Mette Kataja

CYBER THREAT INTELLIGENCE

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2019

TIIVISTELMÄ

Kataja, Mette
Cyber Threat Intelligence
Jyväskylä: Jyväskylän yliopisto, 2019, 28 s.
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja: Kollanus, Sami

Maailman digitalisoitumisen seurauksena kerätyn datan määrä on kasvanut eksponentiaalisesti, tietojärjestelmät ovat monimutkaistuneet ja kyberhyökkäykset kehittyneet teknologian kehityksen mukana, jonka vuoksi perinteiset kyberhyökkäyksen torjuntamenetelmät eivät enää yksinään riitä suojaamaan organisaatioita.

Tutkielmassa tarkastellaan miten Cyber Threat Intelligence edesauttaa organisaatioiden tietoturvaa ja mitä haasteita se tuo. Tutkielma on toteutettu kirjallisuuskatsauksena ja se pyrkii vastaamaan tutkimuskysymykseen: *"Mitä on Cyber Threat Intelligence, miten se vaikuttaa ja mitä haasteita se tuo?"*.

Kehittyneiden kyberhyökkäyksien takia yleistä toimintatapaa täytyisi siirtää hyökkäyksistä aiheutuneiden vahinkojen jälkikäteen korjaamisesta hyökkäyksien ennakoimiseen. Hyökkäyksien ennakoimiseen yksi tapa on hyödyntää Big Dataa, sillä kerätyn datan seasta löytyy vastauksia useampiinkin ongelmiin, mutta datan valtavan määrän ja monimutkaisuuden vuoksi, sitä on hidasta kerätä ja prosessoida sekä vaikea hallita.

Uusi nouseva trendi on Cyber Threat Intelligence, jonka tarkoituksena on hyödyntää erilaisista lähteistä saatua analysoitua informaatiota kyberuhista ja sen perusteella tehdä päätöksiä tietoturvan parantamiseksi.

Asiasanat: Big Data, tietoturva, Cyber Threat Intelligence, Threat Intelligence

ABSTRACT

Kataja, Mette
Cyber Threat Intelligence
Jyväskylä: University of Jyväskylä, 2019, 28 pp.
Information Systems Science, bachelor's thesis
Supervisor: Kollanus, Sami

Due the world's digitalization, the volume of data has grown exponentially, information systems are more complicated and cyber threats are more advanced than ever. These have led to that traditional cyberthreat prevention methods are no longer as effective as they used to be in securing organizations' information security.

The research is made as a literature review and its attempts to answer question: *"What is Cyber Threat Intelligence, how does it affect and what challenges does it bring?"*.

Because of the advanced cyberattacks organizations should change their methods from managing damages to preventing attacks. One way to help prevent attacks is to use data that has been already collected and most likely contains the answers for multiple questions, however because of the volume and the complexity of data, it cannot be collected, processed as quickly as needed and it's hard to manage.

New rising trend is Cyber Threat Intelligence, which purpose is to exploit analyzed data about cyberattacks, which is gathered from different sources, and with the help of the analyzed data to make informed decisions for developing one's information security.

Keywords: Big Data, Information Security, Cyber Threat Intelligence, Threat Intelligence

SISÄLLYS

1 Johdanto

Viime vuosien aikana kyberhyökkäykset ovat lisääntyneet ja monimutkaistuneet, jonka vuoksi niiden havaitsemiseen, analysointiin ja suojaukseen tarvitaan yhä enemmän reaaliaikaista tietoa (Conti, Dargahi & Dehghantanha, 2018). Samaan aikaan kerätyn tiedon määrä jatkaa kasvamistaan. Osaa tiedosta on vaikea luokitella, jonka vuoksi se jää osittain tai kokonaan jäsentymättömäksi, jolloin myös sen analysointi ja käyttöönottaminen on hidasta. (Chen, Mao & Liu, 2014; Khan, Yaqoob, Hashem, Inayat, Ali, Kamaleldin, Alam, Shiraz & Gani, 2014; Tien, 2013.) Tämän vuoksi tarvitaan uusia tekniikoita ja menetelmiä, joista yksi ratkaisu on vapaaehtoiseen datan jakamiseen ja keräämiseen perustuva Cyber Threat Intelligence. Cyber Threat Intelligence käsitteelle ei löydy vakiintunutta suomalaista käsitettä, joten tässä tutkielmassa siihen viitataan tästä lähin lyhenteellä CTI.

CTI:n tarkoituksena on hyödyntää erilaisista lähteistä saatua analysoitua informaatiota kyberuhista ja sen perusteella tehdä päätöksiä tietoturvan parantamiseksi. Sen keskeinen tarkoitus on saada organisaatiot vapaaehtoisesti jakamaan itse keräämiään tietojaan ja vaihtoehtoisesti saada uutta tietoa muilta organisaatioilta sekä kolmansilta osapuolilta.

Tämä kandidaatintutkielma tutkii kirjallisuuskatsauksen avulla, miten CTI vaikuttaa organisaation tietoturvaan ja mitä haasteita se tuo. Tutkimuksen tavoitteena on saada vastaus tutkimuskysymykseen *”Mitä on Cyber Threat Intelligence, miten se vaikuttaa ja mitä haasteita se tuo?”*. CTI on uusi ilmiö, jonka vuoksi siitä löytyy vielä vain rajallisesti tietoa. Tämän tutkimuksen praktinen tarkoitus oli siis täyttää havaittu tutkimusaukko CTI:n tietämyksessä. Tutkielman keskeiset käsitteet ovat Big Data, tietoturva sekä Cyber Threat Intelligence.

Aineisto on hankittu pääosin käyttämällä Google Scholar -hakupalvelua ja aikaisempien tutkimusten lähdeluetteloita sekä niiden tukena julkaisufoorumia julkaisujen laatutasojen varmistamiseksi. Hakusanoina pääosin toimi ”big data”, ”information system security”, ”information system security policy”, ”Cyber Threat Intelligence” ja ”Threat Intelligence”.

Tutkielma koostuu johdannosta, kolmesta sisältöluvusta ja yhteenvedosta. Toisessa luvussa käsitellään tietoturvallisuutta varsinkin organisaatiotasolla. Kolmas luku käsittelee Big Dataa, sen haasteita ja miten se vaikuttaa kyberturvallisuuteen. Neljännessä luvussa käsitellään CTI, sen määritelmää ja ominaisuuksia, mitä hyötyä siitä on organisaatioille ja minkälaisia haasteita se tuo. Lopuksi viidennessä luvussa esitetään tutkielman yhteenveto ja pohditaan sen vaikutusta yhteiskuntaan.

2 Tietoturvallisuus organisaatiossa

Viime vuosien aikana kyberhyökkäykset ovat lisääntyneet ja monimutkaistuneet, jonka vuoksi niiden havaitsemiseen, analysointiin ja suojaukseen tarvitaan yhä enemmän reaaliaikaista tietoa. (Conti ym., 2018; Wheelus, Bou-Harb & Zhu, 2016.)

2.1 Tietoturvallisuus vs. kyberturvallisuus

Tietoturvallisuus tarkoittaa tiedon suojaamista sen luomisen, prosessin, tallentamisen, transformaation ja poistamisen aikana. Käyttämällä apuna loogisia, teknillisiä, fyysisiä ja organisaatiollisia toimenpiteitä, jotka estävät tiedon luotettavuuden, eheyden ja saatavuuden vähenemistä (Ključnikov, Mura & Sklenár, 2019). Tietoturvallisuuden on siis tarkoitus estää ja/tai minimoida hyökkäyksestä aiheutunut vahingot koko tiedon elinkaaren aikana.

Tiedon luotettavuudella tarkoitetaan tiedon jakamisen estämistä luvattomille henkilöille tai järjestelmille (Jang-Jaccard & Nepal, 2014), kuten jonkun nettisivun käyttäjien salasanojen vuotaminen julkisuuteen. Eheydellä pyritään tiedon muokkaamisen tai poistamisen estämiseen luvattoman tahon toimesta, kun taas saatavuudella pyritään varmistamaan, että järjestelmät ottavat vastuun tiedon toimittamisesta, tallentamisesta ja prosessoinnista saataville silloin kun organisaatio tarvitsee kyseistä tietoa (Jang-Jaccard & Nepal, 2014).

Kyberturvallisuus ja tietoturvallisuus määritellään hyvin usein samaksi asiaksi, mutta todellisuudessa tietoturvallisuus suojelee niin fyysistä kuin digitaalista aineistoa, kun taas kyberturvallisuuden tarkoitus on estää juuri verkon välityksellä toimivat hyökkäykset eli kyberhyökkäykset. (Von Solms & Van Niekerk, 2013; Safa, Von Solms & Furnell, 2016.) Kyberturvallisuus on balanssi hyökkääjän ja puolustajan välillä (Schneier, 2012). Hyökkääjä on yleensä voitolla, sillä hänen tarvitsee löytää vain yksi heikko kohta järjestelmässä, kun taas puolustajan pitää puolustaa kaikkia kohtia.

Kyberhyökkäykset ovat lisääntyneet huomattavasti ja niiden odotetaan lisääntyvän, sillä ne ovat halvempia, kätevämpiä sekä vähemmän riskialttiita

(Jang-Jaccard & Nepal, 2014) kuin esimerkiksi perinteinen pankin ryöstö aseen kanssa. Hyvin usein kyberhyökkäykseen tarvitsee vain tietokoneen, toimivan nettiyhteyden ja hieman taitoa. Tietokoneen ansiosta hyökkäyksen voi tehdä mistä päin maailmaa tahansa sekä internetin anonyymisyyden tuomasta identiteettisuojasta johtuen kiinnijäämisriski on pienempi, mikäli osaa peitellä jälkensä oikein. Suosittuja kyberhyökkäyksien toimintatapoja ovat esimerkiksi virukset, tietojenkalastelu, roskapostien jatkuva lähettäminen sekä kohdistettu hyökkäys.

2.2 Tietoturvallisuus organisaatiossa

Tietoturvallisuuden johtaminen on työkalujen ja toimenpiteiden hallitsemista jouhevan tietoturvallisuuden varmistamiseksi. Nykyaikana se on osa koko organisaation johtamisjärjestelmää ja sen tavoitteena on asettaa, operoida, monitoroida, tarkastaa, ylläpitää ja kehittää organisaation tietoturvallisuutta. (Ključnikov ym., 2019.)

Digitalisoitumisen myötä myös organisaatiot ovat hyvin riippuvaisia teknologiasta, jolloin myös teknologiaan mukana tulevat riskit aiheuttavat monia haasteita, sillä riskien toteutumisella voi olla suoria seurauksia organisaatiolle (Bulgurcu, Cavusoglu & Benbasat, 2010; Safa ym., 2016), kuten maineen menettäminen tai hetkellinen vahinko liiketoiminnassa.

Tietojärjestelmien turvaamiseksi mm. hyväksi- ja väärinkäytöltä sekä tuhoamiselta organisaatiot käyttävät hyödyksi erilaisia työkaluja ja toimenpiteitä (Ifinedo, 2012; Safa ym., 2016), kuten palomuureja, viruksentorjuntaohjelmia, varmuuskopiointeja, hyökkäyksien havannointiohjelmia ja salattujen avainten käyttämistä.

Tietoturvauhka voi tulla joko organisaation sisä- tai ulkopuolelta. Ulkopuolelta tulevia hyökkäyksiä voivat olla mm. luonnonkatastrofit tai kyberhyökkäykset. Kyberhyökkäyksiä on monenlaisia, mutta yleistymässä ovat varsinkin keihästetty (spear) tietojenkalastelu, sivustoharhautus (Brockett, Golden & Wolman, 2012) sekä kohdistettu hyökkäys (Advanced Persistent Threat) (Hudson, 2014; Rot & Olszewski, 2017; Tounsi & Rais, 2018).

Tietojenkalastelussa työntekijälle lähetetään sähköposti, jossa ilmoitetaan esimerkiksi arvonnan voitosta ja pyydetään täyttämään viestin sisältämän linkin kautta henkilökohtaisia tietoja, kuten pankkitunnuksen tiedot. Todellisuudessa työntekijä ei ole voittanut mitään vaan tiedot ovat menneet kyberhyökkääjälle, joka käyttää niitä omiin tarkoituksiinsa, esimerkiksi identiteettivarkauteen. Keihästetyssä tietojenkalastelussa kyberhyökkääjä on valmiiksi tutkinut kohteensa tietoja, jolloin hän pystyy paremmin kohdentamaan huijausviestiään uskottavammaksi (Brockett ym., 2012).

Huijaussivustoissa tekijät ovat kaapanneet jonkun virallisen organisaation nettisivujen ulkoasun ja tehneet sen perusteella uuden sivuston, joka voi sisältää esimerkiksi haittaohjelmia (Brockett ym., 2012). Näin alkuperäisen sivuston asiakkaat vierailevat huijaussivustolla ja tietämättään lataavat haittaohjelmia omille koneilleen tai täyttävät henkilökohtaisia tietoja tekijöiden

käytettäväksi. Pahimmassa tapauksessa tekijät saavat siis useamman henkilön tiedot olematta kuitenkaan kontaktissa kyseisiin henkilöihin (Brockett ym., 2012).

Kohdistettu hyökkäys on kyberuhka, jossa organisoituneet ja taloudellisesti tuetut järjestöt pyrkivät hyökkäämään valmiiksi valittujen kohteiden järjestelmiin ja pysymään sisällä mahdollisimman pitkän ajan, jonka aikana he keräävät tai poistavat tietoa (Hudson, 2014; Rot & Olszewski, 2017; Tounsi & Rais, 2018). Kohdistetut hyökkäykset ovat kehittyneitä (advanced), sillä ne käyttävät uusimpia teknologioita sekä tekijät ovat kouluttautuneita IT-osaajia ja osana suurempaa joukkoa (Rot & Olszewski, 2017). Kohdistettuja hyökkäyksiä kutsutaan kohdistuneiksi (Hudson, 2014), sillä uhri valitaan tarkoin ja heitä tutkitaan ennen varsinaista hyökkäystä. Ne eroavat myös muista hyökkäyksistä siten, että ne hyödyntävät monia eri tekniikoita (Hudson, 2014), esimerkiksi sähköpostia tietojenkalastelun toivossa, jolloin virustorjuntaohjelma ei pysty estämään hyökkäystä.

Riskien vähentämiseksi organisaatiot hyvin usein investoivat vain teknologisiin ratkaisuihin, kun todellisuudessa paremman tietoturvallisuuden voi saavuttaa investoimalla niin teknologisiin kuin ei-teknologisiin resursseihin (Bulgurcu ym., 2010; Ifinedo, 2012), kuten henkilöstön kouluttamiseen. Sillä hyvin harvoin hyökkäys on suoritettu vain teknologian kautta, vaan hyökkääjä on käyttänyt hyödyksi ihmistä kuten organisaation työntekijää tai hyökkääjänä on toiminut työntekijä.

Usein sanotaankin, että työntekijät ovat organisaation heikoin lenkki (Brockett ym., 2012; Bulgurcu ym., 2010) kun puhutaan tietoturvallisuudesta. Työntekijöitä pidetään tietoturvariskinä heidän tietämättömyytensä, virheidensä sekä tahallisten tekojensa takia. Myös IT-resurssien hyväksi- ja väärinkäyttö on tutkimusten mukaan todettu olevan iso ongelma työntekijöiden taholta. (Bulgurcu ym., 2010.) Työntekijä voi siis esimerkiksi tietämättömyyttään langeta tietojenkalasteluyritykseen tai tahallaan varastaa arvokasta tietoa ja myydä tai jakaa sitä ulkopuolisille henkilöille.

Työntekijät ovat organisaation suurin tietoturvauhka, joten minimoidakseen riskejä organisaatiot hyvin usein luovat koko organisaation kattavan tietoturvapolitiikan.

Tietoturvapolitiikka sisältää ohjeita siitä, miten tietoturvallisuutta pidetään yllä samalla kun työntekijät käyttävät järjestelmiä työnsä puolesta (Bulgurcu ym., 2010). Se siis määrittelee säännöt mihin tarkoitukseen ja miten työntekijät saavat käyttää organisaation teknologiaresursseja. Tietoturvapolitiikan pitäisi olla selkeästi samalla linjalla organisaation tavoitteiden kanssa sekä tukea, sitouttaa ja informoida koko organisaation henkilöstöä (Ključnikov ym., 2019).

Tietoturvapolitiikka myös informoi työntekijöitä mahdollisista riskeistä (Safa ym., 2016). Joten jos tietoturvapolitiikka on laadittu hyvin, myös työntekijöiden tietoturvatietoisuus paranee. Parempi tietoturvatietoisuus vaikuttaa työntekijän käytökseen ja sitä kautta mahdollisesti minimoi organisaation tietoturvariskejä.

3 Big Data

Viimeisen parinkymmenen vuoden aikana datan määrä on lisääntynyt huomattavasti usealla eri alalla (Abu, Selamat, Ariffin & Yusof, 2018; Chen ym., 2014; Ernst & Young, 2014). Tämä on tuonut niin lisää hyötyä kuin myös uusia haasteita eri toimijoille.

3.1 Big Data käsitteenä

Big Data on abstrakti käsite, jonka määrittelystä tutkijat eivät ole päässeet vielä yhtenäiseen päätökseen (Chen ym., 2014) mitä se oikein sisältää. Yleisin määrittely Big Datasta on aineisto, mitä ei voida kerätä, hallita ja prosessoida tavallisilla tietokoneilla tarpeeksi nopeasti (Chen ym., 2014; Kaisler ym., 2013; Tien, 2013) ja tämän lisäksi McKinsey & Company määritteli Big Datan olevan tulevaisuuden innovaation, kilpailun ja tehokkuuden edelläkävijä (Chen ym., 2014).

Edellisestä määritelmästä poiketen International Data Company määrittelee Big Datan olevan Big Data -teknologiaa, jonka tarkoituksena on poimia arvokasta tietoa suurista ja erilaisista data-aineistoista hyödyntämällä nopeaa datan keräämistä, löytämistä ja analysointia (Chen ym., 2014).

Lisäksi National Institute of Standards and Technology määrittelee Big Datan olevan dataa, jonka määrä, keräämisnopeus ja/tai datan esiintyminen rajoittavat perinteisen prosessointimetodien kykyjä suorittaa tehokasta analyysia. Vaihtoehtoisesti Big Data on dataa, joka voidaan tehokkaasti prosessoida käyttämällä tärkeitä horisontaalisia tarkennusteknologioita. (Chen ym., 2014.) Mutta yksinkertaisesti Big Data voidaan määritellä valtavaksi määräksi erimuotoista dataa.

Toinen tapa määritellä Big Data on ilmaista sen perusominaisuudet eli niin sanotut ”3V’s”, jotka ovat määrä (Volume), vauhti (Velocity) ja moninaisuus (Variety) (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Kaisler ym., 2013; Katal, Wazid & Goudar, 2013; Khan ym., 2014; Mtsweni, Mutemwa & Mkhonto, 2016; Sagiroglu & Sinanc, 2013). Myöhemmin mukaan tuli kahdesta kolmeen uutta ominaisuutta, joiden käyttäminen riippuu

määrittelystä. Ne ovat totuudenmukaisuus (Veracity) (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Kaisler ym., 2013; Katal ym., 2013), arvo (Value) (Alguliyev & Imamverdiyev, 2014; Kaisler ym., 2013; Katal ym., 2013) sekä monimutkaisuus (Katal ym., 2013; Kaisler ym., 2013). Määrällä tarkoitetaan valtaisaa datan määrää, jota ei voida prosessoida perinteisin keinoin. Vauhti taas viittaa datan keräysnopeuteen ja kuinka sitä pitää myös prosessoida mahdollisimman nopeasti. Moninaisuus sisältää kerätyn datan tyypit: data voi olla jäsenneltyä, osittain jäsennelty tai jäsentymätöntä, kuten tekstejä, videoita ja äänitallenteita sekä monimutkaisuus viittaa ongelmaan yrittää yhdistellä, poistaa ja muokata erilaista ja eri tyyppistä dataa.

3.2 Hyödyt ja haasteet

Viime aikoina eri organisaatiot ovat kiinnostuneet Big Datan potentiaalista ja siitä, mitä lisäarvoa se voisi tuoda heille. Sosiaalisen median, multimedian ja esineiden internetin (Internet of Things) yleistymisen ansiosta organisaatiot pystyvät keräämään yhä enemmän dataa (Chen ym., 2014) niin asiakkaistaan kuin myös muista alan organisaatioista. Parhaimmillaan Big Data tuo mahdollisuuden löytää uutta tietoa jo kerätystä tiedosta sekä auttaa ymmärtämään sitä syvällisemmin. Se voi myös parantaa organisaation tehokkuutta ja kilpailukykyä (Chen ym., 2014; Sagiroglu & Sinanc, 2013; Tien, 2013), sillä hyvin usein jo kerätystä datasta voi prosessoida informaatiota, jota organisaatio voi hyödyntää oman liiketoimintansa parantamisessa.

Big Datan nousu on tuonut mukanaan myös haasteita ja ongelmia, kuten miten niin valtaisaa määrää dataa voidaan kerätä, organisoida, hallita ja analysoida riittävän nopealla tahdilla (Chen ym., 2014; Kaisler ym., 2013; Katal ym., 2013; Khan ym., 2014; Mtsweni ym., 2016; Sagiroglu & Sinanc, 2013). Liian hitaasti saatu informaatio voi vanhentua ennen kuin sitä olisi voinut hyödyntää.

Kriittisin ongelma Big Datan käsittelyssä on se, miten löytää piilossa olevaa arvokasta tietoa niin valtavan ja moninaisen määrän sisältä, joka lisääntyy hetki hetkeltä (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Kaisler ym., 2013; Khan ym., 2014), sekä miten muokata se sellaiseen muotoon, jota organisaatiot voivat käyttää hyödykseen (Katal ym., 2013). Vaikka tutkijat ovat kehittäneet useitakin tekniikoita ja työkaluja tätä varten, datan nopea prosessointi on vielä tänä päivänä haaste.

Saman datan toistumisen vähentäminen ja tietomäärän supistaminen on myös osoittautunut ongelmaksi (Chen ym., 2014). Tehokas toistuvuuden poistaminen ja sen kautta tiedon supistuminen vähentäisi välillisesti koko järjestelmän kustannuksia ilman, että datan arvo vähentyisi.

Tiedon tallennusjärjestelmien hidas kehittyminen ja datamäärän nopea kasvu aiheuttavat ongelmia datan elinkaaren hallinnassa, sillä nykyajan järjestelmät eivät kykene hallitsemaan valtaisaa datan määrää (Chen ym., 2014; Kaisler ym., 2013; Katal ym., 2013; Khan ym., 2014; Sagiroglu & Sinanc, 2013; Wheelus ym., 2016). Sen lisäksi valtavan datamäärän siirtäminen

tallennuspaikasta toiseen (Kaisler ym., 2013; Katal ym., 2013; Sagiroglu & Sinanc, 2013) sekä prosessointi (Kaisler ym., 2013) vie liian paljon aikaa. Datamäärän hallitsemiseksi on ehdotettu ratkaisua, jossa valittaisiin mitä tietoa tallennetaan ja mitä ei (Chen ym., 2014; Katal ym., 2013; Mtsweni ym., 2016). Tämä luo kuitenkin ongelman siitä, millä perusteilla valittaisiin, että mikä tieto ei ole tärkeää (Katal ym., 2013) sekä mistä voidaan tietää, etteikö kyseinen tieto olisi tärkeää tulevaisuudessa.

Tulevaisuuden analyysijärjestelmien pitäisi pystyä laajentumaan ja skaalautumaan datamäärän lisääntyessä, jonka järjestelmien hidas kehittyminen on johtanut pilvipalveluihin siirtymiseen. mutta pilvipalvelut yhdistelevät analysoitavaa aineistoa isommiksi kasoiksi kuin työmäärä oikeasti tarvitsisi. Isommat kasat taas voivat johtaa järjestelmän kaatumiseen. (Katal ym., 2013; Khan ym., 2014; Sagiroglu & Sinanc, 2013; Wheelus ym., 2016.)

Datan muuttaminen visuaaliseen muotoon tekee datasta entistä merkittävämpää ihmisille ja analysoivalle tietojärjestelmälle, mutta jos dataa esitetään väärin, se alentaa alkuperäisen datan arvoa ja saattaa jopa häiritä tehokasta datan analysointia. (Chen ym., 2014; Wheelus ym., 2016.)

Myös järjestelmien hinta (Katal ym., 2013; Sagiroglu & Sinanc, 2013) sekä energian kulutuksen (Katal ym., 2013) todetaan olevan ongelmana, kun puhutaan Big Datasta organisaatiotasolla. Järjestelmien monimutkaistuessa ja datamäärän kasvaessa organisaatiot joutuvat vuosi vuodelta investoimaan yhä enemmän tarvittaviin työkaluihin ja ohjelmistoihin, jotta he pystyisivät hallitsemaan valtavaa aineistoa. Tämä myös johtaa suurempaan energiankulutukseen.

Valtaisa datamäärä aiheuttaa ongelmia sen hallinnassa (Kaisler ym., 2013; Sagiroglu & Sinanc, 2013). Niin suurta määrää on vaikea validoida, jolloin sekaan jää myös huonoa tai merkityksetöntä tietoa. Kerätty tieto myöskin harvoin sisältää tarvittavaa metadataa kuten milloin ja mistä se on kerätty, jolloin on vaikeaa myöskään poistaa sitä, kun ei tiedetä voiko se olla arvokasta.

Jäsentymättömän datan analysointi ja prosessointi on hankalaa ja kallista eikä sen muuttaminen jäsennetyksi dataksi ole mahdollista (Katal ym., 2013). Jos kaikki kerätty data olisi jäsenneltyä myös järjestelmävaatimukset muuttuisivat todennäköisesti helpommiksi.

Big Data on alana hyvin nuori, jonka vuoksi myös siihen erikoistuneita asiantuntijoita on hyvin vähän (Alguliyev & Imamverdiyev, 2014; Katal ym., 2013; Sagiroglu & Sinanc, 2013). Tämä on ongelmallista, sillä Big Datan käyttäminen organisaatioissa on kuitenkin koko ajan yleistymässä.

3.3 Big Data ja kyberturvallisuus

Big Data luo useita ongelmia myös kyberturvallisuuden saralla. Datan määrän kasvaessa, myös järjestelmät kasvavat ja monimutkaistuvat, mikä taas luo yhä enemmän kyberturvallisuusriskejä.

Merkittävimmät ongelmat ovat suojaus ja yksityisyys (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Mtsweni ym, 2016; Wheelus ym.,

2016).Yksityisyys sisältää kaksi näkökulmaa, joista ensimmäinen on henkilötietojen yksityisyys datan keräämisen aikana (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Katal ym., 2013; Kaisler ym., 2013; Khan ym., 2014; Mtsweni ym., 2016; Wheelus ym., 2016), jolloin henkilöstä saatetaan kerätä hänen tietämättään hyvinkin yksityistä tietoa ja jopa sellaista tietoa, mitä hän ei haluaisi itsestään kerättävän. Toinen näkökulma on kerättyjen henkilötietojen suojaaminen (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Katal ym., 2013; Khan ym., 2014; Mtsweni ym., 2016; Wheelus ym., 2016), esimerkiksi tietovuodon sattuessa.

Big Data tuo ongelmia tiedon salaamisen (encryption) suhteen (Alguliyev & Imamverdiyev, 2014; Chen ym., 2014; Khan ym., 2014), sillä sen suuren mittakaavan ja monimuotoisuuden vuoksi vanhat traditionaaliset salausmenetelmät eivät vastaa enää niin suuren määrän salausvaatimuksia.

Pilvipalveluiden yleistyessä Big Datan tallentaminen luo ongelmia niin datan eheyden säilyttämisessä (Khan ym., 2014) kuin myös kolmannen osapuolen tuomat riskit tietoturvallisuudessa (Katal ym., 2013).

Big Data ei tuo pelkästään haasteita vaan myös uusia ratkaisuja kyberturvallisuuteen (Chen ym., 2014). Big Datan avulla on mahdollista mm. havaita sekä ennustaa tulevia erilaisia hyökkäyksiä.

4 Cyber Threat Intelligence

Teknologioiden monimutkaistuessa myös niiden tietoturvallisuus heikkoudet lisääntyvät (Schneier, 2012) ja lisäksi kyberrikolliset sekä heidän toimintatapansa kehittyvät teknologian mukana. Myös kerätyn datan määrä on kasvanut niin valtaisaksi, että sitä on hyvin vaikea kerätä, hallita ja prosessoida tavallisilla tietokoneilla tarpeeksi nopeasti (Chen ym., 2014; Kaisler ym., 2013; Tien, 2013).

Uuden sukupolven kehittyneet tietoturvahyökkäykset ovat 1) dynaamisia, sillä ne pystyvät kohteesta riippuen mukautumaan sille sopivaksi 2) monitasoisia, sillä ne hyödyntävät eri resursseja esimerkiksi sähköpostia, sovelluksia tai nettisivuja ja 3) monesti myös useamman tehtävän sisältäviä (Tounsi & Rais, 2018) eli ensimmäisenä hyökkäys vain pyrkii pääsemään sisään, sitten leviämään mahdollisimman laajalle järjestelmässä ja viimeiseksi lataamaan arvokasta informaatiota järjestelmästä hyökkääjän järjestelmään. Näiden kehittyneiden hyökkäyksien takia vanhat ”allekirjoituksiin” perustuvat tietoturvatoimet eivät kykene suojautumaan täysin uusilta hyökkäyksiltä.

Kyberhyökkäyksiltä puolustautuminen ei ole helppoa. Tutkijat ovat kehittäneet uusia toimintatapoja, jossa siirrytään hyökkäyksien vahinkojen jälkikäteen korjaamisesta hyökkäyksien ennakoimiseen. Yksi nousevista trendeistä on Cyber Threat Intelligence (CTI), jota vuonna 2018 SANS Instituten tuottaman tutkimuksen mukaan, jopa 68% vastanneista organisaatioista hyödynsivät (Shackleford, 2018).

4.1.1 Käsite

CTI on vasta kehittymäisillään (Ernst & Young, 2014) ja monista yrityksistä huolimatta akateemisessa yhteisössä ei ole selkeää selitystä (Abu ym., 2018; Qamar, Anwar, Rahman, Al-Shaer & Chu, 2017), mitä CTI tarkoittaa.

Suosituimman määritelmän CTI:lle on antanut Gartner, jossa CTI määritellään olevan todisteisiin perustuva tieto, joka sisältää kontekstin, mekanismit, indikaattorit, osallisuuden ja toimintatavan, olemassa olevasta tai ilmaantuvasta uhasta. Tätä tietoa voidaan käyttää avuksi organisaation

päättäessä, mitä kyseiselle uhalle kannattaa tehdä. (Abu ym., 2018; Tounsi & Rais, 2018.)

Vaihtoehtoinen määritelmä on, että CTI on prosessi ja tuote, joka muodostuu raa'asta datasta analysoitua informaatiota, mikä sisältää tietoa vastustajista, joilla on aikomus, tilaisuus sekä valmiudet aiheuttaa harmia (Abu ym., 2018).

Hyvin usein CTI:tä viitataan nimellä Threat Intelligence (TI), mutta osa tutkijoista määrittelee TI:n olevan enemmänkin tehtävä, jossa kerätään todisteisiin perustuvaa tietoa. CTI taas sisältää myös strategista toimintaa. (Brown, Gommers & Serrano, 2015; Zibak & Simpson, 2019.)

Yritys Ernst & Young (2014) määrittelee CTI:n olevan edistyksellinen prosessi, joka mahdollistaa arvokkaan tiedon keräyksen. Tämä tieto sisältää asiayhteyteen ja tilanteeseen liittyvää tietoa ja sitä voidaan muokata organisaation uhkatilanteen mukaan (Ernst & Young, 2014).

CTI:n tarkoitus on pienentää epävarmuutta (Zibak & Simpson, 2019) eli määrittää faktoja ja myöhemmin toimittaa luotettavia päätelmiä ja ennustuksia, jotka auttavat tietoturvaan liittyvässä päätöksenteossa.

Olkoon määritelmä mikä vain, Abu ym., (2018) toivat tutkimuksessaan esille, että lopullinen määrittely tulisi sisältää CTI:n pääelementit eli sen pitäisi olla relevanttia, täsmällistä, ajatonta sekä toimintakelpoista (Abu ym., 2018; Ernst & Young, 2014).

CTI voidaan jakaa erilaisiin alakategorioihin eri kriteereiden mukaan. Moni tutkija mieltää jakamisen Steelen (2007) mukaan, jossa CTI jaetaan 1) strategiseen, 2) operationaaliseen, 3) taktiseen sekä 4) teknilliseen kategorioihin (Mtsweni ym., 2016; Tounsi & Rais, 2018).

Strateginen CTI on korkeatasoista informaatiota, jota organisaation päätöksentekijät, esimerkiksi johto, käsittelevät ja sen avulla pyrkivät tekemään hyvin informoituja päätöksiä. Strategisen CTI :n tarkoitus on auttaa päätöksentekijöitä ymmärtämään paremmin nykyisiä sekä identifioimaan tulevia riskejä. (Borum, Felker, Kern, Dennesen & Feyes, 2015; Ernst & Young, 2014; Tounsi & Rais 2018.)

Operationaalinen CTI on yksityiskohtaista informaatiota tietystä hyökkäyksestä, joka pyrkii hyökkäämään organisaatioihin. Tämän tyyppistä informaatiota käsittelee vain arvovaltaiset IT-johtajat. Steelen (2007) mukaisessa mallissa operationaalinen CTI on harvinaista, sillä yksittäisellä organisaatiolla ei yleensä ole oikeutta päästä näkemään muiden organisaatioiden järjestelmiä, jonka vuoksi kerätty informaatio jää merkityksettömäksi. Sen vuoksi tämän kaltaista informaatiota keräävät usein vain valtiot. (Steele, 2007; Tounsi & Rais, 2018.)

Taktinen CTI on informaatiota siitä, miten kyberrikolliset suorittavat hyökkäyksen. Taktista TI:tä käyttävät yleensä tietoturvajärjestelmistä vastaavat tahot, jonka avulla he pystyvät varmistamaan, että suojaukset ja tutkimukset ovat ajan tasalla. (Ernst & Young, 2014; Tounsi & Rais 2018.)

Teknillinen CTI informaatio kerätään normaalisti teknillisten lähteiden kautta ja se yleensä syötetään suoraan organisaation tutkiviin tai monitorioiviin funktioihin, esimerkiksi palomuuriin (Tounsi & Rais, 2018).

Muita jakoja ovat esimerkiksi CTI:n jakaminen virallisiin ja epävirallisiin käytäntöihin, jotka auttavat hyödyntämään hiljaista tietämystä yhteistyökumppaneiden välillä. Epävirallinen käytäntö tarkoittaa, että yhteistyökumppanit jakavat tietämystä, työkaluja ja neuvoja keskenään, kun taas virallinen käytäntö on jakaa enemmän teknistä tietoa kuten Indicators of Compromise (IOC). (Tounsi & Rais 2018.)

CTI:n voidaan jakaa myös strategiseen ja operationaaliseen CTI:hin, missä strateginen CTI tarvitsee korkeasti koulutettuja analyytikkoja ja operationaalinen TI valmistetaan vain teknologiaa käyttäen (Tounsi & Rais, 2018).

Vaihtoehtoisesti CTI:n voi jakaa strategiseen ja taktiseen tietämykseen, jossa strateginen CTI on korkealaatuista ja voidaan hyödyntää strategisissa päätöksenteoissa ja kun taas taktinen CTI on matalempaa ja helpompi mitata ja täten käyttää esimerkiksi tietoturvaohjelmissa (Tounsi & Rais 2018).

4.1.2 Historia

CTI tuli näkyvään asemaan 2000-luvun puolivälissä uutena ratkaisuna alati kehittyville tietoturvauhkille. Tarkemmin sanottuna joulukuussa 2007 kun markkinoille ilmestyi ensimmäinen TI-ratkaisu Object Description and Exchange Format (IODEF) (Elmellas, 2016), joka on XML-kieleen pohjautuva määrittelyformaatti, jonka avulla kyberhyökkäystilanteiden reagointiryhmät jakavat tietoa hyökkäyksestä.

2009 the Research and Education Network Information Sharing and Analysis Centre kehitti viitekehyksen (Collective Intelligence Framework, CFI), mihin sisältyi serveri, joka kerää ja tallentaa TI dataa, kuten esimerkiksi IP-osoitteita, sähköpostiosoitteita, domainien nimiä ja muita attribuutteja. Se myös sisälsi tietoa esimerkiksi siitä, minkälainen hyökkäys on ollut kyseessä ja kuinka vakava se oli. (Elmellas, 2016.)

2010 Verizon vapautti markkinoille The Vocabulary for Event Recording and Incident Sharing (VERIS) viitekehyksen, joka tarjoaa standardin toimintatavan hyökkäyksien raporttien määrittelyyn sekä jakamiseen. VERIS oli myös ensimmäinen virstanpylväs, jolloin TI:tä alettiin käyttämään myös strategisissa päätöksenteoissa, kun aikaisemmin sitä pidettiin vain taktisena tietona. (Elmellas, 2016.)

2012 AlienVault (nykyinen AT&T Cybersecurity) kehitti Open Threat Exchange (OTX) ohjelmiston, jonka tarkoitus on puhdistaa, yhdistää, validoida ja julkaista dataa yli 18 000 laitteesta, mikä teki siitä myös keskeisimmän ohjelmiston TI datan keräämiseen, tallentamiseen ja analysoimiseen. (Elmellas, 2016.)

2013 Mitre kehitti Structured Threat Information Expression (STIX) viitekehyksen (Elmellas, 2016), joka vielä tähän päivään asti on pysynyt

suosituimpana CTI:n määrittelykielenä (Zibak & Simpson, 2019). STIX määrittelee hyökkäyksen tiedot ei pelkästään teknillisinä tietoina (IP-osoite jne.) vaan liittää mukaan myös kontekstin. STIX:in suosion ansiosta Mitre kehitti samana vuonna myös Trusted Automated Exchange of Indicator Information (TAXII) standardin, joka määrittelee vaatimukset palveluille kyberuhkien tietojen jakamiseen. (Elmellas, 2016.)

Tieteellisessä yhteisössä CTI alkoi saamaan paljon huomiota vasta vuoden 2010 jälkeen, varsinkin vuonna 2013 oli huomattavissa trendi kiinnostumisesta ja jo vuonna 2014 tutkimuksia julkaistiin ennätysmäärä. Viimeisen parin vuoden aikana kiinnostunut on kuitenkin taas hieman hiipunut. (Wagner, Mahbub, Palomar & Abdallah, 2019.)

4.1.3 Datan keräys

Tiedon määrä on nykyaikana valtava, joten sitä on enää yhden tahon vaikea prosessoida yksinään. Tämän vuoksi olisi tehokkaampaa, että tiedon keräämiseen, analysointiin ja jakamiseen osallistuisi useampi organisaatio.

Organisaatiot keräävät dataa jo valmiiksi analysoituna tietona tai raakana datana, joko sisäisesti erilaisten mittareiden ja havaitsemisprosessien avulla tai ulkoisista avoimista lähteistä, esimerkiksi blogeista, tai ostamalla sitä kolmansilta osapuolilta (Abu ym., 2018; Veerasamy, 2017). Organisaatiot voivat myös hyödyntää erilaisia suljettuja Threat Intelligence Management Platforms (TIMP) julkaisualustoja, joissa vapaaehtoiset organisaatiot jakavat keskenään tietoa luotettavien kanavien välityksellä (Abu ym., 2018; Brown ym., 2015; Mtsweni ym., 2016; Sauerwein, Sillaber, Mussmann & Breu, 2017; Zibak & Simpson, 2019).

CTI informaatiota kerätään useasta eri lähteestä, joka voi olla teknistä, esimerkiksi logi tietystä hyökkäyksestä, tai inhimillistä, esimerkiksi puhelinkeskustelulla tai lukemalla keskusteluita TI:hen liittyvillä julkaisualustoilla. Tyypillisesti organisaatiot eivät kerää tietoa vain yhdestä lähteestä vaan yhdistelevät niitä (Sauerwein, Sillaber & Breu, 2018).

TIMP:n toivat esille ensin Dandurand ja Serrano (2013), jotka määrittelevät julkaisualustoille vaatimuksiksi 1) jakamisen helpottaminen 2) automaation mahdollistaminen sekä c) datan tuottamisen, hienomuokkauksen sekä tarkastuksen helpottaminen. Sen lisäksi useampi standardi on kehitetty helpottamaan informaation yhdenmukaisuutta ja niiden kautta sen hyödyntämistä. (Dandurand & Serrano, 2013; Sauerwein ym., 2017.) Standardeista huolimatta tiedeyhteisössä on erimielisyysiä siitä, mitä näiltä julkaisualustoilta odotetaan.

Suurin osa julkaisualustoista on täysin suljettuja (Sauerwein ym., 2017), jotta jaettu tieto pystytään pitämään salassa kyberhyökkääjiltä ja täten pysymään edes pari askelta heitä edellä.

Dalziel (2014) määrittelee kirjassaan, että analysoitu ja prosessoitu CTI informaation täytyy täyttää kolme vaatimusta. Ensimmäiseksi informaation täytyy olla relevanttia eli liittyä jotenkin organisaatioon, joka kyseistä

informaatiota hyödyntää. Toisekseen informaation täytyy olla arvokasta eli siitä on jotain hyötyä organisaatiolle. Viimeiseksi informaation pitäisi olla toimintakelpoista eli tarpeeksi täsmällistä, jotta organisaatio pystyy sen perusteella reagoimaan, joko tekemällä uusia päätöksiä tai jopa suoraan muuttamalla toimintatapojaan.

Organisaatiot jakavat kerättyä dataa yleensä Indicators of Compromise (IOC) avulla, jotka ovat siis jälkeenjääneitä todisteita kyberhyökkäyksestä (Liao, Yuan, Wang, Li, Xing & Beyah, 2016), kuten IP-osoitteita tai viruksien allekirjoituksia. IOC ovat siitä suosittuja, että kun ne ovat kerätty oikein, ne voidaan automaattisesti muuntaa ja syöttää kyberhyökkäyksen torjunta ohjelmille (Liao ym., 2016).

4.2 Hyödyt

SANS Instituten vuonna 2018 tehdyn tutkimuksen mukaan tutkimukseen vastanneista organisaatioista 81% olivat sitä mieltä, että CTI on parantanut organisaation tietoturvaa (Shackleford, 2018; Veerasamy, 2017; Zibak & Simpson, 2019;). CTI:n on myös todistettu tuovan muita hyötyjä organisaatioille. Esimerkiksi se vähentää toistuvaa tietoa ja hyökkäyksistä johtuvia vahinkoja, tukee vastatoimenpiteitä hyökkäyksille, pyrkii tuomaan markkinoille yhä enemmän asiantuntijoita, säästää organisaation menoja, vähentää tietoturvainvestointien epätietoisuutta sekä vahvistaa organisaatioiden ja valtion suhdetta (Zibak & Simpson, 2019). Näiden lisäksi CTI:n avulla organisaatio kykenee laajentamaan professionaalisia verkostojaan (Zibak & Simpson, 2019; Veerasamy, 2017) sekä parantamaan organisaation tilannetietoisuutta (Ernst & Young, 2014; Zibak & Simpson, 2019).

Tutkijat Ahrend, Jirotka ja Jones (2016) tunnistivat CTI:n lyhytaikaisiksi hyödyiksi mm., että se täydentää muualta hankittua tietoa ja auttaa identifioimaan nousevia trendejä (Veerasamy, 2017), mitä ei muuten välttämättä tunnistettaisi. CTI myös helpottaa löytämään tuntemattomia tapahtumia tai tapauksia omista järjestelmistä ja auttaa olemaan varuillaan samankaltaisia hyökkäyksiä varten, joita on jo tapahtunut toisessa organisaatiossa. (Ahrend ym., 2016.)

Pitkäkestoisia hyötyjä ovat mm. tietoturvan kehittäminen, tulevaisuuden hyökkäyksien estäminen tai ainakin merkittävä vähentäminen, haittojen kurissa pitäminen sekä auttaminen ennakoivien prosessien käyttöönotossa ja paremman käsitys globaalista tilanteesta (Ahrend ym., 2016).

4.3 Haasteet

Tieteellisestä näkökulmasta suurin haaste liittyen CTI:hen tällä hetkellä on epäyhtenäinen määrittely siitä, mitä sen on tarkoitus saavuttaa ja miten (Veerasamy, 2017; Zibak & Simpson, 2019).

Myös taitavien osaajien pula on todettu yhdeksi haasteeksi (Veerasamy, 2017; Zibak & Simpson, 2019), sillä jopa 62% SANS Instituten tutkimukseen vastanneista pitivät suurena ongelmana, että koulutettuja ammattilaisia on niin vähän (Shackleford, 2018). Toisaalta tämä ei ole ihme, sillä ilmiö on niin uusi ja eikä siitä ja sen ominaisuuksista ole edes yhteisymmärrystä, joten siihen on myös vaikea erikoistua.

Big Datasta tuttu datan ylikuormitus on myös ilmaistu olevan CTI:ssä ongelmana, sillä uhkiin perustuvaa dataa kerätään lisää jokainen hetki ja sitä voi kerätä useasta eri lähteestä. Tämä on johtanut siihen, että kerättyä dataa tuntuu olevan liikaa samalla kun tietoa taas liian vähän (Abu ym., 2018; Ernst & Young, 2014). Data on myös hyvin moninaista ja monimutkaista (Qamar ym., 2017). Tämä on iso ongelma, sillä hyvin usein kaikkea dataa ei ehditä analysoimaan, jolloin on todennäköistä, että oleellinen tieto voi jäädä huomioimatta. Ongelmaksi muodostuu myös analysoidun datan ajankohtaisuus, jos sitä ei saada prosessoitua tarpeeksi nopeasti

Tiedon jakamisen prosessi on vielä kehittymäisillään ja rajoittunutta (Chandel, Yan, Chen, Jiang & Ni, 2019; Dandurand & Serrano, 2013), sillä suuren datamäärän vuoksi olisi ihanteellista, että datan prosessointi olisi täysin automaattista, mitä se ei vielä ole vaan mm. uhkien identifiointi, korjaaminen ja ennaltaehkäisy vaatii vielä ainakin ihmisoperoijan (Wagner ym., 2019).

Kielimuuri on iso ongelma monikansallisilla julkaisualustoilla joko niin, että tapahtumista ilmoitetaan jollain muulla kielellä kuin englanti tai että kyberhyökkäysilmoitukset tehdään ilman yhteistä standardia (Qamar ym., 2017; Veerasamy, 2017; Wagner ym., 2019; Zibak & Simpson, 2019), jolloin ilmoituksesta saattaa jäädä pois jotain olennaista tai se ei ole yleisesti ymmärrettävissä.

4.3.1 Miksi organisaatiot eivät osallistu avoimeen tiedon jakamiseen

CTI perustuu hyvin vahvasti avoimeen tietoon hyökkäyksistä, joten se on hyvin riippuvainen siitä, että organisaatiot jakaisivat tietojaan. Valitettavasti moni organisaatio on epäilevä siitä, haluavatko he varmasti antaa julkiseksi tietoja hyäkkäyksen kohteeksi joutumisesta, sillä he pelkäävät saavansa negatiivista julkisuutta (Tounsi & Rais, 2018; Veerasamy, 2017; Wagner ym., 2019; Zibak & Simpson, 2019). Tieto hyökkäyksestä voi luoda mielikuvan, että organisaatiolla ei ole tietoturvatoimet tämän päivän tasolla, jolloin myös asiakkaiden tiedot ovat vaarassa. Tämän vuoksi myös esimerkiksi pelkästään jo huhu tietomurrosta voi johtaa asiakkaiden menetykseen.

Toinen syy siihen, miksi organisaatiot saattavat olla vastahakoisia jakamaan on laki- ja yksityysasiat (Abu ym., 2018; Tounsi & Rais, 2018; Veerasamy, 2017; Wagner ym., 2019; Zibak & Simpson, 2019). CTI on kansainvälistä tiedonvaihtoa, jolloin organisaation kotimaan lait saattavat estää tietynlaisen tiedon jakamisen. Organisaatio voi olla epävarma mitä tietoa hän saa jakaa, joten helpompaa on olla jakamatta.

Jaetun informaation pitäisi olla relevanttia, ajankohtaista, paikkaansa pitävää, vertailtavaa, yhtenäistä ja selkeää, jonka vuoksi yksi syy jakamatta jättämiseen on ongelmat datan laadussa (Dandurand & Serrano, 2013; Tounsi & Rais, 2018; Zibak & Simpson, 2019). Jos informaatio on esimerkiksi vanhaa, eikä se auta päätöksen tekemisessä, organisaatio jättää sen hyvin usein jakamatta. Tämä voi olla hyvä asia, sillä informaatiota tulee muutenkin niin paljon, että sitä on vaikea hallita.

Neljäs ongelma on epäluotettavat osallistujat (Al-Ibrahim, Mohaisen, Kamhoua, Kwiat & Njilla, 2017; Chandel ym., 2019; Tounsi & Rais, 2018; Veerasamy, 2017; Wagner ym., 2019; Zibak & Simpson, 2019). TI perustuu vapaaehtoisuuteen ja yhteistyöhön, jossa kaikkien osallistujien on tarkoitus jakaa tietoa ja hyötyä siitä. Jos jakamisverkostossa on osallistujia, jotka vain ottavat tietoa vastaan, mutta eivät koskaan jaa sitä, luo se epävarmuutta ja epäluotettavuutta yhteistyössä, jolloin aktiiviset osallistujat tuntevat, että he eivät saa vastineeksi mitään. Tämä voi johtaa osallistujien määrän pienentymiseen ja tiedon laadun alenemiseen.

Maailma pyörii rahan ympärillä, joten myös budjetointi on luokiteltu yhdeksi syyksi sille miksi organisaatiot eivät jaa tietojaan muiden organisaatioiden kanssa (Tounsi & Rais, 2018; Veerasamy, 2017; Zibak & Simpson, 2019). Tämä vaikuttaa varsinkin pienyrityksiin, joilla ei ole tällä hetkellä varaa tehdä isoja investointeja. SANS Instituutin kyselyyn vastanneista melkein 40% vastaajista olivat sitä mieltä, että budjetti oli iso ongelma (Shackleford, 2018). Budjetointi voi tulla ongelmaksi mm., kun pitäisi palkata uusia henkilöitä, jotka ovat erikoistuneet TI:hen tai esimerkiksi silloin, kun informaatiota ostetaan kolmannelta osapuolelta.

Joskus organisaatiot olettavat, että heillä ei ole tietoa jaettavaksi (Tounsi & Rais, 2018). Tiedon vähäisyys voi johtua siitä, ettei hyökkäyksiä ole huomattu esimerkiksi vanhentuneiden tietoturvatoimenpiteiden vuoksi, vaikka todellisuudessa organisaatio on ollut hyökkäyksen alaisena. Toinen syy vähäisyyteen voi olla organisaatiossa vallitseva syyttämiskulttuuri (Tounsi & Rais, 2018). Virheen sattuessa tapahtuneeseen etsitään syyllistä ja häntä pyritään rankaisemaan. Tämä johtaa siihen, että työntekijät eivät ole innokkaita jakamaan tietoa mahdollisista hyökkäyksistä.

Viimeisenä Tounsi ja Rais (2018) määrittelevät jakamatta jättämisen ongelmaksi organisaation uskomuksen, että asialle ei ole mitään tehtävissä. Esimerkiksi epävarmuus siitä voiko poliisin puoleen kääntyä hyökkäyksen tapahtuessa johtaa siihen, että organisaatio ei raportoi hyökkäystä mihinkään vaan pyrkii korjaamaan sen sisäisesti ja piilossa julkisuudesta.

4.3.2 Informaation laadun hallintaongelmat CTI julkaisualustoilla

Edellisessä osiossa todettiin, että yksi syy sille, miksi organisaatiot ovat vastahakoisia jakamaan tietojaan muiden organisaatioiden kanssa oli kerätyn informaation laatuongelmat. Sillaber, Sauerwein, Mussmann ja Breu (2016) löysivät tutkimuksessaan, että mitä enemmän organisaatioita osallistuu tiedon

jakamiseen CTI julkaisualustoilla, sitä enemmän yhdistettyjä tietoja luodaan, mikä taas johtaa monimutkaisempaan ja virheellisempään sisältöön. Sen takia olisi tärkeää, että tiedon alkuperä olisi tiedossa ja se on tarvittaessa jäljitettävissä. Myös tahojen, jotka ylläpitävät CTI julkaisualustoja, täytyisivät ottaa enemmän vastuuta huolehtiakseen, että eri informaatiolähteiden tiedot integroidaan oikealla tavalla heidän julkaisualustalleen.

He myös löysivät, että ajantasaisuus ja merkityksellisyys voivat kärsiä, kun informaatiota tulee paljon (Sillaber ym., 2016), jolloin on myös hankalaa löytää haluttua tietoa suuren datamäärän keskeltä (Dandurand & Serrano, 2013). Merkityksellisyys siinä mielessä, että esimerkiksi finanssialan organisaatio ei halua tietää, mitä hyökkäyksiä farmasia-alan organisaatiolle on tapahtunut, sillä ne hyvin harvoin liittyvät toisiinsa. Ja ajantasaisuus siten, että julkaisualustat sisältävät ja sinne lähetetään myös vanhaa tietoa, joka ei ole enää ajankohtaista.

Kolmanneksi julkaisualustat ovat hyvin harvoin muokattavissa (Sillaber ym., 2016), jolloin organisaatio voisi määritellä filttereiden ja hakutoimintojen avulla näkyville vain heitä kiinnostavia tietoja. Esimerkiksi organisaatio voisi halutessaan valita kielen, lähteen tai kategorian mukaan.

Viimeiseksi julkaisualustat ovat hyvin harvoin täysin automatisoituja, jolloin organisaation työntekijän täytyy lisätä tiedot manuaalisesti järjestelmään (Sillaber ym., 2016). Tämä jättää tilaa ihmisvirheille (Sauerwein ym., 2017), joita on myöhemmin valtavan datan määrän vuoksia vaikea havaita (Shackleford, 2018).

5 Yhteenveto ja pohdinta

Tämän tutkimuksen tarkoituksena oli selvittää, miten CTI vaikuttaa organisaation tietoturvaan ja mitä haasteita se tuo. Tässä tutkielmassa ensin selviteltiin CTI:n taustalla liittyvää Big Dataa kyberturvallisuuden näkökulmasta ja tietoturvaa organisaation näkökulmasta. Sen jälkeen selvitettiin mitä on CTI, mitä ominaisuuksia sillä on ja mitä haasteita se tuo. CTI on vielä kohtalaisin uusi ilmiö, jonka vuoksi siitä löytyy vielä vain rajallisesti tietoa. Tämän tutkimuksen praktinen tarkoitus oli siis täyttää havaittu tutkimusaukko CTI:n tietämyksestä.

Tutkielma toteutettiin kirjallisuuskatsauksena ja sen tavoitteena oli vastata kysymykseen *"Mitä on Cyber Threat Intelligence, miten se vaikuttaa ja mitä haasteita se tuo?"*. CTI:n tarkoituksena on hyödyntää erilaisista lähteistä saatua analysoitua informaatiota kyberuhista ja sen perusteella tehdä päätöksiä tietoturvan parantamiseksi.

Maailman digitalisoitumisen seurauksena kerätyn datan määrä on kasvanut eksponentiaalisesti, tietojärjestelmät ovat monimutkaistuneet ja kyberhyökkäykset kehittyneet teknologian kehityksen mukana, jonka vuoksi perinteiset kyberhyökkäyksen torjuntamenetelmät eivät enää yksinään riitä suojaamaan organisaatioita. CTI on uusi nouseva trendi, jossa ajattelu siirretään hyökkäyksien vahinkojen korjaamisesta hyökkäyksien ennakoimiseen.

Tiedon määrän runsauden vuoksi sitä on enää yhden tahon vaikea prosessoida yksinään, vaan olisi tehokkaampaa, että tiedon keräämiseen, analysointiin ja jakamiseen osallistuisi useampi organisaatio, jonka vuoksi CTI:n tarkoituksena on jakaa ja kerätä tietoa vapaaehtoisesti muiden organisaatioiden kanssa. Tiedon kerääminen ja jakaminen on vapaaehtoista ja siitä on monia hyötyjä organisaatioille mm. se edesauttaa organisaatiota saamaan paremman tilannekuvan, mitä kybermaailmassa tapahtuu ja esimerkiksi minkälaiset kyberuhat ovat nousussa. Haasteita taas ovat mm. epäluotettavat osallistujat, budjetin suuruus, valtavan datan määrän hallinta ja prosessointi sekä datan laatuongelmat.

Niin kuin tutkielmat yleensä myös tämä tutkielma on rajallinen, sillä se ei syventynyt CTI:n teknisempään puoleen, esimerkiksi siihen, miten sitä varten kehitetyt standardit sekä IOC oikein toimivat.

CTI:n rajallisen tutkimusaineiston takia jatkotutkimusaiheeksi esitetään sen ominaisuuksien tarkempaa tarkastelua, kuten julkaisualustojen vaatimuksista tai miten standardit vaikuttavat siihen, miten organisaatiot jakavat tietoa sekä totta kai yleisen määrittelyn yhtenäistämistä. Tutkielmassa kävi ilmi, että tutkijat ovat eri mieltä, mitä CTI oikein on ja mitä se sisältää. CTI on nouseva trendi, jonka monet johtajat ovat jo ottaneet käyttöön omassa organisaatiossaan, joten olisi tärkeää myös tietää sen eri vaikutuksista niin organisaatioon kuin myös yhteiskuntaan, mutta ennen kuin sen vaikutuksia voidaan tutkia, täytyy sille olla yhteneväinen määrittely.

LÄHTEET

Abu, M. S., Selamat, S. R., Ariffin, A. & Yusof, R. (2018). Cyber threat intelligence: Issue and challenges. Indonesian *Journal of Electrical Engineering and Computer Science*, 10(1), 371-379. Haettu osoitteesta http://ijeecs.iaescore.com/index.php/IJEECS/article/view/11065

Ahrend, J. M., Jirotka, M. & Jones, K. (2016). On the collaborative practices of cyber threat intelligence analysts to develop and utilize tacit threat and defence knowledge. Teoksessa *2016 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA)*. London, June 13-14, 2016. Haettu osoitteesta https://doi.org/10.1109/CyberSA.2016.7503279

Alguliyev, R. & Imamverdiyev, Y. (2014). Big data: Big promises for information security. Teoksessa *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*. Astana, October 15-17, 2014. Haettu osoitteesta https://doi.org/10.1109/ICAICT.2014.7035946

Al-Ibrahim, O., Mohaisen, A., Kamhoua, C., Kwiat, K. & Njilla, L. (2017). Beyond free riding: Quality of indicators for assessing participation in information sharing for threat intelligence. *Cryptography and Security*. Haettu osoitteesta https://arxiv.org/abs/1702.00552v1

Borum, R., Felker, J., Kern, S., Dennesen, K. & Feyes, T. (2015). Strategic cyber intelligence. *Information & Computer Security, 23(3)*, 317-332. Haettu osoitteesta https://doi.org/10.1108/ICS-09-2014-0064

Brockett, P. L., Golden, L. L. & Wolman, W. (2012). Enterprise cyber risk management. Teoksessa Emblemsvåg, J. (toim.), *Risk Management for the Future - Theory and Cases (319-340)*. IntechOpen.

Brown, S., Gommers, J. & Serrano, O. (2015). From cyber security information sharing to threat management. Teoksessa *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security, WISCS '15 (43-49)*. Denver, Colorado, October 12, 2015. Haettu osoitteesta https://doi.org/10.1145/2808128.2808133

Bulgurcu, B., Cavusoglu, H. & Benbasat, I. (2010). Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness. *MIS Quarterly*, 34(3), 523-548. Haettu osoitteesta http://doi.org/10.2307/25750690

Chandel, S., Yan, M., Chen, S., Jiang, H. & Ni, T. (2019). Threat intelligence sharing community: A countermeasure against advanced persistent threat. Teoksessa *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (353-359)*. San Jose, California, March 28-30, 2019. Haettu osoitteesta https://doi.org/10.1109/MIPR.2019.00070

Chen, M., Mao, S. & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19(2)*, 171-209. Haettu osoitteesta https://doi.org/10.1007/s11036-013-0489-0

Conti, M., Dargahi, T. & Dehghantanha, A. (2018). Cyber threat intelligence: Challenges and opportunities. Teoksessa A. Dehghantanha, M. Conti & T. Dargahi (toim.), *Cyber threat intelligence (1-6)*. Springer. Haettu osoitteesta https://doi.org/10.1007/978-3-319-73951-9_1

Dalziel, H. (2014). *How to define and build an effective cyber threat intelligence capability*. Elsevier.

Dandurand, L. & Serrano, O. S. (2013). Towards improved cyber security information sharing. Teoksessa K. Podins, J. Stinissen & M. Maybaum (toim.), *2013 5th International Conference on Cyber Conflict, (CYCON 2013) (1-16)*. Tallinn, June 4-7, 2013. Haettu osoitteesta https://ieeexplore.ieee.org/abstract/document/6568369

Elmellas, J. (2016). Knowledge is power: The evolution of threat intelligence. *Computer Fraud & Security, 2016(7)*, 5-9. Haettu osoitteesta https://doi.org/10.1016/S1361-3723(16)30051-3

Ernst & Young, (2014). *Cyber threat intelligence − how to get ahead of cybercrime.* Haettu osoitteesta https://www.ey.com/Publication/vwLUAssets/EY-cyber-threat-intelligence-how-to-get-ahead-of-cybercrime/$FILE/EY-cyber-threat-intelligence-how-to-get-ahead-of-cybercrime.pdf

Hudson, B. (2014). *Advanced persistent threats: Detection, protection and prevention*. Sophos Ltd., US, Haettu osoitteesta https://au.insight.com/content/dam/insight/en_US/pdfs/sophos/sophos-advanced-persistent-threats-detection-protection-prevention-whitepaper.pdf

Ifinedo, P. (2012). Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security, 31(1)*, 83-95. Haettu osoitteesta https://doi.org/10.1016/j.cose.2011.10.007

Jang-Jaccard, J. & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences, 80(5)*, 973-993. Haettu osoitteesta https://doi.org/10.1016/j.jcss.2014.02.005

Kaisler, S., Armour, F., Espinosa, J. A. & Money, W. (2013). Big data: Issues and challenges moving forward. Teoksessa R. H. Sprague Jr. (toim.), *2013 46th Hawaii International Conference on System Sciences, HICCS (995-1004)*, Wailea, Maui, Hawaii, January 7-10, 2013. Haettu osoitteesta https://doi.org/10.1109/HICSS.2013.645

Katal, A., Wazid, M. & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. Teoksessa M. Parashar, A. Zomaya, J. Chen, J.-N. Cao, P. Bouvry & S. K. Prasad (toim.), *2013 Sixth international conference on contemporary computing (IC3) (404-409)*, Noida, August 8-10, 2013. Haettu osoitteesta https://doi.org/10.1109/IC3.2013.6612229

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, W. K. M., Alam, M., Shiraz, M. & Gani, A. (2014). Big data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal, 2014*. Haettu osoitteesta http://dx.doi.org/10.1155/2014/712826

Ključnikov, A., Mura, L. & Sklenár, D. (2019). Information security management in SMEs: Factors of success. *Entrepreneurship and Sustainability Issues, 6(4)*, 2081-2094. Haettu osoitteesta https://doi.org/10.9770/jesi.2019.6.4(37)

Liao, X., Yuan, K., Wang, X., Li, Z., Xing, L. & Beyah, R. (2016). Acing the IOC game: Toward automatic discovery and analysis of open-source cyber threat intelligence. Teoksessa *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, CCS '16 (755-766)*, Vienna, October 24-28, 2016. Haettu osoitteesta https://doi.org/10.1145/2976749.2978315

Mtsweni, J., Mutemwa, M. & Mkhonto, N. (2016). Development of a cyber-threat intelligence-sharing model from big data sources. *Journal of Information Warfare, 15(3)*, 56-68. Haettu osoitteesta https://www.jstor.org/stable/26502744

Qamar, S., Anwar, Z., Rahman, M. A., Al-Shaer, E. & Chu, B. (2017). Data-driven analytics for cyber-threat intelligence and information sharing. *Computers & Security, 67*, 35-58. Haettu osoitteesta https://doi.org/10.1016/j.cose.2017.02.005

Rot, A. & Olszewski, B. (2017). Advanced persistent threats attacks in cyberspace. threats, vulnerabilities, methods of protection. Teoksessa M. Ganzha, L. Maciaszek & M. Paprzycki (toim.), *Position Papers of the 2017 Federated Conference on Computer Science and Information Systems vol. 12 (113-117)*. Prague, September 3-6, 2017. Haettu osoitteesta http://dx.doi.org/10.15439/2017F488

Safa, N. S., Von Solms, R. & Furnell, S. (2016). Information security policy compliance model in organizations. *Computers & Security, 56*, 70-82. Haettu osoitteesta https://doi.org/10.1016/j.cose.2015.10.006

Sagiroglu, S. & Sinanc, D. (2013). Big data: A review. Teoksessa W. W. Smari & G. C. Fox (toim.), *2013 International Conference on Collaboration Technologies and Systems (CTS) (42-47)*. San Diego, California, May 20-24, 2013. Haettu osoitteesta https://doi.org/10.1109/CTS.2013.6567202

Sauerwein, C., Sillaber, C. & Breu, R. (2018). Shadow cyber threat intelligence and its use in information security and risk management processes. Teoksessa P. Drews, B. Funk, P. Niemeyer & L. Xie (toim.), *Multikonferenz Wirtschaftsinformatik (MKWI 2018) (1333-1344)*. Lüneburg, March 6-9, 2018.

Sauerwein, C., Sillaber, C., Mussmann, A. & Breu, R. (2017). Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. Teoksessa J. M. Leimester & W. Brenner (toim.), *Proceeding of 13th International Conference on Wirtschaftsinformatik (837-851)*. St. Gallen, February 12-15, 2017.

Schneier, B. (2012). How changing technology affects security. *IEEE Security & Privacy, 10(2),* 104. Haettu osoitteesta https://doi.org/10.1109/MSP.2012.39

Shackleford, D. (2018). CTI in security operations: SANS 2018 cyber threat intelligence survey. SANS Institute.

Sillaber, C., Sauerwein, C., Mussmann, A. & Breu, R. (2016). Data quality challenges and future research directions in threat intelligence sharing practice. Teoksessa *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security, WISCS '16 (65-70)*. Vienna, October 24, 2016. Haettu osoitteesta https://doi.org/10.1145/2994539.2994546

Steele, R. D. (2007). Open source intelligence. Teoksessa L. K. Johnson (toim.), *Handbook of Intelligence Studies (129-147)*. London: Routledge.

Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering, 22(2)*, 127-151. Haettu osoitteesta https://doi.org/10.1007/s11518-013-5219-4

Tounsi, W. & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security, 72*, 212-233. Haettu osoitteesta https://doi.org/10.1016/j.cose.2017.09.001

Veerasamy, N. (2017). Cyber threat intelligence exchange: A growing requirement. Teoksessa M. Scanlon. & N. Le-Khac (toim.), *Proceedings of*

the 16th European Conference on Cyber Warfare and Security, ECCWS17 (513-518). Dublin, June 29-30, 2017.

Von Solms, R. & Van Niekerk, J. (2013). From information security to cyber security. *Computers & Security, 38*, 97-102. Haettu osoitteesta https://doi.org/10.1016/j.cose.2013.04.004

Wagner, T. D., Mahbub, K., Palomar, E. & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security, 87*, (1-13). Haettu osoitteesta https://doi.org/10.1016/j.cose.2019.101589

Wheelus, C., Bou-Harb, E. & Zhu, X. (2016). Towards a big data architecture for facilitating cyber threat intelligence. Teoksessa M. Badra, G. Pau & V. Vassiliou (toim.), *2016 8th IFIP International Conference on New Technologies, Mobility and Security, (NTMS) (1-5)*. Larnaca, November 21-23, 2016. Haettu osoitteesta https://doi.org/10.1109/NTMS.2016.7792484

Zibak, A. & Simpson, A. (2019). Cyber threat information sharing: Perceived benefits and barriers. Teoksessa *Proceedings of the 14th International Conference on Availability, Reliability and Security, ARES '19 (1-9)*. Cantebury, August 26-29, 2019. Haettu osoitteesta https://doi.org/10.1145/3339252.3340528