jamk.fi

Kolmannen osapuolen uhkaindikaattorit SIEM-järjestelmässä

Jani Järvinen

Opinnäytetyö
Lokakuu 2020
Tekniikan ala
Insinööri (AMK), tieto- ja viestintätekniikka

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

jamk.fi

Kuvailulehti

Tekijä(t)	Julkaisun laji	Päivämäärä
Järvinen, Jani	Opinnäytetyö, AMK	Lokakuu 2020
	Sivumäärä 53	Julkaisun kieli Suomi
		Verkkojulkaisulupa myönnetty: x

Työn nimi
Kolmannen osapuolen uhkaindikaattorit SIEM-järjestelmässä

Tutkinto-ohjelma
Tieto- ja viestintätekniikka

Työn ohjaaja(t)
Rantonen, Mika

Toimeksiantaja(t)
Qvantel Finland Oy

Tiivistelmä

SIEM-järjestelmän toiminnallisuuden päivittämistä ulkopuolisten uhkaindikaattoreiden avulla käsittelevässä opinnäytetyössä perehdyttiin yleisesti SIEM-järjestelmiin ja uhkaindikaattoreihin sekä pyrittiin parantamaan olemassa olevia ratkaisuja. SIEM-järjestelmien toiminnan ymmärtämiseksi oli tärkeää perehtyä myös lokien luontiin ja hallintaan. Lokienhallinnassa kohdattiin lakiasetuksia ja säädöksiä, jotka vaikuttavat mm. säilytysaikoihin. Näiden lisäksi pintaa raapaistiin tietoturvan osalta, mutta vähänlaisesti koska se ei ollut tärkeää tavoitteiden kannalta.

Tavoitetta lähdettiin tutkimaan etsimällä valmiita, avoimen lähdekoodin ratkaisuja, joita muokattiin ja täydennettiin toimeksiantajan vaatimusten täyttämiseksi. Tämän prosessin aikana löydettiin kolme olemassa olevaa avoimen lähdekoodin ratkaisua. Näihin ratkaisuihin perehdyttiin ja ne todettiin potentiaalisiksi vastauksiksi toimeksiantajan antamaan tehtävään.

Kolmen ratkaisun lisäksi toimeksiantajan olemassa olevaa ratkaisua päivitettiin lataamaan useampia indikaattoreita kerralla.

Yhdessä nämä neljä ratkaisua luovat kattavan otannan erilaisista indikaattoreista, joita toimeksiantajan käyttämä SIEM-järjestelmä tukee.

Löydettyjä ratkaisuja on mahdollista jatkokehittää siistimällä koodia sekä etsimällä uusia osoitteita, jotka tarjoavat indikaattoreita.

Avainsanat (asiasanat)
SIEM, SIM, SEM, lokitiedosto, tietoturva, kyberturvallisuus, uhkaindikaattori

Muut tiedot (Salassa pidettävät liitteet)

jamk.fi

Description

Author(s)	Type of publication	Date
Järvinen, Jani	Bachelor's thesis	October 2020
		Language of publication: Finnish
	53	Permission for web publication: x

Title of publication
Third-party Indicators of Compromise in SIEM-system

Degree programme
Information- and Communications Technology

Supervisor(s)
Rantonen, Mika

Assigned by
Qvantel Finland Oy

Abstract

Bachelor's thesis about adding third-party Indicators of Compromise to SIEM-system focuses on learning about SIEM-systems in general and indicators, while improving existing solutions. In order to understand the mechanics and basic principles of how a SIEM works, it was necessary to study log creation and management. When gathering information about log management, regulations and national laws came across, which affect e.g. how long logfiles should be stored. In addition to these, the surface of Information Security was merely scratched as it was not the goal of the thesis.

The goal was approached by searching for existing open source solutions, which after finding were modified and complemented by the author. Three solutions were discovered during this process. After familiarizing with the solutions, they were found to be potential answers for the assigners' requirements.

The existing solution from the assigner was updated to download more and different indicators from more sources.

None of the solutions excel on their own, each with their own faults, but together these four solutions create a comprehensive ground for updating the SIEM-system with additional indicators.

Found solutions also give an opportunity to be developed further by tidying up the code and finding new sources of indicators.

Keywords/tags (subjects)
SIEM, SIM, SEM, log, cyber security, indicator of compromise

Miscellaneous (Confidential information)

Sisältö

Kuviot

Taulukot

Lyhenteet

AD	Active Directory
API	Application Programming Interface
CIA	Confidentiality, Integrity and Availability
CSV	Comma Separated Values
CTI	Cyber Threat Intelligence
DHCP	Dynamic Host Configuration Protocol
IDS	Intrusion Detection System
IP	Internet Protocol
JSON	JavaScript Object Notation
LDAP	Lightweight Directory Access Protocol
OTX	Open Threat Exchange
SEM	Security Event Management
SIM	Security Information Management
SIEM	Security Information and Event Management
TLS	Transport Layer Security
UEBA	User and Entity Behavior Analytics
XML	eXtensible Markup Language

1 Johdanto

1.1 Toimeksianto

Toimeksiantona oli tutustua toimeksiantajan valitsemaan SIEM-ratkaisuun sekä antaa parannusehdotuksia siihen. Tavoitteena oli tutkia keinoja parantaa olemassa olevaa SIEM-ratkaisua lisäämällä siihen ulkopuolisia komponentteja, mahdollisesti automatisoiden niitä. Opinnäytetyön tekijän tavoite oli perehtyä SIEM-teknologiaan ja työkaluihin ja näiden tarjoamiin mahdollisuuksiin ja haittoihin. Toimeksianto toteutettiin pääasiassa teoreettisin menetelmin sekä jo valmiiksi olemassa olevia avoimen lähdekoodin ratkaisuja hyödyntäen, koska toimeksiantajalla oli jo SIEM-ratkaisu käytössä. Toimeksiantajana tutkimukselle toimi Qvantel Finland Oy.

1.2 Toimeksiantaja

Qvantel Finland Oy on Suomessa vuonna 1995 perustettu, teleoperaattoreiden palveluiden digitalisoinnin vallankumouksen pioneeri. Qvantel tarjoaa uniikkia, pilvipohjaista tulevaisuuden kestävää full-stack ratkaisua, jolla houkutella asiakkaita, innovoida palveluita ja tuotteita sekä hallita liikevaihtoa. Qvantel auttaa omia asiakkaitaan hallitsemaan näiden asiakkaita läpi koko asiakkuuden elinkaaren toiminnallisilla työkaluilla, jotka ovat luotettavia ja helppoja käyttää. Qvantelin asiakkaisiin kuuluu teleoperaattoreita ympäri maapallon. (Qvantel company n.d.)

1.3 Tutkimusmenetelmät

Opinnäytetyön tutkimus on tehty fenomenologisella analyysilla, koska tutkimuksen tavoite oli löytää ja verrata erilaisia keinoja tai työkaluja, joilla ladata ja työstää uhkaindikaattoreita toimeksiantajan tietoturvakeskukseen eli tutkimus perustui välittömiin havaintoihin ja kokemuksen pohdintaan. Fenomenologinen analyysi on yksi laadullisen analyysin menetelmistä, missä painotetaan välittömiä havaintoja sekä

tutkimuksen kohteesta saatua kokemuksen pohdintaa ja reflektointia. Avoimuus ja tutkimuskohteen lähestyminen ilman ennakko-olettamuksia on lähtökohta fenomenologiselle analyysille. Fenomenologisella reduktiolla kuvataan tutkijan havaintoja tai kokemuksia haittaavien ulkoisten tekijöiden eliminointia. Tutkimuksen kohteen lisäksi analyysi voi rakentua tutkijan itsensä ja hänen kokemustensa ympärille. Fenomenologinen analyysi kattaa laajasti erilaisia tutkimusorientaatioita fenomenologian tieteenfilosofiasta. Tämä tarkoittaa sitä, että fenomenologista analyysia voidaan yhdistää useisiin muihin analyysitapoihin. (Fenomenologinen analyysi, n.d.)

Tutkimusta varten valikoitui kolme olemassa olevaa työkalua, joiden toiminnallisuuteen perehdyttiin tavoitteet mielessä. Opinnäytetyö keskittyy kyseisten työkalujen toimintaan ja muokkaamiseen toimeksiantajan tarpeisiin sopivaksi.

Näiden työkalujen arvioinnissa käytettiin löyhästi regressioanalyysia, kun eri työkalujen tehokkuutta ja toimintoja vertailtiin. Regressioanalyysi tarkoittaa analyysimenetelmää, jossa selvitetään yhden tai useamman muuttujan vaikutusta toiseen muuttujaan. (Riippuvuussuhteiden analyysit, n.d.)

2 Security Information and Event Management

2.1 Yleistä

Yrityksien siirtyessä toimimaan verkossa on niiden palveluiden toimivuuden takaamiseksi tärkeää sisällyttää kyberturvallisuustyökaluja sekä uhkien havainnointiin erikoistuneita ohjelmistoja. Security Information and Event Management (SIEM)-järjestelmät on kehitetty uhkien havainnointia ja seurantaa varten.

Security Information Management

Security Information Management (SIM)-järjestelmä keskittyy lokien keräämiseen, monitorointiin ja analysointiin turvallisuuteen liittyvien merkintöjen kautta. SIMin keräämä data normalisoidaan ihmisluettavaan muotoon. SIM-tuotteet ovat ohjelmisto agentteja, jotka pitävät yhteyttä keskitettyyn palvelimeen, ilmoittaen tälle turvallisuuteen liittyvistä tapahtumista. SIM näyttää raportteja, tilastoja sekä graafeja näistä tiedoista. (Security Information Management, 2015.)

Security Event Management

Security Event Management (SEM) on prosessi, jossa tunnistetaan, kerätään, monitoroidaan ja raportoidaan turvallisuuteen liittyviä tapahtumia tietojärjestelmissä. SEM sallii tapahtumien tallentamisen ja tarkastelun, sekä auttaa turvallisuustiimiä tai järjestelmäarkkitehtejä tietojärjestelmien ylläpidossa ja suunnittelussa. (Security Event Management, 2015.)

Security Information and Event Management

SIEM-järjestelmä yhdistää SIM ja SEM-ratkaisut, luoden teoriassa kaiken kattavan monitoroinnin aina verkkoliikenteestä yksittäisten henkilöiden toimiin. (Keary, 2020.)

2.2 Historia

2.2.1 Yleistä

Termi SIEM keksittiin vuonna 2005 Mark Nicolettin ja Amrit Williamsin toimesta Gartnerin raporttia varten. He ehdottivat uutta turvallisuusinformaatio-järjestelmää, perustuen kahteen edeltävään sukupolveen, SIMiin ja SEMiin. (What is SIEM, n.d.)

Ensimmäisen sukupolven SIEM-järjestelmät antoivat tietoturva-asiantuntijoille näkyvyyttä järjestelmiin, mutta toisen sukupolven SIEM-järjestelmät tuottivat niin paljon dataa, että sitä ei ollut mahdollista käsitellä manuaalisesti. Nykyiset kolmannen sukupolven SIEM-järjestelmät ovat ottaneet koneoppimisen hyödyksi ja ovat tehneet SIEM-ratkaisuista jälleen relevantteja. (Gailey, 2019.)

2.2.2 Ensimmäinen sukupolvi

Ensimmäiset SIEM-järjestelmät tarkoittivat uuden aikakauden alkua tietoturvallisuuden alalla, yhdistäen SIM- ja SEM-ratkaisut. Näiden ensimmäisten järjestelmien heikkous oli vertikaalinen skaalautuvuus, mikä tarkoitti, että ne tarvitsivat koko ajan enemmän ja enemmän resursseja pystyäkseen tekemään työnsä. Skaalautuvuus ei myöskään ollut ainoa heikkous ensimmäisessä sukupolvessa; datan siirtämisessä SIEM-järjestelmään ja sieltä pois oli myös oma haasteensa. Kojelaudat, raportit ja hälytykset olivat alkukantaisia luonteeltaan. (Mt.)

2.2.3 Toinen sukupolvi

Viisi vuotta ensimmäisen sukupolven jälkeen toinen sukupolvi saapui juuri ajoissa, mutta ei ilman ongelmia sekään. Suurin muutos edelliseen oli skaalautuvuuden muuttaminen horisontaaliseksi. Tämä salli vähentää SIEM-järjestelmälle osoitettuja resursseja. Luonnollisesti kehityksen edistyttyä SIEM 2.0 salli paremmat työkalut kojelautoihin ja raportointiin, kuten myös historiallisen datan käsittelyä ja vielä järkevässä ajassa. Skaalautuvuus, joka oli alkujaan suurin muutos verrattuna SIEM 1.0:aan paljastui lopulta myös sen suurimmaksi heikkoudeksi. Tämä uusi toimintatapa tuotti liian paljon dataa, mikä johti samanlaiseen lopputulokseen kuin dataa ei olisi ollenkaan. Myös hälytysominaisuudet olivat jääneet vähemmälle huomiolle, mikä johti siihen, että järjestelmät joutuivat edelleen luottamaan esikonfiguroituihin hälytyspisteisiin. (Mt.)

2.2.4 Kolmas sukupolvi

SIEM-järjestelmien kolmas ja tällä hetkellä uusin sukupolvi keskittyi radikaalisti enemmän operatiivisiin toimintoihin kuin teknologiaan. Tämä uusi järjestelmä toi mukanaan analytiikan koneoppimisen avulla. Analytiikka tekee tästä sukupolvesta huomattavasti erilaisen verrattuna aiempiin, koska esikonfiguroitujen hälytysten sijaan se keskittyy riski-painotteiseen analytiikkaan. Hälytykset ovat edelleen olemassa, mutta riskejä tutkimalla on helpompi havaita nollapäivä-haavoittuvuuksia.

Analytiikka perustuu käyttäjien ja laitteiden toiminnan seuraamiseen. Tätä kutsutaan termillä user and entity behaviour analytics eli UEBA. (Mt.)

2.3 Mikä on SIEM?

SIEM on yleinen nimitys ohjelmistoratkaisulle, joka kokoaa yhteen ja analysoi dataa yhtiön koko IT-infrastruktuurista reaaliajassa. SIEM-järjestelmä kerää turvallisuuteen painottuvaa dataa verkkolaitteista, palvelimista, toimialueen ohjauskoneista ja monesta muusta. Se tallentaa, normalisoi, kokoaa yhteen ja liittää analytiikan siihen dataan löytääkseen trendejä, havaitakseen uhkia ja salliakseen yhtiöiden tutkia kaikkia hälytyksiä. (Keary, 2020.)

Perinteiset tietoturvaohjelmistot keskittyvät yksittäisiin uhkiin huomaamatta kokonaiskuvaa. Intrusion Detection System (IDS)-järjestelmät harvoin pystyvät tekemään muuta kuin seuraamaan paketteja ja Internet Protocol (IP)-osoitteita. Lokit näyttävät pelkästään käyttäjäsessiot sekä konfiguraatiomuutokset. SIEM yhdistää nämä kaikki yhdeksi kokonaisuudeksi, jonka avulla voi seurata hyökkäyksen koko polun huomattavasti helpommin kuin manuaalisesti käyden kaiken läpi. (Mt.)

Nämä palvelut ovat vasta hiljattain saavuttaneet jalansijaa huomattavien kustannusten takia. Yleensä SIEM-järjestelmä myös tarvitsee pari osoitettua jäsentä tarkkailemaan sitä. InsightIDR korjaa tämän kattavalla automatisoinnilla. (Mt.)

2.4 SIEM-järjestelmien puutteita

Kuten kaikki tietoturvaan tarkoitetut ohjelmistot, myös tehokas SIEM-järjestelmä tarvitsee asiansa osaavan henkilön käyttämään sitä. Teknologia, jota SIEM-järjestelmät käyttävät, tuottaa erittäin paljon dataa ja tästä syystä myös paljon

hälytyksiä sekä turhia hälytyksiä. Työkalun ja järjestelmän tuntevan ihmisen on kuitenkin käytävä läpi kaikki hälytykset, joten pahimmillaan kokonaisia työpäiviä voi kulua pelkästään turhien hälytysten tarkastamiseen. (Miller, 2019.)

Vaikka SIEM-järjestelmä kerää paljon dataa ja aiheuttaa hälytyksiä, niillä ei ole merkitystä jos SIEMin keräämä data ei ole hyödyllistä. SIEMiä asentavan henkilön täytyy siis osata ja muistaa sallia oikeanlaisen datan keräys valvottavista järjestelmistä.

SIEM-järjestelmälle täytyy myös kertoa käyttäjätapauksia, että SIEM osaa tarkkailla tiettyjä liikkeitä SIEMiä käyttävän yrityksen infrastruktuurissa. (Henderson, 2018.)

2.5 Lokit

Lokitiedosto on automaattisesti luotu tiedosto, johon on kerätty tapahtumia ja viestejä esimerkiksi ohjelmistoista, käyttöjärjestelmästä ja käyttäjien tekemistä toimista. (Gavin, 2018).

Lokitiedostot ovat tärkeitä, koska ne ovat paras keino lähteä suorittamaan ongelmanratkaisua, jos esimerkiksi jokin lakkaa yhtäkkiä toimimasta. Kaikenlaiset virheilmoitukset tallentuvat automaattisesti lokitiedostoihin, joten jos debuggaus ei tuota ratkaisua, kannattaa tarkistaa ohjelman tuottamat lokit (Milecia, 2019.)

Lokitiedostoilla on erilainen merkitys riippuen käyttäjästä. Ohjelmistokehittäjä esimerkiksi ei ole kiinnostunut palvelimella vierailleiden käyttäjien puuhista vaan haluaa lukea koodaamansa ohjelmiston lokeja. Samalla tavalla tietoturvaihmiset eivät ole kiinnostuneita yksittäisen ohjelmiston tuottamista lokeista, vaan haluavat tarkkailla ennemmin koko järjestelmän toimintaa ja käyttäjien tekemisiä (Carstensen, 2018.)

2.6 Lokienhallinta

2.6.1 Yleistä

Lokien määrä, volyymi sekä moninaisuus on suurentunut valtavasti ja tämä on johtanut tarpeeseen ottaa käyttöön lokienhallinnan käytäntöjä. Lokienhallinnalla tarkoitetaan lokitiedoston koko elämänkaaren prosessia; lokitiedoston generoitumisesta sen poistamiseen asti. Tämä on tärkeää koska joitain tiedostoja voi joutua säilömään vuosikausia, kun taas joitakin tiedostoja voi poistaa jo viikon jälkeen. Lokin säilömiseen vaikuttavat monet asiat, kuten lainsäädäntö, sopimukset kolmansien osapuolien kanssa tai levytilan kapasiteetti. Pohjimmainen ongelma lokienhallinnassa onkin tasapainoilu levytilan kanssa, kun uusia lokeja tulee jatkuvasti. Lokienhallinnassa tulee ottaa myös huomioon CIA-periaate, eli Confidentiality, Integrity ja Availibilty eli Luotettavuus, Eheys ja Saatavuus. Lokien hallitsijan tulee toimillaan varmistaa, että kukaan, jolla ei ole oikeutta koskea lokeihin ei niin pääse tekemään. (Kent & Souppaya, 2006.)

2.6.2 Luotettavuus, eheys ja saatavuus

Luotettavuus, eheys ja saatavuus (LES tästä eteenpäin) -kolmikko on malli, joka on luotu antamaan ohjeistusta ja käytäntöjä yrityksille tietoturvan saralla. Luotettavuudella tarkoitetaan pääsyn rajoittamista dataan, eheydellä viitataan datan koskemattomuuteen ja tarkkuuteen, sekä saatavuudella varmistetaan että data on aina saatavilla tarpeellisille henkilöille. (Rouse, 2020.)

Luotettavuus

Luotettavuus tarkoittaa periaatteessa yksityisyyttä. Luotettavuutta edistävillä toimilla on tarkoitus rajata dataan pääsevien henkilöiden määrä ainoastaan heihin, joille data on tarpeellista. Yleinen käytäntö on jaotella data vahingon määrän ja tyypin mukaan, mikäli ei-toivottu osapuoli pääsisi käsiksi kyseiseen dataan. (Mt.)

Joskus erityinen koulutus on tarpeen henkilöille, jotka käsittelevät arkaluontoista dataa ja dokumentteja. Yleinen keino on kouluttaa kyseisiä henkilöitä tunnistamaan riskejä ja vaaroja, sekä kuinka suojautua niiltä. Oleellinen osa riskiensuojautumisessa on riittävän vahvat salasanat sekä tieto Social Engineering-uhista, eli kohdennetuista hyökkäyksistä henkilöä kohtaan joko phishing-sähköposteilla tai fyysisellä kanssakäymisellä. (Mt.)

Luotettavuutta tarvitsevan datan yleinen suojauskeino on kryptaus, eli tiedostojen salaaminen. Yksinkertaisin keino tähän ovat salasanat, mutta ne yksistään eivät tuo riittävää suojaa korkeampaa suojaustasoa vaativille tiedostoille. Pelkästä salasanasta seuraava askel on kaksivaiheinen todennus, josta on tulossa normi tietoturvallisen tunnistautumisen saralla. Näiden lisäksi on olemassa myös fyysisyyteen perustuvia salausmekanismeja, kuten biometrinen tunnistaminen esim. sormenjäljen tai iiriksen avulla. (Mt.)

Eheys

Eheydellä tarkoitetaan toimia, joilla varmistetaan datan yhtenäisyys, paikkansapitävyys sekä luotettavuus koko sen elinkaaren ajan. Tämä tarkoittaa, että datan sisältö ei saa muuttua siirtojen aikana ja on varmistettava, että ulkopuolinen henkilö ei ole muokannut dataa esim. Luotettavuuden pettäessä. Näihin keinoihin kuuluu muun muassa tiedosto-oikeudet sekä käyttäjänhallinta, jolla varmistetaan kuka voi lukea tai muokata tiedostoa. Versionhallinnalla voidaan välttää vahingollisia muutoksia tai tiedostojen poistoa sallittujen henkilöiden toimesta. Ihmisten aiheuttamien uhkien lisäksi myös ihmisen toiminnasta riippumattomiin tapauksiin on varauduttava, esim. sähkökatkoksiin tai palvelimien kaatumisiin. Edellä mainittuun tapaukseen varmuuskopiot ja redundanssisuus ovat hyvä suojauskeino. (Mt.)

Saatavuus

Saatavuudella tarkoitetaan pääasiassa laitteistoon liittyviä keinoja. Näihin kuuluu viallisten komponenttien korvaaminen, generaattoreiden käyttö sähkökatkosten varalta sekä uusimpien ohjelmistopäivitysten asentaminen. Redundanssi, vikasieto sekä

RAID-klusterit voivat ehkäistä vakavaa datan menetystä komponenttivian sattuessa. Varmin keino pitää data aina saatavilla on ottaa huomioon myös luonnonkatastrofit kuten tulvat ja tulipalot. Näiden välttämiseksi paras keino on pitää kahta erillistä datakeskusta kaukana toisistaan. (Mt.)

2.6.3 Lainsäädäntö lokienhallinnassa

Jokaisella maalla voi olla omia laissa esitettyjä vaatimuksia yleisesti tietojenkäsittelyyn, ja lokitiedostojen säilytys on yksi näistä asioista. Esimerkiksi seuraava vaatimus on esitetty sekä Suomen laissa että EU-laissa

> *Tietojen luvattoman muuttamisen ja muun luvattoman tai asiattoman tietojen käsittelyn havaitsemiseksi tietojenkäsittely-ympäristössä toteutetaan luotettavat menetelmät turvallisuuteen liittyvien tapahtumien jäljitettävyyteen.* (L 681/2010, 5§ 1 mom. 6 kohta, 20 §).

Tähän vaatimukseen on annettu toteutusesimerkkejä Katakrissa, eli Tietoturvallisuuden auditointityökalussa viranomaisille.

> *Suojaustasolla IV vaatimus voidaan täyttää siten, että toteutetaan alla mainitut toimenpiteet:*
>
> *2) Keskeiset tallenteet säilytetään vähintään 6 kk, ellei lainsäädäntö tai sopimukset edellytä pitempää säilytysaikaa.* (Puolustusministeriö, s. 46)

Toteutusesimerkissä mainittu Suojaustaso IV on yksi neljästä tasosta, jotka on määritetty Suomen laissa. Suojaustasot on tarkoitettu käytettäväksi salassa pidettävien asiakirjojen luokittelussa. (L 681/2010, § 9)

Suojaustaso I

jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa erityisen suurta vahinkoa salassapitosäännöksessä tarkoitetulle yleiselle edulle. (L 681/2010, § 9, 1 kohta)

Suojaustaso II

jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa merkittävää vahinkoa salassapitosäännöksessä tarkoitetulle yleiselle edulle. (L 681/2010, § 9, 2 kohta)

Suojaustaso III

jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa vahinkoa salassapitosäännöksessä tarkoitetulle yleiselle tai yksityiselle edulle. (L 681/2010, § 9, 3 kohta)

Suojaustaso IV

jos asiakirjaan sisältyvän salassa pidettävän tiedon oikeudeton paljastuminen tai oikeudeton käyttö voi aiheuttaa haittaa salassapitosäännöksessä tarkoitetulle yleiselle tai yksityiselle edulle. (L 681/2010, § 9, 4 kohta)

Lokienhallinnan ongelmat voidaan karkeasti jakaa kolmeen osaan: Lokien luonti ja säilytys, Lokien suojelu sekä Lokien Analysointi.

2.6.4 Lokien luonti ja säilytys

Helpottaakseen lokien analysointia on järkevää muuntaa kaikki lokitiedostot yhdenlaiseen formaattiin samanlaisilla datakentillä. Koska melkein kaikki tietoverkkolaitteet yrityksen järjestelmässä tuottavat useita erilaisia lokitiedostoja, lokien koko ja määrä yrityksen sisällä voi nopeasti käydä sietämättömäksi. Jotkut komponentit tuottavat vielä valtavan kokoisia lokitiedostoja päivittäin, pahimmillaan jopa satojen gigatavujen kokoisia. Tämä vaikuttaa suoraan järjestelmän vaatimiin resursseihin. (Kent & Souppaya, 2006.)

2.6.5 Lokien suojelu

Koska lokit saattavat sisältää arkaluontoista tietoa järjestelmistä tai henkilötietoja, niiden luotettavuuden ja eheyden takaaminen on tärkeä osa lokienhallintaa. Riittämättömästi suojellut lokitiedostot ovat alttiita tahalliselle ja tahattomalle väärinkäytölle, muokkaamiselle tai jopa poistamiselle. Tämä voi aiheuttaa pahimmillaan hyökkääjän havaitsemattoman pääsyn järjestelmään. Esimerkiksi monet rootkit-haittaohjelmat on suunniteltu muokkaamaan lokitiedostoja kiinnijäämisen välttämiseksi.

Yrityksien täytyy pitää myös huolta lokitiedostojen saatavuudesta. Jos loki on konfiguroitu ylikirjoittamaan x-määrän tapahtumien jälkeen, tämä johtaa datan häviämiseen eli saatavuus on mahdotonta toteuttaa. Tiedon säilytyksen vaatimusten täyttämiseksi yritykset saattavat joutua pitämään varmuuskopioita lokitiedostoista, mikä voi pitkällä aikavälillä johtaa tilanpuutteeseen, jos lokitiedostot ovat suuria. Tätä voi lieventää suodattamalla dataa, jota tallennetaan lokeihin. (Mt.)

2.6.6 Lokien analysointi

Ennen SIEM-järjestelmiä lokien analysointi on jäänyt verkko- ja järjestelmänvalvojille. Lokien analysoinnin vaikeuden ja tylsyyden takia sitä on kohdeltu alemman tärkeys-

luokan prioriteettina, mikä on johtanut reaktiopohjaiseen toimintaan proaktiivisuuden sijaan. SIEM-järjestelmä on tuonut merkittäviä helpotuksia tähän automatisoiden suuren osan prosessista hälytysten ja oleellisen informaation näyttämisellä. (Mt.)

2.7 SIEM-järjestelmän toiminta

SIEM-järjestelmä tarjoaa kaksi pääasiallista kyvykkyyttä Incident Response -tiimille: raportointi ja rikostekninen tutkimus tietoturvapoikkeamista sekä analytiikkaan ja sääntöihin perustuvat hälytykset. Pohjimmiltaan SIEM-ratkaisut ovat datan keräys, haku ja raportointi järjestelmiä (Petters, 2019).

Karkeasti sanoen SIEM-järjestelmän toiminta voidaan jakaa kolmeen prosessiin (ks. kuvio 1). Tapahtumatietojen keräyksellä tarkoitetaan varsinaista lokitiedoston luomista, jossa itse data kerätään. Tämän jälkeen SIEM-järjestelmä analysoi saamaansa dataa ja löytäessään jotain poikkeuksellista aiheuttaa hälytyksen. Hälytyksen tultua vastuullinen henkilö tai tiimi lähtee selvittämään hälytystä ja edeltänyttä tapahtumavirtaa. (Kinnunen, 2017.)

Kuvio 1. SIEM-järjestelmän toimintaprosessi (Kinnunen, 2017).

3 Tietoturva

3.1 Mitä on tietoturva?

Tietoturva tarkoittaa hallussa pidettävän datan saatavuudesta, eheydestä ja luottamuksellisuudesta huolehtimista. Data voi olla saatavilla sekä digitaalisessa että fyysisessä muodossa. Saatavuudella taataan, että tieto on saatavilla aina tarvittaessa. Eheydellä pidetään huolta siitä, että data on pysynyt elinkaarensa ajan luotettavana ja muuttumattomana epätoivottujen tahojen toimesta. Luottamuksellisuudella tarkoitetaan, että dataa käsittelevät ainoastaan henkilöt, joilla on valtuudet tehdä niin (Muurinen, 2019.)

3.2 Tietoturvauhat

Koska tietoturva on niin laaja-alainen käsite, ovat myös uhat sitä kohtaan moninaiset. Näihin uhkiin sisältyy muun muassa kyberhyökkäykset, identiteettivarkaudet, laitteiden tai informaation varastaminen, sabotaasi ja informaatiolla kiristäminen. Uhka voi olla mitä tahansa mikä voi hyödyntää haavoittuvuutta järjestelmässä ja tämän avulla päästä muokkaamaan, poistamaan tai kopioimaan dataa. (Garg, n.d.)

Kyberhyökkäyksellä tarkoitetaan esimerkiksi tietokoneviruksella, madolla tai troijalaisella hevosella suoritettua hyökkäystä tietojärjestelmää kohtaan. Kaikki edellä mainitut sisältyvät haittaohjelmien kirjoon. Haittaohjelmalla tarkoitetaan pahantekoon tarkoituksellisesti luotua ohjelmaa tai koodinpätkää. Haittaohjelmat voidaan jakaa kahteen kategoriaan seuraavilla tavoilla (mt.):

1. Tarttumistapa
2. Haittaohjelman toiminnot

Tarttumistavan mukaan jaettaviin metodeihin kuuluu seuraavat haittaohjelmat (mt.):

1. Virukset
    - Virusten toimintaperiaate on monistautua liittämällä itsensä kohdekoneella tiedostoihin ja tämän jälkeen siirtyä verkon ylitse mahdollisimman laajalle.
2. Madot
    - Myös matojen toimintaperiaate on monistautua, mutta ne eivät liittäydy tiedostoihin uhrikoneella. Madot hyödyntävät suoraa verkkoyhteyttä levitäkseen laajalle.
3. Troijalaiset
    - Troijaisten toimintaperiaate on piiloutua ohjelmistojen sisälle ja aktivoitua kun kyseinen ohjelmisto ajetaan.
4. Botit
    - Botteja voidaan pitää matojen kehittyneempänä muotona, ne eivät välttämättä tarvitse ihmisinteraktiota koska ne voidaan luoda automatisoiduiksi prosesseiksi. Haitallisten bottien tarkoitus on levitä mahdollisimman laajalle luoden bottiverkko, jota pahantekijä voi käyttää pahoihin tarkoituksiin.

Toiminnon mukaan jaettaviin metodeihin kuuluu seuraavat haittaohjelmat (mt.):

1. Mainosohjelmat
    - Mainosohjelmat eivät välttämättä ole pahantahtoisia luonnostaan, mutta hyökkääjän niin halutessa voivat olla sitäkin. Pääasiassa ne rikkovat käyttäjän yksityisyyttä ja näyttävät mainoksia työpöydällä tai muissa ohjelmistoissa.
2. Vakoiluohjelmat
    - Vakoiluohjelma, kuten nimikin antaa ymmärtää, on haittaohjelma, joka vakoilee uhrin tekemisiä ja välittää tiedot eteenpäin. Yleinen esimerkki vakoiluohjelmistosta on Keylogger, joka tallentaa käyttäjän kaikki näppäimistön painallukset. Kyseisiä Keyloggereita saa myös laillisina ohjelmistoina, jos jostain syystä haluaa itse tallentaa omalla tietokoneellaan tehdyt näppäimistöpainallukset.
3. Kiristysohjelmat
    - Kiristysohjelmat voivat kryptata tärkeitä tiedostoja tai koko kovalevyn näin tehden tietokoneesta käyttökelvottoman.
4. Säikytysohjelmat
    - Säikytysohjelmat esittäytyvät hyödyllisinä työkaluina mutta todellisuudessa tartuttavat tai tuhoavat järjestelmän.
5. Piilohallintaohjelmat
    - Piilohallintaohjelmat on suunniteltu antamaan hyökkääjälle root-pääsy tai järjestelmänhaltijan oikeudet uhrin järjestelmään. Tämä tarkoittaa käytännössä rajatonta toimintavaltuutta uhrin järjestelmässä.
6. Zombit
    - Periaatteeltaan toimii samalla tavalla kuin vakoiluohjelmat, mutta vakoilun sijaan ne odottavat komentoja hyökkääjältä.

4 Insight

4.1 Insight-alusta

Vuonna 2015 Rapid7 julkaisi Insight-alustansa, joka kokosi yhteen Rapid7n ajan myötä karttuneen tiedon ja taidon haavoittuvuuksien etsinnässä, Internetin-laajuisen datan skannauksen, altistus-analytiikat ja tosiaikaisen raportoinnin tarjotakseen työkalun muuttaa saatavilla olevaa dataa tiedoksi ja vastauksiksi.

InsightAppSec

InsightAppSec on osa Rapid7: n turvallisuusohjelmistoa, mikä on kehitetty ohjelmistokehittäjiä varten. InsightAppSec skannaa ohjelmistoja ja sitä voi muokata tuhansilla eri vaihtoehdoilla vastaamaan omiin tarpeisiin. AppSec keskittyy tietoturvaongelmiin ohjelmistoissa. (Welcome to InsightAppSec, n.d.)

InsightConnect

InsightConnect on Rapid7: n Security Orchestration and Automation Response -ratkaisu, jonka avulla voidaan automatisoida aikaa vieviä turvallisuus prosesseja, kuten esimerkiksi käyttäjän AD-tunnuksen käytöstä poisto. (What is InsigtConnect? n.d.)

InsightIDR

InsightIDR on tietoturvakeskus tietoturvapoikkeamien havainnointiin ja vastaamiseen, käyttäjäautentikointien seuraamiseen ja työpisteiden näkyvyyteen. Se havaitsee väärät kirjautumisyritykset ulkoisilta ja sisäisiltä uhilta ja korostaa epäilyttävää käyttäjäkäytöstä. InsightIDR yhdistää työasemien forensiikan, lokien etsinnän ja edistyneet kojelaudat yhteen näkymään. Se on Software as a Service eli SaaS-palvelu, jota kehittää ja myy Rapid7. (InsightIDR Overview, n.d.)

InsightVM

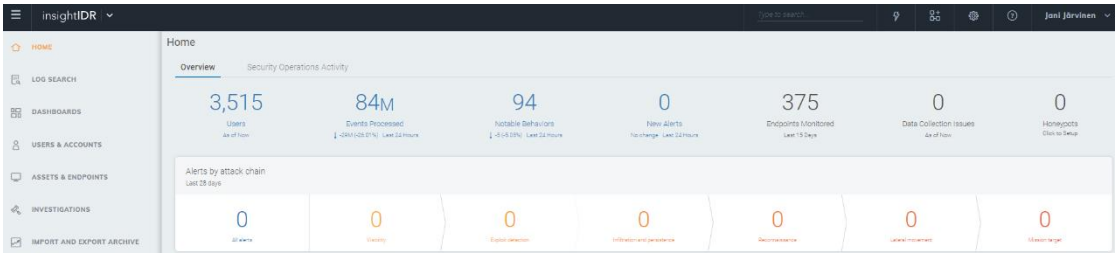
InsightVM nostaa Insight-alustan reaaliaikaiseen haavoittuvuuksien ja työasemien analysointiin, keskittyen enemmän ohjelmistopohjaisiin tietoturvariskeihin. InsightVM pohjautuu Rapid7 edeltävään palkintoja voittaneeseen Nexposeen. (InsightVM FAQ, n.d.)

4.2 InsightIDR

Asennuksen yhteydessä Insightille syötetään käyttäjät ja tämän jälkeen niiden tietoja rikastetaan Lightweight Directory Access Protocollalla (LDAP), Active Directorylla (AD), Dynamic Host Configuration Protocollalla (DHCP) sekä omalla Insight Agent

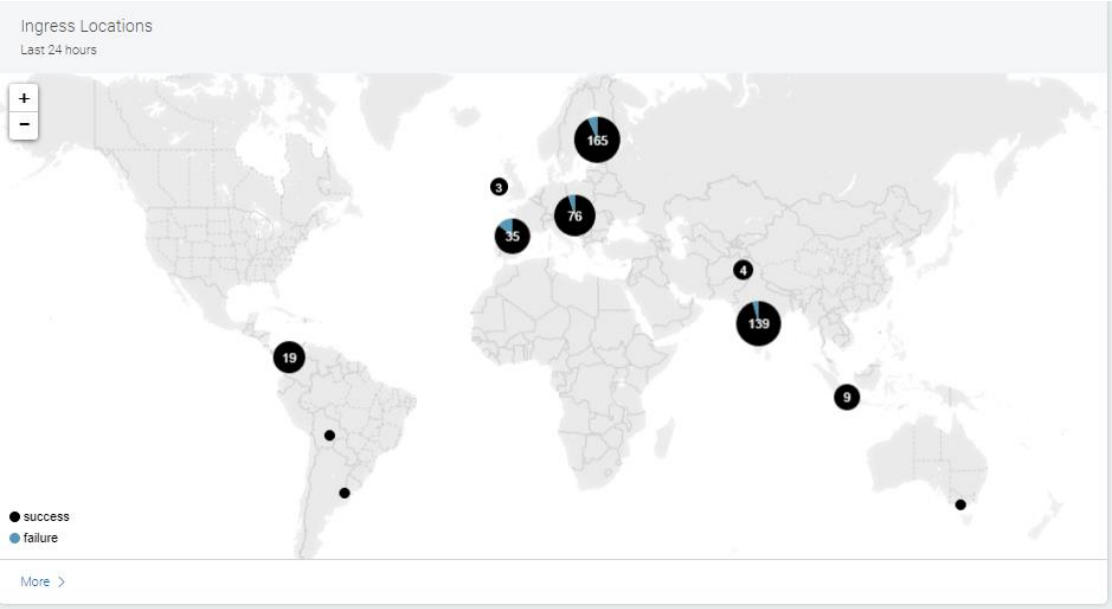
ohjelmistolla tarkastellaan näiden käyttäytymistä. Tämän jälkeen huolellisen suunnittelun myötä Insightin avuksi etsitään vähintään yksi palvelin, jolle asennetaan Kerääjä. Kerääjän tehtävä on siis nimensä mukaisesti kerätä kaikki lokitapahtumat, joita yrityksen verkossa tapahtuu. Tämän jälkeen se normalisoi, määrittelee, analysoi ja lopulta esittää havaintonsa Insightin kojelaudalla. (InsightIDR Quick Start Quide, n.d.)

Kuviossa 2 on esitetty InsightIDRn oletusnäkymä sisäänkirjautumisen jälkeen, mistä näkee nopealla vilkaisulla aktiivisten hälytysten määrän, valvottavien kohdekoneiden määrän, sekä muita tärkeitä asioita.



Kuvio 2. InsightIDR kojelauta

Aktiivisten hälytysten lisäksi oletusnäkymä tarjoaa myös pikanäkymän aktiivisiin käyttäjiin ja viimeisimpiin prosesseihin, joita pyörii valvottavilla koneilla, sekä maailmankartan, josta näkyy sekä onnistuneet että epäonnistuneet kirjautumisyritykset käyttäjien tileille (ks. Kuvio 3). Hälytysten lisäksi InsightIDR erittelee myös poikkeuksellista toimintaa osoittaneet henkilöt, joiden toiminta ei kuitenkaan ole hälytyksen arvoista. Jos esimerkiksi henkilö, joka ei normaalisti kirjaudu useille palvelimille yhtäkkiä kirjautuukin, vaikka päivitystöitä tehdäkseen, InsightIDR merkitsee tämän toiminnan epänormaalina, mutta ei kuitenkaan nosta hälytystä.

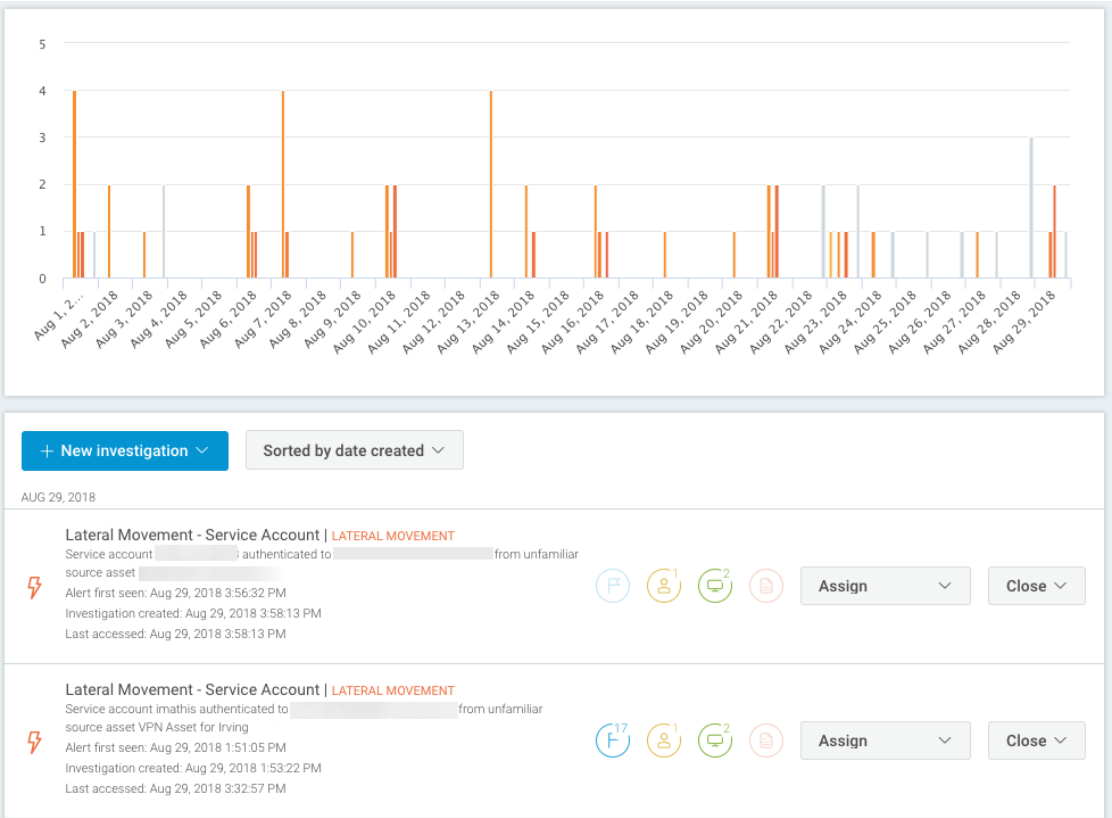


Kuvio 3. Sisäänkirjautumiset viimeisen vuorokauden aikana

4.2.1 Hälytyspolku InsightIDR:ssä

Epäilyttävää toimintaa havaitessaan Insight luo tutkimuksen ja lähettää sähköposti-ilmoituksen säädetyille vastaanottajille. Vastuullinen henkilö tai tiimi kirjautuu sisään Insightiin ja aloittaa tutkimukset. Hälytykset sisältävät todistusaineistoa, joilla InsightIDR kertoo millä perusteella hälytys on aktivoitunut. Tämä todistusaineisto sisältää hälytyksestä riippuen erilaisen määrän tietoa. Yleisesti kuitenkin siitä voi nähdä koskeeko hälytys käyttäjää, ulkoista hyökkääjää vai jotain muuta. Näiden tietojen perusteella hälytyksen tutkija lähtee selvittämään onko kyseessä vakava ongelma vai väärä hälytys. Jos hälytys on aiheutunut käyttäjätoimista, tämä käyttäjä tulisi välittömästi siirtää tarkkailulistalle. Tarkkailulista InsightIDR:ssä tarkoittaa sitä, että käyttäjä on tarkemman valvonnan alla ja aiheuttaa herkemmin hälytyksiä.

Kuviossa 4 on esitetty kaksi esimerkkihälytystä, jotka eivät ole kyseisellä hetkellä aktiivisen työskentelyn alla.



Kuvio 4. Viimeisimmät hälytykset Insightissa

4.2.2 Uhkaindikaattorit

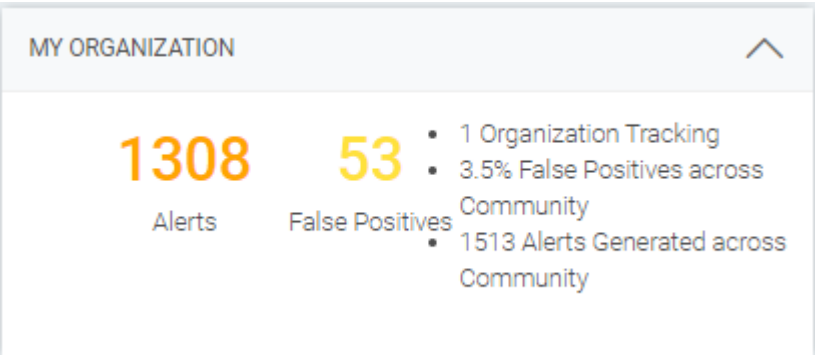
Vaikka InsightIDR sisältää sisäänrakennettuna lukuisia erilaisia hälytyksiä, näiden lisäksi voi hyödyntää Uhkasyötettä, joka sisältää muiden käyttäjien tai ryhmien tekemiä hälytyksiä, joita ei oletuksena InsightIDR:ssä näy. Uhkasyötteitä voi myös luoda itse. Tämä ominaisuus on hyödyllinen, jos haluaa keskittyä tietyntyyppisiin uhkiin. Indikaattorit, joita InsightIDR oletuksena tukee uusissa uhkalistoissa ovat: Domainit, MD5-Hashit, IP-osoitteet sekä URL-osoitteet. (Threats, n.d.)

Ulkoisten uhkaindikaattoreiden lisäys InsightIDR:ään tapahtuu Threat Application Programming Interfacella (API) eli ohjelmointirajapinnalla (Use the Threat API n.d). API tarkoittaa ohjelmointirajapintaa, jonka avulla sovellukset voivat kommunikoida toisten sovellusten kanssa, hakien esimerkiksi dataa näistä. (API – Mikä on API, n.d).

Tällä hetkellä InsightIDR tukee kahta eri API-rajapintaa: Threat API ja InsightIDR REST API. Threat API on tällä hetkellä osittain vanhentunut, mutta sitä voi silti käyttää manuaaliseen indikaattoreiden hakuun ulkoisilta sivuilta. InsightIDR REST API on uudempi ja tarjoaa automatisoidun tavan lisätä ja korvata vanhoja indikaattoreita. (Use the Threat API, n.d.)

5 Tutkimussuunnitelma

Qvantel Finland Oy on ottanut InsightIDR:n käyttöön turvallisuuskeskukseksi tietoturvapoikkeamien havainnointiin ja hoitamiseen sekä tunnistautumisten ja päätepisteiden valvontaan. Työkalun on tarkoitus helpottaa automaattisella valvonnallaan epäilyttävän toiminnan havainnointia tuhansista datalähteistä. InsightIDR ei siis korvaa täysin perinteisiä menetelmiä, se on vain yksi työkalu muiden joukossa. Turvallisuustiimin silmissä InsightIDR on osoittautunut todella tehokkaaksi ajansäästäjäksi. Kuviossa 5 esitetään kaikkien hälytysten määrä koko siltä ajalta kun InsightIDR on ollut käytössä Qvantelilla.



Kuvio 5. Hälytysten ja turhien hälytysten määrä

Tutkimuksen tavoite on perehtyä erilaisiin keinoihin etsiä ja ladata omia uhkaindikaattoreita InsightIDR:ää varten ja kehittää jo olemassa olevaa ratkaisua. Tämän on

tarkoitus johtaa ajan tasalla olevaan SIEM-järjestelmään, jonka manuaalinen päivitys tai hoitaminen on minimaalista, mikä antaa turvallisuustiimille mahdollisuuden keskittyä oikeisiin hälytyksiin ja tärkeisiin asioihin.

6 Toteutus teoriassa

6.1 Yleistä

Rapid7 on julkaissut kattavat dokumentaatiot InsightIDR:n APIn käytöstä, eli kuinka saada laajennettua uhkaindikaattoreita omatoimisesti. Toteutuksessa tutustutaan useaan eri vaihtoehtoon tämän toteuttamiseksi. Vaihtoehdot sisältävät PowerShell skriptin, Qvantelin edustajan luoman skriptin, jota aiotaan jatkokehittää sekä kaksi avoimen lähdekoodin ratkaisua GitHubista. Kaikki toteutuksessa esitetty koodi on ajettu Windows 10 käyttöjärjestelmällä, osa Windows Subsystem Linuxilla (WSL), jossa on käytössä Ubuntu 20.04 viimeisimmillä päivityksillä.

6.2 Kuinka uhkaindikaattoreiden lisäys onnistuu?

InsightIDR tarjoaa useita keinoja uhkaindikaattoreiden manuaaliseen lisäämiseen. Lisäämisen lisäksi on myös mahdollista korvata olemassa olevat indikaattorit uusilla, poistaen vanhat. Indikaattoreita itsejään voivat olla Domainit, MD5-Hashit, IP-osoitteet sekä URL-osoitteet. Näitä voidaan lisätä manuaalisesti kirjoittamalla, lataamalla indikaattorit sisältävä tiedosto InsightIDR:ään tai lähettämällä HTTP POST pyyntö. Tuettuja tiedostomuotoja ja formaatteja ovat JSON, stix-xml sekä CSV. Suurin sallittu tiedostokoko on 50 MB. (InsightIDR API (v1), n.d.)

JSON

JavaScript Object Notation (JSON) on syntaksi datan vaihtamiseen ja tallettamiseen. Se on luotu helppolukuiseksi sekä ihmisille että tietokoneille. Pohjimmiltaan se on

tekstiä, joka on muotoiltu tietyllä tavalla. Kuviossa 6 on esitetty esimerkki JSON-muotoilusta sekä InsightIDR:n hyväksymästä JSON-indikaattorista. Kuvion 6 esimerkillä lisättäisiin yksi objekti jokaiseen indikaattoriin. (Introducing JSON, n.d.)

Kuvio 6. JSON-esimerkki

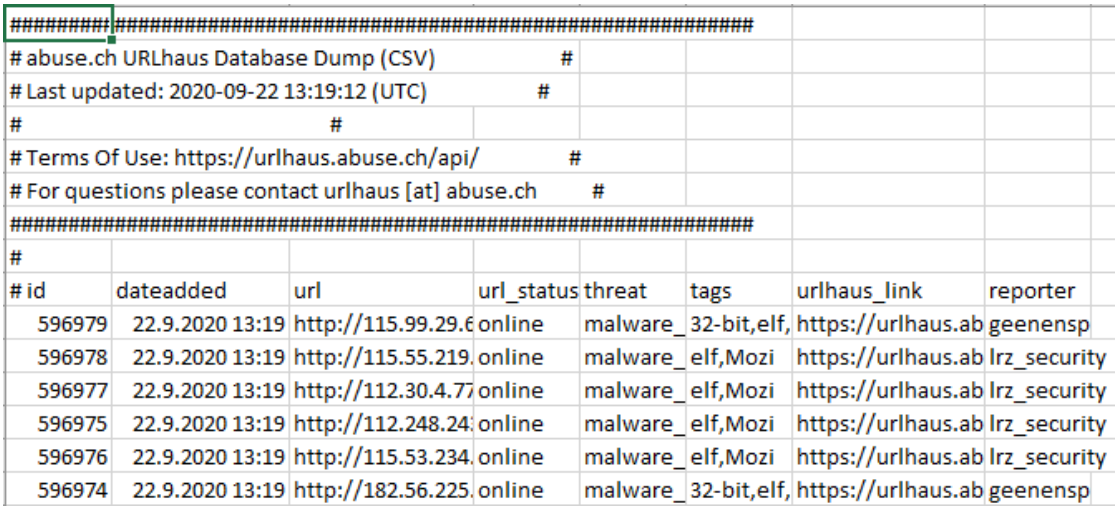
STIX-XML

Structured Threat Information Expression (STIX) on Cyber Threat Intelligencen (CTI) jakamiseen ja vaihtamiseen luotu kieli ja sarjallistettu formaatti. Tämä tiedostotyyppi sisältää huomattavasti enemmän tietoa, mutta on myös raskaampi prosessoida sekä ihmislukea. Kuviossa 7 on esimerkki muokkaamattomasta STIX-XML indikaattorista.

Kuvio 7. Ote STIX-XML tiedostosta

CSV

Comma-Separated Value on tiedostomuoto, jossa taulukkomuotoista dataa voidaan pilkulla ja rivinvaihdolla erotettuna tallentaa tekstitiedostoon (CSV-tiedosto, n.d.). Kuviossa 8 on esitetty CSV-tiedosto Excelissä.



##							
# abuse.ch URLhaus Database Dump (CSV) #							
# Last updated: 2020-09-22 13:19:12 (UTC) #							
# #							
# Terms Of Use: https://urlhaus.abuse.ch/api/ #							
# For questions please contact urlhaus [at] abuse.ch #							
##							
#							
# id	dateadded	url	url_status	threat	tags	urlhaus_link	reporter
596979	22.9.2020 13:19	http://115.99.29.6	online	malware_	32-bit,elf,	https://urlhaus.ab	geenensp
596978	22.9.2020 13:19	http://115.55.219.	online	malware_	elf,Mozi	https://urlhaus.ab	lrz_security
596977	22.9.2020 13:19	http://112.30.4.77	online	malware_	elf,Mozi	https://urlhaus.ab	lrz_security
596975	22.9.2020 13:19	http://112.248.24:	online	malware_	elf,Mozi	https://urlhaus.ab	lrz_security
596976	22.9.2020 13:19	http://115.53.234.	online	malware_	elf,Mozi	https://urlhaus.ab	lrz_security
596974	22.9.2020 13:19	http://182.56.225.	online	malware_	32-bit,elf,	https://urlhaus.ab	geenensp

Kuvio 8. Esimerkki CSV-tiedostosta Excelistä

6.3 Keinot indikaattoreiden lisäämiseen

6.3.1 Windows PowerShell

PowerShell on nykyisin järjestelmäriippumaton tehtävien automatisointi ja konfiguroinnin hallinnan kehys, koostuen komentorivitulkista ja skriptauskielestä. PowerShell on rakennettu .NET Common Language Runtimen päälle ja hyväksyy ja palauttaa .NET objekteja, tehden siitä merkittävästi erilaisen verrattuna perinteisiin komentorivitulkkeihin. Objekti on jäsennettyä informaatiota, joka sisältää enemmän tietoa kuin mitä komentorivi näyttää tulosteessa. (What is PowerShell, 2020)

6.3.2 OTX Alienvault

Open Threat Exchange (OTX) on maailman ensimmäinen ja suurin oikeasti avoimen uhkadatan vaihtopalvelu. OTX tarjoaa pääsyn maailmanlaajuiseen yhteisöön, jossa

uhkatutkijat ja tietoturva ammattilaiset raportoivat päivittäin yli 19 miljoonaa uhkaindikaattoria. OTX vaatii ilmaisen rekisteröitymisen dataan pääsemiseksi. (Open Threat Exchange FAQ, n.d.)

Tavoitteita varten OTX:n kanssa kokeiltiin kahta eri lähestymistapaa, molemmat avoimen lähdekoodin skriptejä.

6.3.3 Olemassa olevan ratkaisun parantelu

Qvantel Finland Oyn työntekijä on luonut Pythonilla skriptin, joka on kirjoitushetkellä käytössä ja jota on tarkoitus kehittää. Kyseinen skripti hakee ”urlhaus.abuse.ch” osoitteesta viimeisimpiä haitallisia URL-osoitteita CSV-tiedostona ja muuntaa ne JSON-muotoon ennen lähetystä InsightIDR:ään. Skripti on kopioitu GitHubista ja siihen on lisätty omaa koodia. Alkuperäinen skripti on esitetty kuviossa 9.

```
#!/usr/bin/env python3
import requests
import ssl
import csv
import json
ssl._create_default_https_context = ssl._create_unverified_context

ioc_url = "https://urlhaus.abuse.ch/downloads/csv_recent/"
threat_key = ""
api_key = ""
idr_url = "https://eu.api.insight.rapid7.com/idr/v1/customthreats/key/" + threat_key + "/indicators/replace?format=json"

headers = {
    'X-Api-Key': api_key,
    'Content-Type': 'application/json',
}

ioc_list = []
ioc_urls = {
    'domain_names':[''],
    'hashes':[''],
    'ips':[''],
    'urls':['']
    }

try:
    r = requests.get(ioc_url)
except r.status_code as e:
    if e.code == 404:
        print('404')
    else:
        print('jotain')
except r.status_code as e:
    print('Cant connect')
else:
    # 200
    print('Data retrieval complete!\n')
    body = r.text

lines = body.splitlines()
reader = csv.reader(lines, delimiter=',')

for row in reader:
    ioc_list.append(row)

# Remove the URLhause header information
del ioc_list[0:9]

for value in ioc_list:
    ioc_urls['urls'].append(value[2])

r = requests.post(idr_url, headers=headers, data=json.dumps(ioc_urls))
```

Kuvio 9. Alkuperäinen apicall-skripti

7 Toteutus

7.1 PowerShell

Rapid7:n blogikirjoituksessa ”Import External Threat Intelligence with the InsightIDR Threats API” on annettu PowerShell skripti, joka päivittää InsightIDR:ään haluttuja uhkaindikaattoreita. PowerShell toteutus ei vaadi ylimääräisiä asennuksia.

Skripti toimii siten, että se lataa indikaattorit käyttäjän määrittelemästä osoitteesta, muuntaa ne CSV-tiedostoksi ja sen jälkeen lähettää ne määriteltyyn Custom Threatiin InsightIDR:ään. Sellaisenaan skriptiin tarvitsee tehdä vain viisi muutosta. ***$IOCURL***, jolla määritetään osoite, josta haluttu indikaattorilista ladataan, ***$threatkey***, jolla määritetään mihin Custom Threatiin kyseinen lista halutaan lisätä, ***$headers***, jossa annetaan henkilökohtainen alustan API-avain sekä ***$Url***, josta pitää muuttaa InsightIDR:n aluekoodi oikeaksi, esimerkissä tämän ollessa **us**. Edellä mainitut ovat muuttujia, jotka annetaan skriptin alussa. (Copple, 2019)

$IOCURL = https://feodotracker.abuse.ch/downloads/ipblocklist.txt

$ThreatKey = " 12345esimerkkiavain "

$headers["X-Api-Key"] = " 54321avainesimerkki "

*$Url = "https://**eu**.api.insight.rapid7.com/idr/v1/customthreats/key/" + $ThreatKey + "/indicators/replace?format=csv"*

Kun skripti ajetaan, se luo ensimmäisenä kaksi väliaikaista tiedostoa: indicators.txt ja indicators.csv. Edellä mainitut tiedostot luodaan samaan sijaintiin kuin mistä skriptiä ajetaan. Sijainti määritetään komennolla **$path = Get-Location**.

$IOCOutputFileName = "indicators.txt"

$CSVOutputFileName = "indicators.csv"

$path = Get-Location

$IOCFilePath = "$path\" + "$IOCOutputFileName"

$CSVFilePath = "$path\" + "$CSVOutputFileName"

Tietoturva on otettu huomioon käyttämällä Transport Layer Security (TLS) 1.2 protokollaa lataukseen sekä tarkistamalla kohdeosoitteen varmenne.

[System.Net.ServicePointManager]::ServerCertificateValidationCallback = {$true}

[Net.ServicePointManager]::SecurityProtocol = 'Tls12'

TLS-protokollan ensisijainen tavoite on varmistaa yksityisyys ja datan eheys tiedonsiirrossa (RFC 5246:2008, 3).

Ennen indikaattoreiden lataamista skripti tarkastaa ja tarvittaessa poistaa vanhat samannimiset tiedostot. Jos ladattava tiedosto on tyhjä, skripti keskeytyy.

```
if (Test-Path $IOCFilePath) {
   Write-Host "Deleting existing indicator file: $IOCFilePath"
   Remove-Item $IOCFilePath
   }
if (Test-Path $CSVFilePath) {
   Write-Host "Deleting existing CSV file: $CSVFilePath"
   Remove-Item $CSVFilePath
   }
```

Esimerkkiskriptin lopussa on kommentoitu rivi koodia, jolla voidaan tarvittaessa poistaa tyhjät rivit tiedostosta. Kommentoinnilla tarkoitetaan sitä, että kyseistä kohtaa skriptistä ei ajeta, kun koko skripti ajetaan. (Copple, 2019.)

```
#(Get-Content $IOCFilePath) | ? {$_.trim() -ne "" } | set-content $CSVFilePath
```

Tavoitteita varten skriptiin lisättiin kaksi riviä, jolla tiedosto muunnetaan JSON-muotoon sekä formatoidaan poistamalla ylimääräistä dataa.

$JsonOutput = Get-Content $CSVOutputFileName

$JsonOutput -replace "^*?:" -replace ",", "" | ConvertTo-Json | Set-Content $JSONOutputFileName

7.2 OTX TAXII

Trusted Automated Exchange of Intelligence Information (TAXII) on ohjelmistoprotokolla, joka on luotu CTI-datan lähettämiseen HTTPS:n yli (Introduction to TAXII, 2020). Github käyttäjä Kirtar22 on luonut avoimen lähdekoodin Shell-skriptin, jolla haetaan Taxii-protokollalla STIX-muotoinen paketti OTXn sivuilta. Tämä skripti vaatii ulkopuolisen ohjelmiston – Cabbyn- käyttämistä ja/tai asentamista. Asennus tapahtuu helposti PIPin avulla.

pip install cabby

Vaihtoehtoisesti Cabbya voi ajaa Docker konttina.

docker run cabby

Cabby toimii Taxii-asiakkaana OTX palvelimen toimiessa Taxii-palvelimena. Cabby tarjoaa kolme pääasiallista toimintatapaa datan löytämiseen ja lataamiseen. Sitä ajetaan komennolla

taxii-x --path 'lisää argumentteja'

missä "x" voi olla discovery, collection tai poll. Discovery palauttaa kohdeosoitteen päätepisteet, eli mitä palveluita on käytettävissä. Esimerkiksi OTX palauttaa seuraavanlaista dataa:

janij@DESKTOP-VCPG00U:~/testi$ taxii-discovery --path https://otx.alienvault.com/taxii/discovery

2020-09-16 10:52:00,451 INFO: Sending Discovery_Request to https://otx.alienvault.com/taxii/discovery

2020-09-16 10:52:00,782 INFO: 3 services discovered

=== Service Instance ===

Service Type: POLL

Service Version: urn:taxii.mitre.org:services:1.1

Protocol Binding: urn:taxii.mitre.org:protocol:https:1.0

Service Address: https://otx.alienvault.com/taxii/poll

Message Binding: urn:taxii.mitre.org:message:xml:1.1

Available: True

Message: OTX Taxii Polling

Collection palauttaa kokoelmat, joita käyttäjä seuraa. Tämä ominaisuus vaatii henkilökohtaisen API-avaimen käyttöä.

janij@DESKTOP-VCPG00U:~/testi$ taxii-collections --path https://otx.alienvault.com/taxii/collections --username 'API-AVAIN' --password foo

2020-09-16 11:01:39,024 INFO: Sending Collection_Information_Request to https://otx.alienvault.com/taxii/collections

=== Data Collection Information ===

Collection Name: user_Metadefender

Collection Type: DATA_FEED

Available: True

Collection Description: Data feed for user: Metadefender

Supported Content: All

=== Polling Service Instance ===

Poll Protocol: urn:taxii.mitre.org:protocol:https:1.0

Poll Address: https://otx.alienvault.com/taxii/poll

Message Binding: urn:taxii.mitre.org:message:xml:1.1

Ilman avainta collection palauttaa pelkästään julkiset kokoelmat.

janij@DESKTOP-VCPG00U:~/testi$ taxii-collections --path https://otx.alien-vault.com/taxii/collections

2020-09-16 11:01:10,604 INFO: Sending Collection_Information_Request to https://otx.alienvault.com/taxii/collections

=== Data Collection Information ===

Collection Name: user_AlienVault

Collection Type: DATA_FEED

Available: True

Collection Description: Data feed for user: AlienVault

Supported Content: All

=== Polling Service Instance ===

Poll Protocol: urn:taxii.mitre.org:protocol:https:1.0

Poll Address: https://otx.alienvault.com/taxii/poll

Message Binding: urn:taxii.mitre.org:message:xml:1.1

Toteutusta varten poll on tärkein toiminto, koska sillä suoritetaan varsinainen datan lataaminen. Skripti vaatii haluttavan kokoelman määrittelyn, tässä tapauksessa user_Alienvault.

taxii-poll --path https://otx.alienvault.com/taxii/poll --collection user_AlienVault --username 'API-avain' --password foo --begin 2020-09-01 --dest-dir /home/janij/testi/

Kohdekansion määritys on tärkeä osa skriptiä, koska ilman sitä tuloste piirtyy raakadatana näytölle. Kuviossa 10 on esimerkki näytölle tulostetusta raakadatasta, joka kattaa ainoastaan yhden indikaattorin.

Kuvio 10. Raakadatan tuloste

Erillisellä Python-skriptillä XML-tiedostot saadaan muunnettua JSON-muotoon, tehden datasta ihmisystävällisempää luettavaa. Kuvio 11 sisältää saman datan kuin Kuvio 10, mutta eri muodossa. Skripti on luotu siten, että se vaatii käyttäjää antamaan muunnettavan tiedoston nimen, esimerkki:

python xml_to_json2.py user_AlienVault_6d156300cf4e0598be2372842638928a

Josta tulee skriptin ajon jälkeen uusi tiedosto vanhan jäädessä paikalleen.

janij@DESKTOP-VCPG00U:~/testi/threat_feeds$ ls | grep user_AlienVault_6d156300cf4e0598be2372842638928a

user_AlienVault_6d156300cf4e0598be2372842638928a

user_AlienVault_6d156300cf4e0598be2372842638928a.json

Vaihtoehtoinen käyttötapa skriptille on muokata se ottamaan haluttu tiedostonimi.

```
{
    "@id": "alienvault-otx:indicator-80e4e300-31bc-ce54-7c2d-cffcc8282c13",
    "@timestamp": "2020-09-15T18:49:27",
    "@xsi:type": "indicator:IndicatorType",
    "ns6:Description": null,
    "ns6:Observable": {
        "@id": "alienvault-otx:Observable-80e4e300-31bc-ce54-7c2d-cffcc8282c13",
        "ns7:Object": {
            "@id": "alienvault-otx:Address-973ce1d1-23f3-4388-93e0-46891aca9466",
            "ns7:Properties": {
                "@category": "ipv4-addr",
                "@xsi:type": "AddressObj:AddressObjectType",
                "ns9:Address_Value": "51.81.104.17"
            }
        },
        "ns7:Title": "IPv4 - 51.81.104.17"
    },
    "ns6:Title": "51.81.104.17 from https://otx.alienvault.com/pulse/5f610cb62458e403adeca72d"
},
```


Kuvio 11. Raakadata muunnettu JSON-muotoon

7.3 Signature-base

Github käyttäjä Neo23x0 on luonut avoimen lähdekoodin Python skriptin, joka lataa uhkaindikaattoreita OTX sivustolta, mutta antaen valinnaisia lisäoptioita ohjelman käyttämiseen. OTX API-avain on pakollinen tämän skriptin käyttämiseen. Tämä skripti hyödyntää OTX:n virallista Python Software Development Kitiä (SDK), OTXv2:sta. OTXv2 täytyy asentaa erikseen.

pip install OTXv2

Tähän skriptiin on myös sisällytetty niin sanotut valkolistat, joilla voidaan suodattaa luotettavaksi luokiteltuja sivustoja, jotka syystä tai toisesta saattavat löytyä raportoiduista haitallisista sivuista tai osoitteista.

HASH_WHITELIST = [

 # Empty file

'd41d8cd98f00b204e9800998ecf8427e',

'da39a3ee5e6b4b0d3255bfef95601890afd80709',

'e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855',

Skriptin alussa luodaan tyhjät merkkijonot täytettäville indikaattoreille sekä määritetään formatointi ja tiedostomuoto.

 # IOC Strings

hash_iocs = ""

filename_iocs = ""

c2_iocs_ipv4 = ""

c2_iocs_ipv6 = ""

c2_iocs_domain = ""

 # Output format

separator = ";"

use_csv_header = False

extension = "txt"

hash_upper = True

filename_regex_out = True

Skriptillä saadaan ulos myös CSV-päätteistä dataa, jos halutaan valmista dataa SIEM-järjestelmiä varten. CSV-muotoisen latauksen ero on erottimessa: .txt muodossa se on puolipilkku ja CSV-muodossa pilkku.

```
if siem_mode:
    self.separator = ","
    self.use_csv_header = csvheader
    self.extension = extension
    self.hash_upper = True
    self.filename_regex_out = False
```

Ladattavan datan määrää on mahdollista rajoittaa muokkaamalla **days_to_load** arvoa, tällöin skripti hakee dataa vain määrättyjen päivien verran.

```
mtime = (datetime.now() - timedelta(days=days_to_load)).isoformat()
```

Itse lataamiseen käytetään virallista OTX kirjastoa ja sen funktiota **getall()** (ks. Kuvio 12). Sitä kutsutaan komennolla

```
self.events = self.otx.getall()
```

```
def getall(self, modified_since=None, author_name=None, limit=20, max_page=None, max_items=None, iter=False):
    """
    Get all pulses user is subscribed to.
    :param modified_since: datetime object representing earliest date you want returned in results
    :param author_name: Name of pulse author to limit results to
    :param limit: The page size to retrieve in a single request
    :return: the consolidated set of pulses for the user
    """
    args = {'limit': limit}
    if modified_since is not None:
        if isinstance(modified_since, (datetime.datetime, datetime.date)):
            modified_since = modified_since.isoformat()

        args['modified_since'] = modified_since
    if author_name is not None:
        args['author_name'] = author_name

    return self.walkapi(
        self.create_url(SUBSCRIBED, **args), iter=iter,
        max_page=max_page, max_items=max_items
    )
```

Kuvio 12. Funktio getall

Ladattujen tapahtumien määrä ilmoitetaan ja sen jälkeen aloitetaan ladattujen tietojen prosessointi, jossa ladattua tekstiä formatoidaan.

print("Download complete - %s events received" % len(self.events))

self.hash_iocs += "{0}{3}{1} {2}\n".format(

hash,

description,

" / ".join(event["references"])[:80],

self.separator)

Formatoitu data tallennetaan tiedostoon.

with open(hash_ioc_file, "w") as hash_fh:

if self.use_csv_header:

hash_fh.write('hash{0}'.format(self.separator) + 'source\n')

hash_fh.write(self.hash_iocs)

print("{0} hash iocs written to {1}".format(self.hash_iocs.count('\n'), hash_ioc_file))

7.4 Apicall

Qvantel Finland Oy:n työntekijä on luonut apicall.py nimisen Python-skriptin, jonka lähdekoodi on vapaasti nähtävissä Githubissa. Skripti lataa ***urlhausista*** viimeisimmät haitalliset URL-osoitteet sisältävän CSV-tiedoston.

```
ioc_url = https://urlhaus.abuse.ch/downloads/csv_recent/

r = requests.get(ioc_url)
```

Datan lataamisen jälkeen jokainen rivi jaetaan omaksi listaobjektiksi. Näihin objekteihin lisätään erottimeksi pilkku. Tämän jälkeen skripti käy läpi jokaisen rivin ja lisää ne listaan.

```
lines = body.splitlines()

reader = csv.reader(lines, delimiter=',')

for row in reader:

    ioc_list.append(row)
```

Tästä listasta arvot siirretään dictionaryyyn ja tiettyyn “avaimeen”, tässä tapauksessa urleihin.

```
for value in ioc_list:

    ioc_urls['urls'].append(value[2])
```

Toteutusta varten päivitetyssä skriptissä edellä mainitut asiat toistetaan muille dictionaryn avaimille, käyttäen eri latauslähteitä.

8 Tulosten arviointi

8.1 Yleiskatsaus

Tarkastaltaessa toteutuksessa käytettyjä skriptejä sekä niiden toimivuutta on tärkeää huomioida, että toimivatko skriptit kuten pitäisi, lataavatko ne ajantasaista dataa ja onko tallennusformaatti oikea. Pidemmällä aikavälillä tulee ottaa huomioon myös tallennuskapasiteetti, jos ladattuja tiedostoja halutaan jättää talteen myöhempää katselua varten. Toteutusta varten ei määritelty mitään tiettyä lähdettä, mistä indikaattoreita pitäisi ladata, joten jokainen skripti hakee niitä eri tavalla ja eri osoitteista. Tästä syystä jokaista toteutustapaa pitää tarkastella omana kokonaisuutena ennen kuin niitä vertaa keskenään.

8.2 Powershell toteutus

PowerShell-skriptillä saatiin onnistuneesti ladattua uhkaindikaattoreita muutamasta abuse.ch sivuston palvelusta. Tekemällä vain pieniä muutoksia tai ei muutoksia lainkaan, PowerShell-skriptillä saadaan ladattua seuraavia indikaattoreita: IP-osoitteita, MD5-hasheja sekä Domaineja.

IP-osoitteet tulevat feodotracker.abuse.ch:sta ja sisältävät IP-osoitteita jotka ovat kytköksissä botnetteihin. MD5-hashit tulevat bazaar.abuse.ch osoitteesta ja nimensä mukaisesti sisältää haitalliseksi todettuja hasheja. Kaksi edellä mainittua eivät vaadi muutoksia skriptiin, muuta kuin lähdeosoitteen muutos. Domainien lataus sen sijaan vaatii myös pienen lisäyksen koodiin, jolla saadaan toistuvat 127.0.0.1 osoitteet poistettua (ks. Kuvio 13). Kuviossa 14 esitetään miltä ladattu data näyttää lataamisen ja JSON-muotoon muuttamisen jälkeen.

```
################################################################
# abuse.ch URLhaus Host file                                   #
# Last updated: 2020-09-18 08:49:20 (UTC)                      #
#                                                              #
# Terms Of Use: https://urlhaus.abuse.ch/api/                  #
# For questions please contact urlhaus [at] abuse.ch           #
################################################################
#
127.0.0.1       0-24bpautomentes.hu
127.0.0.1       01.shgrasp.vip
127.0.0.1       0931tangfc.com
127.0.0.1       1.17110.com
127.0.0.1       1314.ren
```

Kuvio 13. Haitalliset domainit lähdeosoitteessa

```
[
    "\t0-24bpautomentes.hu",
    "\t01.shgrasp.vip",
    "\t0931tangfc.com",
    "\t1.17110.com",
    "\t1314.ren",
```

Kuvio 14. Domainit ladattuna ja muunnettuna

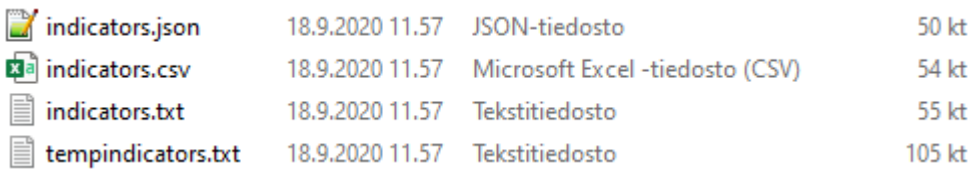
Eri latauslähteitä katsoessa nousee esille myös niiden päivittämistahti, joka on suoraan verrannollinen siihen miten ajantasaista dataa skriptillä saadaan ulos. Abuse.ch on voittoa tavoittelematon projekti joka toimii lahjoitusten varassa sekä osittain yhteisön avulla. Edellämainituista sivuista Feodotracker saa uusia havaintoja ainoastaan ylläpitäjien toimesta. URLHaus sallii uusien havaintojen lisäämisen myös tunnistautuneilta käyttäjiltä. Malwarebazaar on näistä kolmesta selkeästi kehittynein, sisältäen aktiivisen käyttäjäkunnan ja ylläpitäjät. Malwarebazaar tarjoaa kattavat statistiikka tiedot sivullaan (https://bazaar.abuse.ch/statistics/), josta voi nähdä esimerkiksi yleisimmät tiedostotyypit jotka ovat yhteydessä näytteisiin.

PowerShell skriptin tehokkuutta mitattiin lisäämällä skriptin alkuun ja loppuun komento **(Get-Date).Millisecond**. Lopuksi loppuarvosta vähennettiin alkuarvo ja näin saatiin suuntaa antava arvo siitä, miten kauan skriptillä kesti suorittaa toimintonsa. Mittaustapa ja näin ollen tulokset eivät ole täysin luotettavia, mutta antavat jonkinlaista indikaatiota siitä miten nopea skripti on (ks. taulukko 1). Testaus suoritettiin lataamalla samaa tiedostoa 10 kertaa putkeen, ensimmäisessä testissä koskematta

olemassa oleviin tiedostoihin ja toisessa testissä poistamalla ne testin jälkeen. Tulosten perusteella voidaan havaita, että skriptin suoritusnopeuden kannalta ei ole merkitystä kosketaanko ladattuihin tiedostoihin, sillä 4 millisekuntia on käytännössä olematon ero. Säilytystilan kannalta tiedostojen olemassaololla ei myöskään ole merkitystä, koska pelkkää tekstiä sisältäen tiedostot ovat todella pieniä (ks. kuvio 15) ja edelliset tiedostot poistetaan aina kun skriptiä ajetaan uudelleen.

Taulukko 1. Nopeustesti

Latauskerta	Lataus ilman poistoa (ms)	Lataus ja poisto (ms)
1	150	144
2	154	148
3	152	143
4	150	155
5	153	154
6	147	154
7	157	210
8	163	156
9	155	148
10	152	165
Keskiarvo	**153,3**	**157,7**



Kuvio 15. Skriptin lataamat ja luomat tiedostot

8.3 OTX TAXII

Ensimmäinen OTX:ää käyttävä skripti käytti TAXII-tekniikkaa, jonka ansiosta itse skripti on todella lyhyt. Periaatteessa se ei tarvitse edes omaa skriptiä, koska komentoa voi ajaa suoraan komentoriviltä. Myöhempää käyttöä ja mahdollista au-

tomaatiota varten skripti on kuitenkin hyvä olla. Tämä skripti kuitenkin vaatii kolmansien osapuolien hyödyntämistä alkaen tunnuksen luomisesta otx.alienvaultin sivustolle, jos haluaa laajentaa saatavan datan määrää. Ilman tunnustakin pystyy kyllä lataamaan kaikille avoimen “AlienVault”-nimisen käyttäjän indikaattoreita. Määritellyn tavoitteen saavuttamiseksi erillinen skripti vaaditaan ladatun datan muuntamiseksi JSON-muotoon. Tätä muunnosta varten on luotu Python skripti. SIEM-käyttöä varten tätä muunnosta ei kuitenkaan tarvitse tehdä, koska InsightIDR tukee STIX-paketteja.

TAXII-skriptin etuina on määriteltävä aikajakso, kuinka pitkältä ajalta halutaan ladata indikaattoreita, mahdollisuus laajentaa eri indikaattoreihin luomalla tunnus ja seuraamalla eri käyttäjiä. Suurin etu lienee kuitenkin saatavan datan määrä; siinä missä PowerShell-skripti lataa yhden tietyn tiedoston, joka sisältää esim. pelkkiä IP-osoitteita, TAXII-skripti lataa STIX-paketteja, jotka kertovat huomattavasti enemmän indikaattorista (ks. Kuvio 8).

Skriptin ajoaika riippuu merkittävästi siitä, kuinka pitkältä aikajaksolta halutaan ladata indikaattoreita. Esimerkiksi ladaten viimeiseltä kolmelta päivältä indikaattoreita, skriptillä kestää ajaa vain 6 sekuntia, mutta ladatessa viimeiseltä 17 päivältä skriptillä kestää jo 16 sekuntia. Jos skriptiä halutaan automatisoida ajamaan säännöllisin väliajoin, tällöin suoritusajalla ei ole niin suurta merkitystä. Huomionarvoista on kuitenkin se, että tämä skripti lataa samoja tiedostoja uudestaan, mutta eri tunnisteella, eli päällekkäisyyksiä voi syntyä ja ajan myötä tiedostomäärä kasvaa valtavaksi jos vanhoja tiedostoja ei poisteta.

8.4 Signature-base

Toinen OTX:ää hyödyntävä ohjelma on toteutettu Pythonilla ja se on huomattavasti monimutkaisempi kokonaisuus. Kyseistä ohjelmaa ei ole päivitetty kahteen vuoteen. Se myös vaatii pakollisen OTX-tunnuksen luomisen, sillä kestää useita minuutteja

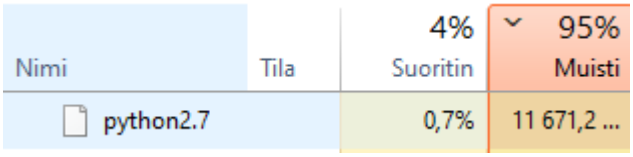
ajaa sekä vaatii erillisen skriptin JSON-muotoa varten. JSON-muunnos skriptin voisi mahdollisesti myös implementoida suoraan ohjelmaan. Useista haittapuolista huolimatta se tuottaa eniten dataa, jakaen sen useisiin tiedostoihin helpottaen jatkokäsittelyä huomattavasti. Kuviot 16 ja 17 sisältävät esimerkin tiedostonimistä ladatusta indikaattorilistasta ennen ja jälkeen JSON-muutosta.

Kuvio 16. Ladattu tekstitiedosto tiedostonimistä

Kuviosta 17 on karsittu tulostetta pois järkevämmän kuvan koon vuoksi.

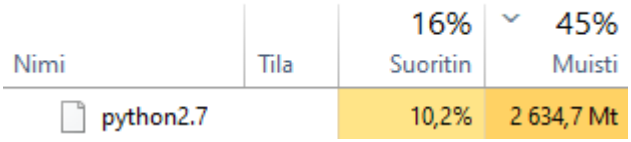
Kuvio 17. Osa kuvion 13 datasta JSON-muodossa

Skriptin ajonopeutta testatessa havaittiin, että ladattavien päivien määrän muuttamisella ei ole vaikutusta toimintaan eikä ladattavien tuloksien määrään. Oletuksena ohjelma kuitenkin lataa yhden päivän ajalta indikaattoreita. Koska päiviä ei voi muokata, muodostuu suoritusnopeutta muuttavaksi tekijäksi otx.alienvault sivuston seurattujen käyttäjien määrä, koska skripti hakee datan kaikilta seuratuilta käyttäjiltä. Skriptin suoritusajan kasvaessa myös sen vaatimat resurssit alustalta kasvavat. Hakiessa indikaattoreita kolmelta käyttäjältä ajoaika nousee yli tuntiin ja keskusmuistin käyttö on erittäin korkea (ks. Kuvio 18), johtaen huomattavaan hidasteluun työkoneella.



Kuvio 18. Skripti vie paljon muistia

Skriptin siirtyessä lataamisesta prosessointi vaiheeseen muistinkäyttö tippuu puolella (ks. Kuvio 19).



Kuvio 19. Muistinkäyttö lataamisen jälkeen

Lopulta kun skripti oli pyörinyt liki 18 tuntia eikä osoittanut merkkejä lopettamisesta, se keskeytettiin. Tuloksena erittäin suuri rajoitus ladattavien uhkaindikaattorien tarjoajien määrään otx.alienvault sivustolta.

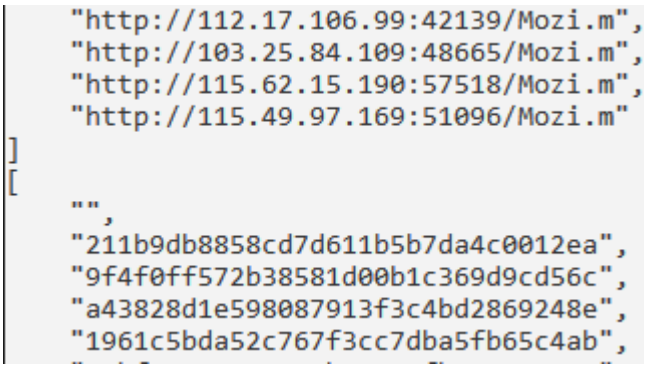
Taulukko 2. Signature-base testimittaus

Seurattujen käyttäjien lkm	Skriptin ajoaika	Ladattujen tapahtumien määrä
1	7m29s	2864
2	27m31s	15935
3	18h/keskeytetty	31156

8.5 Apicall

Olemassa olevan skriptin päivitysten myötä se lataa onnistuneesti indikaattoreita useammasta lähteestä, mutta levylle tallentamisen sijasta tulostaa ne vain ruudulle. Alkuperäisen skriptin toimintaperiaate pyrittiin pitämään ennallaan, eli skripti lähettäisi tiedostot vain suoraan InsightIDR:ään. Hyödyntämällä aiemmista Python

skripteistä tuttuja komentoja tämänkin skriptin saisi helposti muunnettua tallentamaan datan levylle, josta sitä voisi ihminenkin lukea. Tämän skriptin lataamat tiedostot ovat kuitenkin yksinkertaisia tekstitiedostoja, jotka eivät sisällä muuta kuin osoitteita tai hasheja (ks. Kuvio 20), joten tallentamista levylle ei koettu tarpeelliseksi. Skripti toimii muutosten jälkeenkin todella nopeasti, eikä toiminnallisuus ole merkittävästi muuttunut jo olemassa olevasta ratkaisusta, joten suoritustestejä ei koettu tarpeelliseksi.



Kuvio 20. Apicall skriptin tuloste

9 Yhteenveto

9.1 Työn toteutus ja tulokset

Työtä varten tutustuttiin useisiin avoimen lähdekoodin ohjelmiin ja skripteihin eri kielillä, joista valikoitui testattavaksi ne joiden havaittiin täyttävän kriteerit tai yksinkertaisesti vain toimivan. Löydettyjen skriptien joukossa oli ratkaisuja joita ei saatu toimimaan mitenkään, tämä saattoi johtua tekijän kokemattomuudesta ohjelmoinnin saralla. Työn tavoite ja testausmahdollisuudet muuttuivat projektin loppuvaiheessa, kun tunnukset InsightIDR:ään menetettiin työsuhteen loppumisen myötä ja toimeksiantaja ei nähnyt tarpeelliseksi antaa erillisiä tunnuksia sitä varten. Täten päädyttiin etsimään ratkaisua indikaattoreiden löytämiseen ja lataamiseen. Eniten aikaa työn

parissa kului ohjelmiin ja skripteihin tutustumisessa sekä muokkaamisessa, johtuen osaamisen heikosta lähtötasosta niiden parissa.

Toteutuksen tulosten ja testien perusteella suositellaan ottamaan käyttöön useampi esitetyistä ratkaisuista, koska jokaisella niistä on omat heikkoutensa, oli kyse sitten suppeasta indikaattorien määrästä tai skriptin ajamisen pitkästä kestosta ajallisesti. On myös tärkeää ottaa huomioon, että kaikki ratkaisut ovat kolmannesta osapuolesta riippuvia, esimerkiksi abuse.ch-projektista on viime vuonna lakkautettu **ransomware tracker**, johon oli luotu esimerkki skripti PowerShellillä (Copple, 2019).

OTX-sivuston kanssa kannattaa myös perehtyä käyttäjiin ja näiden indikaattoreihin mitä seurata, koska signature-basen skripti on äärimmäisen hidas ja resurssisyöppö jo pelkästään kolmella käyttäjällä.

Esitettyihin toteutustapoihin on mahdollista ja kenties kannattavaakin kehitellä parannuksia, esimerkiksi signature-baseen saisi melko varmasti implementoitua toteutuksessa esitetyn erillisen python skriptin JSON-muotoa varten. Signature-basen skripti on kaksi vuotta vanha ja tehty Python 2.7:lle, eli senkin puolesta sen päivitys olisi paikallaan. Myös skripti TAXII-toteutuksen JSON-kääntäjään tarvitsee lisäkehitystä, koska nykymuodossaan se muuntaa yhden tiedoston kerrallaan ja senkin käyttäjän syötteestä.

Kaikki työssä käytetyt ohjelmat ja skriptit löytyvät osoitteesta https://github.com/jiikoojii/Thesis-insight. Kyseisestä osoitteesta löytyy myös linkit alkuperäisiin skripteihin ja ohjelmiin, joita on käytetty pohjana.

Lähteet

API - Mikä on API? N.d. Viitattu 17.5.2020. https://www.visma.fi/epasseli/kirjanpidon-sanakirja/a/api/

Carstensen, N. 2018. Why is log management important? Verkkojulkaisu. Viitattu 26.4.2020. https://www.graylog.org/post/why-is-log-management-important

Copple, T. 2019. Import External Threat Intelligence with the InsightIDR Threats API. Blogikirjoitus. Viitattu 7.9.2020. https://blog.rapid7.com/2019/10/16/import-external-threat-intelligence-with-the-insightidr-threats-api/

CSV-tiedosto – Mikä on CSV-tiedosto? N.d. Verkkoartikkeli. Viitattu 18.9.2020. https://www.visma.fi/epasseli/kirjanpidon-sanakirja/c/csv-tiedosto/

Fenomenologinen analyysi. N.d. Verkkoartikkeli. Viitattu 23.9.2020. https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/aineiston-analyysimenetelmat/fenomenologinen-analyysi

Gailey, S. 2019. A brief history of SIEM. Verkkojulkaisu. Viitattu 26.3.2020. https://techerati.com/features-hub/opinions/a-brief-history-of-siem/

Garg, R. N.d. Threats to information security. Viitattu 3.2.2020. https://www.geeksforgeeks.org/threats-to-information-security/

Gavin, B. 2018 What Is a Log File (and How Do I Open One)? Verkkojulkaisu. Viitattu 16.1.2020. https://www.howtogeek.com/359463/what-is-a-log-file/

Henderson, J. 2018. Why a SIEM won’t solve all your problems: 5 common issues and how to avoid them. Verkkoartikkeli. Viitattu 5.4.2020. https://red-canary.com/blog/common-siem-issues/

InsightIDR API (v1). N.d. Verkkoartikkeli. Viitattu 15.9.2020. https://help.rapid7.com/insightidr/en-us/api/v1/docs.html#operation/addIndicators

InsightIDR Overview. N.d. Verkkoartikkeli. Viitattu 16.5.2020. https://insightidr.help.rapid7.com/docs

InsightIDR Quick Start Quide. N.d. Verkkoartikkeli. Viitattu 16.5.2020. https://insightidr.help.rapid7.com/docs/quick-start-guide

InsightVM FAQ. N.d. Verkkoartikkeli. Viitattu 26.4.2020. https://www.rapid7.com/info/introducing-insightvm/faq/

Introducing JSON. N.d. Verkkoartikkeli. Viitattu 15.9.2020. https://www.json.org/json-en.html

Introduction to TAXII. 8.8.2020. Verkkoartikkeli. Viitattu 16.8. https://oasis-open.github.io/cti-documentation/taxii/intro.html

Keary, T. 2019. 8 Best SIEM Tools: A Guide to Security Information and Event Management. Verkkoartikkeli. Viitattu 10.1.2020. https://www.comparitech.com/net-admin/siem-tools/

Kent, K. & Souppaya, M. 2006. Guide to Computer Security Log Management. Verkkojulkaisu. Viitattu 26.3.2020. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf

Kinnunen, Y. 2017. SIEM-järjestelmä on organisaation kyberturvallisuuden hermokeskus. Viitattu 13.1.2020. https://www.insta.fi/ajankohtaista/siem-j%C3%A4rjestelm%C3%A4-on-organisaation-kyberturvallisuuden-hermokeskus

L 681/2010. Valtioneuvoston asetus tietoturvallisuudesta valtionhallinnossa. Viitattu 26.4.2020. https://www.finlex.fi/fi/laki/alkup/2010/20100681

Milecia, McG. 2019. The Importance of Log Files. Verkkojulkaisu. Viitattu 16.1.2020. https://dev.to/flippedcoding/the-importance-of-log-files-37d6

Miller, J. 2019. How can SIEM improve your organizations cyber security? Verkkoartikkeli. Viitattu 5.4.2020. https://www.bitlyft.com/how-siem-can-improve-your-organizations-cyber-security/

Muurinen, M. 2019. Tietosuoja vai tietoturva? Blogikirjoitus. Viitattu 3.2.2020. https://www.visma.fi/blog/tietosuoja-tietoturva/

Open Threat Exchange FAQ. N.d. Verkkoartikkelil. Viitattu 15.9.2020. https://otx.alienvault.com/faq

Petters, J. 2019. What is SIEM? A Beginner's Guide. Blogikirjoitus. Viitattu 13.1.2020. https://www.varonis.com/blog/what-is-siem/

Puolustusministeriö. PDF 2015. Katakri 2015. Viitattu 16.5.2020. https://www.defmin.fi/files/3165/Katakri_2015_Tietoturvallisuuden_auditointity-okalu_viranomaisille.pdf

Qvantel Company. N.d. Viitattu 10.1.2020. https://www.qvantel.com/company/

RFC 5246:2008. The Transport Layer Security (TLS) Protocol Version 1.2. Viitattu 16.9.2020. https://tools.ietf.org/html/rfc5246

Riippuvuussuhteiden analyysit. N.d. Verkkoartikkeli. Viitattu 23.9.2020. https://koppa.jyu.fi/avoimet/hum/menetelmapolkuja/menetelmapolku/aineiston-analyysimenetelmat/riippuvuussuhteiden-analyysit

Rouse, M. 2020. Confidentiality, integrity and availability. Verkkoartikkeli. Viitattu 9.5.2020. https://whatis.techtarget.com/definition/Confidentiality-integrity-and-availability-CIA

Security Event Management. 2015. Verkkoartikkeli. Viitattu 26.4.2020. https://www.techopedia.com/definition/25763/security-event-management

Security Information Management. 2015. Verkkoartikkeli. Viitattu 26.4.2020. https://www.techopedia.com/definition/4098/security-information-management

Threats. N.d. Verkkoartikkeli. Viitattu 17.5.2020. https://insightidr.help.rapid7.com/docs/threats

Use the Threat API. N.d. Verkkoartikkeli. Viitattu 17.5.2020. https://insightidr.help.rapid7.com/docs/use-the-threat-api

Welcome to InsightAppSec. N.d. Verkkoartikkeli. Viitattu 15.9.2020. https://docs.rapid7.com/insightappsec/

What is InsightConnect? N.d. Verkkoartikkeli. Viitattu 15.9.2020. https://docs.rapid7.com/insightconnect/

What is PowerShell? 22.5.2020. Verkkoartikkeli. Viitattu 7.9.2020. https://docs.microsoft.com/en-us/powershell/scripting/overview?view=powershell-7

What is SIEM? N.d. Verkkoartikkeli. Viitattu 16.5.2020. https://www.exabeam.com/siem-guide/what-is-siem/