Sampo Bredenberg

Uhkatietojen rikastus julkisten lähteiden tiedustelutiedoilla

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tutkinto-ohjelman nimi

Insinöörityö

31.12.2024

Tiivistelmä

Tekijä: Sampo Bredenberg

Otsikko: Uhkatietojen rikastus julkisten lähteiden tiedustelutiedoilla

Sivumäärä: xx sivua + x liitettä

Aika: 31.12.2024

Tutkinto: Insinööri (AMK)

Tutkinto-ohjelma: Tieto- ja viestintätekniikka

Ammatillinen pääaine: Ohjelmistotuotanto

Ohjaajat: Tehtävänimike Etunimi Sukunimi  
Tehtävänimike Etunimi Sukunimi

**Tiivistelmä**

Avainsanat: avainsana, avainsana

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tämän opinnäytetyön alkuperä on tarkastettu Turnitin Originality Check -ohjelmalla.

Abstract

Author: First Name Last Name

Title: Title of the Thesis

Number of Pages: xx pages + x appendices

Date: 1 March 2021

Degree: Bachelor of Engineering

Degree Programme: Name of the degree programme

Professional Major: Name of the professional major

Supervisors: First name Last name, Title (e.g., Project Manager)

First name Last name, Title (e.g., Principal Lecturer)

Tiivistelmän tekstiosuus kirjoitetaan niin, että se mahtuu sivulla käytössä olevaan tilaan. Tekstiosuudessa käytetään Leipäteksti ilman välistystä-tyyliä.

Keywords: Keyword, Keyword

Sisällys

Lyhenteet

[1 Johdanto 1](#_Toc68862316)

[2 Pääluvun otsikko 1](#_Toc68862317)

[2.1 Alaluvun otsikko 1](#_Toc68862318)

[2.2 Alaluvun otsikko 2](#_Toc68862319)

[3 Asiakirjan rakenteet 3](#_Toc68862320)

[3.1 Taulukot 3](#_Toc68862321)

[3.1.1 Taulukon tekeminen Wordissä 4](#_Toc68862322)

[3.1.2 Alaluvun alaotsikko 5](#_Toc68862323)

[3.2 Lainaukset 5](#_Toc68862324)

[3.3 Luetelmat 5](#_Toc68862325)

[3.4 Ohjelmakoodi 6](#_Toc68862326)

[3.5 Kaavat 6](#_Toc68862327)

[4 Kuvan ja taulukon vaihtoehtoinen teksti 7](#_Toc68862328)

[4.1 Vaihtoehtoisen tekstin lisääminen kuvalle 7](#_Toc68862329)

[4.2 Vaihtoehtoisen tekstin lisääminen taulukolle 8](#_Toc68862330)

[5 Saavutettava asiakirja 9](#_Toc68862331)

[5.1 Asiakirjan ominaisuuksien viimeistely 9](#_Toc68862332)

[5.2 Insinöörityön saavutettavuuden tarkistus 9](#_Toc68862333)

[5.3 Word-tiedoston tallentaminen saavutettavaksi PDF-tiedostoksi 10](#_Toc68862334)

[Lähteet 12](#_Toc68862335)

Liitteet

Liite 1: Liitteen nimi

Liite 2: Liitteen nimi

Lyhenteet

ORM: *Object-relational mapping*. Ohjelmiston oliomallin mukaisen esityksen kuvaaminen relaatiotietokantamallin mukaiseksi esitykseksi ja kääntäen.

TKHJ: Tietokannan hallintajärjestelmä. Ohjelmisto tiedon tehokkaan hakemisen, säilyttämisen ja päivittämisen toteuttamiseksi.

# Johdanto

Uhkatiedot ovat todisteita haitallisesta toiminnasta IT-ympäristöissä. Uhkatietoja keräämällä ja niitä tutkimalla pyritään paljastamaan aktiivinen väärinkäyttö, sekä väärinkäyttöyritykset.

* Aidon datan kerääminen rajoitettua, yleensä vain oikeista ympäristöistä dataa
* Työlästä tunnistaa suuria määriä haitallista toimintaa
* Hyötyä muiden tehdystä ja automatisoidusta työstä

# Lähtökohdat

Uhkatietojen rikastamisella tunnistetaan ja luokitellaan tietojenkäsittelyyn liittyviä tunnisteita. Tietoturvaloukkauksen tapahtuessa, hyökkääjän toiminnasta jää tunnisteita, kuten haittaohjelmatiedostoja ja yhteyksiä verkkotunnuksiin. Tunnisteita keräämällä voidaan löytää viitteitä tunnettuihin hyökkääjiin tai hyökkäystyyppeihin, ja näin nopeuttaa tietoturvaloukkaukseen reagointia sekä tarvittavia toimenpiteitä. Vuonna 2023 Red Canary-yritys, joka erikoistuu uhkien havaitsemiseen, havaitsi 37 miljoonaa mahdollisesti haitallista tapahtumaa asiakkaiden ympäristöissä joista 58 000 varmistettiin oikeiksi uhkiksi. Uhkatietojen keräykseen, niiden kategorisointi ja analysointiin tarvittava aika kasvaa yli käytettävissä olevien resurssien. Erilaiset organisaatiot, kaupalliset toimijat sekä yhteisöt suorittavat aktiivisesti uhkatietojen keräämistä ja kategorisointia. Näitä avoimia ja kaupallisia lähteitä käyttämällä uhkatietojen tunnistaminen helpottuu, kun ne voidaan yhdistää haitalliseen toimintaan aikaisempien tapahtumien tai analysoinnin perusteella. Eri lähteiden antamia tietoja yhdistämällä, saavutetaan kokonaisvaltaisempi johtopäätös mahdollisten tapahtumien arvioinnissa. **Jotain jotain yhdistämistä automatisoimalla / tarjoamalla alustan joka mahdollistaa tietojen nopean hakemisen, nopeutetaan toimintaa kun uhkatietoon yhdistetään suoraan arviointeja eri lähteistä.**

[RedCanary 2023 stats] <https://resource.redcanary.com/rs/003-YRU-314/images/2024ThreatDetectionReport_RedCanary.pdf?version=0> 17.11

## Uhkatiedot

Uhkatiedot ovat IT-ympäristöjen tunnisteita ja kaavamaisia tapoja, joita on käytetty tietoturvaloukkauksissa ja väärinkäyttöyrityksissä. **Tässä opparissa** rajataan uhkatiedot pelkästään tunnisteisiin, joita ovat:

* IPv4- ja IPv6-osoitteet
* Verkkotunnisteet (domain)
* Verkko-osoitteet (URL)
* Tiedostojen hajautusarvot (hash)

Tunnisteita voidaan havaita palomuuri- ja kirjautumislokeilta lähtö- tai kohdeosoitteina, sekä laitteen toiminnasta ja hakemistoista verkkoliikenteenä sekä tiedostoina. Uhkatiedot ovat ennestään tunnistettuja uhka toimijoiden käyttämiä haitallisia tiedostoja, sekä tietoverkko liikennettä uhka toimijoiden hallitsemiin osoitteisiin ja tunnisteisiin.

Haittaohjelmia käytetään saavuttamaan jokin tavoite. Ennen lopullista tavoitetta, haittaohjelma pyrkii usein tunkeutumaan syvemmälle ympäristöön, saamaan salasanoja haltuun ja korottamaan käyttöoikeuksia, kunnes tavoite on mahdollista toteuttaa. Tavoitteen edellytyksenä voi olla jokin tietty laite, verkkoympäristö tai saamaan tietyn tasoiset käyttöoikeudet.

Verkkotunnisteet hostaavat haitallisia tiedostoja tai muita resursseja, joiden kautta saadaan ensimmäinen pääsy ympäristöön. Ne voivat toimia myös paikkana, joihin voidaan luvattomasti viedä tietoaineistoa ympäristöstä, tai komentokeskuskina, joiden kautta voidaan komentaa haittaohjelmia toimimaan halutulla tavalla ja jopa saamaan ”hands-on” yhteys laitteelle, joka on saastutettu haittaohjelmalla.

### Uhkatietotyypit

IP-osoitteet toimivat tietoliikenteen perustana. Kun jotain halutaan lähettää, vastaanottaa tai hakea toiselta laitteelta, täytyy laitteen osoittaa mistä liikenne liikkuu mihin. IP-osoitteet toimivat lähtö- ja kohdeosoitteina, joiden avulla voidaan tunnistaa mikä laite tai verkko lähettää minne. IP-osoitteet voidaan jakaa neljään tyyppiin, julkisiin-, yksityisiin-, dynaamisiin- ja staattisiin IP-osoitteisiin. Uhkatietotyypeille oleelliset IP-osoite tyypit ovat julkiset IP-osoitteet. Julkiset IP-osoitteet ovat pysyviä, ainutlaatuisia ja ne toimivat tunnisteina julkisessa internetissä. Uhkatietojen kontekstissa IPv4 ja IPv6 eivät juurikaan eroa. IP-osoitteet yleensä aina osana tietoturvaloukkauksia, sillä kaikki verkkoliikenne tarvitsee aina IP-osoitteen.

[IP-osoite tyypit] <https://nordvpn.com/fi/blog/types-of-ip-addresses/> 17.11

Verkkotunnisteet tarvitsevat aina IP-osoitteen, jonka takana jotain palvelua hostataan. Verkkotunniste on usein ihmisymmärrettävä merkkijono, joka vastaa IP-osoitetta. Verkkotunnisteet toimivat hierarkialla, jossa tasot erotetaan pisteellä. Korkeinta tasoa hallitsee aina jokin organisaatio, joka jakaa toisen tason verkkotunnisteita organisaatioille sekä yksityishenkilöille. Uhkatiedot koskevat toisen tai jopa kolmannen tason verkkotunnisteita, joissa uhka toimija hostaa haitallista toimintaa. Usein koko toisen tason verkkotunniste osallistuu haitalliseen toimintaan.

Verkko-osoitteet ovat aina osana verkkotunnistetta. Verkko-osoitteilla tarkoitetaan tarkkaa osoitetta verkkotunnisteella, joka johtaa tiettyyn resurssiin. Verkko-osoitteessa voi olla mukana myös kysely parametrejä.

Tiedostojen hajautusarvot ovat matemaattisia laskentatuloksia tiedoston sisällöstä. Tiedostojen nimet ovat mielivaltaisia ja niitä pystytään muuttamaan ilman että tiedoston sisältöön tarvitsee tehdä muutoksia. Tiedostonimi ei tällöin toimi tunnisteena tiedostoille. Hajautusarvot ovat uniikkeja merkkijonoja, jotka tehdään erilaisten laskenta algoritmien avulla ja kuvaavat tiedoston sisältöä. Vertaamalla tiedoston hajautusarvoa voidaan varmistaa, onko tiedoston sisältö muuttunut, jolloin voidaan varmistaa, onko ennestään tunnettua tiedostoa muutettu esimerkiksi hyökkääjän toimesta. Tiedoston toiminta voidaan analysoida, sen haitallisuus arvioida ja tulokset julkaista tai tallentaa. Hajautusarvot pysyvät samana, jos tiedoston sisältö pysyy täysin saman, joten tallennettuja tuloksia voidaan käyttää arvioimaan tiedostoa, jos se huomataan uudestaan. Useat virustentorjuntaohjelmat käyttävät osana toimintaansa hajautusarvoja tunnistaakseen ennestään löydettyjä haittaohjelmia.

[Microsoft PowerShell Get-FileHash, hyvää dokumentaatiota mihin hasheja käytetään] <https://learn.microsoft.com/en-us/powershell/module/microsoft.powershell.utility/get-filehash?view=powershell-7.4> 16.11

### Uhkatietojen kerääminen

Uhkatiedot ovat aina reaktiivisia luonteeltaan. Uhkatietoja kerätään ja löydetään joko osana teknistä tutkimusta tapahtuneen tietoturvaloukkauksen jälkeen, tai todettujen väärinkäyttöyritysten perusteella tunnistamalla yritykseen osallistuneet tunnisteet.

Uhkatieto-tunnisteet tallentuvat lokitiedostoihin valvottavista ympäristöistä. Uhkatietoja voidaan löytää lokeilta monella tapaa.

Jäsennellyillä lokeilta, joilla on jokin standardi tiedostomuoto, tunnisteet kuten IP-osoitteet ovat usein aina tietyssä kohtaa, jolloin tunnisteet voidaan erottaa helposti lokeilta esimerkiksi ohjelmallisesti. Lokeilta, joilla ei ole struktuuria voidaan uhkatiedot erottaa käyttämällä säännöllisiä lausekkeita (regex).

Tekninen tutkimus seuraa tietoturvaloukkausta. Sen tarkoituksena on käydä systemaattisesti läpi digitaalista todistusaineistoa, kuten tietoliikennelokeja, järjestelmälokeja ja haittaohjelmia. Sen tavoitteena on usein jäljittää mitä tietoturvaloukkauksessa on tapahtunut, mitä hyökkääjä on saanut aikaan ja miten laajalle toiminta on ulottunut. Tutkimuksen sivutuotteena on uhkatietoja, mitä tietoturvaloukkauksessa on käytetty.

Kehittyneemmissä ympäristöissä, lokeja viedään usein SIEM-alustoille, jotka toimivat keskitettyinä lokien hallinta alustoina. SIEM-alusta mahdollistaa tietoturva tapahtumien keskitetyn monitoroinnin, analysoinnin ja hallinnan. Alustalle viedään tietoturvatapahtumien valvontaan tarvittavia lokeja, kuten palomuuri-, käyttäjienhallinta alusta- ja laite tapahtumalokia. Useat SIEM-alustat jäsentelevät lokit automaattisesti sellaiseen muotoon, joka mahdollistaa tehokkaan uhkatietojen keräämisen sekä niiden analysoinnin.

Ei kuulu scopeen, nopea maininta? Yksi tapa kerätä uhkatietoja, on jättää ympäristöön näkyvälle laite, joka houkuttelee hyökkääjiä yrittämään tunkeutua siihen sisään. Näitä laitteita kutsutaan hunajapurkeiksi. Niiden tarkoituksena on esittäytyä helppona kohteena, jonka annetaan tahallaan joutua väärinkäytön uhriksi, mutta jotka ovat eristettynä muusta verkosta ja laitteen toimintoja tallennetaan ja kerätään ylös. Hyökkääjän toimintoja voidaan seurata ja saadaan tietoja aitojen uhkien käyttämistä tekniikoista ja tunnisteista.

[What is a honeypot] <https://www.fortinet.com/resources/cyberglossary/what-is-honeypot> 19.11

### Uhkatietojen käyttö

En koske miten tunnisteita nostetaan ylös. Ympäristöissä on normaalia toimintaa, johon liittyy paljon verkkoliikennettä. Vertaamalla tunnettuja uhkatietoja ympäristön toiminnassa esiin tuleviin tunnisteisiin, voidaan aktiiviset tietoturvaloukkaukset ja hyökkäys yritykset tunnistaa ajoissa. Nostot mahdollisista väärinkäyttöyrityksistä vaativat yleensä manuaalista analysointia, jossa lokien pohjalta pyritään tunnistamaan haitallinen toiminta ja löytöjen perusteella tekemään vastatoimenpiteitä. Analysoinnissa käsitellään usein lokeja, jotka sisältävät useita erilaisia tunnisteita, jolloin niiden vertaaminen uhkatietoihin nopeuttaa haitallisen toiminnan löytämistä sekä tunnistamista.

## Julkisiin lähteisiin perustuva tiedustelutieto

Koska uhkatietojen kerääminen on reaktiivista, tunnisteet usein lyhytkestoisia **(kappale/osaksi uhkatietotyyppejä miksi?)** ja uhka tekijöitä on useita, yksittäisen tahon on hankala pitää ajantasaista uhkatieto tietokantaa yksin pystyssä. Tietoturva yhteisössä toimii monia tahoja, joiden tavoitteena on kerätä sekä jakaa uhkatietoja vapaasti tai kaupallisesti käyttöön, tavoitteena lieventää tietoturvaloukkausten vaikutusta, havaita aktiiviset hyökkäykset nopeammin ja ennaltaehkäistä mahdollisia hyökkäyksiä. Näiden julkisten lähteiden uhkatietojen käyttäminen mahdollistaa ajantasaisen uhkien seurannan. Vertaamalla monia lähteitä keskenään saadaan tarkka kuva uhkatietojen todenmukaisuuksista ja nopeutetaan analysointia, kun kategorisointi on jo osakseen tehty.

Monet alustat tarjoavat myös API ratkaisun alustan ominaisuuksien, kuten uhkatietojen hakemisen käyttöön. Tyypillisesti vapaasti käytettävä, ilmainen API rajoittaa hakujen määrän, kun taas kaupallinen käyttö, jossa rajoitukset ovat laajemmat, edellyttää maksullisen API:n käyttöä.

### Avoimet toimijat

Avoimet toimijat julkaisevat sekä mahdollistavat uhkatietojen hakemisen alustoiltaan vapaasti ilman maksua.

VirusTotal on yksi julkisiin lähteisiin perustuvien tiedustelutietojen johtaja alustoista. Se tarjoaa yhden ominaisuus rikkaimmista kokemuksista ja sen toiminnallisuus on integroitu moneen tuotteeseen. VirusTotal kokoaa monien toimijoiden tietoturva-tuotteet yhteen paikkaan, ja mahdollistaa kenen tahansa syöttää tiedosto, verkkosivusto tai IP-osoite arviointiin. VirusTotal skannaa syötteen tietoturva-tuotteilla ja tallentaa tulokset vapaasti saataville. Havainnot eri toimijoiden tuotteista antaa kattavan näkymän tunnisteen toiminnasta. **jatka…**

### Kaupalliset toimijat

GreyNoise

Mandiant

### Joukkoistettu tiedustelutieto

Moni alusta mahdollistaa yksittäisten käyttäjien antamaan syötteensä uhkatietoihin. Satunnaisen käyttäjän sanaan tunnisteen haitallisesta toiminnasta ei ole luotettavaa, mutta kymmenien käyttäjien erillinen ilmoitus tunnisteen haitallisesta toiminnasta antaa suuntaa, jota verrata omiin löydöksiin.

AbuseIPDB tarjoaa alustan, jossa rekisteröitynyt ja varmistettu käyttäjä pystyy ilmoittamaan IP-osoitteita haitallisesta toiminnasta. IPv4- tai IPv6-osoitteen lisäksi täytyy väärinkäyttö kategorisoida ainakin yhteen valmiista kategorioista ja vapaa muotoinen kommentti liittää ilmoitukseen. Yhteisön ilmoitukset kerätään IP-osoitteen sivulle, jossa yksittäiset ilmoitukset ja niiden tiedot näkyvät. Tämän lisäksi alusta laskee "Confidence of Abuse" arvon IP:lle, joka kuvaa kuinka varmasti IP-osoite suorittaa väärinkäyttöä asteikolla 0–100.

[AbuseIPDB FAQ] <http://www.abuseipdb.com/faq.html> 17.11

VirusTotal mahdollistaa rekisteröityneiden käyttäjien äänestää (-1 tai +1) uhkatietoja, sekä jättää vapaan kommentin uhkatiedolle.

OpenThreatExchange

## Uhkatietojen analysointi

Uhkatiedot luokitellaan usein kolmeen kategoriaan:

* Benign (hyvänlaatuinen)
* Suspicious (epäilyttävät)
* Malicious (haitallinen)

**Kategoriat auki**

Mahdollisten haitallisten tapahtumien noustessa, täytyy tapahtumat analysoida ja varmistaa mitä ympäristössä on käynyt. Analyysin tuloksena voidaan varmistaa toiminta asialliseksi tai haitalliseksi ja aloittaa asiaan kuuluva häiriötilanteisiin reagointi.

Haitallisen toiminnan varmistamiseen kuuluu usein uhkatietoja. Verkkoyhteyksiä tuntemattomasta IP-osoitteesta, käyntejä erikoisilla verkkotunnuksilla tai epätavallisia tiedostoja. Uhkatietoja voi yhteen tapahtumaan liittyä monia, koska laitteiden normaaliin toimintaan kuuluu useita erilaisia tiedostoja ja verkkoyhteyksiä. Kaikkien tunnisteiden läpikäyminen vie aikaa. Analyytikon tavoitteena on löytää oleelliset uhkatiedot, joiden seuraaminen johdattaa haitallisten toimintojen tutkimiseen.

### Työkalut

# Teknologiat

Työssä käytetyt teknologiat.

## Python

Työn ohjelmointi kielenä käytetään python 3.11. Python tarjoaa laajan valikoiman kehittyneitä kirjastoja, joiden käyttäminen helpottaa kehittämistä. **töissä käytetään, tuttu kieli, helppo kirjoittaa ja ylläpitää, sopii yhteen käyttötapauksen kanssa ja muihin käytettäviin teknologioihin.**

### Flask

Flask on web-applikaatio framework kirjasto, joka mahdollistaa yksinkertaisten http-pohjaisten applikaatioiden rakentamisen. Flaskia käytetään työssä määrittämään API-rajapinnat, joita kutsumalla sovelluksen toimintaa määritetään.

Flask ”route” dekoraattoreita käyttämällä, rekisteröidään URL osoitteet, jotka käynnistävät sovelluksen eri toimintoja. Tekemällä kutsuja URL endpointteihin, sovellus käynnistää background taskeja ja palauttaa formatoituja vastauksia http-protokollalla.

<https://flask.palletsprojects.com/en/stable/>

### Gunicorn

Reliable web-server. Tarjoaa alustan web-sovellukselle.

### Celery

Background workers.

Tasks are given to workers through a broker (redis). Workers take tasks and complete them.

### Asyncio

Asynchronous. Async/await, needed for web applications.

# 