojen luomien lokien kanssa kuin koodikirjaston käyttäminen. Datadogin lokiagentin asentaminen ja määrittely on monimutkaisempi toimenpide kuin LogDNA:n kohdalla, mutta toisaalta sen konfigurointimahdollisuudet ovat monipuolisempia. Lokiagentin asentamisohjeet löytyvät Help-välilehden Quick Start-osiosta. Sieltä valitaan kohta ”Install an integration”, jonka jälkeen palvelu näyttää kaikki asennetut integraatiot ja joitakin suosituksia. Koska vertailussa Mac-tietokone toimii palvelimena, yläpalkista valitaan kohta Agent ja alustaksi valitaan Mac OS X. Lokiagentti on mahdollista asentaa Macille joko komentokehotteen avulla tai DMG-pakettina. Näistä valitaan ensimmäinen vaihtoehto, sillä se on yksinkertaisempi tapa asentaa lokiagentti. Asennukseen käytetty komento jätetään pois opinnäytetyöstä, sillä se on ainoastaan rekisteröityjen käyttäjien saatavilla.

Kun lokiagentti on asennettu, lokien kerääminen tulee kytkeä erikseen päälle. Lokiagentin kansioon conf.d luodaan uusi hakemisto komennolla ”mkdir testapp.d”, jonka jälkeen kyseiseen hakemistoon luodaan conf.yaml -tiedosto nano-tekstieditorin avulla. Seuraavaksi tiedostoon lisätään monitoroitavan lokitiedoston tiedostopolku, tyyppi sekä palvelun ja lähteen nimet (Kuva 6).

Kuva 6. Tiedosto, jolla määritellään, mitä hakemistoa Datadogin agentin tulee monitoroida.

Lopuksi lokien kerääminen kytketään päälle lokiagentin konfigurointitiedostossa, jossa ”logs_enabled” -parametrin arvoksi asetetaan ”true”, ja parametrin edestä poistetaan kommenttimerkki ”#”. Kun tämä toiminto on suoritettu, lokiagentti alkaa keräämään lokeja ja lähettämään niitä palveluun.

Esivalmistelut ja käyttö

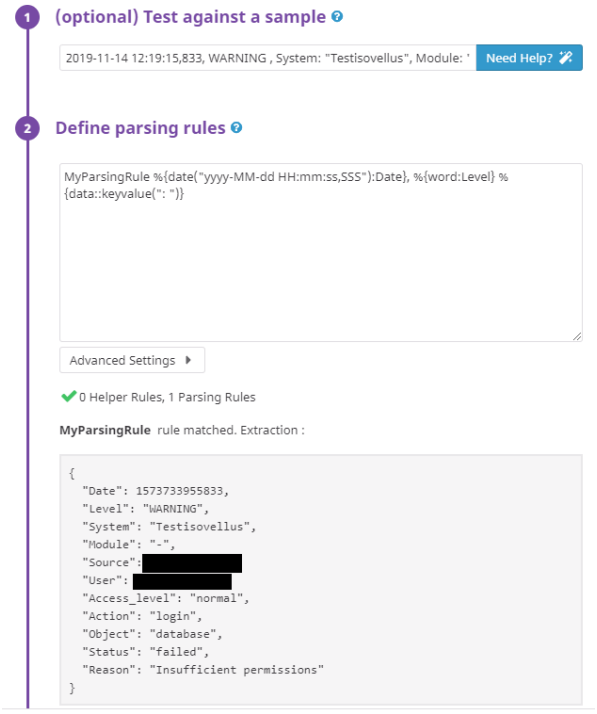
Ensivaikutelma Datadogin käyttöliittymästä on moderni. Koska palvelu tarjoaa lokienhallinnan ohella useita muita toimintoja, käyttöliittymä vaikuttaa aluksi sekavalta. Tässä opinnäytetyössä ei tarkastella muita Datadogin ominaisuuksia, vaan keskitytään ainoastaan lokien kannalta olennaisiin toimintoihin. Kaikki palvelun toiminnot lokienhallinnasta infrastruktuurin monitorointiin on järjestetty käyttöliittymän vasemmassa reunassa sijaitsevaan palkkiin, ja jokainen ominaisuus sisältää useita eri alaotsikoita. Aluksi palvelusta on hankalaa löytää haluamiaan ominaisuuksia niiden suuren määrän vuoksi. Palvelun dokumentaatio on kuitenkin kattava ja selkeä, ja siihen voi turvautua tarvittaessa.

Logs-näkymästä löytyy kaikkien eri lähteiden lähettämät lokit. Koska komentosarjojen luomat lokit ovat räätälöidyssä formaatissa ja sisältävät ylimääräisiä kenttiä, niitä ei voi hyödyntää kuvaajissa tai suodattimissa oletusarvoisesti. Kuten LogDNA:nkin kohdalla, palvelulle tulee määrittää, miten sen tulee jäsennellä komentosarjojen luomia räätälöityjä sovelluslokeja. Koska komentosarjojen luomat lokit eivät ole JSON-formaatissa, ne pitää lähettää niin sanotun pipelinen kautta. Kun uusi pipeline on luotu ja se on määritelty suodattamaan ainoastaan komentosarjojen lisäämiä lokeja, sille lisätään niin sanottuja prosessoreita. Prosessorien tehtävä on muun muassa jäsentää eli parsia lokimerkinnöistä tiettyjä attribuutteja.

Ensimmäiseksi prosessoriksi valitaan Grok Parser -jäsennin, sillä sen avulla on mahdollista jäsennellä tekstimuotoisten lokimerkintöjen eri kenttiä. Jäsentimelle lisätään näyteloki, johon jäsennyssääntöjä sovelletaan. Ensimmäiseksi jäsennyssäännöksi määritellään sääntö, joka muuntaa lokimerkinnän aikaleiman millisekunneiksi.

Koska komentosarjan luomissa lokimerkinnöissä tapahtuman vakavuutta ei esitetä nimi/arvo-parina, kentälle täytyy luoda nimi, jotta palvelu osaa jäsennellä sen oikein. Tämä tapahtuu säännöllä ”%{word:Level}”, joka luo kentälle nimen ”Level” ja käyttää lokimerkintöjen sisältämää vakavuutta sen arvona.

Koska loput lokimerkinnän kentät esiintyvät nimi/arvo-parina, ne voidaan automaattisesti jäsennellä säännöllä ”%{data::keyvalue(”: ”)}”. Mikäli sääntö on määritelty oikein, palvelu näyttää jäsennetyn lokimerkinnän (Kuva 7).



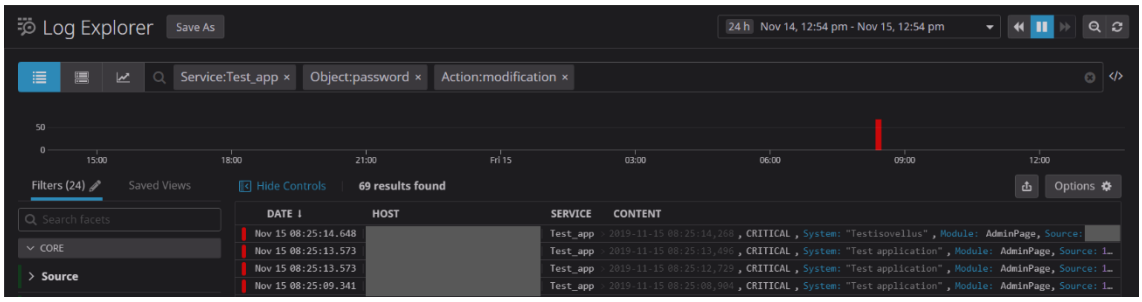
Kuva 7. Datadogin Grok Parser.

Lokeja tarkastellessa huomattiin, että Datadog ei ota huomioon räätälöityjen lokien vakavuuden tasoa, vaan se määrittelee kaikkien merkintöjen tasoksi Info-tason. Tämän vuoksi pipelineen lisätään toinen prosessori, Status Remapper, joka korjaa kyseisen ongelman automaattisesti.

Jotta räätälöityjen lokien attribuutteja voidaan käyttää kuvaajissa ja hauissa, ne pitää vielä lisätä niin sanottuihin näkökohtiin (engl. facets). Näkökohdat näyttävät kaikki lokien attribuutit yhdessä sarakkeessa ja niiden avulla voidaan suodattaa lokimerkintöjä eri attribuuttien avulla. Lisäksi näkökohdat näyttävät lokien ja niiden attribuuttien lukumäärät. (Datadog 2019b.) Attribuuttien lisääminen näkökohtiin tapahtuu Log Explorer -näkymässä, josta valitaan haluttu räätälöity lokimerkintä. Tämän jälkeen valitaan haluttu attribuutti ja suoritetaan toiminto ”Create facet for”. Näkökohdalle voidaan määrittää tämän jälkeen nimi, tyyppi, ryhmä ja kuvaus. Kun kaikki attribuutit ovat lisätty näkökohtiin, palvelun kokeilu voidaan aloittaa.

Lokeja on mahdollista tarkastella Logs-välilehdeltä, jota klikkaamalla aukeaa näkymä, jossa esitetään kaikkien lähteiden lokimerkinnät kronologisessa järjestyksessä. Lokeja on mahdollista suodattaa sivupalkissa olevien suodattimien tai käyttöliittymän yläosassa sijaitsevan hakukentän avulla. Hakukentän avulla lokitapahtumia voidaan suodattaa esimerkiksi tapahtuman objektin mukaan, mutta yksinkertaisemmin tämä onnistuu valitsemalla sivupalkista niin sanotun näkökohdan. Lokeja on myös mahdollista suodattaa ajankohdan mukaan yläpalkista löytyvällä alasvetovalikolla, joka tarjoaa ennalta määriteltyjä ajankohtia, ja mahdollisuuden suodattaa käyttäjän valitseman ajanjakson mukaan.

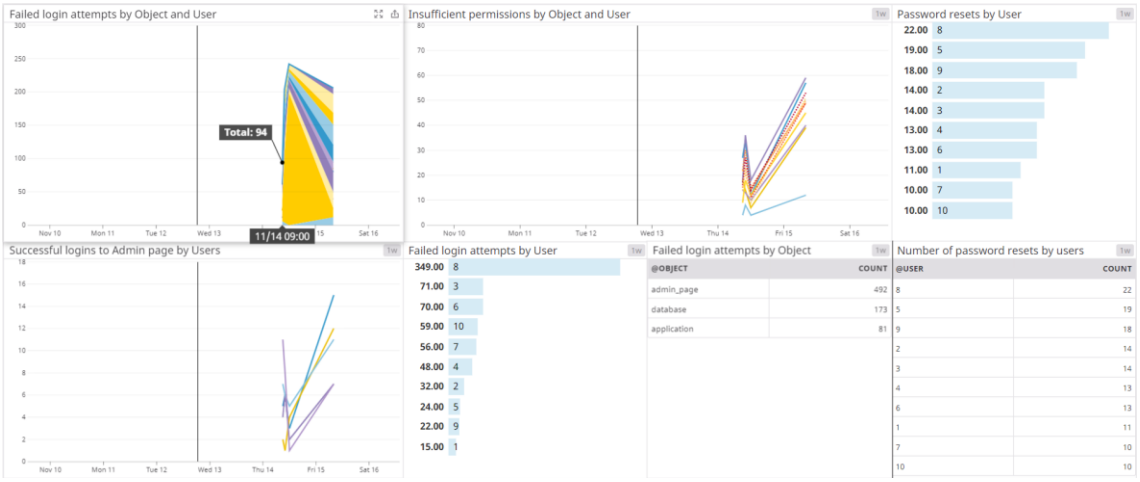
Datadogin hakutoiminto toimii hyvin, ja kuten LogDNA:nkin kohdalla, siinä on mahdollista käyttää Boolen perusoperaattoreita. Aluksi hämmennystä aiheutti se, että räätälöityjä lokikenttiä ei ole mahdollista hakea ilman että hakutermin eteen laittaa @-merkin. Dokumentaatiosta löytyy kuitenkin maininta tästä. Mikäli käyttäjä haluaa esimerkiksi hakea kaikki tapahtumat, joissa salasana on vaihdettu, haku tulee suorittaa seuraavalla hakusanalla: ”@Object:password @Action:modification” (Kuva 8).



Kuva 8. Datadogin hakutoiminto.

Datadogin ominaisuuksissa on panostettu erityisesti lokien visualisointiin. Erilaisia kuvaajia on mahdollista luoda sivupalkista löytyvän Dashboard-painikkeen avulla.

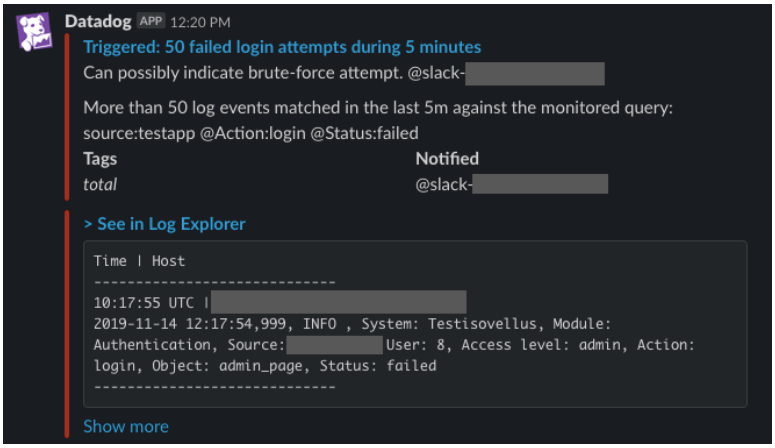
Palvelussa on mahdollista luoda muun muassa timeseries-kuvaajia, tauluja ja kärkiluetteloita. Kuvaajia voidaan myös suodattaa erinäisillä hakusanoilla, ja tuloksia voidaan ryhmitellä eri attribuuttien perusteella.



Kuva 9. Datadogin Screenboard.

Palvelu sisältää niin sanotun Screenboard-ominaisuuden, jonka avulla kuvaajia ja tilastoja voidaan näyttää esimerkiksi TV-ruudulla. Tämän ominaisuuden testausta varten luodaan useita erityyppisiä kuvaajia, joiden avulla saadaan nopea yleiskuva tietoturvan kannalta olennaisista tapahtumista (Kuva 9).

Datadog mahdollistaa myös erilaisten hälytysten luomisen. Testausta varten hälytykselle määritellään hakukysely, joka suodattaa kaikki epäonnistuneet sisäänkirjautumisyritykset. Seuraavaksi määritellään, että hälytys lähetetään, kun Datadog havaitsee 50 tapahtumaa viiden minuutin aikana. Tämän jälkeen palvelu lähettää hälytyksen Slackiin sille tarkoitetulle kanavalle (Kuva 10).



Kuva 10. Datadogin hälytys.

Datadog sisältää roolipohjaisen pääsynhallinnan, johon kuuluu kolme käyttäjäryhmää: pääkäyttäjät, tavalliset käyttäjät ja vain luku -käyttäjät. Pääkäyttäjien ja tavallisten käyttäjien ero on se, että pääkäyttäjät voivat muokata laskutustietoja, hallita muita käyttäjiä ja poistaa API avaimia. Muuten molemmilla käyttäjäryhmillä on lähes samat oikeudet. Vain luku -käyttäjillä sen sijaan on ainoastaan lukuoikeus erinäisiin lokeihin ja näkymiin. (Datadog 2019d.) Käyttövaltuuksien hallitsemista kokeillaan luomalla uusi käyttäjä, jolle annetaan eri käyttövaltuustasoja. Käyttövaltuuksien hallitseminen tapahtuu sivupalkin Team-painikkeesta, josta näkee kaikki organisaatioon lisätyt jäsenet. Käyttövaltuustasoa on mahdollista vaihtaa klikkaamalla jäsenten avatarta, ja samasta painikkeesta on mahdollista poistaa käyttäjätili käytöstä. Valitettavasti palvelusta ei löytynyt tapaa rajoittaa organisaation käyttäjien katseluoikeuksia vain tiettyihin lokeihin, näkymiin tai kuvaajiin. Kuvaajia on kuitenkin mahdollista jakaa julkisten linkkien avulla esimerkiksi organisaation ulkopuolisille jäsenille.

Datadogin yksi erityispiirteistä on ”Log Rehydration”, joka käytännössä tarkoittaa sitä, että arkistoidut lokit on mahdollista palauttaa järjestelmään myöhempää analysointia varten. Tämä ominaisuus on hyödyllinen erityisesti tilanteissa, joissa organisaatio on havainnut tietoturvapoikkeaman, ja sen täytyy analysoida esimerkiksi kuukausien takaisia lokeja. (Mooney 2019.) Valitettavasti tämä ominaisuus on kirjoitushetkellä julkisessa beetatestauksessa ja avoinna vain Datadogin US-sivustoa käyttäville, joten ominaisuutta ei päästä testaamaan käytännössä.

Viimeinen kokeiltava ominaisuus on sensitiivisen datan piilottaminen lokeista. Kokeilua varten luodaan sääntö, joka piilottaa käyttäjien IP-osoitteet. Tämä tapahtuu samassa tiedostossa, jossa määritellään, mistä hakemistosta lokiagentin tulee kerätä lokeja. Tiedostoon määritellään säännöllinen lauseke (engl. regular expression, regex), joka tarkistaa lokimerkinnän IP-osoitteiden varalta, ja korvaa ne arvolla ”[masked_ip_address]”. Toiminto on vaikeakäyttöinen, mikäli säännölliset lausekkeet eivät ole ennestään tuttuja. Lisäksi toimeksiantajan käyttötapauksessa kyseiselle ominaisuudelle ei ole todennäköisesti tarvetta, mutta tästä huolimatta ominaisuus on hyvä lisä.

Arvio

Datadog on ominaisuuksiltaan kattava, ja se sisältää lokienhallinnan ohella useita muita ominaisuuksia, joita ei tarkasteltu tässä soveltuvuusselvityksessä. Osittain

ominaisuuksien suuresta määrästä johtuen, palvelu vaikuttaa aluksi monimutkaiselta, ja palvelusta saattaa olla aluksi hankalaa löytää haluamiaan toimintoja. Palvelun dokumentaatio on kuitenkin kattava, ja sieltä löytyy yksityiskohtaiset ohjeet suurimpaan osaan ominaisuuksista.

Palvelun Grok Parser -jäsennintä on yksinkertaista käyttää ja sen avulla voidaan helposti jäsennellä tekstipohjaisten lokien nimi/arvo-parit. Palvelun suodattimet ja hakutoiminto toimivat hyvin, ja niiden avulla voidaan suodattaa lokimerkintöjä yksityiskohtaisesti. Datadogin hälytysominaisuudet toimivat hyvin, mutta eri hälytysvaihtoehtoja voisi olla useampia. Lisäksi palvelussa ei kirjoitushetkellä ollut mahdollista kokeilla lokien siirtämistä arkistoista takaisin palveluun, sillä ominaisuus oli varattu US-sivuston käyttäjille. Ominaisuus on kuitenkin kirjoitushetkellä julkisessa beetatestauksessa, joten se oletettavasti julkaistaan kaikkien käyttäjien saataville lähitulevaisuudessa.

Datadogin visualisointiominaisuudet ovat erinomaiset, ja lokeista on mahdollista luoda useita erityyppisiä kuvaajia. Visualisoinnissa on kuitenkin huomioitava, että lokeja ei voi käyttää kaikissa erityyppisissä kuvaajissa, vaan osa kuvaajista on varattu vain metriikoiden visualisoinnille. Datadogilla on myös valtava määrä erilaisia integraatioita eri palveluiden kanssa, joita ei tässä soveltuvuusselvityksessä päästy kokeilemaan.

Palvelussa on selkeästi otettu huomioon tietoturvallisuus. Liikkeessä oleva data salataan TLS- ja HSTS-protokollilla (HTTP Strict Transport Security), kun taas levossa oleva data suojataan AES 256 -salauksella (Datadog 2019f). Koska lokitiedot ovat salattu sekä liikkeessä että levossa, luvattomat henkilöt eivät voi tarkastella tai peukaloida lokeja. Palvelu tarjoaa myös mahdollisuuden käyttää EU-alueen datakeskusta tietojen säilyttämiseen, joka on GDPR:n suhteen hyvä asia. Valitettavasti palvelun roolipohjaisesta pääsynhallinnasta ei löytynyt mahdollisuutta rajoittaa lokeja tai näkymiä vain tietyille käyttäjille. Kokonaisuudessaan Datadogin ominaisuudet ovat hyvät, ja palvelu täyttää toimeksiantajan asettamat kriteerit. Hyvänä lisänä pidetään toimintoa, jonka avulla sensitiivistä dataa sisältävät kentät voidaan piilottaa, vaikkakin kyseinen ominaisuus voisi olla helppokäyttöisempi.

4.5 Palveluiden soveltuvuus

Jo esiselvityksessä tiedostettiin, että kumpikaan testatuista palveluista ei ole ensisijaisesti suunnattu tietoturvatapausten monitorointiin, eikä kumpikaan palveluista sisältänyt

testausvaiheessa tietoturva-analytiikkaa. Tästä huolimatta kummassakin palvelussa on tarvittavat ominaisuudet tietoturvatapausten havaitsemiseen ja selvittämiseen, ja näin ollen kumpikin palvelu soveltuu toimeksiantajan käyttötapaukseen.

Palveluiden ominaisuuksissa on jonkin verran eroja, kuten myös ominaisuuksien toteutustavoissa. Kumpikin palvelu kuitenkin tarjoaa samat lokienhallinnan perusominaisuudet, joihin kuuluvat lokien kerääminen, hakeminen, suodattaminen, jäsentäminen ja arkistointi. Kumpikin palvelu mahdollistaa lokien visualisoinnin ja hälytysten luomisen.

Testatuista palveluista Datadogilla on huomattavasti enemmän erilaisia toimintoja. Tämä johtuu pitkälti siitä syystä, että Datadog tarjoaa lokienhallinnan ohella muitakin toimintoja. Lisäksi Datadog tukee useampia integraatioita kuin LogDNA. Siinä missä Datadog loistaa ominaisuuksien ja integraatioiden määrällä, LogDNA loistaa helppokäyttöisyydellään sekä selkeydellään.

Kummankin palvelun hakutoiminto ja suodattimet ovat melko samankaltaisia, kuten myös hälytystoiminto. Kummassakin palvelussa on mahdollista hakea ja suodattaa lokimerkintöjä eri kenttien ja hakusanojen avulla, joten poikkeustilanteissa on helppoa selvittää esimerkiksi tietyn käyttäjän suorittamat toimet. Erilaisten näkymien ja kuvaajien avulla on mahdollista saada nopea yleiskuva esimerkiksi Vuosikello-sovelluksen käyttäjien suorittamista salasanojen vaihdoista tai muista tietoturvan kannalta olennaisista tapahtumista.

Palveluiden hälytysominaisuudet ovat myös hyvin samankaltaisia, ja hälytyksiä on mahdollista lähettää usealle eri henkilölle useita eri kanavia pitkin. Esimerkiksi poikkeustilanteessa kummankin palvelun kautta on mahdollista lähettää hälytys tietoturvasta vastaaville henkilöille esimerkiksi Slackin ja sähköpostin välityksellä. Hälytykset ovatkin olennainen osa lokienhallintajärjestelmää, sillä niiden avulla voidaan havaita lähes reaaliajassa esimerkiksi toistuvat epäonnistuneet sisäänkirjautumisyritykset tai mahdolliset ongelmat infrastruktuurin lokituksessa.

Kummankin palvelun tietoturvaominaisuudet ovat myös kunnossa, sillä kumpikin palvelu salaa kaikki lokitiedot sekä liikkeessä että levossa. Kumpikaan palveluista ei myöskään mahdollista lokimerkintöjen muokkaamista tai poistamista, joten käyttäjät eivät voi peukaloida lokeja. Datadog mahdollistaa myös arkaluontoisen datan piilottamisen lokimerkinnöistä ennen kuin niitä lähetetään palveluun.

Sekä LogDNA että Datadog tukevat räätälöityjä lokimerkintöjä, ja palveluissa on mahdollista määritellä, miten lokeja tulee jäsennellä. Tämä on erityisen hyödyllinen ominaisuus etenkin sovelluslokien kohdalla, sillä ne ovat usein eri formaateissa ja sisältävät räätälöityjä kenttiä. Lisäksi tietokantojen, kuormantasaajien ja muiden komponenttien lokit saattavat olla eri muodoissa, jolloin lokeja voidaan joutua jäsentelemään eri säännöillä. Kumpikin palvelu mahdollistaa myös lokien arkistoinnin, mutta kirjoitushetkellä kummassakaan palvelussa ei ollut mahdollista kokeilla arkistoitujen lokien siirtämistä takaisin palveluun. Arkistointi on erityisen hyödyllinen ominaisuus, sillä sen avulla on mahdollista säilyttää lokeja edullisesti pitkiäkin aikoja. Arkistoituja lokeja voidaan tarkastella myöhemminkin, mikäli esimerkiksi tietoturvapoikkeamia havaitaan.

Soveltuvuusselvityksen alkuperäinen tarkoitus ei ollut suositella erityisesti kumpaakaan palvelua, vaan ennemminkin tuoda esille niiden ominaisuuksia, käytettävyyttä ja mahdollisia puutteita, sekä arvioida palveluiden soveltuvuutta toimeksiantajan käyttötapaukseen. Alun perin soveltuvuusselvityksen johtopäätös oli, että toimeksiantajan tulee huomioida mahdollisessa valinnassaan muun muassa kummankin palvelun ominaisuudet tai niiden puute, sekä integraatioiden määrä ja tarpeellisuus. Kumpikin palvelu sisältää lähes kaikki olennaiset toiminnot ja integraatiot, mutta koska palveluita kyettiin ainoastaan testaamaan komentosarjojen luomien lokien avulla, integraatioiden määrä ja etenkin tarpeellisuus korostuu.

Aivan opinnäytetyöprosessin loppuvaiheilla Datadog julkisti tietoturvatapahtumien monitorointiin liittyviä ominaisuuksia, kuten reaaliaikaisen uhkien havaitsemisen ja tietoturvapoikkeamien havaitsemissäännöt (Tremsal & Brown 2019). Valitettavasti tässä vaiheessa opinnäytetyöprosessia, uusia ominaisuuksia ei ollut enää mahdollista testata, sillä kyseiset ominaisuudet olivat kirjoitushetkellä yksityisessä beetavaiheessa ja avoinna ainoastaan US-sivuston käyttäjille. Koska kyseiset ominaisuudet tulevat todennäköisesti kaikkien käyttäjien saataville lähitulevaisuudessa, Datadog on suositeltavampi vaihtoehto toimeksiantajan käyttötapaukseen, sillä kyseisten ominaisuuksien avulla tietoturvatapahtumia on mahdollista havaita ja selvittää huomattavasti tehokkaammin.

5 POHDINTA

Opinnäytetyön tavoitteena oli tutkia, minkälaisista tapahtumista toimeksiantajan SaaS-verkkopalvelussa tulee kerätä lokeja, jotta tietoturvatapahtumia voidaan havaita ja selvittää. Lisäksi tavoitteena oli kokeilla kahden pilvipohjaisen lokienhallintapalvelun eri ominaisuuksia ja selvittää, olisiko jompikumpi näistä sopiva toimeksiantajan käyttötapaukseen. Opinnäytetyön tuloksena syntyi ohjeistus tietoturvan kannalta olennaisista lokeista sekä soveltuvuusselvitys kahdesta lokienhallintapalvelusta.

Opinnäytetyössä käytettiin konstruktiivista tutkimusotetta. Opinnäytetyön tuloksena syntyneen ohjeistuksen lähteinä käytettiin muun muassa alan kirjallisuutta sekä erinäisten asiantuntijaorganisaatioiden, kuten OWASPin, NISTin ja VAHTIn, ohjeistuksia. Osana soveltuvuusselvityksen esiselvitystä luotiin taulukko, joka sisälsi tietoa viidestä potentiaalisesta lokienhallintapalvelusta ja niiden ominaisuuksista. Toimeksiantaja valitsi esiselvityksen pohjalta kaksi palvelua, joiden ominaisuuksia ja käytettävyyttä arvioitiin. Testatuista palveluista kumpikin soveltuu toimeksiantajan käyttötapaukseen, sillä kumpikin palvelu sisältää tarvittavat ominaisuudet. Tästä huolimatta, Datadog on soveltuvampi toimeksiantajan käyttötapaukseen, sillä se sisältää enemmän integraatioita, ja lisäksi opinnäytetyöprosessin loppuvaiheilla palvelu julkisti useita tietoturvallisuuden valvontaan liittyviä toimintoja.

Opinnäytetyöprosessin aikana aihe muuttui ja tarkentui muutamaan otteeseen. Alkuperäinen suunnitelma oli luoda yksityiskohtaiset ohjeet Vuosikello-verkkopalvelusta kerättävistä lokeista, mutta tiukan aikataulun ja muiden syiden vuoksi päädyttiin luomaan mahdollisimman yleisellä tasolla kirjoitettu ohjeistus, jotta siitä hyötyy toimeksiantajan lisäksi muutkin henkilöt ja organisaatiot. Soveltavaa osuutta jouduttiin myös rajaamaan pari kertaa. Alkuperäinen suunnitelma oli käyttää oikeiden palvelimien ja sovellusten lokitietoja, mutta aikataulullisten ongelmien vuoksi tästä ajatuksesta luovuttiin, ja lokeja päädyttiin luomaan kolmen eri komentosarjan avulla. Näistä ensimmäinen loi sovelluslokeja, kun taas toinen loi Apache-verkkopalvelimen käyttölokeja. Viimeisen komentosarjan luomat lokit sen sijaan pyrkivät simuloimaan väsytyshyökkäystä.

Opinnäytetyössä palvelimena toimi Mac-tietokone, joten tuotantoympäristössä palveluiden käyttöönotto ja konfigurointi eroaa hieman tässä opinnäytetyössä esitetystä prosessista. Koska palveluita ei päästy testaamaan oikeassa tuotantoympäristössä, palveluiden soveltuvuus -luvussa on huomautettu siitä, että toimeksiantajan on mahdollista

valintaa tehdessä huomioitava kummankin palvelun eri integraatiot ja niiden tarpeellisuus. Lisäksi on huomioitava, että esimerkiksi Apache-verkkopalvelimien lokien kerääminen on oikeassa ympäristössä yksinkertaisempaa, sillä kumpikin testatuista palveluista osaa tulkita niiden luomia lokeja automaattisesti.

Opinnäytetyön aihe oli haastava siinä mielessä, että aihe oli minulle ennestään melko tuntematon. Lokienhallinta on aiheena todella laaja ja opinnäytetyön rajaaminen oleellisiin asioihin oli paikoitellen haastavaa. Myös verkkosovelluksen ja sen infrastruktuurin tietoturvallisuuden kannalta olennaisista lokeista ja tapahtumista oli aluksi hankalaa löytää kattavaa ja luotettavaa tietoa. Sopivia lokienhallintapalveluita ei meinannut aluksi löytyä, sillä lähes kaikissa niissä oli joitain ratkaisevia puutteita. Mielestäni opinnäytetyön tavoitteeseen kuitenkin päästiin, ja koen, että opin työskentelyni aikana paljon lokienhallinnasta ja sen merkityksestä tietoturvallisuuden hallinnassa.

Opinnäytetyön tulokset luovat hyvän pohjan lokienhallinnan suunnittelulle. Toimeksiantaja voi hyödyntää opinnäytetyön tuloksia Vuosikello-verkkopalvelun lokien sisällön suunnittelussa, kun taas soveltuvuusselvityksen tuloksia voidaan mahdollisesti käyttää palvelun valinnan perusteena. Jatkokehitysehdotuksena on Vuosikello-verkkopalvelun infrastruktuurin metriikoiden monitorointi, sillä sitä ei ehditty aikataulujen puitteissa kokeilemaan. Opinnäytetyössä ei käsitellä lainkaan SIEM-palveluita, joten jatkokehitysehdotuksena on ottaa sellainen tulevaisuudessa käyttöön lokienhallintapalvelun rinnalle. Myöskään Datadogin tietoturvallisuuden valvontaan liittyviä ominaisuuksia ei päästy opinnäytetyöprosessin aikana testaamaan, joten jatkokehitysehdotuksena on kokeilla kyseisiä ominaisuuksia.

LÄHTEET

Apache HTTP Server Project 2019. Viitattu 12.11.2019 https://httpd.apache.org/.

AWS 2019. Amazon S3. Object storage built to store and retrieve any amount of data from anywhere. Viitattu 27.11.2019 https://aws.amazon.com/s3/.

Bateman, R. 2018. GDPR and Log Data. TermsFeed. Viitattu 10.10.2019 https://www.termsfeed.com/blog/gdpr-log-data/.

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.

Datadog 2019a. Archives on AWS S3. Datadog Docs. Viitattu 5.10.2019 https://docs.datadoghq.com/logs/archives/s3/?tab=ussite.

Datadog 2019b. Log Explorer. Datadog Docs. Viitattu 15.11.2019 https://docs.datadoghq.com/logs/explorer/?tab=logsearch.

Datadog 2019c. See it all in one place. Viitattu 23.11.2019 https://www.datadoghq.com/product/.

Datadog 2019d. Team. Datadog Docs. Viitattu 24.11.2019 https://docs.datadoghq.com/account_management/team/.

Datadog 2019e. Sign up for your free Datadog Trial. Monitor and analyze metrics, traces, and logs in minutes. Viitattu 26.11.2019 https://www.datadoghq.com/free-datadog-trial/.

Datadog 2019f. Security. Viitattu 26.11.2019 https://www.datadoghq.com/security/.

Doyle, S. TCP vs. UDP: Understanding the Difference. Privacy News Online. Private Internet Access. Viitattu 4.12.2019 https://www.privateinternetaccess.com/blog/2018/12/tcp-vs-udp-understanding-the-difference/.

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

Finder 2019. Perjantai Markkinointiviestintä Oy. Viitattu 8.11.2019 https://www.finder.fi/Mainostoimisto/Perjantai+Markkinointiviestint%C3%A4+Oy/Turku/yhteystiedot/532763.

Fisher, T. 2019. What Is a Hostname? Definition of hostname and how to find it in Windows. Lifewire. Viitattu 12.11.2019 https://www.lifewire.com/what-is-a-hostname-2625906.

Gebel, G 2018. Why you need both authorization and authentication. CSO. Viitattu 20.10.2019 https://www.csoonline.com/article/3269302/why-you-need-both-authorization-and-authentication.html.

Graylog 2018. The data explosion and its effect on security. Viitattu 15.10.2019 https://www.graylog.org/post/the-data-explosion-and-its-effect-on-security.

Graylog 2019. Introduction to correlation engine. Viitattu 9.10.2019 https://www.graylog.org/features/correlation-engine.

IBM 2019a. Brute force attacks. IBM Knowledge Center. Viitattu 12.11.2019 https://www.ibm.com/support/knowledgecenter/en/SSB2MG_4.6.0/com.ibm.ips.doc/concepts/wap_brute_force.htm.

IBM 2019b. IBM Study Shows Data Breach Costs on the Rise; Financial Impact Felt for Years. Breaches Pose Growing Risk for Small Businesses, Costing up to 5% of Annual Revenue. Viitattu 17.10.2019 https://newsroom.ibm.com/2019-07-23-IBM-Study-Shows-Data-Breach-Costs-on-the-Rise-Financial-Impact-Felt-for-Years.

Infosec 2018. CIA Triad. General Security. Viitattu 3.10.2019 https://resources.infosecinstitute.com/cia-triad/.

Jensen, P. 2018. Mikä on IP-osoite? Kotimikro. Viitattu 12.11.2019 https://kotimikro.fi/internet/verkko/mika-on-ip-osoite.

Kent, K. & Souppaya, M. 2006. Guide to Computer Security Log Management. Suositus. Gaithersburg: National Institute of Standards and Technology. Viitattu 4.10.2019 https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf.

Kyberturvallisuuskeskus 2016. Palvelunestohyökkäykset ovat internetin arkipäivää. Viitattu 12.11.2019 https://legacy.viestintavirasto.fi/kyberturvallisuus/tietoturvanyt/2016/04/ttn201604291231.html.

Kyberturvallisuuskeskus 2019. Näin keräät ja käytät lokitietoja. Viitattu 4.10.2019 https://kyberturvallisuuskeskus.fi/fi/nain-keraat-ja-kaytat-lokitietoja.

LogDNA 2017. Choosing The Right Ingestion Client. Viitattu 3.10.2019 https://logdna.com/blog/choosing-the-right-ingestion-client/.

LogDNA 2018. Apache and Nginx Logging – Get the Most Out of Web Server Logs. Viitattu 14.10.2019 https://logdna.com/apache-nginx-logs/.

LogDNA 2019a. What is Log Management? The Complete Logging Guide. Viitattu 3.10.2019 https://logdna.com/what-is-log-management/.

LogDNA 2019b. About LogDNA. Viitattu 1.11.2019 https://docs.logdna.com/docs/getting-started.

LogDNA 2019c. Create Views & Alerts. Viitattu 5.11.2019 https://docs.logdna.com/docs/views.

LogDNA 2019d. Archiving Log Files. Viitattu 12.11.2019 https://docs.logdna.com/docs/archiving.

LogDNA 2019e. RBAC (Role Based Access Control). Viitattu 24.11.2019 https://docs.logdna.com/docs/rbac.

LogDNA 2019f. Pricing. Viitattu 26.11.2019 https://logdna.com/pricing/

Loggly 2019. Archiving Logs to Amazon S3. Support. Viitattu 5.10.2019 https://www.loggly.com/docs/amazon-s3-archive/.

LogSentinel 2018. GDPR Logging Requirements. Viitattu 10.10.2019 https://logsentinel.com/gdpr-logging-requirements/.

Lokiohje 2009. Valtiovarainministeriö. Helsinki: Edita Prima Oy. Viitattu 4.10.2019 https://www.vahtiohje.fi/c/document_library/get_file?uuid=775179cb-6c54-4dfb-b65d-e925d47c61d2&groupId=10229.

Mell, P. & Grance, T. 2011. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology. Viitattu 4.10.2019 https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

Mooney, M. 2019. Introducing Metric from Logs and Log Rehydration™. Datadog. Viitattu 18.11.2019 https://www.datadoghq.com/blog/logging-without-limits-new-features/.

Mozilla 2019. User-Agent. MDN web docs. Viitattu 15.11.2019 https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers/User-Agent.

MySQL 2019. 13.7.1.6 GRANT Statement. MySQL 8.0 Reference Manual. Viitattu 26.11.2019 https://dev.mysql.com/doc/refman/8.0/en/grant.html.

National Institute of Standards and Technology 2017. About NIST. Viitattu 8.10.2019 https://www.nist.gov/about-nist.

Onelogin 2019a. What is Multi-Factor Authentication (MFA)? How MFA prevents attacks from cybercriminals. Viitattu 12.11.2019 https://www.onelogin.com/learn/what-is-mfa.

Onelogin 2019b. How does single sign-on work? How single sign-on works, step by step. Viitattu 12.11.2019 https://www.onelogin.com/learn/how-single-sign-on-works.

OWASP 2013. Data Validation. Viitattu 27.11.2019 https://www.owasp.org/index.php/Data_Validation#Sanitize.

OWASP 2017. OWASP Top 10 – 2017 The Ten Most Critical Web Application Security Risks. Viitattu 17.10.2019 https://www.owasp.org/images/7/72/OWASP_Top_10-2017_%28en%29.pdf.pdf.

OWASP 2019a. OWASP™ Foundation. Viitattu 12.11.2019 https://www.owasp.org/index.php/Main_Page.

OWASP 2019b. Logging Cheat Sheet. The Cheat Sheet Series. Viitattu 17.10.2019 https://cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html.

OWASP 2019c. Log Injection. Viitattu 13.11.2019 https://www.owasp.org/index.php/Log_Injection.

Papertrail 2019a. Automatic S3 archive export. Knowledge Base. Viitattu 4.10.2019 https://help.papertrailapp.com/kb/how-it-works/automatic-s3-archive-export/.

Papertrail 2019b. LaaS – Logging as a Service. Viitattu 28.10.2019 https://www.papertrail.com/solution/laas/.

Petters, J. 2019. What is SIEM? A Beginner's Guide. Varonis. Viitattu 20.11.2019 https://www.varonis.com/blog/what-is-siem/.

Privacy Shield Framework 2019. Privacy Shield Program Overview. Viitattu 20.11.2019 https://www.privacyshield.gov/Program-Overview.

Rouse, M. 2019a. Syslog. WhatIs.com. Viitattu 4.10.2019 https://whatis.techtarget.com/definition/syslog.

Rouse, M. 2019b. Software as a Service. SearchCloudComputing. Viitattu 7.10.2019 https://searchcloudcomputing.techtarget.com/definition/Software-as-a-Service.

SecurityMetrics 2014. 5 Minimum Necessary HIPAA PHI Tips. Viitattu 12.11.2019 https://www.securitymetrics.com/blog/5-minimum-necessary-hipaa-phi-tips.

Sematext 2019. Logagent. Viitattu 4.10.2019 https://sematext.com/logagent/.

SFS-EN ISO/IEC 27002 2017. Information technology. Security techniques. Code of practice for information security controls (ISO/IEC 27002:2013 including Cor 1:2014 and Cor 2:2015). Helsinki: Suomen Standardisoimisliitto SFS ry.

Study.com 2019. What is Data Tampering? – Definition & Prevention. Viitattu 17.11.2019 https://study.com/academy/lesson/what-is-data-tampering-definition-prevention.html.

Techopedia 2019a. Regular Expression. Viitattu 25.11.2019 https://www.techopedia.com/definition/25843/regular-expression.

Techopedia 2019b. Uniform Resource Locator (URL). Viitattu 12.11.2019 https://www.techopedia.com/definition/1352/uniform-resource-locator-url.

Techopedia 2019c. Authentication. Viitattu 20.10.2019 https://www.techopedia.com/definition/342/authentication.

Tietosuojatyöryhmä 2018. Asetuksen 2016/679 mukaista läpinäkyvyyttä koskevat suuntaviivat. Viitattu 10.10.2019 https://tietosuoja.fi/documents/6927448/8316711/L%C3%A4pin%C3%A4kyvyys+fi/c102605b-e386-4661-9b51-bf427875c8db/L%C3%A4pin%C3%A4kyvyys+fi.pdf.

Tietosuojavaltuutetun toimisto 2019a. Mikä on henkilötieto? Viitattu 10.10.2019 https://tietosuoja.fi/mika-on-henkilotieto.

Tietosuojavaltuutetun toimisto 2019b. Usein kysyttyä EU:n tietosuoja-asetuksesta. Viitattu 10.10.2019 https://tietosuoja.fi/gdpr.

Tietosuojavaltuutetun toimisto 2019c. Pseudonymisoidut ja anonymisoidut tiedot. Viitattu 12.11.2019 https://tietosuoja.fi/pseudonymisointi-anonymisointi.

Tietosuojavaltuutetun toimisto 2019d. Rekisterinpitäjän oikeutettu etu. Viitattu 10.10.2019 https://tietosuoja.fi/rekisterinpitajan-oikeutettu-etu.

Tremsal, M. & Brown, M. 2019. Announcing Datadog Security Monitoring. Datadog. Viitattu 25.11.2019 https://www.datadoghq.com/blog/announcing-security-monitoring/.

Vuosikello 2019. Työkalu suunnitelmien hallintaan. Vuosisuunnittelu helpommin, nopeammin ja tehokkaammin. Viitattu 24.10.2019 https://www.vuosikello.net/.

W3Schools 2019. What is http? Viitattu 12.11.2019 https://www.w3schools.com/whatis/whatis_http.asp.

Wikipedia 2019a. syslog. Viitattu 17.10.2019 https://en.wikipedia.org/wiki/Syslog.

Wikipedia 2019b. MAC address. Viitattu 12.11.2019 https://en.wikipedia.org/wiki/MAC_address.

Wikipedia 2019c. Datadog. Viitattu 23.11.2019 https://en.wikipedia.org/wiki/Datadog.

Zhang, E. 2018. What is Log Analysis? Use Cases, Best Practices, and More. Data Protection 101. Data Insider. Digital Guardian. Viitattu 2.10.2019 https://digitalguardian.com/blog/what-log-analysis-use-cases-best-practices-and-more.

Liite 1. Lokienhallintapalveluiden esiselvitys

	Datadog	**LogDNA**	**PaperTrail**	**Scalyr**	**Sematext**
Hinnoittelumalli	Lokirivien määrän ja säilytysajan mukaan.	Säilytysajan ja käyttäjien määrän mukaan.	Lokien määrän mukaan. Sisältää kahden viikon säilytysajan ja vuoden arkistoinnin.	Säilytysajan ja sen mukaan, kuinka paljon lokeja lähetetään keskimäärin päivässä.	Ominaisuuksien, tallennustilan ja säilytysajan mukaan.
Palvelimien sijainti	EU tai US.	US.	US.	EU tai US.	EU tai US.
GDPR ja Privacy Shield	Privacy Shield -sertifioitu ja noudattaa GDPR:ää.	Privacy Shield -sertifioitu ja noudattaa GDPR:ää.	Privacy Shield -sertifioitu ja noudattaa GDPR:ää.	Noudattaa GDPR:ää.	Noudattaa GDPR:ää.
Salaus liikkeessä/levossa	Kyllä.	Kyllä.	Mahdollista salata liikkeessä. Levossa oleva data ilmeisesti salattu.	Liikkeessä ja ilmeisesti myös levossa.	Ilmeisesti vain liikkeessä.
MFA/SSO	SSO.	SSO.	Ei.	Ei.	Kyllä.
RBAC	Administrators, standard users ja read only users.	Owner, admin, member ja read.	Mm. full access, read-only ja specific group access.	Full, readLog, limited.	Owner, admin, billing admin ja user.
Siirtämismetodit	Agentti, koodikirjastot, integraatiot, jne.	Agentti, alustat, syslog, API ja koodikirjastot.	Mm. remote_syslog2, Python SysLogHandler, rsyslog, jne.	Agentti, integraatiot, syslog, jne.	Logstash, Filebeat, rsyslog, Logagent, koodikirjastot, jne.
Visualisointi	Kyllä.	Kyllä.	Ei.	Kyllä.	Kyllä.
Hälytykset	Slack, PagerDuty ja Webhooks.	Mm. email, Slack, Webhook ja PagerDuty.	Mm. email, Slack, PagerDuty ja Campfire.	Mm. email, Slack, PagerDuty ja OpsGenie.	Mm. email, PagerDuty ja Slack.
Arkistointi	S3-säiliöön ja Google Cloud Storageen.	Mm. S3-säiliöön, Azure Blob Storageen ja Google Cloud Storageen. Arkistointi päivittäin.	Palvelu arkistoi lokit palvelun omaan S3-säiliöön. Lisäksi mahdollista määrittää oma S3-säiliö.	S3-säiliöön tunneittain.	S3-säiliöön.
Sensitiivisen datan piilottaminen	Kyllä.	Ei.	Ei.	Kyllä, esim. säännöllisillä lausekkeilla. Mahdollista myös hajauttaa MD5-algoritmillä.	Kyllä: hash-suodatin, AES-salaus, kenttien poistaminen ja mahdollisuus muuttaa IP-osoitteen viimeinen oktetti nollaksi.