Esa Jokinen

TUTKIMUS RFC 9116:N KÄYTTÖÖNOTOSTA SUOMALAISTEN ORGANISAATIOIDEN FI-VERKKOTUNNUKSISSA

Opinnäytetyö

Tekniikan ylempi ammattikorkeakoulututkinto

Kyberturvallisuuden koulutus

2023

Kaakkois-Suomen ammattikorkeakoulu

Kaakkois-Suomen ammattikorkeakoulu

Tutkintonimike	Insinööri (ylempi AMK)
Tekijä	Esa Jokinen
Työn nimi	Tutkimus RFC 9116:n käyttöönotosta suomalaisten organisaatioiden fi-verkkotunnuksissa
Toimeksiantaja	Liikenne- ja viestintävirasto Traficom
Vuosi	2023
Sivut	80 sivua, liitteitä 3 sivua
Työn ohjaajat	Vesa Kankare, Kimmo Kääriäinen

TIIVISTELMÄ

RFC 9116 määrittelee verkkopalveluille polun "/.well-known/security.txt", jota organisaatiot voivat käyttää tietoturvayhteystietojensa ja -käytäntöjensä julkaisemiseen, jotta tietoturvatutkijoiden on helpompi ilmoittaa löytämistään haavoittuvuuksista. Haavoittuvuuksista on pystyttävä viestimään luottamuksellisesti, jotta ne ehditään korjata ennen kuin ne pääsevät julkisuuteen. Liikenne- ja viestintävirasto Traficomin Kyberturvallisuuskeskuksen haavoittuvuuskoordinaatio avustaa haavoittuvuuden löytäjää parempaan yhteistyöhön ohjelmistovalmistajien ja järjestelmäintegroijien kanssa. Tässä tutkimuksessa selvitettiin, onko RFC 9116:sta konkreettista apua haavoittuvuuksista viestimisessä.

Tutkimuskysymyksinä oli selvittää, miten yleistä security.txt:n julkaiseminen on suomalaisissa organisaatioissa, millaisia tietoja niissä on julkaistu, sekä miten hyvin toteutuneet käytännöt noudattavat RFC 9116:n asettamia määrittelyjä ja ottavat huomioon sen nostamia turvallisuusnäkökohtia. Tutkimusjoukoksi valittiin suomalaisten organisaatioiden hallinnassa olevat fi-verkkotunnukset (n=366990 lokakuussa 2022 ja n=367942 helmikuussa 2023). Tutkimusaineisto kerättiin kaikkien verkkotunnusten juuren ja www-aliverkkotunnuksen kaikista sallituista HTTPS- ja HTTP-poluista. Ilmiöön liittymättömien havaintojen hylkäämissyyt tilastoitiin ja ilmiöön liittyvien havaintojen sisältöä analysoitiin. Näin tuotettiin kattavaa ja luotettavaa tietoa security.txt:n käytöstä, mistä on apua kyberturvallisuuden tilannekuvan muodostamisessa ja aiheeseen liittyvän opastuksen tuottamisessa.

RFC 9116:n käyttöönottoa voidaan tutkimusjoukon osalta pitää vielä marginaalisena ilmiönä, sillä security.txt:n on julkaistu vain 2,2 ‰:ssa kaikista verkkotunnuksista. Tiedosto ladattiin luotettua varmennetta käyttävässä yhteydessä 85 %:ssa verkkotunnuksista ja oli tarjolla vain salaamattomalla yhteydellä 6 %:ssa. Tiedostojen sisällöissä on suurta vaihtelevuutta, mikä hankaloittaa niiden koneluettavuutta. Kymmenen suurinta toimijaa on julkaissut tiedoston 85 %:ssa sen julkaisseista verkkotunnuksista. Näiden joukossa on myös palveluntarjoajia ja ohjelmistotuottajia, jotka ovat julkaisseet tiedoston asiakkaidensa verkkotunnuksissa. Koska yksittäisillä toimijoilla on täten suuri vaikutus verkkotunnuskohtaisiin tuloksiin, analysoitiin tuloksia myös saman toimijan julkaisemat tiedostot yhdistäen. Vaikka OpenPGP:tä tarjotaan salausmenetelmänä yhteydenottoihin, on sen käyttö tiedostojen allekirjoittamiseen hyvin harvinaista. Ilmiön seurantaa on syytä jatkaa, mikä tutkimusta varten kehitettyjen työkalujen avulla on jatkossa pitkälti automatisoitua.

Asiasanat: haavoittuvuus, tietoturva, verkkotunnukset, verkkopalvelut, kvantitatiivinen tutkimus, standardointi

South-Eastern Finland University of Applied Sciences

Degree title	Master of Engineering
Author	Esa Jokinen
Thesis title	Research on the implementation of RFC 9116 in .fi domain names registered to Finnish organisations
Commissioned by	Finnish Transport and Communications Agency Traficom
Time	2023
Pages	80 pages, 3 pages of appendices
Supervisors	Vesa Kankare, Kimmo Kääriäinen

ABSTRACT

RFC 9116 defines a web service path "/.well-known/security.txt" organisations can use for publishing their vulnerability disclosure contacts and policies for security researchers. This communication should be confidential as the organisations must be able to address the vulnerability before it becomes public. The NCSC-FI vulnerability coordination helps the finders of vulnerabilities to co-operate with the software manufacturers and system integrators. This study examines whether RFC 9116 concretely provides a useful tool for this communication with Finnish organisations.

The research questions were how common publishing security.txt is in Finnish organisations, what kind of information is published, and how well the published files follow the specifications and security considerations of RFC 9116. The research group was the .fi domain names registered to Finnish organisations (n=366990 in October 2022 and n=367942 in February 2023). The data was collected from the domain apex and www subdomain using all the relevant HTTPS and HTTP paths. Statistics from the observations not related to the phenomenon as well as the reason to disregard them were made and the related observations were analysed further. This provided comprehensive and reliable information on the use of security.txt, which is helpful for situation awareness in cyber security and in producing guidance related to the topic.

The use of RFC 9116 within the research group was quite marginal as only 2.2 ‰ of the domains published the security.txt file in at least one of the paths examined. The connection used for retrieving the file was secured with a trusted certificate in 85 % of the domain names, and only available with plain text protocol in 6 %. The contents of the files varied decreasing machine readability. Ten largest actors published the file for 85 % of the domain names, allowing a single actor to affect the results significantly. Therefore, the combined domains from a single actor were analysed, too. Despite OpenPGP was offered as an encryption method for the communication, signing the files with OpenPGP was rare. It is recommended continuing monitoring the phenomenon, which will be largely automated in the future with the help of the tools developed for this research.

Keywords: vulnerability, data security, domain names, web services, quantitative research, standardisation

SISÄLLYS

1 JOHDANTO

Tietoturvatutkijoille on tuottanut haasteita löytää niitä yhteydenottokanavia, joiden kautta organisaatiot toivoisivat heidän ilmoittavan haavoittuvuuksista ja muista tietoturvaongelmista. Heidän työtään helpottaisi, mikäli käytännöt olisivat yhtenäisempiä. Organisaatiot ovat erilaisia, eikä aina ole ilmeistä, keitä tietoturvaan liittyvät asiat koskevat. Tällaisten henkilöiden tai organisaatioyksiköiden yhteystiedot voisivat löytyä aina samasta paikasta. Tämän ohella tietoturvatutkijat kaipaavat tietoa siitä, millaisia ongelmia heidän on luvallista etsiä, sekä miten niistä voidaan ilmoittaa turvallisesti ja luottamuksellisesti.

Yhteystietojen ja käytäntöjen löytämistä pyrkii helpottamaan huhtikuussa 2022 *Internet Engineering Task Force*n (IETF):n *Requests for Comments* -arkistosarjassa julkaistu RFC 9116 (Foudil & Shafranovich 2022), joka määrittelee verkkosivulle standardin polun "/.well-known/security.txt" (sekä taaksepäin yhteensopivan sijainnin "/security.txt"), jota organisaatiot voivat käyttää tietoturvayhteystietojen julkaisemiseksi. Samalla tietoturvatutkijat pystyvät käyttämään niitä haavoittuvuuksien etsimisen ja niistä viestimisen menetelmiä sekä viestin salausmenetelmiä, joita organisaatiossa toivotaan käytettävän. Lisäksi käytäntö mahdollistaa sen arvioimisen, miten luotettavia tätä kautta saadut yhteystiedot ovat.

Liikenne- ja viestintävirasto Traficomin Kyberturvallisuuskeskuksen lakisääteisiin tehtäviin kuuluu tukea, ohjata ja valvoa tietoturvallisuutta ja yksityisyyden suojan toteutumista sähköisessä viestinnässä, ylläpitää kansallisen kyberturvallisuuden tilannekuvaa sekä edistää ja varmistaa tietojärjestelmien ja tietoliikennejärjestelyiden tietoturvallisuutta (Laki Liikenne- ja viestintävirastosta, 3. §). Tehtäviä toteutetaan muun muassa antamalla toimialaa koskevia määräyksiä, asetuksia, ohjeita ja suosituksia sekä julkaisemalla ajankohtaisia tietoturvauutisia. Tämä tutkimus mahdollistaa tutkimustietoon perustuvaa tiedottamista tuottamalla uutta tietoa tuoreen tietoturvallisuuden edistämiseen tähtäävän teknologian käyttöönotosta suomalaisissa organisaatioissa. Tutkimuksen tuottaman analyysin perusteella on mahdollista selvittää, missä laajuudessa käytäntö on yleistymässä. Siten se auttaa myös arvioimaan, miten paljon Kyberturvallisuuskeskuksen tulisi edistää sen yleistymistä. Lisäksi se tarjoaa mo-

nipuolista ja kattavaa tietoa siitä, minkälaisten yksityiskohtien osalta organisaatiot kaipaavat lisää opastusta, jotta käytännöstä olisi mahdollisimman paljon hyötyä sen tarkoitusten saavuttamiseksi.

Aiheesta tehdyt aiemmat tutkimukset ovat kohdistuneet lähinnä suosituimpien verkkosivustojen listoihin ja niiden käyttökelpoiset tulokset ovat olleet pistemäisiä. Tämä tutkimus tarjoaa kattavan poikkileikkauksen kaikkiin yhden maakohtaisen ylätason verkkotunnuksen alla oleviin verkkotunnuksiin, jotka ovat jonkin suomalaisen organisaation hallinnassa. Tutkimusjoukon kattavuuden vuoksi aineiston perusteella voidaan tehdä myös ajallista vertailua. Myöhemmin, käytännön mahdollisesti yleistyessä, verkkotunnuksen haltijasta saatavilla olevien julkisten tietojen perusteella on mahdollista tehdä vertailua eri toimialojen välillä. Tutkimusta varten kehitetyillä aineistonkeruu- ja analyysimenetelmillä voidaan lisäksi jatkaa ilmiön seurantaa ja selvittää, miten Kyberturvallisuuskeskuksen aiheeseen liittyvät julkaisut vaikuttavat ilmiöön. Opinnäytetyön toisena tavoitteena onkin tuottaa toimeksiantajan käyttöön työkaluja, joilla ilmiön seuraaminen on jatkossa mahdollisimman helppoa ja automatisoitua.

2 TUTKIMUSASETELMA

2.1 Tutkimusongelma, tutkimuskysymykset ja tutkimuksen tavoitteet

Tämän tutkimuksen tutkimusongelmana on tietoturvaongelmista ilmoittamisen haasteet, jotka liittyvät siihen, että tietoturvaongelman löytäneiden tietoturvatutkijoiden on ollut hankala löytää sekä niiden henkilöiden yhteystietoja, jotka vastaavat organisaation tietoturvasta, että niitä käytäntöjä, joita organisaatio haavoittuvuuksien etsinnässä ja niistä ilmoittamisessa toivoo käytettävän. Ongelmassa on keskeistä aiemmin löytämättömiin haavoittuvuuksiin liittyvien tietojen arkaluontoisuus, sillä haavoittuvuuksista on pystyttävä viestimään luottamuksellisesti, jotta ne ehditään korjata ennen kuin ne pääsevät julkisuuteen. Myös haavoittuvuuden korjauksen julkaisevat päivitykset täytyy ehtiä toimittaa ja ottaa käyttöön ennen kuin niitä ehditään väärinkäyttää.

Kyberturvallisuuskeskuksen haavoittuvuuskoordinaatio vastaanottaa ilmoituksia haavoittuvuuksista ja avustaa haavoittuvuuden löytäjää parempaan yhteistyöhön ohjelmistovalmistajien ja järjestelmäintegroijien kanssa (Traficom

2021a). Haavoittuvuuden löytäjä saattaa haluta olla ilmoittamatta haavoittuvuudesta suoraan siinä pelossa, ettei vastaanottaja ymmärtäisi hänen olevan hyvällä asialla, vaan tulkitsisi hänen murtautuneen organisaation järjestelmiin. Tämän riskiä saattaa lisätä, mikäli ilmoituksen vastaanottaa organisaatiossa henkilö, jolla ei ole riittävää asiantuntemusta sen käsittelyyn.

Tietoturvayhteystietojen ja ilmoittamiskäytäntöjen löytämisen helpottaminen on yksi tapa vastata näihin haasteisiin. Yksittäisten tietoturvatutkijoiden ohella tähän tähtäävät yleiset käytännöt, kuten RFC 9116:n esittämä yhdenmukainen ja koneluettava ilmoituskäytäntöjen julkaisukanava, voivat helpottaa myös Kyberturvallisuuskeskuksen haavoittuvuuskoordinaation toimintaa. Tutkimusongelmasta johdettuina tutkimuskysymyksinä on analysoida tämän yhdeksi ongelman ratkaisuksi ehdotetun käytännön käyttöönottoa selvittämällä suomalaisten organisaatioiden hallitsemien fi-verkkotunnusten osalta:

1. Miten yleistä security.txt:n julkaiseminen on?
2. Millaisia tietoja niissä on julkaistu?
3. Miten hyvin ne noudattavat RFC 9116:n asettamia määrittelyjä?
4. Miten hyvin niissä on otettu huomioon RFC 9116:n nostamia turvallisuusnäkökohtia?

Näihin kysymyksiin vastaamisen tavoitteena on tuottaa kattavaa ja luotettavaa tietoa security.txt:n käytöstä kyberturvallisuuden tilannekuvan muodostamiseksi ja aiheeseen liittyvän relevantin opastuksen tuottamiseksi. Tutkimusta varten valittujen menetelmien ja kehitettyjen työkalujen mahdollistaman ajallisen seurannan tavoitteena on lisäksi saada tietoa siitä, miten aiheeseen liittyvä opastus ja tiedottaminen vaikuttavat ilmiön kehittymiseen. Tämä pidempiaikainen seuranta ja Kyberturvallisuuskeskuksen viestinnän vaikutus ilmiöön on kuitenkin rajattu tämän opinnäytetyön ulkopuolelle, sillä ensimmäisen aineistonkeruun tulosten ohjaamana havainnointiväliä päätettiin harventaa, ja tutkimusajanjaksolla ei myöskään julkaistu uusia aiheeseen liittyviä tiedotteita.

2.2 Käytettävät tutkimusmenetelmät

Opinnäytetyö on hyvin tutkimuksellinen, sillä se tuottaa perustutkimukselle tyypillisesti uutta, luotettavaa ja jaettavaa tietoa, joka täydentää ja laajentaa aiempaa aiheesta tehtyä tutkimusta. Tavoitteena on kuitenkin antaa vastauksia, joita tullaan käyttämään konkreettisesti toiminnan ja mahdollisesti koko

toimialaan vaikuttavien käytäntöjen kehittämiseen. Näiden ominaisuuksien yhdistelmä tekee opinnäytetyöstä tutkimus- ja kehittämistoimintaa, joka luo systemaattisesti uutta tietoa ja käyttää sitä tavoitteellisesti uusiin sovelluksiin (Tilastokeskus s.a.). Tutkimuksen aikana ei kuitenkaan vielä aktiivisesti vaikutettu tutkittavaan ilmiöön, joten tutkimuksen tulosten voidaan katsoa olevan objektiivisia ja seuraavan ilmiön luonnollista kehitystä ilman sellaisia interventioita, joita olisi tehty, mikäli tutkimusstrategiaksi olisi valittu toimintatutkimus (Heikkinen 2010).

Ilmiön aktiivisen edistämisen sijaan tutkimus asemoituu seuraamaan kehitystä toteutunutta käyttöönottoa analysoimalla. Tutkimusote on pääasiassa kvantitatiivinen, vaikkakin osa havainnoista on niin harvinaisia, että niistä voidaan lopulta tehdä vain laadullista, kuvailevaa analyysiä. Kvantitatiivisessa tutkimuksessa ilmiön parametrit tai muuttujat on tunnettava, ja useimmiten ne on ensin selvitetty kvalitatiivisella tutkimuksella (Kananen 2011, 15–17). Tässä tapauksessa tutkittavan ilmiön taustalla on tekninen määritelmä, jonka perusteella on mitattavissa kustakin muuttujasta, täyttääkö se määritelmän vai ei, tai mikä muuttujan arvo on. Tämä tekee tutkittavasta ilmiöstä riittävän täsmentyneen ja määritellyn, jotta sitä voidaan mitata kvantitatiivisin menetelmin (Kananen 2011, 18). Tutkimuskysymyksistä johdettuja muuttujia tarkastellaan lähemmin analyysimenetelmien käsittelyn yhteydessä.

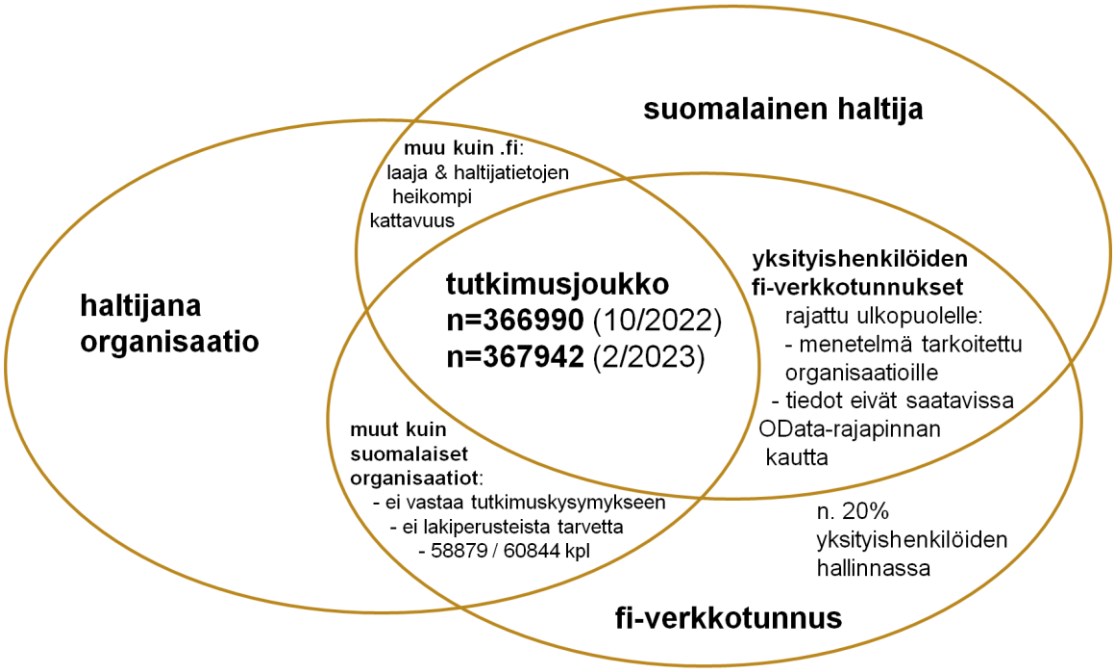
Toisaalta myös osa määritelmän mukaan virheellisistä sisällöistä on selitettävissä aiempien tutkimusten (Poteat & Li 2021; Findlay & Abdou 2022) ja määritelmän luonnosversioiden perusteella. Luonnosversioiden eroavaisuuksien tarkastelu on tässä opinnäytetyössä sisällytetty yhdeksi alaluvuksi osaksi kvantitatiivisen tutkimuksen perusteena olevaa teoreettista viitekehystä. Yhtä hyvin se voitaisiin nähdä myös kvalitatiivisin menetelmin tehtynä esitutkimuksena, joka antaa lisää tietoa niistä tekijöitä, joista ilmiö koostuu, ja joka toimii siten kvantitatiivisen tutkimuksen mittareiden täsmentäjänä (Kananen 2011, 16).

Monimenetelmäisyyden piirteitä näkyy myös tulosten analyysissä, jossa kvalitatiivinen havainto siitä, että sisällöltään samankaltaisia tiedostoja esiintyi paljon, nosti tarpeen analysoida erikseen kokonaan uutta havaintoyksikköä. Verkkotunnuksen ohelle havaintoyksiköksi nousi toimija, joka oli julkaissut

keskenään samankaltaiset tiedostot. Tällä perspektiivin laajennoksella oli merkittävää vaikutusta niin tutkimuksen tuloksiin kuin pohdintoihin niiden merkityksestäkin. Havaintoyksiköllä eli tilasto- tai otantayksiköllä tarkoitetaan tutkimuksen kohdetta, jota tutkimuksessa analysoidaan (Kananen 2011, 58). Käytän termiä havaintoyksikkö, koska se kuvaa hyvin aineiston luonnetta ja sopii valittuun kokonaistutkimukseen eli tutkimukseen, jossa kaikki perusjoukon havaintoyksiköt mitataan (Holopainen & Pulkkinen 2008, 29; Kananen 2011, 65). Tutkimusjoukon sekä siihen kohdistettavan kokonaistutkimuksen valintaa perustellaan tarkemmin seuraavassa alaluvussa.

2.2.1 Tutkimusjoukon valintaperuste ja kattavuus

Tutkimusjoukkona on suomalaisten organisaatioiden hallinnassa olevat fi-verkkotunnukset. Kohteen rajaaminen maakohtaiseen ylätason verkkotunnukseen kohdistaa tutkimuksen suomalaiseen ympäristöön, vaikka suomalaiset organisaatiot voivat varata myös muita verkkotunnuksia. Huolimatta siitä, että käytännön voi kohdistaa myös IP-osoitteeseen, ovat verkkotunnuksiin liittyvät verkkosivustot ja -palvelut security.txt:n tyypillinen käyttötapa (Foudil & Shafranovich 2022, 9–10). Näistä syistä valittu tutkimusjoukko on tutkimuskysymyksiin vastaamisen kannalta relevantti. Tutkimusjoukon valintaa ja rajausta on havainnollistettu kuvassa 1.



Kuva 1. Tutkimusjoukon rajaus suomalaisten organisaatioiden fi-verkkotunnuksiin.

Tiedot organisaatioiden hallinnassa olevista fi-verkkotunnuksista haltijan maan ja Y-tunnuksen kera ovat myös julkisesti saatavilla ilman tunnistautumista (Traficom 2021b, 38; Viestintävirasto 2017, 6–9), mikä tekee tutkimuksesta toistettavan sekä mahdollistaa rajauksen pelkästään suomalaisiin organisaatioihin. Julkisen rajapinnan kautta ei tarjota tietoa yksityishenkilöiden verkkotunnuksista, mutta 80 % verkkotunnuksista on organisaatioiden tai yhteisöjen käytössä (Viestintävirasto 2017, 6). Lisäksi RFC 9116 on tarkoitettu pääasiassa organisaatioiden käyttöön (Foudil & Shafranovich 2022, 3–4).

Kvantitatiivisessa tutkimuksessa käytetään usein otantaa, jossa perusjoukkoa edustavasta otoksesta saatuja tuloksia yleistetään koskemaan koko perusjoukkoa, sillä otannan kasvaessa luotettavuus ei kasva lineaarisesti, mutta otantaan liittyvät kustannukset kasvavat (Kananen 2011, 65–68). Nummenmaa ym. (1996, 35) jopa väittävät, ettei empiirisessä tutkimuksessa voida analysoida koko perusjoukkoa, vaan analyysi joudutaan suorittamaan otannan perusteella. Tässä tutkimuksessa kustannukset liittyvät aikaan, joka aineiston keräämiseen ja analysoimiseen kuluu. Kustannusten selvittämiseksi tehtiin kokeiluja suurimmasta mahdollisesta keräysnopeudesta, joka ei vaikuttanut havaintojen luotettavuuteen. Tästä oli laskettavissa, että aineiston kerääminen koko perusjoukosta on mahdollista alle viikossa. Kerätyn aineiston analyysissä puolestaan yksittäisiin kysymyksiin saatiin vastauksia aineiston luokittelun osalta minuuteissa ja ilmiöön liittyvien havaintojen osalta sekunneissa.

Aineiston koko on hyvä myös silloin, jos olennainen saadaan selville keräämättä turhaa aineistoa, mutta harvinaisemman ilmiön havainnoinnissa voidaan tarvita laajempaa aineistoa, jotta ilmiö saadaan ylipäätänsä näkyviin (Nummenmaa ym. 1997, 23). Findlayn & Abdoun (2022, 5–6) aiemmassa tutkimuksessa menetelmän käyttö oli osoittautunut sitä harvinaisemmaksi, mitä pidemmälle suosittujen verkkotunnusten listalla edettiin. Koska tietoa haluttiin nyt kokonaisesta maakohtaisesta ylätason verkkotunnuksesta, oli aiemman tutkimustiedon valossa oletettavissa ilmiön mahdollisesti olevan harvinainen. Tästä syystä ja kustannusten ollessa kohtuullisia päädyttiin kokonaistutkimukseen. Koska tutkimusjoukko on täten rajattu tutkimuskysymysten mukaiseksi ja sisältää koko perusjoukon otantamenetelmien käyttämisen sijaan, voidaan tuloksia pitää hyvin kattavina ja luotettavina.

2.2.2 Aineistonkeruumenetelmät ja -välineet

Lista tutkimusjoukkona olevista verkkotunnuksista muodostettiin hyödyntämällä fi-verkkotunnusten OData-palvelurajapintaa, josta on mahdollista saada tutkimukseen tarvittavat tiedot kaikista verkkotunnuksista ja niiden haltijoista Y-tunnuksineen (Viestintävirasto 2017, 6–9). Rajapinta pohjautuu OASIS-konsortion OData JSON V4.0-formaattiin; tietoa voidaan hakea yksinkertaisilla HTTP GET-kyselyillä, ja vastaukset ovat pyydettävissä JSON-muodossa, joka sisältää osajoukon saatavilla olevista vastauksista sekä osoitteen, josta seuraava osajoukko on haettavissa (Handl ym. 2016, 10–12).

Tutkimusjoukko rajattiin hakutuloksista verkkotunnuksen tilan mukaan, sillä vain voimassa olevilla verkkotunnuksilla voi olla verkkosivut. Aktiivisia organisaatioiden hallussa olevia verkkotunnuksia 12.10.2022 oli 425916 kappaletta, josta edelleen haltijan kotimaan mukaan rajaamalla saatiin suomalaisten organisaatioiden fi-verkkotunnusten tutkimusjoukko n=366990. 20.2.2023 verkkotunnuksia oli puolestaan 428786, joista vastaavalla rajauksella saatiin tutkimusjoukko n=367942. Kummallakin kerralla tutkimusjoukko kattoi siis n. 86 % rajapinnasta saaduista aktiivisista organisaatioiden verkkotunnuksista, mutta vastasi valitun rajauksen mukaista perusjoukkoa, joten kokonaistutkimuksen määritelmä täyttyi. Tutkimusjoukon koon muuttumista kesken tutkimuksen voitaisiin kritisoida aineistonkeruukertojen vertailtavuutta heikentävänä tekijänä, mutta valittu kokonaistutkimus edellyttää tämän sallimista, sillä muutos vastaa muutosta tutkimuksen perusjoukossa. Sen sijaan saman joukon säilyttäminen vääristäisi tuloksia, koska verkkotunnuksia rekisteröidään uusia, ne vaihtavat haltijaa sekä niitä jätetään uusimatta, jolloin ne poistuvat käytöstä.

Varsinainen tutkimusaineisto kerättiin The ZMap Projectin (s.a.) tuottaman avoimen lähdekoodin ZGrab 2.0 -työkalun HTTP-moduulilla. Keräyskertojen vertailtavuus toteutettiin työkalun ympärille kirjoitettavalla komentosarjalla, joka teki aineistonkeruun molemmilla kerroilla samalla tavalla. Tiedot kerättiin kaikista niistä poluista, joissa tiedoston julkaiseminen on ollut sallittua myös RFC 9116:n luonnosvaiheissa. Valmiin työkalun ympärille rakennettavaan komentosarjaan kokonaan oman työkalun kirjoittamisen sijaan päädyttiin, jotta vastaavan aineiston keruu olisi myös muiden tutkijoiden toistettavissa.

Yksi muuttuja-ajattelun hyödyistä on, että se auttaa suunnittelemaan ja toteuttamaan aineiston keruun (Kananen 2011, 58). ZGrab 2.0 valittiin, koska se tuottaa tekemistään HTTP-kyselyistä JSON-muotoista aineistoa, joka sisältää käytännössä kaikki yhteyden yksityiskohdat niin HTTP- kuin TLS-protokollienkin osalta. Aineisto sisältää siten jäsennellyssä muodossa kaikki ne yksityiskohdat, joista tutkimuskysymyksistä johdetut muuttujat muodostuvat. Tarkkojen yksityiskohtien perusteella on lisäksi mahdollista saada varmuus epäselvissä tilanteissa ilman, että täytyy kerätä lisää aineistoa. ZGrab 2.0:n tuottamaa JSON-rakennetta (ks. liite 1) analysoimalla saatiin selville kaikki ne yksityiskohdat, joita sekä ilmiöön liittymättömien havaintojen karsimiseksi että yhteyden luotettavuuden arvioimiseksi tarvittiin. Tutkimusaineiston keräämisen ja analysoimisen tekniset yksityiskohdat, kuten työkalujen kehittäminen ja toimintaperiaate, ovat kuvattuina perusteellisemmin tutkimuksen toteutusta käsittelevässä luvussa. Seuraavaksi tarkastellaan kerätyn aineiston analyysimenetelmiä ja -välineitä sekä tutkimuskysymyksistä johdettuja muuttujia yleisellä tasolla.

2.2.3 Analyysimenetelmät ja -välineet

Kvantitatiivisen tutkimuksen peruskäsite on muuttuja eli ominaisuus, jota mitataan (Kananen 2011, 57). Jotta aineistoa voidaan analysoida tavalla, joka tuottaa vastauksia tutkimuskysymyksiin, täytyy määritellä kunkin tutkimuskysymyksen kannalta merkitykselliset muuttujat. Lisäksi muuttujien rakenne on tunnettava eli tiedettävä, minkälaisia arvoja ne voivat saada, jotta voidaan määritellä, millaista asteikkoa käytetään (Kananen 2011, 58–62; Holopainen & Pulkkinen 2008, 15–16). Tässä alaluvussa eritellään, millaisia muuttujia tarkastelemalla kuhunkin tutkimuskysymykseen vastataan, miten muuttujien valintaa perustellaan, sekä millaisia menetelmiä muuttujien käsittelyyn käytetään. Lisäksi kuvataan, miten aineistoa on tarpeen käsitellä, jotta muuttujien arvot saadaan selville.

Kysymys siitä, miten yleistä security.txt-tiedoston julkaiseminen on, pelkistyy havaintoyksikön tasolla muuttujaksi siitä, onko verkkotunnuksessa julkaistu tiedosto vai ei. Tällaisen vain kaksi arvoa saavan muuttujan sanotaan olevan dikotominen eli kahtiajakoinen (Kananen 2011, 59). Tämän kvalitatiivisen

muuttujan mittaamiseen käytetään luokittelu- eli nominaaliasteikkoa, jossa kukin havaintoyksikkö kuuluu vain yhteen luokkaan (Holopainen & Pulkkinen 2008, 15). Kun tarkastelu nostetaan tutkimusjoukon tasolle, saadaan kvantitatiivinen arvo sille, kuinka monessa verkkotunnuksessa tiedosto on julkaistu. Koska tällä arvolla on absoluuttinen nollakohta, on se sijoitettavissa suhdeasteikoille eli sen saamien arvojen suuruuksia on mahdollista vertailla (Kananen 2011, 62; Holopainen & Pulkkinen 2008, 16).

Kysymys siitä, millaisia tietoja tiedostoissa on julkaistu, vaikuttaa muodoltaan kvalitatiiviselta. Jotta sitä voitaisiin analysoida kvantitatiivisin menetelmin, on se saatava muutettua kvantitatiivisesti tarkasteltavissa olevaan muotoon. Koska RFC 9116:ssa on tarkkaan määritelty, millaisia tietoja tiedostoissa voidaan julkaista ja missä muodossa niiden on oltava, onkin kvantitatiivisesti tarkasteltavissa, kuinka paljon mitäkin ominaisuutta on käytetty. Näihin rinnastuvat samalla myös ominaisuudet, jotka ovat olleet tuettuja RFC 9116:n luonnosversioissa sekä yleisesti käytössä olleet, epäviralliset ominaisuudet. Kumpaankin liittyvien lisämuuttujien tunnistamiseksi on teoreettisessa viitekehyksessä tarkasteltu luonnosversioiden välisiä eroavaisuuksia ja muiden tutkimusten tekemiä havaintoja. Lisäksi löydettyjä tiedostoja on esitarkasteltu kvalitatiivisesti käytettyjen ominaisuuksien löytämiseksi ennen niiden laskemista. Menetelmällisesti tämä käsittely on samanlaista kuin avointen kysymysten vastausten kvantifiointi (Kananen 2011, 101), mutta helpommin rajattavissa.

Myös käytetyt ominaisuudet eli tiedoston kentät muodostavat dikotomisia muuttujia siitä, onko ominaisuus käytössä vai ei, ja näiden muuttujien summista saadaan kvantitatiivisia arvoja sille, kuinka yleistä kunkin ominaisuuden käyttö on. Kutakin ominaisuutta on tarkasteltava itsenäisenä, sillä havaintoyksiköitä ei ole mahdollista sijoittaa vain yhtä ominaisuutta käyttävään luokkaan. Samalla ominaisuuden yleisyyttä kuvaavat havaintoyksiköiden määrät ovat keskenään samalla suhdeasteikolla, jolloin niiden kesken on tehtävissä vertailua siitä, onko jokin ominaisuus toista yleisempi.

RFC 9116:n asettamien määrittelyiden noudattamiseen liittyvä kysymys on pilkottava osiin. Yksinkertaisimmillaan olisi tarkasteltavissa, noudattaako tiedosto määrittelyä vai ei, mutta se ei antaisi vielä vastauksia siihen, millaisten

virheiden vuoksi määrittelyä ei ole kyetty täyttämään. Juuri tieto erilaisten virheiden laadusta ja yleisyydestä onkin tutkimuksen tavoitteiden kannalta merkityksellisempää. Tämän vuoksi määrittelyn noudattamisen rinnalle nousee dikotomisia muuttujia siitä, onko jokin yksittäinen virhe tehty vai ei. Samassa havaintoyksikössä on voitu tehdä useampia virheitä, mutta yhdenkin virheen tehnyt havaintoyksikkö ei ole täyttänyt RFC 9116:n määritelmiä. Muutoin virheiden analyysiin pätevät samat lainalaisuudet kuin ominaisuuksien käytön analyysiin.

Turvallisuusnäkökohtien huomioimiseen liittyvä kysymys on edellisiä monisyisempi, sillä tiedoston muodon tavoin ei ole yksiselitteisesti määriteltävissä, että turvallisuusnäkökohdat on joko huomioitu tai niitä ei ole huomioitu. Turvallisuusnäkökohtia on useita ja niitä on perusteltu eri riskeillä. Riskit myös riippuvat tapauskohtaisista riskiarvioista, jotka ovat subjektiivisia: eri organisaatiot arvioivat itsenäisesti sitä, miten suuri jonkin riskin todennäköisyys on, millaisia vaikutuksia organisaatiolle riskin toteutuessa koituu, sekä miten paljon sen välttämiseen halutaan panostaa (Rousku 2017, 14–17). Kustannus turvallisuusnäkökohdan huomioon ottamisesta voi toisinaan olla suurempi kuin hyöty, jota siitä arvioidaan saatavan. Riski tietoturvaongelman ilmoittamatta jättämisestä saatetaan esimerkiksi nähdä suuremmaksi kuin todennäköisyys tiedon joutumisesta vääriin käsiin väärennettyjen yhteystietojen julkaisemisen vuoksi, mutta yhtä hyvin voidaan arvioida päinvastoin. Tällaisia valintoja ei voi laittaa objektiivisesti tärkeysjärjestykseen.

Turvallisuusnäkökohtien osalta on kuitenkin erikseen arvioitavissa muutaman yksittäisen ominaisuuden yleisyyttä tai laatua. Näitä ominaisuuksia käydään yksityiskohtaisemmin läpi teoreettisessa viitekehyksessä, mutta muuttujien määritteleminen on sijoitettu tähän lukuun, jotta se löytyisi samasta paikasta kuin muiden muuttujien määrittelyt. Tiedoston lataukseen käytetyn yhteyden luotettavuus on jaettavissa luokkiin, jotka voidaan laittaa keskenään turvallisuusjärjestykseen: salattu ja luotettu yhteys, salattu yhteys, johon ei voida luottaa, sekä salaamaton yhteys. Lisäksi yhteyden luotettavuuden puutteessa on sen syiden perusteella eri vakavuusasteita. Tämän vuoksi siihen liittyvä muuttuja on luokitteluasteikon lisäksi sijoitettavissa järjestysasteikolle (Holopainen & Pulkkinen 2008, 15). OpenPGP-allekirjoitusten käyttö osoittautui niin harvinaiseksi, että sitä voitiin tarkastella pelkästään laadullisesti, mutta mikäli

käyttö olisi ollut yleisempää, voitaisiin joko ominaisuutta tai sen yksittäisiä piirteitä tarkastella vähintään luokitteluasteikolla ja joissain tapauksissa myös järjestysasteikolla.

Kolmanneksi mitattavissa olevaksi ja siten muuttujana esitettäväksi turvallisuusnäkökohdaksi nousi tiedoston voimassaoloaikojen käyttö ja niiden jakauma. Voimassaolo ilmoitetaan sen päättymisen aikaleimana, mutta absoluuttista aikaa oleellisempi muuttuja on voimassaolon ja havaintohetken välinen aika, sillä turvallisuusnäkökohtien mukaan aikaleiman tulisi olla korkeintaan vuoden tulevaisuudessa (Foudil & Shafranovich 2022, 8, 14–15). Kahden ajan välisen etäisyyden suuruus itsessään on muuttuja, jolle voitaisiin käyttää välimatka-asteikkoa, jossa luokat ovat saman levyisiä eli välimatka luokasta toiseen säilyy samana (Holopainen & Pulkkinen 2008, 15; Kananen 2011, 62). Kaikki etäisyydet eivät kuitenkaan ole saman arvoisia, joten muuttujalle sopii paremmin järjestysasteikko, jossa se voi sijoittua joko suositellulle aikavälille (0–12 kuukautta), olla jo vanhentunut, olla hieman liian kaukana tulevaisuudessa (1–3 vuotta) tai vielä kauempana tulevaisuudessa. Tärkeää on tiedostaa, että kaikki luokat yli vuoden tulevaisuudessa ovat määritelmän kannalta keskenään samanarvoisia ja valittu mielivaltaisesti silloinkin, kun ne kuvaavat hyvin edustamaansa aineistoa.

Aineistonkeruuseen käytetyn ZGrab 2.0 -työkalun tuottama analysoitava tutkimusaineisto oli määrältään huomattava; yhteensä 74 gigatavua. Jotta haut aineistosta olivat nopeampia, JSON-tiedostoja jouduttiin ennen security.txt-tiedostojen sisältöjen ja yhteyden luotettavuuden analysointia karsimaan vain niihin tietoihin, jotka ovat tutkimuksen kannalta oleellisia. JSON-aineiston rakennetta (ks. liite 1) analysoitiin, jotta saatiin eroteltua toisistaan ilmiöön liittyvät ja siihen liittymättömät tiedot. Analyysi jakautui tämän perusteella kahteen osaan: ilmiöön liittymättömät havainnot tilastoitiin, jotta myös aineiston hylkäämisen johtaneet syyt tulivat dokumentoiduiksi tarkkaan ja läpinäkyvästi, ja ilmiöön liittyvät havainnot eroteltiin yksityiskohtaisempaa analyysiä varten.

Osa analyysissä käytettyjen muuttujien arvoista saadaan luettua JSON-rakenteesta tulkitsemalla rakenteen kenttien arvoja tai niiden osia muuttujien arvoiksi. Näitä ovat erityisesti ilmiöön liittyvien havaintojen tunnistamiseen ja yhteyden luotettavuuden arviointiin liittyvät muuttujat. JSON-tiedostoista pystyy

tekemään hakuja ja rajauksia jq-työkalulla (Dolan s.a.), jonka ulostuloa voi käsitellä edelleen tyypillisillä Unix-komennoilla. Näiden tekniseen toimintaan palataan tutkimuksen toteutusta käsittelevässä luvussa. Security.txt-tiedoston sisältö (esimerkkinä liite 2) on kokonaan samassa JSON-rakenteen kentässä, joten siihen liittyvien muuttujien arvojen ratkaisemiseksi sisältöä tulee jäsentää edelleen, jotta voidaan erotella tiedoston kenttiä ja niiden arvoja.

Tutkimusaineiston karsinnassa eri hakujen tulokset tai useampi eri tulos saattoivat johtaa samaan tulkintaan. Muuttujat eivät siis olleet aineistossa valmiiksi oikeassa muodossa, vaan niitä piti muokata eli uudelleen koodata (Nummenmaa ym. 1997, 57), jotta ne vastasivat täsmällisemmin niitä luokitteluja, joita karsimista varten oli tehtävä. Tutkimusaineiston käsittelyssä on noudatettava huolellisuutta sekä tarkasteltava aineiston ja sen muuttujien käyttökelpoisuutta, sillä tässä vaiheessa vältettävissä olevia virheitä ei voida rinnastaa mittausvirheiksi (Nummenmaa ym. 1997, 50–51). Tämän vuoksi eri vaiheissa karsittujen havaintojen tulkintojen eli uudelleen koodattujen muuttujien summia verrattiin sen poissulkemiseksi, että jokin havainto olisi jäänyt tarkastelun ulkopuolelle tai laskettu mukaan useampaan eri tulkintaan. Satunnaisesti esiintyneet poikkeamat olivat alle kymmenen havainnon suuruusluokassa. Koska tarkastuslaskennat johtivat samoihin tuloksiin, selittyvät virheet todennäköisimmin sillä, että yksittäisistä havainnoista puuttui joitakin JSON-rakenteen hauissa käytettyjä avaimia. Karsintavaiheen tulokset pyöristettiin – ei niinkään näiden poikkeamien vaan tulosten suuruusluokan vuoksi – tuhansien havaintojen tarkkuuteen. Ilmiöön liittyvien havaintojen poimimisessa käytetyt komennot osoittautuivat vastaavassa arvioinnissa luotettaviksi, joten epätarkkuus ei vaikuttanut niihin.

Valmiit työkalut itse security.txt-tiedoston sisällön validointiin ja lähempään analyysiin olivat vähäisiä, sillä usean projektin kehitys oli lopetettu jo RFC 9116:n varhaisten luonnosvaiheiden aikana. Tämän vuoksi sisällön jäsentämiseen ja analysoimiseen kehitettiin omia työkaluja. Tekniset analyysimenetelmät ja -työkalut kehitettiin opinnäytetyön toteutuksen aikana, sillä kerätystä aineistosta nousevat havainnot ohjasivat ja täsmensivät tarpeita. Järkevyystarkastelussa arvioitiin, ettei muuttujilla ollut mielekästä mahdollisuutta olla toisistaan riippuvaisia, joten niiden välisten korrelaatioiden laskemista ei nähty hyö-

dylliseksi (Holopainen & Pulkkinen 2008, 228). Ainoastaan tiettyjen ominaisuuksien käytön yleisyys saattaisi korreloida tietoturvanäkökohtien huomioimisen kanssa, mutta näiden ominaisuuksien käyttö ei ollut riittävän yleistä kvantitatiivisesti analysoitavaksi. Siksi työkaluista saatuja muuttujien arvoja ei syötetty uudelleen erilliseen tilastolliseen analyysiin tarkoitettuun ohjelmistoon, kuten SAS, SPSS tai BMDP (Nummenmaa ym. 1997, 44), vaan tulokset olivat valmiiksi siinä muodossa, jossa niitä tutkimuskysymyksiin vastaamisesi tarvittiin. Myös analyysityökalujen rakentamiseen ja käyttöön liittyvät tekniset yksityiskohdat kuvataan tarkemmin tutkimuksen toteutusta käsittelevässä luvussa.

2.3 Tutkimuseettiset näkökohdat

Tutkimuksen aineistonkeruu on luonteeltaan tekninen eikä siihen liity henkilöihin kohdistuvaa tutkimusta tai haastatteluja. Tutkimuksen toteuttamisen eettiseen arviointiin nousevat aineistonkeruumenetelmän tekniset vaikutukset tutkimuksen kohteena olevien organisaatioiden järjestelmiin sekä kerättyjen tietojen mahdollisesti sisältämät henkilötiedot. Lisäksi, kun tiedon keruuta ja säilytystä suoritetaan valtion virastossa, tulee tämän olla perusteltua viraston lakisääteisten tehtävien kannalta. Tutkimus tukee erityisesti Kyberturvallisuuskeskuksen lakisääteistä tehtävää edistää ja varmistaa tietojärjestelmien ja tietoliikennejärjestelyiden tietoturvallisuutta (Laki Liikenne- ja viestintävirastosta, 3. §). Siitä on konkreettista hyötyä sekä haavoittuvuuskoordinaation että tiedotustoiminnan näkökulmista. Toisaalta tämä vaatimus oli myös yksi peruste lisää rajata ulkomaisten organisaatioiden fi-verkkotunnukset tutkimuksen ulkopuolelle.

Aineistonkeruu tuotti jokaisella kerralla kahdeksan HTTP-pyyntöä jokaista verkkotunnusta kohden. Ottaen huomioon kaikille Internet-sivustoille kohdistuvat normaalit liikennemäärät, odotettavissa olevan vastauksen pienen koon sekä pyynnön harmittoman luonteen voidaan vaikutuksia yksittäisen aineistonkeruun kohteena olevan verkkotunnuksen tai organisaation kannalta pitää häviävän pieninä. Pyynnöt lisäksi hajautettiin pitkälle aikavälille siten, että samanaikaisia yhteyksiä oli korkeintaan 75, ja kunkin minuutin aikana muodostettiin yhteensä 300 yhteyttä. Tästä syystä aineistonkeruun ei voida myöskään katsoa kohtuuttomasti kuormittaneen verkkoja havainnoitsijan ja kohteiden välillä.

Aineisto kerättiin julkisista lähteistä ja kerätyt tiedot olivat siten julkisiksi tarkoitettuja, joten tutkimusasetelma ei vaatinut eettisen toimikunnan ennakkoarviota (Kohonen ym. 2019, 16). Aineistonkeruussa yhteen paikkaan koostettava aineisto saattaa välillisesti ja tahattomasti sisältää henkilötiedoiksi luokiteltavissa olevia yksityiskohtia, kuten etunimen ja sukunimen sisältäviä sähköpostiosoitteita. Näitä tietoja ei kuitenkaan kerätty eikä käsitelty henkilötietoina, eikä niiden käyttötarkoituksesta syntynyt henkilörekisteriä. Tutkimusaineistoa itsessään ei saatettu julkisesti saataville, vaan tässä opinnäytetyössä on julkaistu ainoastaan kvantitatiivisen analyysin tuloksia tilastoina ja kvalitatiivisen analyysin tuloksia yleisluontoisina kuvailuina, joista yksittäiset organisaatiot ja yhteystiedot eivät ole tunnistettavissa.

3 TEOREETTINEN VIITEKEHYS

3.1 Requests for Comments (RFC) -julkaisusarja

Internet on löyhästi organisoitua kansainvälistä yhteistyötä, jossa itsenäisiä verkkoja kytketään yhteen noudattamalla vapaaehtoisesti avoimia protokollia ja menettelytapoja, jotka kehittyvät iteratiivisessa vertaisarviointiprosessissa (Bradner 1996, 3). Nimensä mukaisesti *Requests for Comments* (RFC) -julkaisusarja on saanut alkunsa vuodesta 1969 alkaen esitetyistä ideoista ja ehdotuksista, joille on pyydetty kommentteja, eikä siitä ole alun perin ollut tarkoitus rakentaa sellaista julkaisusarjaa, jossa nykyään on tuhansia dokumentteja sisältäen Internet-standardien lisäksi parhaiden käytäntöjen ohjeita, kokeellisia protokollia ja informatiivisia sisältöjä (Flanagan 2019, 2–3). RFC-julkaisusarjasta on tullut osa käytäntöä, jolla Internet-yhteisö, *Internet Engineering Steering Group* (IESG) ja *Internet Architecture Board* (IAB) muodostavat niitä standardeja ja muita käytäntöjä, joilla Internet toimii (Bradner 1996, 6). RFC-dokumentteja julkaisee nykyään *Internet Society*n (ISOC) alainen standardoimisjärjestö *Internet Engineering Task Force* (IETF) (Salz 2022, 5, 9). Myös RFC-julkaisusarjaa itseään on kehitetty useissa RFC-julkaisuissa, ja se kehittyy edelleen vastaamaan yhteisön tarpeita (Flanagan 2019, 2–3).

On tärkeää huomata, että RFC-julkaisusarja ei siis koostu vain normatiivisista standardeista. Verrattuna tyypillisiin standardoimisjärjestöihin IETF on hyvin avoin ja löyhästi organisoitu, sillä sen toimintaan voi osallistua vapaasti kuka tahansa, jota Internetin arkkitehtuurin ja toiminnan kehittäminen kiinnostaa

(Salz 2022, 5), ja kehitystyö tapahtuu avoimilla sähköpostilistoilla. Siihen nykyisin liittyviä organisaatioita ja niiden välistä työnjakoa on dokumentoitu RFC 9281:ssä, joka julkaistiin kesäkuussa 2022 (Salz 2022). Myöskään sanan standardi käyttö ei aina ole vastannut sen virallisia määritelmiä. Esimerkiksi Suomen Standardisoimisliitto SFS määrittelee, että standardi on "konsensukseen perustuva, tunnustetun elimen hyväksymä normatiivinen asiakirja, joka esittää yleistä ja toistuvaa käyttöä varten sääntöjä, ohjeita tai ominaispiirteitä toiminnoille tai niiden tuloksille optimaalisen järjestyksen saavuttamiseksi tietyssä tilanteessa" (SFS-EN 45020: 2007, 17). RFC-julkaisusarjassa standardeiksi kutsutuilta dokumenteilta saattaa toisinaan puuttua konsensukseen pohjautuminen, eivätkä ne ole tunnustetun elimen virallisesti hyväksymiä, vaan ovat pikemminkin olleet tarpeeksi kauan työn alla ollakseen kypsiä julkaistaviksi. Ne eivät myöskään aina pohjaudu vakiintuneisiin tieteen, tekniikan ja kokemuksen avulla saatuihin tuloksiin (SFS-EN 45020: 2007, 17).

Kuka tahansa voi asettaa Internet-luonnoksen (*Internet-Draft*) kommentoitavaksi, ja lopulta tällä tavoin esitetty dokumentaatio voi päästä osaksi virallista RFC-arkistosarjaa (Bradner 1996, 6–8). Siksi on tiedostettava, että RFC-julkaisusarjassa on useita kategorioita (*track*), jotka eivät johda Internet-standardiksi: *experimental* sisältää kokeellisia määritelmiä, *informational* yleisesti Internet-yhteisön tietoon saatettuja määritelmiä, jotka eivät edusta yhteisön konsensusta tai suosituksia, *historic* päivittyneitä tai vanhentuneita määritelmiä ja *best current practice* parhaita vallitsevia ja toisinaan IETF:n itsensä toimintaakin ohjaavia määritelmiä (Bradner 1996, 14–17). Lisäksi lukuisia RFC-dokumentteja on julkaistu pilanpäiten, usein aprillipäivän kunniaksi (Marsan 2005). Näistä syistä on aina syytä tarkastella, minkä kategorian julkaisusta on kyse, mikä on julkaisun kypsyysaste eli missä vaiheessa kategoriansa mukaista prosessia se on, sekä suhtautua kriittisesti siihen, miksi se on julkaistu.

Tämä ei kuitenkaan tarkoita, etteikö Internet-standardien kehitys olisi myös laadukasta, sillä varsinaiset Internet-standardit käyvät läpi useita kypsyysvaiheita ja määritellyn prosessin (Bradner 1996, 11–14, 18–20). Esimerkiksi HTTP/1.1 standardin (Fielding ym. 2022) kehitys aloitettiin jo vuonna 1997, siitä tuli standardiluonnos (*Draft Standard*) 1999, se jaettiin useampaan eri RFC-dokumenttiin (RFC 7230–7235) ehdotetuksi standardiksi (*Proposed*

Standard) 2014, ja siitä tuli lopulta Internet-standardi (RFC 9110 & 9112) kesäkuussa 2022. Koko tämän 25-vuotisen taipaleen ajan protokollaa on kuitenkin käytetty laajasti ajattelematta, että työ on vielä kesken.

McQuistinin ym. (2021, 149) mukaan akateeminen kiinnostus uusia RFC-arkistoituja dokumentteja kohtaan on hiipumassa, minkä osoittaa laskeva trendi viittauksissa, joita uudet RFC-dokumentit saavat kahden vuoden kuluessa niiden julkaisusta. Tutkimus tunnistaa, että RFC:llä on paremmat mahdollisuudet tulla laajalti käyttöön, mikäli se perustuu olemassa olevaan toimintaan, se täyttää jonkin tarpeen, sillä on hyvin määritellyt vaatimukset sekä rajoitettu soveltamisala, eli määrittely voidaan ottaa käyttöön paikallisesti ilman, että esimerkiksi koko Internetin pitää siirtyä siihen kerralla (McQuistin ym. 2021, 148).

Vaikka RFC 9116 luo kokonaan uuden käytännön, se rakentuu teknisesti olemassa olevien rakennuspalikoiden varaan, mikä tekee käyttöönotosta suhteellisen helppoa. Näitä rakennuspalikoita ovat aiemmissa RFC-dokumenteissa määritellyt HTTP (Fielding ym. 2022), TLS (Rescorla 2018), "/.well-known/"-polut (Nottingham 2019), URI-skeemat (Berners-Lee ym. 2005), OpenPGP-salaus ja -allekirjoittaminen (Callas ym. 2007) sekä ISO 8601 -standardin mukaiset aikaleimat, joiden käyttöä Internetin aikaleimoissa suositellaan RFC 3339:ssä (Klyne & Newman 2002). Se myös tunnistaa aiempia yhteystietojen julkaisuun kehitettyjä menetelmiä puutteineen täyttäen uuden tarpeen. Nämä ennustavat laajemman käyttöönoton mahdollisuuksia, mutta toisaalta hieman monimutkaisesti määritellyt ja kehitystyön aikana ristiriitaisestikin muuttuneet vaatimukset voivat puolestaan heikentää niitä. Seuraavaksi tarkastellaan lähemmin RFC 9116:n tarkoitusta, sisältöä ja valmistelutyötä.

3.2 RFC 9116

RFC 9116:n tarkoituksena on tarjota yhdenmukainen paikka niille yhteystiedoille ja käytännöille, joiden mukaan organisaatio haluaa, että havaituista haavoittuvuuksista ilmoitetaan. Verkkosivuille tämä yhdenmukainen sijainti on polkuun "/.well-known/security.txt" sijoitettava määrämuotoinen ja koneluettava tekstitiedosto. Dokumentti listaa aiemmat yleiset käytännöt, jotka ovat tarjonneet yhtenäisiä muotoja sähköpostiosoitteille sekä muita väyliä, joilla tietoturvayhteystietoja on voinut hakea, mutta tunnistaa, ettei niiden yhteydessä ole

ollut mahdollista kertoa tarkempia yksityiskohtia toivotuista ilmoituskäytänteistä. (Foudil & Shafranovich 2022, 3–4, 9.)

Julkaistujen yhteystietojen käyttötarkoituksena on haavoittuvuuksista ilmoittaminen ennen epäilyä siitä, että niitä olisi käytetty väärin. Dokumentin turvallisuusnäkökohdissa ei suositella tietojen käyttämistä tietoturvaloukkauksista ilmoittamiseen, sillä hyökkääjä on voinut kyetä muuttamaan tiedoston sisältöä. Näissä tapauksissa tietojen oikeellisuudesta on ensin varmistuttava lisämenetelmiä, kuten OpenPGP-allekirjoituksen tarkistamista, hyödyntäen. (Foudil & Shafranovich 2022, 14.) Tämä on merkittävää arvioitaessa sitä, voidaanko tällä tavoin julkaistuja yhteystietoja käyttää esimerkiksi tietoturvapoikkeaminen hallinnassa tietoturvaloukkauksista ilmoittamiseen muille organisaatioille. Esimerkiksi nykyisten tietojen vertaaminen aiemmin tallennettuihin tietoihin voisi auttaa luotettavuuden arvioinnissa, mutta tämä vaatisi tietokannan luomista aiemmista versioista ja sen säännöllistä täydentämistä.

Security.txt:n ohella on ollut kehitteillä myös muita käytäntöjä tietoturvayhteystietojen ja -käytäntöjen ilmoittamiseen. Carroll & Ellis (s.a.) ovat julkaisseet maaliskuussa 2021 luonnoksen samankaltaisesta menetelmästä, jossa tiedot ilmoitetaan määrämuotoisina DNS:n TXT-tietueina verkkotunnuksen juuressa. Luonnosta ei ainakaan toistaiseksi ole yritetty saada osaksi RFC-julkaisusarjaa IETF:n menettelyjä käyttäen, vaan kehitystyötä on tehty vielä löyhemmin organisoidusti GitHub-tietovaraston välityksellä. Yhdysvaltain *Cybersecurity and Infrastructure Security Agency* (CISA):n sitova toimintadirektiivi 20-01 määrittelee, että kaikkien Yhdysvaltain virastojen on julkaistava käytäntönsä haavoittuvuuksista ilmoittamiseen heidän pääasiallisen .gov-verkkotunnuksensa polussa "/vulnerability-disclosure-policy" (CISA 2020), mutta tämä ei perustu standardeihin, eikä vastaavaa käytäntöä löydy muista lähteistä.

3.2.1 Valmistelutyö 2017–2022, vastaanotto ja uutisointi

Ennen julkaisua huhtikuussa 2022 RFC 9116 on ollut kehitteillä ja avoin kommenteille syyskuusta 2017 lähtien, mikä on osa edellä kuvattuja prosesseja, joilla IETF muodostaa Internetin käytäntöjä. Sekä luonnosteluun käytetty aika että luonnosten lukumäärä ennen RFC:n julkaisua on ollut kasvussa vuosina 2001–2020, jolloin mediaaniaika ensimmäisestä luonnoksesta julkaisuun on

kasvanut 469 päivästä 1170 päivään (McQuistin ym. 2021, 140–141). Tämän Internet-luonnoksen julkaisi yksityishenkilönä Edwin Foudil, tietoturvatutkija, joka syksyllä 2017 oli juuri aloittanut tietojenkäsittelytieteen opinnot Sveitsin valtiollisessa teknillisessä korkeakoulussa (HackerOne 2017). Julkaisuun kulunut aika oli 1681 päivää eli noin 1,4 vuotta vuoden 2020 mediaania pidempi.

RFC 9116 on julkaistu *informational*-kategoriassa. Kuten RFC-julkaisusarjaa käsitelleessä alaluvussa kuvattiin, tämä tarkoittaa, ettei se ole pyrkimässä varsinaiseksi laajan konsensuksen tai yleisen suosituksen mukaiseksi Internet-standardiksi, vaan kyseessä on kirjoittajan luvalla julkaistu ja samankaltaisen arviointiprosessin läpi käynyt tekninen dokumentaatio (Bradner 1996, 15). Tästä huolimatta näihinkin dokumentteihin viitataan usein standardeina niin mediassa kuin virallisemmissakin yhteyksissä. Tämä käy RFC 9116:n osalta ilmi esimerkiksi Cimpanun (2017) ja Bonderudin (2017) uutisartikkeleista. Tällaiset artikkelit ovat tuoneet samalla julkisuuteen esimerkkejä luonnosversion mukaisesta toteutuksesta, vaikka määrittely on tämän jälkeen vielä muuttunut useita kertoja merkittäviltäkin osin. Samoin uusien verkkopalveluita suojaavien teknologioiden aktiivinen puolestapuhuja ja Microsoftin aluejohtaja Troy Hunt on ottanut security.txt:n esiin blogissaan varhaisessa vaiheessa esimerkin kera (Hunt 2017), mutta käsitellyt aihetta myöhemmin ilman päivitettyjä esimerkkejä (Hunt 2020). Käytännön varhaisessa vaiheessa käyttöön ottaneet saattavat olla tietämättään julkaisseet yhteystietojansa tavoilla, joita ei enää julkaistun RFC:n mukaan tulisi pitää pätevinä ja luotettavina, koska teknisiä muutoksia ei ole pidetty samalla tavoin esillä.

Myös CISA (2020) mainitsee pelkän luonnoksen sitovan toimintadirektiivi 20-01:n saatetekstin usein kysytyissä kysymyksissä ehdotettuna standardina. Toimintadirektiivin luonnosvaiheen uutisoinnissa on lisäksi todettu, että kaikkien Yhdysvaltain virastojen olisi julkaistava security.txt .gov-verkkotunnuksissa 180 vuorokauden kuluessa direktiivin antamisesta (Kuldell 2019; Rashid 2020), vaikka lopulliseen direktiiviin jäikin vain velvollisuus julkaista haavoittuvuuksien paljastamiskäytäntö toisessa polussa (CISA 2020). Internet-standardista tai ehdotetusta Internet-standardista puhutaan lisäksi Kyberturvallisuuskeskuksen tietoturvauutisessa (Traficom 2020), Yhdistyneen kuningaskunnan kyberturvallisuuskeskuksen oppaassa (NCSC UK 2020, 5) sekä oppaaseen

viittaavassa *IoT Security Foundation*in haavoittuvuuksien paljastamisen parhaiden käytäntöjen ohjeessa (Day ym. 2021, 11).

Suhtautuminen RFC 9116:n esittämään käytäntöön ei ole ollut varauksetonta, vaan se on saanut myös kritiikkiä. Krebs (2021) nostaa esiin roskapostin lisääntymisen, kun sähköpostiosoite on helposti saatavilla, sekä erityisesti omatoimisesti tietoturvatutkijoiksi ryhtyneet henkilöt, jotka etsivät haavoittuvuuksia automatisoiduilla menetelmillä ja lähettävät mahdollisesti epäollennaisia havaintoja security.txt-tiedostosta löytyviin osoitteisiin. Toisaalta tämä riski on tiedostettu jo RFC 9116:n turvallisuusnäkökohdissa, joissa organisaatioita kehotetaan punnitsemaan käytännön hyötyjä ja haittoja roskapostin ja harhaanjohtavien ilmoitusten näkökulmasta (Foudil & Shafranovich 2022, 16).

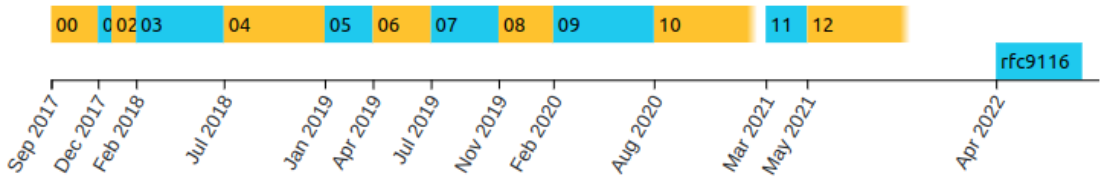
Tämä katsaus RFC 9116:n ja sen valmistelutöiden ympärillä käytyyn julkiseen keskusteluun kertoo ensinnäkin siitä, että käytäntö hakee yhä muotoaan. Toisaalta se paljastaa, miten virheelliset käsitykset kopioituvat helposti lähteistä toisiin, sekä millaista sekaannuksen vaaraa aiheutuu siitä, että teknologiaa yhä kehitetään samalla, kun sitä jo implementoidaan käytäntöön. Seuraavassa alaluvussa käsitellään tarkemmin niitä teknisiä yksityiskohtia, jotka ovat valmistelutyön aikana muuttuneet, jotta voitaisiin ymmärtää, millaisia virheitä tiedostoista saattaa tämän vuoksi löytyä, sekä tunnistaa näiden virheiden liittyvän luonnosvaiheissa sallittuihin sisältöihin.

3.2.2 Julkaistun RFC 9116:n erot aiempiin luonnoksiin

Monet julkaistun RFC 9116:n (Foudil & Shafranovich 2022) yksityiskohdat poikkeavat luonnosversioista merkittävästikin, sillä käytäntö on kehittynyt viiden vuoden aikana saadun palautteen johdosta, kun uusia turvallisuusnäkökulmia on otettu huomioon. Tunnistamalla näitä muutoksia on mahdollista analysoida, mitä lopullisen version mukaan virheellisinä pidettäviä käytäntöjä on voinut jäädä käyttöön, kun käyttöönotto on tehty luonnoksen pohjalta.

Koska luonnosversiot eivät ole RFC-dokumenttien tavoin sellaisessa asemassa, että niihin tulisi viitata tutkimuksissa (Bradner 1996, 8), käsitellään niitä tässä vertailussa tutkimusaineiston tavoin, eikä niitä ole siksi sisällytetty osaksi lähdeluetteloa. Tämä on perusteltua myös siksi, että luonnosversioita

on varsinaisen standardin lisäksi kolmetoista, mikä paisuttaisi lähdeluetteloa aiheettomasti. Luonnosversioiden eroavaisuuksien käsittely on sisällytetty osaksi teoreettista viitekehystä eikä tätä tutkimusta, sillä se muodostaa oleellisen osan sitä teoriapohjaa, joka on ymmärrettävä ensin, jotta aiheeseen kohdistuvaa tutkimusta voidaan tehdä. Siinä jäsennellään uudelleen olemassa olevaa tietoa, jota on käytetty muidenkin aiheesta tehtyjen tutkimusten taustalla. Kaikki versiot ovat saatavissa IETF:n *Datatracker*-palvelun kautta (https://datatracker.ietf.org/doc/rfc9116/) ja niihin viitataan vertailussa numeroinnilla 00–12, jota vastaava aikajana on esitetty kuvassa 2.



Kuva 2. RFC 9116:n valmistelun aikajana IETF:n Datatracker-palvelussa.

Luonnos 00 sijoitti tiedoston verkkosivuston juureen (/security.txt), mutta se siirrettiin jo seuraavassa luonnoksessa alihakemistoon (/.well-known/security.txt). Luonnos 03 palautti mahdollisuuden käyttää verkkosivuston juurta vararatkaisuna yhteensopivuussyistä tai uudelleenohjauksena alihakemistoon, mikä jäi myös RFC 9116:n julkaistuun versioon. Tällä oli tutkimuksen kannalta merkitystä, sillä sen vuoksi myös juurihakemistossa olevat tiedostot täytyi hyväksyä, ja siten myös tämä sijainti tuli ottaa huomioon aineiston keruussa.

Luonnokset 01–04 mahdollistivat security.txt:n ulkoisen allekirjoituksen URL-viittauksella, jolloin allekirjoitus ja allekirjoitukseen käytetty avain tuli kummatkin hakea salatussa HTTPS-yhteydessä. Luonnoksessa 05 ulkoinen allekirjoitus korvataan suosituksella käyttää OpenPGP-viestimuodon selväkielistä allekirjoitusta. Tässä muodossa viesti aloitetaan erityisellä otsakkeella, jota seuraa käytetty tiivistemuoto, tyhjä rivi ja selväkielinen viesti, joiden jälkeen viestin loppuun lisätään ASCII-muotoinen OpenPGP-allekirjoitus (Callas ym. 2007, 59–60).

Luonnoksesta 06 eteenpäin HTTP-palvelimilta edellytetään salattua yhteyttä myös security.txt:n itsensä lataamiseen, ja samassa yhteydessä turvallisuusnäkökohtiin on lisätty vaatimus palvelimen identiteetin todentamisesta RFC

6125:n mukaisesti. Tämä tarkoittaa, että TLS-yhteyden varmenteen tulee olla varmentajan (*certification authority*) allekirjoittama ja HTTPS-protokollan tapauksessa sen *Subject Alternative Name* (SAN) -laajennoksen (tai sen puuttuessa "Common Name" -kentän) tulee vastata pyydettyä isäntänimeä (Saint-Andre & Hodges 2011). Aiemmin menetelmän käyttöön ottaneet ovat tehneet silloisen määritelmän mukaan oikein, vaikka olisivat laittaneet tiedoston saatavilla salaamattomalla yhteydellä. Tutkimuksen kannalta tämän vuoksi aineiston keruussa on haettava tiedostot myös salaamattomilla yhteyksillä, jotta saadaan luotettavaa tilastoa myös sen osalta, missä verkkotunnuksissa security.txt on otettu käyttöön varhaisimpien luonnosten aikaan. Samoin tiedoston lataaminen HTTPS-yhteydellä, jonka varmenne ei täytä asetettuja vaatimuksia, voi olla jäänne aiemman luonnosversion noudattamisesta, ja vähintään sitä voidaan pitää yrityksenä ottaa security.txt käyttöön.

Itse tiedosto koostuu pareista kenttiä ja niiden arvoja; kentän nimi erotetaan arvosta kaksoispisteellä ja välilyönnillä, ja kentät erotetaan toisistaan rivinvaihdolla. Näiden lisäksi tiedostossa saa olla #-merkillä alkavia kommenttirivejä, tyhjiä rivejä sekä varsinaisen sisällön ympäröivä OpenPGP-allekirjoitus. Samassa kentässä ei saa olla useita arvoja peräkkäin, vaan useampaa arvoa varten pitää olla useampi samanniminen kenttä, jollei toisin ole mainittu. Kenttiä voi olla rajaton määrä, ja *Internet Assigned Numbers Authority* (IANA) pitää niistä yllä rekisteriä (kuva 3). Vain "Contact" ja "Expires" ovat pakollisia kenttiä, ja kaikkia muita kenttiä tulee pitää valinnaisina. (Foudil & Shafranovich 2022, 4–5, 17.) Nämä perussäännöt ovat pysyneet pääosin samanlaisina, mutta ainoa useamman arvon listana salliva kenttä "Preferred-Languages" sekä OpenPGP-allekirjoittaminen on lisätty luonnoksessa 05. "Preferred-Languages"-kentän lisäksi ainoastaan "Expires"-kenttä saa esiintyä vain kerran.

Field Name	Description	Multiple Appearances	Status	Change Controller	Reference
Acknowledgments	link to page where security researchers are recognized	yes	current	IETF	[RFC9116]
Canonical	canonical URI for this file	yes	current	IETF	[RFC9116]
Contact	contact information to use for reporting vulnerabilities	yes	current	IETF	[RFC9116]
Expires	date and time after which this file is considered stale	no	current	IETF	[RFC9116]
Encryption	link to a key to be used for encrypted communication	yes	current	IETF	[RFC9116]
Hiring	link to the vendor's security-related job positions	yes	current	IETF	[RFC9116]
Policy	link to security policy page	yes	current	IETF	[RFC9116]
Preferred-Languages	list of preferred languages for security reports	no	current	IETF	[RFC9116]

Kuva 3. IANA:n rekisteri security.txt:n kentistä 2.6.2022 (IANA 2022).

Eri luonnosversioiden välillä kenttiä on lisätty ja poistettu sekä niiden määrittelyjä muutettu. RFC 9116 tuntee kentät "Acknowledgments", "Canonical",

"Contact", "Encryption", "Expires", "Hiring", "Policy" ja "Preferred-Languages", joista pakollisia ovat vain "Contact" ja "Expires". (Foudil & Shafranovich 2022, 6–8.) Näistä "Expires" on lisätty vasta luonnoksessa 09 ja tullut pakolliseksi luonnoksessa 10, joten voidaan olettaa, että "Expires" kenttä puuttuu niiltä, jotka ovat ottaneet security.txt:n käyttöön aiemmin. Luonnoksissa 09–11 aika esitettiin RFC 5322 -muodossa "Thu, 31 Dec 2021 18:37:07 -0800" (Resnick 2009, 14–15), mutta luonnoksessa 12 se vaihdettiin ISO 8601 -muotoon "2021-12-31T18:37:07z", jota tulisi RFC 3339:n mukaan käyttää Internet-protokollien aikaleimoissa (Klyne & Newman 2002), minkä vuoksi molempia saattaa esiintyä. Tähän jäi vielä julkaisuunkin virhe, sillä ISO 8601 -standardin UTC-aikaleiman tulisi päättyä isoon Z-kirjaimeen (Klyne & Newman 2002, 3).

"Expires"-kentässä olevan ajan suositellaan luonnoksesta 10 alkaen olevan korkeintaan vuoden tulevaisuudessa, mutta tämä ei ole pakollista. Kentän mukaan vanhentunutta tietoa ei sen sijaan tulisi enää pitää luotettavana (Foudil & Shafranovich 2022, 8, 14). "Contact" on ollut pakollinen alusta alkaen, mutta kenttien esiintymisen mukainen tärkeysjärjestys on tullut mukaan luonnoksessa 03, ja "mailto:" ja "tel:" URI-skeemojen pakollinen käyttö sähköpostiosoitteen ja puhelinnumeron edellä luonnoksessa 04. Varhain käyttöön otetuissa versioissa voi siten olla yhä puhelinnumeroita ja sähköpostiosoitteita sellaisinaan.

Monet valinnaiset kentät on määritelty siten, että niiden sisällön tulee olla salatun yhteyden yli haettava HTTPS-osoite. Tällaisia ensimmäisestä luonnoksesta asti säilyneitä valinnaisia kenttiä ovat "Acknowledgments" ja "Encryption", joista "Encryption"-kenttään on luonnoksesta 03 sallittu myös DNS URI-skeeman (Josefsson 2006) muotoinen viittaus "OPENPGPKEY"-tietueeseen sekä luonnoksesta 04 avaimen sormenjälki openpgp4fpr URI-skeemassa. Muita valinnaisia lisätietoja tarjoavan HTTPS-osoitteen sisältäviä kenttiä ovat "Hiring" luonnoksesta 03 ja "Policy" luonnoksesta 02.

Myös kenttä "Canonical" sisältää HTTPS-osoitteita, mutta kentällä viitataan tiedoston omaan osoitteeseen, ja sen tarkoituksena on estää tiedoston kopioimista muille sivustoille silloin, kun se on allekirjoitettu OpenPGP:llä; jonkin kentällä annetun osoitteen on vastattava osoitetta, josta tiedosto on haettu, eikä tiedoston muihin osoitteisiin tule luottaa, ellei saman sisältöinen tiedosto

ole saatavissa myös sieltä (Foudil & Shafranovich 2022, 6, 14). Kenttä lisättiin luonnoksiin 05–09 muodossa, jossa se sai esiintyä vain kerran, mutta määriteltiin luonnoksessa 10 uudestaan siten, että useampi rinnakkainen osoite on sallittu. Tämä helpottaa ylläpitäjien työtä, kun sama tiedosto voidaan lisätä useammalle sivustolle yhteisellä OpenPGP-allekirjoituksella.

Joistain kentistä on myös luovuttu. Ainoastaan luonnos 00 sisälsi kentän "Disclosure", joka sai sisältää vain kolme mahdollista arvoa: "Full", "Partial" tai "None". Kenttä "Policy" korvaa tämän sisällöllisesti, mutta mahdollistaa monipuolisempien käytäntöjen laatimisen. Vain luonnoksessa 04 oli mukana kenttä "Permission", joka sai esiintyä vain kerran ja jonka ainoalla mahdollisella arvolla "none" voitiin nimenomaisesti kieltää kaikenlainen tietoturvatestaus, mutta sen puutetta ei voinut pitää lupana testaukselle. Kentän poistoa ei ole erikseen perusteltu, mutta todennäköisesti sen merkitys on koettu vähäiseksi, koska kentän "Policy" avulla voidaan lupaa koskevat ehdot määritellä hyvinkin yksityiskohtaisesti. Ennen tiedostoon sisällytettyä OpenPGP-allekirjoitusta luonnokset 01–04 mahdollistivat tiedoston OpenPGP-allekirjoituksen erillisessä tiedostossa, joka tuli sijoittaa polkuun "/.well-known/security.txt.sig" ja johon tuli viitata "Signature"-kentässä. Luonnos 01 mahdollisti myös tiedostoon sisältyvän allekirjoituksen, jonka tuli olla tyhjän "Signature"-kentän jälkeen rivinvaihdolla eroteltuna.

Findlay & Abdou (2022, 3) tunnistavat tässä alaluvussa kuvatuista muutoksista kuuden olevan yhteensopimattomia aikaisempien versioiden kanssa, ja käyttävät niitä security.txt:n laatimiseen käytetyn luonnoksen tunnistamiseen:

- 12: ISO 8601 -aikaleimat "Expires"-kentässä.
- 11: RFC 5322 -aikaleimat "Expires"-kentässä.
- 09: Viimeinen versio, jossa "Expires"-kenttä on valinnainen.
- 05: Viimeinen versio, joka sallii HTTP URI:t.
- 04: Viimeinen versio, jossa "Signature"-kenttä.
- 02: Viimeinen versio, jossa yhteystietoja ilman URI-skeemaa.

Findlayn & Abdoun tutkimukseen syvennytään tarkemmin seuraavassa alaluvussa, jossa tutustutaan aiempiin tutkimuksiin security.txt-tiedoston käyttöönotosta. Koska aiheesta on tehty vasta kaksi tieteellistä tutkimusta, on vertailuun otettu myös yksi selvitys, jota ei ole julkaistu tieteellisessä julkaisukanavassa, mutta jonka metodologia ja tulokset on kuvattu huolellisesti.

3.3 Aiemmat tutkimukset security.txt:n käyttöönotosta

Security.txt:n käyttöönotosta on aikaisemmin julkaistu kaksi tieteellistä tutkimusta. Ensimmäisen tutkimuksen tekivät Poteat & Li (2021) keräten viidentoista kuukauden tarkastelujakson ajan Alexa Internetin sadantuhannen suosituimman verkkosivuston security.txt-tiedostot wget-työkaluun pohjautuvalla indeksoijalla ja laatien aineiston perusteella tilastoja niin tiedostojen esiintyvyydestä, eri kenttien esiintyvyydestä ja sisällöstä kuin yhteyden salauksestakin. Tiedostoja oli yritetty hakea kummastakin sallitusta polusta sekä HTTP- että HTTPS-yhteyksillä, ja niiden joukosta oli suodatettu vastaukset, jotka eivät olleet muodoltaan tekstitiedostoja, sillä monet vastaukset olivat virhesivuja (Poteat & Li 2021, 527).

Teknologian esiintyvyyttä arvioitaessa kaikki selkeät yritykset julkaista security.txt oli otettu huomioon, ja noin 7500 tiedoston joukosta löytyi noin 2200 uniikkia tiedostoa kertoen, että toisinaan useampi saman organisaation verkkotunnus jakoi samat yhteystiedot (Poteat & Li 2021, 528–529). Tutkimuksessa sisällön validointi tehtiin luonnosversion 11 mukaan (Poteat & Li 2021, 532), mutta teknisistä eroista julkaistuun RFC 9116:n verrattuna ainoastaan "Expires"-kentän päivämäärän muodon vaihtuminen on merkittävä. Keskeisenä tuloksena oli, että yhteensopivuudessa oli huomattavia puutteita: erityisesti pakollinen "Expires"-kenttä löytyi vain 1,7 %:sta tiedoston julkaisseista verkkotunnuksista, kun toinen pakollinen "Contact"-kenttä löytyi 93,1 %:sta (Poteat & Li 2021, 529–530). Myös omia määrittelyn ulkopuolisia kenttiä löytyi: "OpenBugBounty"-kentällä (12,3 %) oli korvattu "Contact"-kenttä 92 %:ssa niistä tapauksista, joissa se puuttui, ja "Signature"-kenttää esiintyi 2,5 %:ssa (Poteat & Li 2021, 529). Jälkimmäinen oli jäänne luonnoksista 01–04, mutta tätä ei ollut tunnistettu, sillä tutkimus ei huomioinut vanhempien luonnosversioiden vaikutusta.

Findlay & Abdou (2022) laajensivat tutkimusta miljoonaan suosituimpaan verkkotunnukseen käyttäen toista verkkotunnuslistaa ja edellisestä tutkimuksesta poikkeavaa metodologiaa, joka on tarkempi sen suhteen, miten julkaistut tiedostot noudattavat RFC 9116:n määritelmiä, sekä tunnistaa, minkä luonnos-

versioiden kanssa julkaistu tiedosto on yhteensopiva. Tutkimus perustuu pistemäiseen otantaan, joten se ei edes pyri vastaamaan siihen, mihin suuntaan käyttöönotto on etenemässä. Käytetty lista, TRANCO, on erityisesti tutkimuskäyttöön kehitetty ja paremmin manipuloinnilta suojattu verkkotunnuslista (Pochat ym. 2019). Vaikka Poteat & Lin (2021, 528) tutkimuksessa tietoja oli kerätty pitkältä seurantajaksolta, siinä tunnistettiin, ettei tiedoista voitu luotettavasti muodostaa trendejä, sillä suosituimpien sivustojen lista haettiin jokaisella kerralla erikseen, jolloin tutkimuksen kohdejoukko muuttui kesken seurannan. Siten kumpikaan tutkimus ei vastaa kysymyksiin siitä, kasvaako käyttöönotto, eivätkä laadi ennusteita sen etenemisestä.

Findlay & Abdoun (2022, 3) tutkimuksessa käytettiin kahta sitä varten erikseen tuotettua Rust-ohjelmointikielellä kirjoitettua työkalua, joista toinen, "SecMap", kerää HTTPS-yhteyksillä security.txt-tiedostot kummastakin tuetusta sijainnista ja karsii joukosta yhteyksiin liittyviä virheitä, kuten virheellisiä varmenteita, virheeseen viittaavia HTTP-tilakoodeja (*response status code*) ja väärän muotoisia vastauksia, ja toinen, "sectxt.rs", analysoi tiedoston sisällön ja kertoo siihen liittyvät puutteet. Kummankin työkalun kerrotaan olevan avointa lähdekoodia, ja alaviitteissä on linkit GitHub-projektiin, mutta projektia ei ollut enää saatavilla, mikä on harmillista, sillä työkalut olisivat olleet mahdollisesti hyödyllisiä myös tässä tutkimuksessa. Työkalujen toiminnan kuvaus antoi kuitenkin arvokkaita näkökulmia omien analyysityökalujen kehittämiseen.

Findlay & Abdou (2022, 5–6) havaitsivat, että kahdesta miljoonasta kyselystä (miljoonan verkkotunnuksen kahteen polkuun) noin 8300 vastasi tekstitiedostolla, joista 2500 tuotti virheitä ja 5800 oli oikean muotoisia. Duplikaattien poiston jälkeen tulos oli, että 4864 verkkotunnusta sisälsi security.txt:n jommassakummassa polussa, mikä vastaa noin 0,49 % kaikista verkkotunnuksista. Suosituimmilla sivustoilla oli tyypillisemmin security.txt kuin vähemmän suosituilla, ja kun tarkastelua rajattiin vain sadantuhannen suosituimman sivuston osajoukkoon, saatiin tulokseksi 1,6 % (Findlay & Abdou 2022, 6). Tämä on huomattavasti pienempi kuin Poteat & Li (2021, 528–529) saama noin 7,5 % kattavuus saman suuruisella joukolla, mutta tulosta selittää tiukempi metodologia, joka ei laskenut mukaan suojaamattomilla yhteyksillä, virheellisillä varmenteilla tai väärällä MIME-tyypillä varustettuja latauksia.

Tutkittaessa security.txt:n käytön suosiota ja kehitystä molempien tutkimuksen tarkastelutavat ovat hyödyllisiä, sillä löyhempi tulkinta kertoo paremmin käyttöönoton halukkuudesta, kun taas tiukempi tulkinta käyttöönoton laadusta ja julkaistujen tietojen luotettavuudesta. Tässä tutkimuksessa on yhdistetty kumpaakin tarkastelutapaa, sillä molempia tarvitaan, jotta kaikkiin tutkimuskysymyksiin voidaan vastata. Halukkuus ottaa teknologia käyttöön kertoo sen yleistymispotentiaalista, ja laatu puolestaan siitä, minkälaisia tiedotus- ja koulutustarpeita käytänteen ympärille tarvitaan, jotta siitä saadaan mahdollisimman hyödyllinen. Myös RFC 9116:n laatijat ottavat kantaa siihen, että implementoijien tulisi olla konservatiivisia omassa toteutuksessaan, mutta vapaamielisempiä sen suhteen, mitä sallivat muilta (Foudil & Shafranovich 2022, 5).

Tietoturvayhteystietoja etsivät ovat realistisesti sen varassa, mitä on tarjolla, jolloin heikompikin julkaisu on todennäköisesti parempi kuin ei julkaisua lainkaan. Luotettavuuden arvioinnin näkökulmasta turvallisuusnäkökohtien tunteminen ja niiden ottaminen huomioon omaa security.txt-tiedostoa laatiessa helpottaa tiedostoa käyttävien työtä, sillä silloin tarvitaan jo lähtökohtaisesti vähemmän niiden peilaamista muihin lähteisiin, kuten saatavissa oleviin historiatietoihin, WHOIS-tietokannassa oleviin yhteystietoihin, kotisivuilla julkaistuihin yhteistietoihin tai Suomessa esimerkiksi Yritys- ja yhteisötietojärjestelmään.

Findlay & Abdou (2022, 6–9) tekivät keräämistään tiedoista myös muita mielenkiintoisia havaintoja ja tilastoja, kuten "Preferred-Languages" kentässä listattujen kielten esiintyvyyttä suhteessa toisiinsa, väärin kirjoitettujen kentännimien yleisyyttä, "Expires"-kentässä ilmoitetun ajan vaihteluväliä sekä "Contact"-kentässä ilmoitettujen ulkopuolisten verkkotunnuksen yleisyyttä: esimerkiksi haavoittuvuuskoordinaatioalusta HackerOne.com esiintyi jopa 6,7 %:ssa yhteystiedoista. Monipuoliset lisähavainnot auttoivat ymmärtämään ilmiöön liittyviä yksityiskohtia ja vaikuttivat siten myös tämän tutkimuksen metodologiaan liittyviin valintoihin.

Findlay & Abdou (2022, 10) esittävät tutkimuksensa johdosta suosituksia sekä käyttöönottoon että käytännön itsensä kehittämiseen. Heidän käyttöönottosuosituksensa perustuvat siihen, miten tietojen julkaisusta saadaan mahdollisimman hyödyllisiä käyttötarkoitukseensa nähden. Niissä kiinnitetään huomiota

RFC 9116:n tuntemukseen, myös sen suositusten noudattamiseen minimivaatimusten lisäksi, suhteellisen lyhyen vanhenemisajan määrittämiseen sekä siihen, että tiedoston tulisi olla koneellisesti luettavissa. Käytännön kehittäjiltä he puolestaan toivovat suosittujen lisäkenttien, kuten "OpenBugBounty" ottamista mukaan standardiin, tärkeiden kenttien "Preferred-Languages" ja "Canonical" muuttamista pakolliseksi, "Preferred-Languages"-kentän poikkeavan muodon yhtenäistämistä muiden kenttien kanssa, "Acknowledgments" useampien kirjoitusasujen sallimista, sekä koneluettavuuden suosimista luettavuuden kustannuksella. Näen näiden suositusten olevan osittain tavoitteiltaan ristiriitaisia, sillä useamman muodon hyväksyminen monimutkaistaa koneellista lukemista, ja RFC 9116:n noudattaminen puolestaan vähentää sellaisten tarvetta. Lisäksi olen samoilla linjoilla kuin Poteat & Li (2021, 50) siitä, etteivät yksittäiset alustat, kuten "OpenBugBounty" tarvitse omia lisäkenttiään, koska useamman, järjestyksellä priorisoidun "Contact"-kentän käyttäminen näiden alustojen URL-osoitteille kattaa jo kaikki tällaiset käyttötarkoitukset.

Näiden lisäksi on tehty epävirallista tutkimusta, jossa ei ole käytetty tieteellistä julkaisukanavaa. Antoine Neuenschwander haki Nuclei-sovelluksen avulla security.txt-tiedostot kaikilta sveitsiläisiltä ch-verkkotunnuksilta, ja Edwin Foudil analysoi hänen keräämänsä tiedot löytäen 1310 tiedostoa, joista 535 oli uniikkeja (Foudil 2022). 2,4 miljoonan ch-verkkotunnuksen joukossa tämä vastasi vain noin 0,06 %:n kattavuutta. Nuclei on haavoittuvuuksien skannaukseen tarkoitettu työkalu, jolla toi tehdä HTTP-, DNS-, TCP- ja tiedostojärjestelmäskannauksia YAML-pohjaisten mallien kautta (ProjectDiscovery Inc. s.a.). Vaikka metodologia on dokumentoitu kevyesti, havainnot perustuvat kahden eri henkilön itsenäiseen työskentelyyn, analyysi on suppeaa ja siltä puuttuu systemaattisuus, on aineistonkeruumenetelmä julkaistu avoimena lähdekoodina ja analyysimenetelmä kuvattu selkeästi, mikä mahdollistaa toistettavuuden. Tutkimusjoukkona oli kokonainen maakohtainen ylätason verkkotunnus, millä oli mahdollisuus tarjota kattavampi poikkileikkaus kuin suosituimpien verkkotunnusten listoihin perustuvilla tutkimusjoukoilla.

Yksinkertaisen esiintymistiheyden lisäksi Foudil (2022) havaitsi, että monet verkkosivustojen julkaisualustat lisäävät oman, yhteisen security.txt-tiedostonsa kaikille alustalla julkaistuille verkkotunnuksille; esimerkiksi Tumblr-käyt-

täjien sivustojen security.txt-yhteystiedot osoittavat haavoittuvuuspalkkio-ohjelmaan HackerOne-sivustolla. Tällöin löydökset eivät tule sen tahon tietoon, jonka sivustosta on kyse, mutta toisaalta julkaisualustan haavoittuvuudet koskevat monesti kokonaista julkaisualustaa ja siten kaikkia sen käyttäjiä.

Suhteessa aiempiin tutkimuksiin tämä organisaatioiden fi-verkkotunnuksia koskeva tutkimus tuottaa uutta tietoa, koska se tarjoaa kattavan poikkileikkauksen kaikkiin yhden maakohtaisen ylätason verkkotunnuksen alla oleviin verkkotunnuksiin, jotka ovat jonkin organisaation hallinnassa. Tutkimusmenetelmäksi valittu kokonaistutkimus mahdollistaa myös ajallisen vertailun, joka jäi puuttumaan Findlayn & Abdoun (2022) tutkimuksesta. Poteatin & Lin (2021) tutkimuksesta poiketen menetelmä sallisi myös trendien ennustamisen, mutta käytännössä havaittu satunnainen vaihtelevuus heikentää sitä toistaiseksi.

3.4 Julkaistun tiedoston validointi

RFC 9116:n luonnosversioiden eroavaisuuksien analysointi ja aiempien tutkimusten havainnot kaikkien vaatimusten yhtäaikaisen täyttämisen harvinaisuudesta osoittavat, että tutkimuksessa on tarpeen pitää mukana kaksi näkökulmaa: Teknologian yleistymisen arvioimiseksi on syytä tilastoida mahdollisimman sallivasti kaikki verkkotunnukset, joissa security.txt on yritetty ottaa käyttöön. Jotta puolestaan saadaan tietoa siitä, minkä vaatimusten täyttämisessä on eniten ongelmia, tulee yksittäisten kriteerien täyttymistä arvioida myös erillään muista kriteereistä sen sijaan, että todettaisiin toteutuksen pelkästään olevan joko RFC 9116:n mukainen tai siitä poikkeava.

3.4.1 Vähimmäisvaatimukset implementointiyrityksen tunnistamiseksi

Jotta voitaisiin tunnistaa selkeät yritykset ottaa security.txt käyttöön joko RFC 9116:n tai sen luonnosvaiheiden tarkoittamalla tavalla, pitää tunnistaa, milloin palvelimen vastaus on oikean suuntainen. Luonnosversioiden poikkeavista määritelmistä johtuen tiedostoa tulee etsiä kahdesta polusta: sekä "/.well-known/security.txt" että yhä yhteensopivuussyistä sallitusta "/security.txt". Tiedostoa tulee hakea myös sekä HTTPS- että HTTP-yhteyksillä. Koska kaikilla verkkotunnuksilla ei välttämättä ole verkkosivustoa suoraan verkkotunnuksen juuressa, tulee jokaisesta verkkotunnuksesta tarkastaa myös "www"-aliverkkotunnus. Kaikkien näiden yhdistelmistä saadaan yhteensä kahdeksan polkua,

jotka ominaisuuksineen on listattuna taulukossa 1. Jatkossa polun mukaan eritellyissä taulukoissa viitataan polkuihin ensimmäisen sarakkeen numeroilla.

Taulukko 1. Vaihtoehtoiset polut, joista security.txt-tiedosto voi löytyä.

	Salaus	**Polku**	**Isäntänimi**	**Esimerkki URL-osoitteesta**
1	TLS-salaus (HTTPS)	ensisijainen	juuri	https://example.com/.well-known/security.txt
2			www	https://www.example.com/.well-known/security.txt
3		yhteensopiva	juuri	https://example.com/security.txt
4			www	https://www.example.com/security.txt
5	Ei salausta (HTTP)	ensisijainen	juuri	http://example.com/.well-known/security.txt
6			www	http://www.example.com/.well-known/security.txt
7		yhteensopiva	juuri	http://example.com/security.txt
8			www	http://www.example.com/security.txt

HTTP-tilakoodin tulisi olla 200 onnistuneen yhteyden merkiksi (Fielding ym. 2022, 127). Vaikka RFC 9116 salliikin edelleenohjaukset (HTTP-tilakoodit 301 ja 302), on niiden seuraamisen kanssa kehotettu ylimääräiseen analyysiin ja erityiseen tarkkaavaisuuteen (Foudil & Shafranovich 2022, 10, 14). Tämä on tutkimusasetelman kannalta ongelmallista, koska jokainen tapaus pitäisi tutkia erikseen, ja on haastavaa määritellä yksiselitteisiä kriteerejä esimerkiksi sille, kuinka moneen peräkkäiseen edelleenohjaukseen raja tulisi vetää. Edelleenohjausta voidaan käyttää hyvin moniin eri tarkoituksiin. Pyyntöjä voidaan ohjata saman verkkotunnuksen juuresta www-aliverkkotunnukseen tai päinvastoin, jotta hakukoneet eivät näkisi sivustoa kahtena (Jones 2008, 71), HTTP-yhteydestä HTTPS-yhteyteen salauksen pakottamiseksi tai kokonaan toiseen verkkotunnukseen. Jokainen polku voidaan ohjata vastaavaan polkuun toisessa isäntänimessä tai kaikki polut toisen isäntänimen juureen, mutta kyseessä voi olla myös monimutkaisempi kokonaisuus, kuten hakukoneoptimoinnissa, kun vanha sivurakenne korvataan uudella (Jones 2008, 84). Kun otetaan kaikki nämä tapaukset huomioon, on lisäksi vaikea käsittää, miten Findlay & Abdou (2022, 6) havaitsivat vain noin 0,1 %:n vastauksista olleen edelleenohjauksia.

Edelleenohjausten moninaisuuden seasta on vaikea havaita, milloin kyseessä on ollut yritys julkaista security.txt-tiedostoa, sillä kaikkien polkujen ohjaaminen vastaavaan polkuun voi osua toisen verkkotunnuksen security.txt-tiedos-

toon myös tahattomasti, tai täysin erilaisen polun takaa saattaisi löytyä tietoturvayhteystiedoksi tarkoitettua sisältöä. Tällaisen kirjon analysointi vaatisi tapauskohtaista tarkastelua, mikä ei ole realistista satojen tuhansien HTTP-pyyntöjen aineistossa, ja toisi esiin lähinnä poikkeuksellisia yksittäistapauksia. Edelleenohjauksiin kohdistuva tarkempi analyysi tuskin auttaisi vastaamaan tämän tutkimuksen tutkimuskysymyksiin, ja on sen vuoksi rajattu tutkimuksen ulkopuolelle. Tästä aiheutuvaa epätarkkuutta poistetaan kuitenkin sillä, että taulukossa 1 listattujen vaihtoehtoisten polkujen läpikäyminen edelleenohjauksia seuraamatta tuottaa saman tuloksen, mikäli saman verkkotunnuksen sisällä edelleenohjaava sivusto julkaisee tiedoston yhdessäkin näistä poluista.

Jotta vastauksen voitaisiin tulkita olevan tekstitiedosto, tulisi sen MIME-tyypin olla "text/plain", mikä on myös RFC 9116:n vaatimus (Foudil & Shafranovich 2022, 10). Tämä tieto löytyy HTTP-otsakkeesta "Content-Type" (Fielding ym. 2022, 57), josta tulisi löytyä kirjainkoosta riippumattomat "text/plain" sekä mahdollinen käytetty merkistö "charset=utf-8". Vaikka RFC 9116 edellyttää UTF-8-merkistön käyttöä (Foudil & Shafranovich 2022, 10), on tässä niin helppo tehdä virhe, että vähimmäisvaatimuksen tasolla voidaan muutkin merkistöt hyväksyä. Sen sijaan muut sisältötyypit on syytä hylätä jo senkin vuoksi, että jotkut sivustot saattavat tarjota esimerkiksi HTML-muotoista sisältöä tai jopa kuvatiedostoja polusta riippumatta, minkä Poteat & Li (2021, 527) havaitsivat.

Jotta kyseessä olisi oikean muotoinen kenttiä ja niiden arvoja sisältävä tiedosto, tulisi sieltä kuitenkin löytyä vähintään yksi rivi, jolla esiintyy kaksoispistettä seuraava välilyönti (": "), jonka edellä on yksi vain kirjaimista koostuva kentän nimi, ja jonka jälkeen on tekstisisältöä tämän kentän arvona. Edes alusta asti pakollisena mukana ollutta "Contact"-kenttää ei voida kuitenkaan käyttää implementaatioyrityksen vähimmäisvaatimuksena, sillä esimerkiksi määrittelemätöntä "OpenBugBounty"-kenttää on selittämättömästä syystä käytetty hyvin usein sen sijasta (Poteat & Li 2021, 529; Findlay & Abdou 2022, 9).

3.4.2 RFC 9116:n noudattaminen tiedoston haussa

RFC 9116 määrittelee, että HTTP-palvelun security.txt on haettava TLS-salatulla yhteydellä, mutta sallii kuitenkin juurihakemiston "/security.txt" käyttämisen "/.well-known/security.txt" polun lisäksi, kunhan sitä käytetään vain silloin,

mikäli varsinaisesta polusta ei löydy tiedostoa (Foudil & Shafranovich 2022, 9–10). Aliverkkotunnukset ("www") on lisätty tutkittavien osoitteiden listalle RFC 9116:n määritelmistä riippumattomasta syystä, sillä security.txt koskee aina sitä isäntänimeä, jonka alta se löytyy, ja joidenkin organisaatioiden verkkosivujen kanoniset osoitteet ovat muotoa "https://www.example.com/", johon verkkotunnuksen juuresta saattaa olla HTTP-edelleenohjaus. Kun tutkitaan RFC 9116:n noudattamista tiedoston haussa, analyysi tulee rajata aineiston osajoukkoon, joka koostuu vain salatulla yhteydellä noudetuista tiedostoista. Näitä vastaavat taulukon 1 polut 1–4, jotka on myös esitetty siinä järjestyksessä, jossa niitä RFC 9116:n mukaisesti tulisi etsiä.

Turvallisuusnäkökohtien mukaan palvelimen identiteetti tulee todentaa RFC 6125:n mukaisesti (Foudil & Shafranovich 2022, 15). Tämä tarkoittaa, että TLS-yhteyden varmenteen tulee olla varmentajan allekirjoittama ja HTTPS-protokollan tapauksessa sen *Subject Alternative Name* (SAN) -laajennoksen (tai sen puuttuessa "Common Name" -kentän) tulee vastata pyydettyä isäntänimeä (Saint-Andre & Hodges 2011). Isäntänimen tulisi kuitenkin vastata nimenomaisesti SAN-laajennoksen mukaista DNS-ID-tunnistetta, ja niin kutsutut "*wildcard*"-varmenteet eli esimerkiksi "*.example.com" tulisi sallia (Foudil & Shafranovich 2022, 15–16), vaikka näillä tosin on tutkimustavasta johtuen merkitystä vain silloin, jos vastaus saadaan pelkästään aliverkkotunnuksesta "www.example.com".

Sisällön tyypin ilmaisevan HTTP-otsakkeen "Content-Type" sisällön tulisi olla "text/plain; charset=utf-8", mutta sekä otsakkeen nimi että sisältö ovat kirjainkoosta riippumattomia, välilyönti puolipisteen jälkeen on valinnainen, ja merkistö saa olla lainausmerkeissä tai ilman (Fielding ym. 2022, 57–58). Tämän vuoksi esimerkiksi seuraavat muodot ovat hyväksyttäviä:

- text/plain; charset="utf-8"
- text/plain;charset=utf-8
- text/plain; charset=UTF-8
- text/plain;charset="UTF-8"
- Text/PLAIN;Charset="utf-8"

Vaikka tiedoston merkistö on tiedoston itsensä ominaisuus, on RFC 9116:ssa määritelty siihen viittaaminen osaksi tiedoston sijainnin otsaketietoja (Foudil & Shafranovich 2022, 10). Otsake onkin luontevampi paikka, koska kyseessä on

tiedostoa itseään koskeva metatieto. Seuraavassa alaluvussa tarkastelu kääntyy siihen, millaisia määritelmiä RFC 9116 asettaa tiedostojen sisällöille.

3.4.3 RFC 9116:n noudattaminen tiedoston sisällössä

Jotta tiedoston sisältö noudattaisi RFC 9116:n vaatimuksia, siinä tulisi olla vähintään yksi "Contact"-kenttä, jonka sisältönä on URI ("https:", "mailto:" tai "tel:"), sekä tasan yksi "Expires"-kenttä, jonka sisältönä on päivämäärä ISO 8601:n mukaisessa muodossa, ja joka ei ole havainnointihetkellä menneisyydessä (Foudil & Shafranovich 2022, 7–8).

Pakollisten kenttien lisäksi tiedoston tulisi sisältää vain muita RFC 9116:ssa erikseen sallittuja elementtejä, jotka on dokumentoitu tarkemmin *augmented Backus–Naur form* (ABNF) -muodossa (Foudil & Shafranovich 2022, 10–14), ja jotka tiivistetysti ovat seuraavia:

- Kommenttirivejä, jotka alkavat "#"-merkillä.
- Tyhjiä rivejä.
- IANA:n (2022) rekisteristä löytyviä kenttiä rekisterin ja RFC 9116:n määrittelemällä sisällöllä.
- Edellä mainittua sisältöä saa ympäröidä OpenPGP:n selväkielisen tekstin allekirjoitus, joka koostuu seuraavista osista:
    1. "-----BEGIN PGP SIGNED MESSAGE-----"
    2. "Hash: algoritmi" ja tyhjä rivi.
    3. Muu security.txt:ssä sallittu sisältö.
    4. "-----BEGIN PGP SIGNATURE-----"
    5. Allekirjoituksessa sallitut otsakkeet ja tyhjä rivi.
    6. BASE64-muotoinen allekirjoitus.
    7. "-----END PGP SIGNATURE-----"

Tiedossa esiintyvien valinnaisten kenttien tulisi noudattaa niille annettuja vaatimuksia: "Preferred-Languages" saisi esiintyä vain kerran ja sen tulisi sisältää vähintään yksi kieli pilkuin eroteltuna listana, jossa kielet on esitetty RFC 5646:ssa (Phillips & Davis 2009) tarkoitetuilla tunnisteilla. Kenttien "Acknowledgments", "Canonical", "Hiring" ja "Policy" tulisi sisältää vain HTTPS-osoitteita. "Encryption"-kentät saisivat sisältää "https:"-, "dns:"- ja "openpgp4fpr:"-osoitteita. Lisäksi mikäli tiedostossa on käytetty "Canonical"-kenttiä, yhden niissä esitetyn osoitteen tulisi vastata osoitetta, josta tiedosto on noudettu, ja tiedoston tulisi olla OpenPGP-allekirjoitettu. (Foudil & Shafranovich 2022, 6–8.)

Koska mahdollisia virheitä on paljon, tutkimuskysymyksiin vastaamiseksi on parempi analysoida ja tilastoida niiden esiintyvyyttä yksittäin kuin vaatia tiedostolta täyttä RFC 9116:n mukaisuutta. OpenPGP-allekirjoituksesta voidaan tällä tasolla vaatia oikeaa syntaksia ja sisällön vastaavuutta allekirjoituksen kanssa, mutta avaimen identiteettiin ei voida ottaa kantaa. Seuraavassa alaluvussa käsitellään tarkemmin sitä, miten OpenPGP-allekirjoitusten luotettavuutta ja muita RFC 9116:n turvallisuusnäkökohtia (Foudil & Shafranovich 2022, 14–16) voidaan ottaa huomioon tiedostoa julkaistaessa, ja mitä piirteitä tiedostosta voidaan tarkastella arvioitaessa, miten tässä on onnistuttu.

3.4.4 RFC 9116:n vaatimuksia korkeampi, turvallinen käytäntö

Tietoturvatutkijoiden työtä voidaan helpottaa laatimalla security.txt-tiedostosta edellä tarkasteltuja sisällöllisesti muotoon perustuvia minimivaatimuksia laadukkaampi. Tällöin tiedoston laatija ottaa käytännössä valmiiksi huomioon niitä turvallisuusnäkökohtia, joita tiedoston käyttäjää kehotetaan tarkastelemaan. Turvallisuusnäkökohdissa kiinnitetään huomiota muun muassa vääristeltyyn ja vanhentuneeseen sisältöön, joista kummatkin heikentävät tiedoston käytettävyyttä (Foudil & Shafranovich 2022, 14–15). Näiden huomioimiseksi turvallisessa käytännössä tiedosto on allekirjoitettu OpenPGP:llä, käytettyyn OpenPGP-avaimeen viitataan useammalla tavalla, joiden voidaan toisistaan riippumattomasti varmentaa liittyvän kohdeorganisaatioon (luotettu HTTPS-yhteys, DNSSEC-varmennettu DNS, avaimen sormenjäljen julkaiseminen), ja "Expires"-kentässä käytetään suosituksen mukaisesti aikaleimaa, joka on korkeintaan vuoden tulevaisuudessa, mikä ohjaa yhteystietojen säännölliseen tarkastamiseen. Myös useamman vaihtoehtoisen yhteydenottokanavan tarjoaminen voi lisätä sekä tavoitettavuutta että tiedon turvallisempaa välittämistä. Minimivaatimuksia turvallisempien käytäntöjen esiintyvyyttä voidaan tarkastella ja tilastoida yksittäisinä yksityiskohtina, muttei juurikaan kokonaisuutena, mikäli ei ole ensin annettu virallista suositusta tai mallia, johon organisaatiot voisivat tukeutua. Suomessa tällaista ei ole vielä laadittu, ja esimerkiksi Kyberturvallisuuskeskuksen ohje (Traficom 2020) mainitsee tiedoston vain lyhyesti.

OpenPGP-allekirjoitusten osalta on huomioitava, että allekirjoitukseen käytetyn avaimen haltijan identiteetin varmentamiseen ei ole saatavissa täysin automatisoitua tapaa. OpenPGP:n ympärillä ei ole keskitettyä julkisen avaimen

infrastruktuuria, vaan varmentaminen perustuu luottamusverkkoon, jonka avaimien välillä ei läheskään aina löydy katkeamatonta polkua (Ulrich ym. 2011, 490, 504). Tämän vuoksi luottamusverkon sijasta OpenPGP-avaimen luotettavuutta arvioidaan nykyisin usein sen lähteen perusteella: mikäli avain on ladattu esimerkiksi salatussa yhteydessä organisaation sivustolta tai sen sormenjälki vastaa henkilökohtaisessa kontaktissa annettua esimerkiksi käyntikorttiin painettua sormenjälkeä, on se yleensä luotettavampi kuin avainpalvelimelta haettu avain, jonka on voinut julkaista kuka tahansa. Tällaisia varmistuksia voidaan myös tehdä useammasta kuin yhdestä lähteestä.

RFC 9116 tarjoaa mahdollisuuden viitata OpenPGP-avaimiin "Encryption"-kentässä. Tiedosto voi olla allekirjoitettu tällä samalla avaimella, mutta tämä tarjoaa parhaimmillaankin kehäpäätelmän, jossa tiedostoon luotetaan, koska se on allekirjoitettu avaimella, johon puolestaan luotetaan, koska se on mainittu tiedostossa. Vaikka näiden tietojen vertaaminen ristiin olisi mahdollista, se ei tuottaisi pätevää tietoa avainten luotettavuudesta, vaan jokainen esiintymä pitäisi tarkastella huomattavasti monimutkaisempana yksittäistapauksena, jotka eivät olisi keskenään vertailukelpoisia. Tämän vuoksi OpenPGP-allekirjoitusten luotettavuuden arviointi rajattiin tutkimuksen ulkopuolelle.

4 TUTKIMUKSEN TOTEUTUS

4.1 Työkalujen valmistelu ja testaus aineistonkeruuta varten

Tutkimusaineiston keräämiseen käytettäviä työkaluja valmisteltiin syys–lokakuussa 2022. Työkalujen kehittämistä ohjasivat sekä valitut tutkimusmenetelmät että teoreettisen viitekehyksen tuoma ymmärrys siitä, minkälaista aineistoa tarvittiin tutkimuskysymyksiin vastaamiseksi. Kehityksen ohella työkaluja testattiin pienellä, ennalta tunnetulla harjoitusaineistolla, jotta eri välivaiheiden tiedostojen muoto ja rakenne olivat selvillä. Tämä mahdollisti sen, että seuraavan vaiheen työkalu kykeni virheittä käyttämään edellisen vaiheen ulostuloa syötteenään. Lisäksi työkaluille tehtiin suorituskyky- ja luotettavuustestausta, joilla voitiin varmistua siitä, että aineisto saadaan kerättyä kohtuullisessa ajassa ilman havainnointivirheitä, jotka vaikuttaisivat tulosten luotettavuuteen.

Luotettavuustestauksessa havaittiin ZGrab 2.0:n tuottavan melko paljon aikakatkaisuja erityisesti HTTPS-yhteyksillä. Aikakatkaisusta johtuvien ongelmien

ratkaisemiseksi ja niistä aiheutuvien vääristymien minimoimiseksi tehtiin vertailuanalyysiä eri mittaisten aikakatkaisujen sekä yhtäaikaisten yhteyksien lukumäärien välillä. Erityisesti HTTPS-kättelyyn ja varmenteiden luotettavuuden arviointiin tuli varata aikaa yksinkertaisia salaamattomia HTTP-yhteyksiä enemmän. Vertailussa aloitettiin pienillä määrillä yhtäaikaisia yhteyksiä ja pitkillä, 90 sekunnin aikakatkaisuilla. Yhtäaikaisten yhteyksien määrää kasvatettiin ja aikakatkaisuja lyhennettiin niihin arvoihin asti, joilla ei aiheutunut vielä muutoksia tuloksiin. Turvallisiksi arvoiksi osoitettiin tällä tavoin 75 yhtäaikaista yhteyttä siten, että HTTPS-yhteyksissä aikakatkaisuksi asetettiin 25 sekuntia ja HTTP-yhteyksissä 5 sekuntia. Lisäksi mahdollisiin verkkokatkoksiin varauduttiin siten, että aineistonkeruuta olisi voitu jatkaa uudestaan siitä kohdasta, jossa katkos olisi tulosten puutteesta tai huomattavasta vähenemisestä voitu päätellä.

Luotettavuustestaus paljasti myös nimipalveluun liittyviä ongelmia silloin, kun ZGrab 2.0:lle annettiin syötteenä pelkkä verkkotunnuslista, jolloin sen oli selvitettävä kohdepalvelimien IP-osoitteet itse ennen yhteyden muodostamista. ZGrab 2.0 yrittää selvittää IP-osoitetta vain kerran, ja nimipalvelun toiminnasta johtuen vastaus ei välttämättä ole heti saatavilla. Ongelma ratkaistiin selvittämällä kaikki isäntänimien IP-osoitteet etukäteen MassDNS-työkalulla (Blechschmidt s.a.), joka jatkaa saman nimen kyselyä, kunnes saa vastauksena joko IP-osoitteen tai varman tiedon siitä, ettei isäntänimelle löydy IP-osoitetta. Oletusasetuksilla työkalu etsi samaa isäntänimeä 50 kertaa, mutta varmuuden vuoksi tämä vielä kaksinkertaistettiin sataan yritykseen.

4.2 Tutkimusaineiston kerääminen

Tutkimusaineisto kerättiin kaksi kertaa samalla menetelmällä ensin lokakuussa 2022 ja sitten helmikuussa 2023. Tarkempi aikataulu ilmenee taulukosta 2. Alkuperäisen tutkimussuunnitelman mukaan aineiston kerääminen oli tarkoitus suorittaa puolen vuoden seurantajaksolla noin kuuden viikon välein, mutta ensimmäisen aineiston analyysin tuloksista ilmeni, ettei näin tiheä tarkasteluväli ollut tarkoituksenmukainen, vaan seurantaa tulisi aluksi tehdä noin puolen vuoden välein ja tihentää vaiheittain käyttöönoton mahdollisesti yleistyessä. Tämä ei ollut opinnäytetyön tavoitteiden vastaista, sillä tarkoituksena oli tutkimustulosten saamisen ohella tuottaa työkalut pidempiaikaiseen ilmiön

seurantaan, jonka alun perinkin tiedostettiin rajautuvan ajallisesti opinnäytetyön ulkopuolelle. Opinnäytetyö tarjosi tähän viitekehyksen, joka valoi tutkimukselle tieteellisen perustan ja varmisti siten työkalujen laadun ja luotettavuuden.

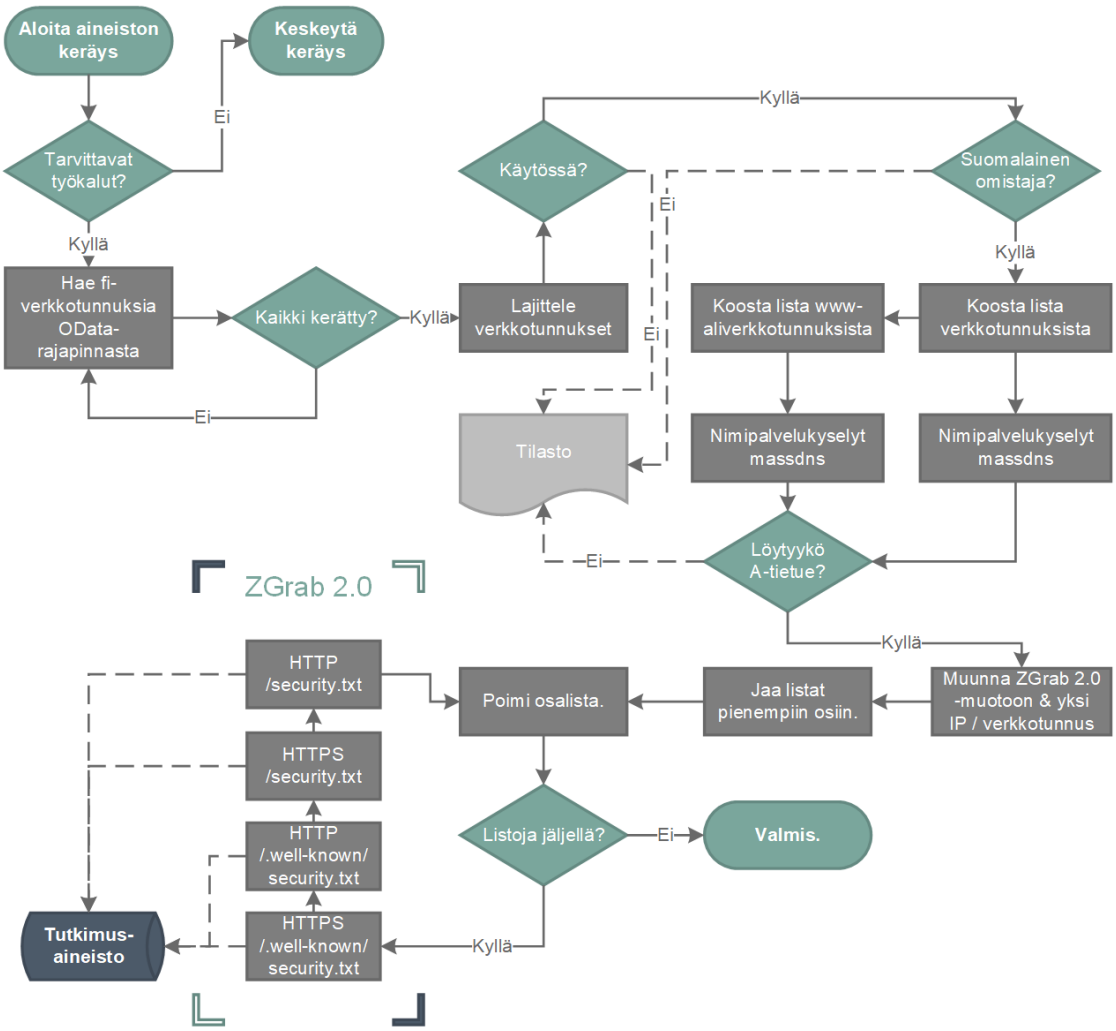
Taulukko 2. Aineistonkeruun toteutunut aikataulu.

	Lokakuu 2022	Helmikuu 2023
Verkkotunnukset OData-rajapinnasta	12.10.2022 09:55	20.2.2023 06:40
Nimipalvelukyselyt	12.10.2022 11:05	20.2.2023 06:55
Aineistonkeruun aloitus	12.10.2022 15:12	20.2.2023 07:04
Siirtyminen www-aliverkkotunnukseen	15.10.2022 06:18	22.2.2023 23:08
Aineistonkeruu valmis	18.10.2022 01:57	25.2.2023 15:20

Verkkotunnusten haku OData-rajapinnasta (Viestintävirasto 2017) toteutettiin Bash-komentosarjalla, joka hyödyntää curl-komentoa tiedon hakuun ja jq-komentoa (Dolan s.a.) linkin seuraamiseen sekä haettujen tietojen yhdistämiseen yhdeksi JSON-tiedostoksi. Komentosarjan suoritus kesti noin kaksikymmentä minuuttia. Kerätyistä verkkotunnuksista rajattiin tämän jälkeen tutkimusjoukon rajausta vastaavasti suomalaisten organisaatioiden aktiiviset verkkotunnukset. Aktiivisia verkkotunnuksia ovat ne, joiden status on joko "Registered" tai "Transfer denied", kun taas "Temporarily removed from fi-root", "In grace period" ja "Validity expired" viittaavat inaktiivisiin verkkotunnuksiin (Viestintävirasto 2017, 8).

Jotta rajapinnasta tuotettua verkkotunnuslistaa voitiin käyttää tiedostojen hakemiseen, täytyi niistä eristää pelkästään verkkotunnus sekä lisätä rajapinnan tarjoamista verkkotunnusten nimistä puuttuva fi-pääte. Lisäksi piti koostaa lista www-aliverkkotunnuksista, sillä kaikilla verkkotunnuksilla ei välttämättä ole verkkosivustoa suoraan verkkotunnuksen juuressa. Tämä toteutettiin Bash-komentosarjalla, joka hyödyntää jq- (Dolan s.a.) ja sed-komentoja. Kummallekin listalle tehtiin nimipalvelukyselyt MassDNS-työkalun (Blechschmidt s.a.) avulla, missä kesti noin kymmenen minuuttia. Nimipalvelukyselyjen tuloksista tuotettiin listat ZGrab 2.0:n tukemassa CSV-muodossa, jossa kullakin rivillä oli pilkuin eroteltuna ensin IP-osoite ja sitten HTTP-pyynnössä käytettävä isäntänimi. Mikäli samalla isäntänimellä oli nimipalvelun mukaan esimerkiksi kuor-

mantasaussyistä useampi IP-osoite, valittiin näistä vain yksi kaksoiskappaleiden välttämiseksi havainnoissa. CNAME-tietueet jätettiin huomiotta, sillä niiden takana olevien A-tietueiden selvittäminen olisi vaatinut MassDNS-työkalun ulostulon lisäjäsentämistä, ja ZGrab 2.0 pystyi tekemään nimipalvelukyselyt luotettavasti uudelleen tällaisten isäntänimien osalta, koska MassDNS oli jo aiheuttanut näiden tietueiden tallentumisen DNS-palvelimen välimuistiin.



Kuva 4. Vuokaavio aineistonkeruuseen käytetyn komentosarjan toiminnasta.

Ensimmäisellä kerralla aineistonkeruu jaettiin useampaan edellä kuvattuun vaiheeseen. Kuhunkin vaiheeseen käytettiin erillistä Bash-komentosarjaa, jonka tuottamia tiedostoja käytettiin seuraavan vaiheen komentosarjan lähteenä. Tämä mahdollisti kunkin vaiheen lopputuloksen tarkastamisen ennen seuraavan vaiheen aloitusta: vaiheen tuottamien tiedostojen tuli muodoltaan vastata harjoitusaineiston tuottamia tiedostoja. Samalla näiden erillisten komentosarjojen toiminta dokumentoitiin tarkasti, ja sen perusteella laadittiin aineistonkeruiden välissä yhtenäinen työkalu eli komentosarja, joka toisti kaikki

vaiheet verkkotunnusten noutamisesta ja rajaamisesta aina valmiiseen aineistoon asti. Tämän komentosarjan toimintaa on havainnollistettu vuokaaviolla kuvassa 4. Toinen aineistonkeruu suoritettiin onnistuneesti tällä työkalulla. Tämä poisti kokonaan viipeet eri vaiheiden välillä, mikä ilmenee myös taulukon 2 ensimmäisten vaiheiden aloitusajoista.

Lokakuussa 2022 n=366990 verkkotunnuksen joukosta löytyi noin 553 000 nimipalvelun A-tietuetta, kun lasketaan yhteen sekä verkkotunnuksen juuri että www-aliverkkotunnus. Helmikuussa 2023 n=367942 verkkotunnuksen joukosta löytyi vastaavasti noin 580 000 tietuetta. Molemmilla kerroilla näistä kuhunkin lähetettiin neljä HTTP-pyyntöä, ja vastausten kaikki tiedot tallennettiin onnistuneiden yhteyksien luotettavuuden arviointia ja epäonnistuneiden yhteyksien epäonnistumisen syiden tilastointia varten.

Lokakuussa 2022 HTTPS-protokollalla saatiin noin 796 000 yhteyttä ja HTTP-protokollalla 1,05 miljoonaa yhteyttä. HTTPS-yhteyksien osalta analysoitavaa tutkimusaineistoa kertyi 28 gigatavua ja HTTP-yhteyksistä 8,1 gigatavua. Helmikuussa 2023 noin 827 000 HTTPS-yhteyttä tuotti 30 gigatavua tutkimusaineistoa ja 1,1 miljoonaa HTTP-yhteyttä 8,4 gigatavua. ZGrab 2.0 tallentaa HTTPS-yhteyksistä myös TLS-varmenneketjut ja TLS-yhteyden salausmenetelmiin liittyviä yksityiskohtia, mikä selittää niiden tuottaman suuremman datamäärän. Epäonnistunut yhteys puolestaan tuottaa vain vähän dataa, koska siitä tallennetaan vain IP-osoite, verkkotunnus, aikaleima ja yhteyden epäonnistumisen syy. Esimerkki laajan JSON-rakenteen sisältämistä avaimista ja niiden saamien arvojen tyypeistä esitellään liitteessä 1. Yhteenvetona tutkimusaineiston määrästä ja rakenteesta voidaan tiivistää, että aineistoa oli huomattava määrä, mutta se oli analyysin kannalta hyvin jäsenneltyä.

Tutkimusprosessin automatisointia jatkettiin edelleen myös toisen aineistonkeruun jälkeen yhdistämällä seuraavassa alaluvussa tarkemmin käsiteltävien ilmiöön liittymättömien havaintojen karsimiseen ja tilastointiin käytetyt komentosarjat osaksi automatisoitua kokonaisuutta. Myös tämä toteutti opinnäytetyön tavoitetta tuottaa työkaluja, jotka helpottavat ilmiön seurantaa. Jatkossa on mahdollista päästä heti uuden aineistonkeruun päätyttyä tutkimaan ilmiöön liittyviä havaintoja sekä vertailemaan niitä aiempiin havaintoihin.

4.3 Ilmiöön liittymättömien havaintojen karsiminen

Jotta voidaan vastata tutkimuskysymykseen, miten yleistä security.txt:n julkaiseminen on, pitää ensin pystyä karsimaan tuloksista sellaiset havainnot, jotka eivät liity ilmiöön. Läpinäkyvyyden vuoksi ja toistettavuuden turvaamiseksi havaintojen hylkäyssyyt on perusteltu ja tilastoitu yksityiskohtaisesti.

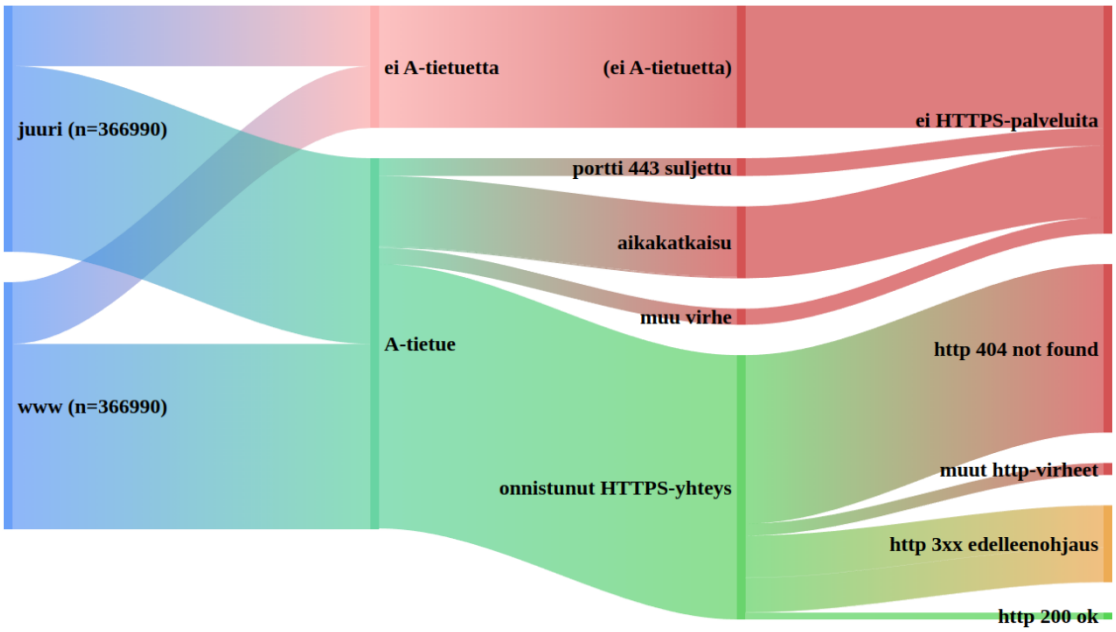
Kumpikin polku, "/.well-known/security.txt" ja "/security.txt", analysoitiin sekä salatun HTTPS-protokollan että salaamattoman HTTP-protokollan osalta erikseen. Verkkotunnuksen juuri ja www-aliverkkotunnus on yhdistetty epäonnistuneiden yhteyksien osalta samaan tarkasteluun, koska tuloksista tulkittujen karsimissyiden jakaumat olivat niiden osalta hyvin lähellä toisiaan. Tämä on odotettua, sillä mikäli RFC 9116:ta ei ole yritetty ottaa käyttöön, vastaa palvelin useimmiten samalla tavalla kummastakin polusta. Eroja syntyy lähinnä siitä, että joissain verkkotunnuksissa on tehty edelleenohjauksia verkkotunnuksen juuresta www-aliverkkotunnukseen ja toisissa taas päinvastoin.

DNS-tietueiden puuttumisessa on yhdistetty sekä DNS-kyselyjen tuloksia että niihin liittyviä ZGrab2:n JSON-rakenteen (ks. liite 1) ".data.http.error" tuloksia. Karkean tason luokittelu on tehty rakenteen ".data.http.status" mukaan, jonka jälkeen virheen tarkemmat syyt on tulkittu rakenteen ".data.http.error" perusteella. Palveluiden puuttumiseksi on todettu ".data.http.error" esiintyneet muut virheet, kuten aikakatkaisut ja yhteyden hylkäys. HTTP-protokollan antamat tilakoodit, joiden perusteella on todettu tiedoston löytymättömyys (404), edelleenohjaukset (3xx), muut virheet (4xx, 5xx) sekä onnistunut yhteys (200) on puolestaan saatu JSON-rakenteesta ".data.http.result.response.status_code". Muiksi virheiksi laskettiin myös epästandardit HTTP-tilakoodit, kuten 999. Onnistuneet yhteydet luokiteltiin puolestaan sisältötyypeittäin. Lopuksi poistettiin kaksoiskappaleet, jotta kustakin verkkotunnuksesta jäisi tarkasteltavaksi vain yksi tiedosto, ja verkkotunnukset saisivat siten saman painoarvon tuloksissa.

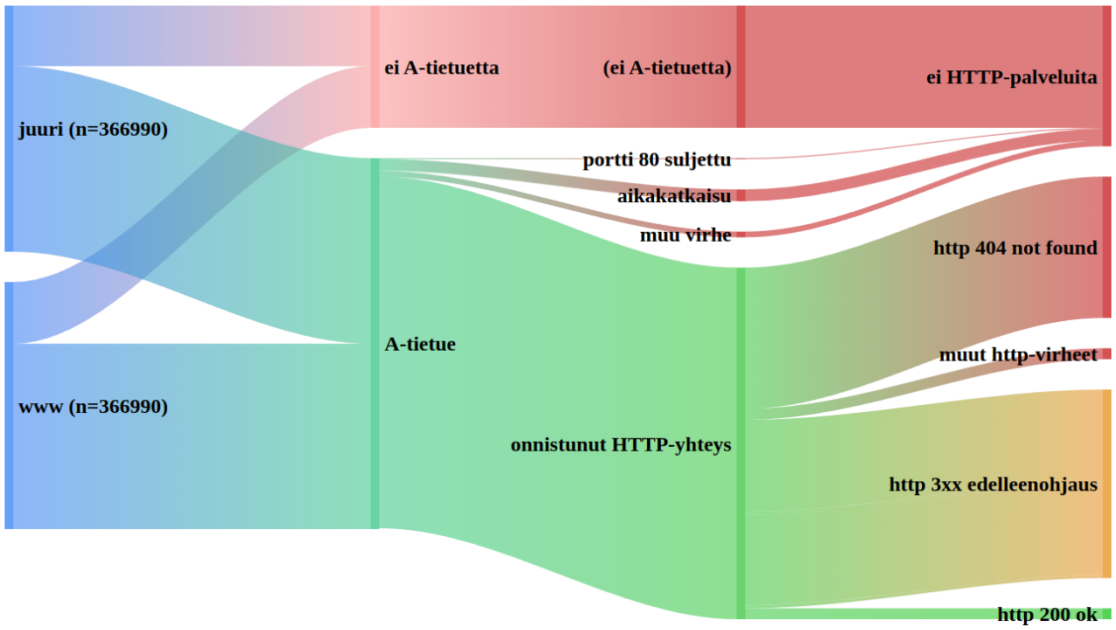
4.3.1 Epäonnistuneet yhteydet ja virheelliset HTTP-tilakoodit

Tehtyihin HTTP-pyyntöihin liittyvät havainnot karsittiin sen perusteella, saatiinko pyyntöön vastaus, ja mikä oli vastauksen HTTP-tilakoodi. Myös DNS-tietuen puuttuminen tilastoitiin palvelun puuttumisena. Karsintaprosessia on

havainnollistettu kuvissa 5 ja 6, jotka on laadittu lokakuun 2022 aineistonkeruun perusteella. Lokakuun 2022 ja helmikuun 2023 aineistonkeruissa ei ollut keskenään niin suurta vaihtelua, että siitä olisi koitunut silmin havaittavaa eroa näihin kaavioihin, joten kaaviot on esitetty vain kerran. Tulosten karsimisessa pyyntöihin vastaavien HTTPS- ja HTTP-palveluiden ja niiden palauttamien virhekoodien perusteella ei ollut merkittävää eroa myöskään polkujen välillä, joten kuvissa 5 ja 6 eritellyt polun "/.well-known/security.txt" tulokset kuvaavat myös vaihtoehtoisen polun "/security.txt" tuloksia.



Kuva 5. Tulosten karsiminen: HTTPS-yhteydet 10/2022.



Kuva 6. Tulosten karsiminen: HTTP-yhteydet 10/2022.

Merkittävin ero HTTPS- ja HTTP-palveluiden välillä oli siinä, että salaamattomalla yhteydellä vastaavia verkkotunnuksia oli enemmän. Toisaalta tulokset ovat linjassa sen kanssa, että tyypillisesti salattua yhteyttä käyttävät palvelut edelleenohjaavat salaamattomasta yhteydestä salattuun käyttäen siten molempia protokollia. Taulukko 3 esittää vastaavat tulosten jakaumat kunkin polun ja protokollan yhdistelmän osalta pyöristettynä tuhannen havainnon tarkkuuteen. Kunkin solun ylempi rivi kuvaa lokakuun 2022 ja alempi rivi helmikuun 2023 tilannetta.

Taulukko 3. Havaintojen jakaumat tuhansina osoitteina.

	HTTPS		HTTP	
10/2022 2/2023	/.well-kown/.. Polut 1 + 2	/security.txt 3 + 4	/.well-kown/.. 5 + 6	/security.txt 7 + 8
Kaikki (juuri & www)	734 736	734 736	734 736	734 736
Ei palvelua	340 326	332 319	210 187	208 185
Sivua ei löydy (HTTP 404)	251 256	241 247	211 217	182 188
Muut HTTP-tilakoodit	18 23	18 21	16 21	15 17
Edelleenohjaus (HTTP 3xx)	115 119	131 138	281 295	313 329
Onnistunut (HTTP 200)	10 11	12 12	16 17	17 17

Edelleenohjauksiin liittyvien HTTP-tilakoodien runsas esiintyminen vastasi ennakkoarviota hakukoneoptimointiin ja salatun yhteyden pakotuksiin liittyvien edelleenohjauksien mahdollisesta runsaudesta, sillä 77 % (sekä 10/2022 että 2/2023) pysyi saman verkkotunnuksen sisällä. Avoimeksi jää kuitenkin kysymys, miksi Findlay & Abdou (2022, 6) saivat ristiriitaisia havaintoja niiden harvinaisuudesta. Heidän menetelmässään käytettiin itse kirjoitettua Rust-sovellusta (Findlay & Abdou 2022, 3), jonka lähdekoodeja ei ollut enää saatavilla, minkä vuoksi sovelluksen toimintaa ei voi tarkastaa. Samasta syystä ei ole myöskään mahdollista toistaa tutkimusta siten, että vertailisi samalla tutkimusjoukolla heidän sovelluksellaan saatavia tuloksia ZGrab 2.0:lla, Nucleilla tai muulla julkisesti saatavissa olevalla työkalulla saataviin tuloksiin.

4.3.2 Väärät sisältötyypit

Onnistuneet yhteydet (HTTP 200) karsittiin HTTP-otsakkeen "Content-Type" mukaan eli rakenteen ".data.http.result.response.headers.content_type[]" alkuosan perusteella, jossa sisältötyypin muut arvot kuin "text/plain" eivät viittaa käyttöönottoyritykseen. Taulukko 4 esittää nämä tulokset yksittäisten havaintojen tarkkuudella kussakin taulukossa 1 esitellyssä kahdeksassa tapauksessa erikseen. Verkkotunnuksen juuren ja www-aliverkkotunnuksen tarkastelu erikseen mahdollistaa kaksoiskappaleiden poiston, joka kuvataan seuraavassa alaluvussa.

Taulukko 4. HTTP-otsakkeen "Content-Type" jakaumat.

	HTTPS				HTTP			
10/2022	/.well-kown/…		/security.txt		/.well-kown/…		/security.txt	
2/2023	Polku 1	2	3	4	5	6	7	8
text/plain	493	563	361	422	196	184	85	87
	646	759	497	602	212	224	90	101
text/html	2961	2964	3240	3348	7008	7185	7254	7473
	2841	3105	3237	2500	7089	7435	7324	7698
muut	12	13	7	7	8	10	6	5
	14	15	8	8	10	10	6	5

Sisältötyypiksi "text/html" ilmoittavien tulosten suuren lukumäärän selittää se, että osa palvelimista vastaa kaikista poluista samalla HTML-sisällöllä riippumatta siitä, onko tällaista polkua tai tiedostoa oikeasti olemassa. Lisäksi osassa "sivua ei löydy" -virheilmoituksista palvelin palauttaa virheellisesti tilakoodin 200, vaikka tällöin tulisi käyttää tilakoodia 404. Merkistöt "Content-Type" headerissa jakautuivat siten, että noin puolessa (380 10/2022 ja 415 2/2023) merkistöä ei ollut ilmoitettu lainkaan ja noin puolessa merkistöksi oli ilmoitettu sallittu "UTF-8" (388 10/2022 ja 391 2/2023). Kummallakin kerralla vain kahdessa tiedostossa oli käytetty väärää merkistöä "ISO-8859-1".

4.3.3 Kaksoiskappaleiden ja selkeästi liittymättömien sisältöjen poisto

Poistamalla tuloksista kaksoiskappaleet saadaan selville, kuinka monessa verkkotunnuksessa tiedosto sijaitsee ensisijaisessa tai vähintään toissijaisessa polussa suojatulla yhteydellä. Lisäksi havaitaan, milloin käyttöönottoa on yritetty missä tahansa polussa myös suojaamattomat yhteydet huomioon otettuna.

Taulukko 5. Oikeat sisältötyypit uniikeissa verkkotunnuksissa.

	HTTPS				HTTP			
10/2022	/.well-kown/…		/security.txt		/.well-kown/…		/security.txt	
2/2023	Polku 1	2	3	4	5	6	7	8
kumulatiivinen	492	635	724	760	809	822	835	838
	644	820	909	941	991	1010	1022	1026
lisäys	–	143	89	36	49	13	13	3
	–	176	89	32	50	19	12	4

Taulukossa 5 esitetään sekä uniikit verkkotunnukset kumulatiivisesti, kun uusia sijainteja otetaan mukaan tarkasteluun, että lisäys edelliseen tarkastelujoukkoon nähden. Osa oikean sisältötyypin ilmoittavista palvelimista palauttaa kuitenkin vielä muuta kuin RFC 9116:n tarkoitukseen liittyvää sisältöä, kuten tyhjän merkkijonon, HTML-sivun, virheilmoituksen tai Python-sovelluksen "It works!" -ilmoituksen. Helmikuun 2023 aineistonkeruussa oli näiden lisäksi ilmestynyt 120 verkkotunnusta, jotka palauttivat merkkijonon "ok" ja liittyivät kaikki samaan edelleenohjauspalveluun. Tuloksia esittelevässä luvussa on vastaava, korjattu taulukko 7, josta kaikki tällaiset sisällöt on karsittu pois.

4.4 Analyysiin käytettävien työkalujen valmistelu ja toiminta

Ensimmäisen aineistonkeruun yhteydessä analysoitiin aineiston rakennetta ja asetettiin rajaukset sille, miten aineiston joukosta erotellaan ne tiedot, jotka analysoidaan tarkemmin niistä tiedoista, jotka vain tilastoidaan ilmiöön liittymättöminä. Tilastointiin käytetyt haut olivat raskaita, joten yksittäisen tiedon saamisessa saattoi kestää useita minuutteja. Karsittu, ilmiöön liittyneiden havaintojen aineisto oli kuitenkin enää 54 megatavua lokakuussa 2022 ja 76 megatavua helmikuussa 2023, joten siihen kohdistuneet haut kestivät vain muutamia sekunteja.

Molemmissa vaiheissa hakujen tekemiseen käytettiin Jq-työkalua, joka on JSON-muotoisen datan analysointiin ja prosessointiin tarkoitettu komentorivityökalu (Dolan s.a.). Sen palauttamia, kunkin haun ehdot täyttäviä tuloksia järjesteltiin, normalisoitiin ja yhdisteltiin tilastoiksi tyypillisillä Unix-komennoilla, kuten sort, uniq, tr, sed ja awk. Tilastolliset tulokset kerättiin laskentataulukkoon, josta ilmeni komentosarja, jolla tulos oli saatu, yhdistettyjen tulosten lukumäärä sekä tulokselle annettu sanallinen tulkinta. Menettelyn tarkoituksena

oli helpottaa tulosten varmentamista, osoittaa päättelyketjun oikeellisuus ja mahdollistaa tarkastuslaskenta.

Ilmiöön liittymättömien tulosten karsinnan jälkeen kustakin verkkotunnuksesta oli jäljellä 1–8 yhteyden tiedot kaikkine yksityiskohtineen. Koska havaintoyksiköiksi oli yhteyden sijaan valittu verkkotunnus sekä myöhemmin mukaan lisätty toimija, oli kustakin verkkotunnuksesta käytettävä vain yhden yhteyden tietoja. Tätä varten kehitettiin Bash-komentosarja, joka käy läpi karsitun aineiston taulukon 1 mukaisten polkujen järjestyksessä ja säilyttää vain ensimmäisen samaa verkkotunnusta koskevan osuman. Samaan työkaluun yhdistettiin aineistonkeruiden välissä ominaisuus, joka mahdollisti todennäköisesti ilmiöön liittymättömien tiedostojen hylkäämisen sillä perusteella, ettei niissä esiintynyt lainkaan rivejä, joiden alussa olisi ollut jonkin RFC 9116:n mukaisen kentän nimen alkuosaa vastaava välimerkitön ja kaksoispisteeseen päättyvä merkkijono. Siten toiseen aineistonkeruuseen liittyvässä karsinnassa myös edellä esitelty kaksoiskappaleiden ja selkeästi liittymättömien sisältöjen poisto oli automatisoitu. Työkalua käytettiin myös ensimmäisen aineistonkeruun tarkastuslaskentaan, jotta tulokset olisivat vertailukelpoisia.

ZGrab 2.0:n tuottama JSON-rakenne (ks. liite 1) on muodoltaan puumainen, ja tutkimuksen muuttujien kannalta olennaiset tiedot löytyvät eri puolilta rakennetta pitkien polkujen päästä, mikä sellaisenaan tekee analyysityökalujen kehittämisestä vaivalloista. Tämän vuoksi edellä mainitun kaksoiskappaleiden poistoon rakennetun työkalun yhteyteen lisättiin myös ominaisuus, joka poimi tarvittavat yksityiskohdat ja lyhensi polut luettavampaan muotoon. Loput kehitetyistä analyysityökaluista käyttävät tämän työkalun ulostuloa syötteenään, mikä tekee niiden rakenteesta yksinkertaisemman. Tällöin myös niiden toimintaa on helpompi ymmärtää luotettavuuden varmistamiseksi.

Taulukosta 6 ilmenee työkalun tuottaman lyhennetyn avaimen lisäksi alkuperäisen ZGrab 2.0:n tuottaman JSON-rakenteen mukaisen avaimen polku sekä rivinvaihdoin eroteltuja esimerkkejä mahdollisista avaimien arvoista eli sisällöistä. Lainausmerkeissä olevat arvot tarkoittavat merkkijonoarvoja, "true" ja "false" yksinkertaisia kyllä/ei-arvoja ja "null" sitä, että avain on joko sisältänyt alkuperäisessä rakenteessa tyhjän arvon tai puuttunut rakenteesta kokonaan.

Taulukko 6. JSON-rakenteen yksinkertaistaminen ja avainten esimerkkiarvoja.

Lyhennetty avain	Alkuperäinen avain	Avaimen arvo
.domain	.domain | sub("^www."; "")	"example.com"
.hostname	.domain	"www.example.com"
.scheme	.data.http.result.response.request.url.scheme	"https" / "http"
.path	.data.http.result.response.request.url.path	"/.well-known/security.txt" "/security.txt"
.content_type	.data.http.result.response.headers.content_type[0]	"text/plain; charset=UTF-8"
.body	.data.http.result.response.body	*Tiedoston sisältö.*
.browser_trusted	.data.http.result.response.request.tls_log .handshake_log.server_certificates .validation.browser_trusted	true false *null*
.matches_domain	.data.http.result.response.request.tls_log .handshake_log.server_certificates .validation.matches_domain	true false *null*
.cn	.data.http.result.response.request.tls_log .handshake_log.server_certificates .certificate.parsed.subject.common_name[0]	"www.example.com"
.san	.data.http.result.response.request.tls_log .handshake_log.server_certificates .certificate.parsed.extensions .subject_alt_name.dns_names	"example.com", "*.example.com" *null*
.tls	.data.http.result.response.request.tls_log .handshake_log.server_key_exchange .signature.tls_version.name	"TLSv1.2" *null*
.validity_start	.data.http.result.response.request.tls_log .handshake_log.server_certificates .certificate.parsed.validity.start	"2023-02-21T11:13:04Z" *null*
.validity_end	.data.http.result.response.request.tls_log .handshake_log.server_certificates .certificate.parsed.validity.end	"2023-05-22T11:13:03Z" *null*
.timestamp	.data.http.timestamp	"2023-02-23T05:17:03Z"

Osa valituista tiedoista vastasi suoraan tarvittavan muuttujan sisältöä, osasta sisältö piti vielä käsitellä toisella työkalulla, ja osa oli valittu mukaan ristiintarkastusta varten. Avaimen "matches_domain" kyllä/ei arvo oli varmennettavissa vertaamalla avaimen "hostname" arvoa avainten "cn" ja "san" varmenteesta poimittuihin arvoihin. Avain "browser_trusted" täsmentyi varmenteen voimassaolon sisältävien avainten "validity_start" ja "validity_end" sekä avaimessa "timestamp" olevan havaintohetken avulla. ZGrab 2.0 tarkastelee näitä toisistaan irrallisena, eli avain "browser_trusted" voi olla "true", vaikka avaimessa "matches_domain" olisi "false". Tämä ei olisi ilmennyt ilman ristiintarkastusta, mikä olisi saattanut johtaa väärään tulkintaan. Jos taas "scheme" oli saanut arvon "http", eivät TLS-salaukseen liittyvät avaimet olleet odotettuja, joten niissä tuli olla arvo "null". Siten ristiintarkastukset auttoivat varmistamaan sekä havaintojen luotettavuuden että niiden oikean tulkinnan.

Tiedoston sisällön vastaavuutta RFC 9116:n määritelmiin tarkasteltiin työkalulla, joka merkitsi tiedostosta rivit, jotka eivät vastanneet määritelmiä. Työkalu kehitettiin ensin muodossa, joka poistaa tiedostosta virheelliset rivit ja lisää sitten OpenPGP-allekirjoituksen, ja Kyberturvallisuuskeskuksen oma security.txt-

tiedosto (ks. liite 2) onkin varmennettu ja allekirjoitettu sen avulla. Koska tiedostojen sisällöissä esiintyi koneluettavuuden estävää vaihtelua, käytiin tämän työkalun tuottamat tulokset läpi manuaalisesti yksi toimija kerrallaan. Toimijat tunnistettiin etsimällä ensin tiedostoista ensimmäinen "Contact:"-esiintymä ja ottamalla talteen tämän kentän arvo. Tämän jälkeen kuhunkin toimijaan liittyvien kaikkien verkkotunnusten tiedostot haettiin työkalulla, joka mahdollisti vapaan tekstihaun tiedoston sisällöstä, kuten "mailto:security@example.com". Mikäli tiedostot näyttivät saman muotoisilta – kuten lähestulkoon aina olikin – eli esimerkiksi vain aikaleima vaihteli, kirjattiin havainnot ja yhteenlaskettujen verkkotunnusten määrä taulukkoon. Samassa yhteydessä kirjattiin ylös toimijan käyttämän "Expires"-kentän aikaleima, josta laskettiin sen ja havaintohetken välinen aika.

Tilastot kenttien esiintyvyydestä ja "Contact"-kentässä käytetyistä URI-skeemoista laskettiin työkalulla, joka etsi tiedostoista rivejä, jotka alkoivat annetulla hakusanalla, jota seurasi kaksoispiste. Löytyneistä tiedostoista laskettiin yhteen ne, joissa ensimmäisen "Contact"-kentän sisältö oli sama. Lisäksi kirjattiin löytyneiden tiedostojen eli verkkotunnusten lukumäärä sekä rivien lukumäärä, josta puolestaan laskettiin, kuinka monta kertaa kenttä esiintyi yhteensä, sillä kaikki muut kentät paitsi "Expires" ja "Preferred-Languages" saavat esiintyä samassa tiedostossa useamman kuin yhden kerran. Yhteydenotoissa toivottujen kielten järjesteltyjen luetteloiden lukemiseen käytettiin samankaltaista työkalua, joka etsi ensin kaikki esiintyneet variantit ja tuotti niistä sitten kummankin havaintoyksikön mukaiset tilastot.

Opinnäytetyön toteutusta kuvanneen luvun tarkoituksena oli kuvata valittujen aineistonkeruu- ja analyysimenetelmien teknistä implementaatiota siten, että sen perusteella on mahdollista toistaa tutkimuksen keskeiset vaiheet. Työkaluina tuotetut komentosarjat jäävät opinnäytetyön toimeksiantajan käyttöön, eikä niitä julkaista osana opinnäytetyön raporttia. Niiden toimintaa on kuitenkin esitelty riittävällä tarkkuudella, jotta kunkin analyysimenetelmissä määritellyn muuttujan saamien arvojen alkuperä on selvillä. Tällöin voidaan varmistua siitä, että tutkimuksen tulokset on tuotettu menetelmän mukaisesti ja ne vastaavat muuttujien kriteerejä. Siten myös tuloksista johdetut tutkimuskysymysten vastaukset perustuvat teoreettiseen viitekehykseen ja käytettyyn metodologiaan. Tämä mahdollistaa sen, että seuraavassa luvussa tulokset voidaan

esitellä tiiviissä muodossa palaamatta uudelleen niiden alkuperään. Se on tarpeen erityisesti siksi, että lokakuun 2022 ja helmikuun 2023 tulokset esitellään erikseen ja niitä vertaillaan vain silloin, kun se on tarkoituksenmukaista.

5 TULOKSET

5.1 Security.txt-tiedoston käytön laajuus

Ensimmäisenä tutkimuskysymyksenä oli selvittää, miten yleistä security.txt:n julkaiseminen on. Lokakuussa 2022 ilmiöön liittyviä havaintoja oli n=366990 verkkotunnuksen joukossa 2,1 ‰ (770), joista RFC 9116:n suositusten mukaisissa poluissa 1,7–2,0 ‰ (613–727). Helmikuussa 2023 ilmiöön liittyvät havainnot olivat kasvaneet hienoisesti, ja niitä oli n=367942 verkkotunnuksen joukossa 2,2 ‰ (808), joista suositusten mukaisissa poluissa 1,8–2,1 ‰ (649–758).

Taulukko 7 esittelee yksityiskohtaisemmin todelliset käyttöönottoyritykset uniikeissa verkkotunnuksissa sen jälkeen, kun niistä on poistettu kaksoiskappaleet ja sisältönsä perusteella ilmiöön liittymättömät havainnot, jotka olivat mukana taulukossa 5. Koska tässä ei ole vielä tarkasteltu tiedostojen sisältöjä tarkemmin, myös onnistuneita käyttöönottoja käsitellään käyttöönoton yrityksinä. Ellei toisin mainita, kaikissa tulosluvun taulukoissa kunkin solun ylemmällä rivillä on lokakuun 2022 ja alemmalla rivillä helmikuun 2023 tilanne.

Taulukko 7. Todelliset käyttöönottoyritykset uniikeissa verkkotunnuksissa.

	HTTPS				HTTP			
10/2022	/.well-kown/…		/security.txt		/.well-kown/…		/security.txt	
2/2023	Polku 1	2	3	4	5	6	7	8
kumulatiivinen	474 509	**613** **649**	696 732	**727** **758**	747 781	754 792	767 804	**770** **808**
lisäys	– –	139 140	83 83	31 26	20 23	7 11	13 12	3 4

Tämän tutkimuksen tulokset asettuvat Findlayn & Abdoun (2022, 5–6) miljoonan suosituimman sivuston 4,9 ‰:n ja Neuenschwanderin (Foudil 2022) Sveitsin maatunnuskohtaisen 0,6 ‰:n välimaastoon. Tulokset ovat linjassa aiempien tutkimusten kanssa, ja ilmiötä voidaan siten myös suomalaisten organisaatioiden hallitsemien fi-verkkotunnusten osalta pitää edelleen hyvin marginaalisena. Aineistonkeruukertojen välinen 0,1 ‰-yksikön kasvu reilun

neljän kuukauden aikana vaikuttaa pieneltä. Se tarkoittaa kuitenkin noin 5 %:n kasvua havaittujen tiedoston julkaisseiden verkkotunnusten määrässä ja noin 4 %:n kasvua verkkotunnuksissa, jotka julkaisevat tiedoston suositusten mukaisissa poluissa. Seuraavassa alaluvussa analysoidaan tarkemmin tiedostojen sisältöä, jolloin selviää tiedostojen sisällön laadun lisäksi myös se, mihin tämä kasvu kohdistuu.

5.2 Julkaistut security.txt-tiedostot

Tämän alaluvun tulokset vastaavat tutkimuskysymyksiin 2 ja 3 siitä, millaisia tietoja security.txt-tiedoston avulla on julkaistu, ja miten hyvin tiedostot noudattavat RFC 9116:n asettamia määrittelyjä. Tämän ohella saatiin kuvaa myös siitä, millaiset toimijat ovat julkaisseet tiedostoja asiakkaidensa puolesta.

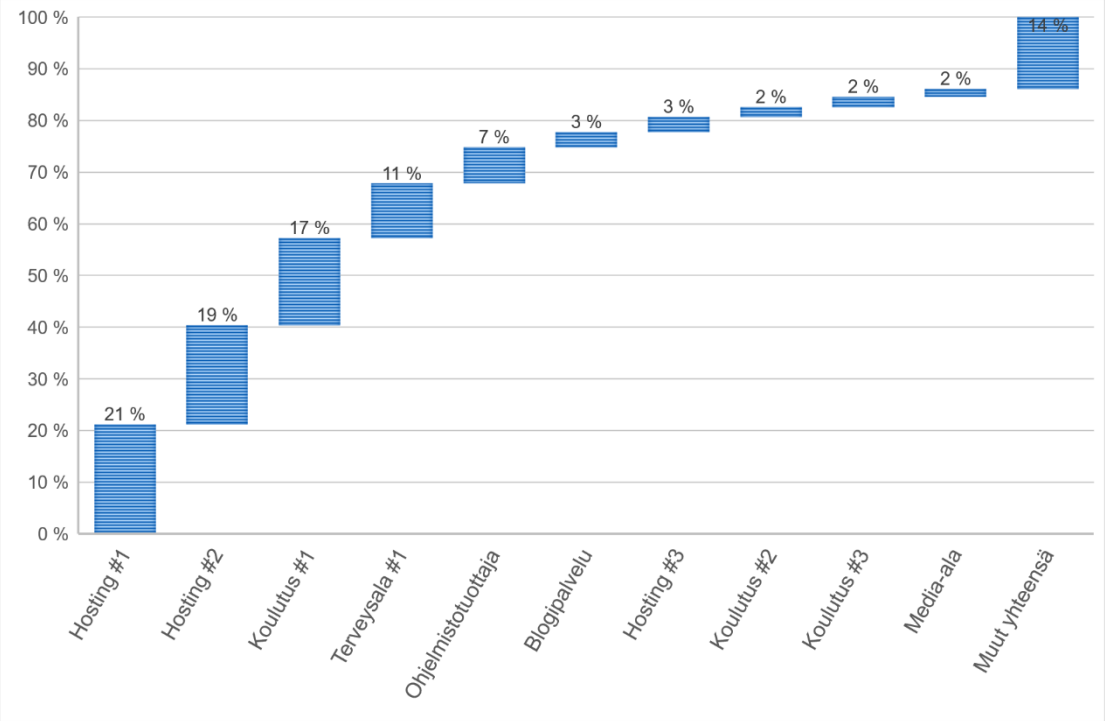
Julkaistujen security.txt-tiedostojen sisällön analyysiä varten kustakin verkkotunnuksesta poimittiin se esiintymä, jonka polku parhaiten vastasi RFC 9116:n määritelmää. Samaa yhdistämistä käytettiin myös luvun 5.3 turvallisuusnäkökohtien analysoinnissa. Polkujen järjestys vastaa taulukoiden 5 ja 7 kumulatiivisia uniikkeja verkkotunnuksia, ja se esiteltiin tarkemmin taulukossa 1.

5.2.1 Useamman verkkotunnuksen samankaltaiset tiedostot

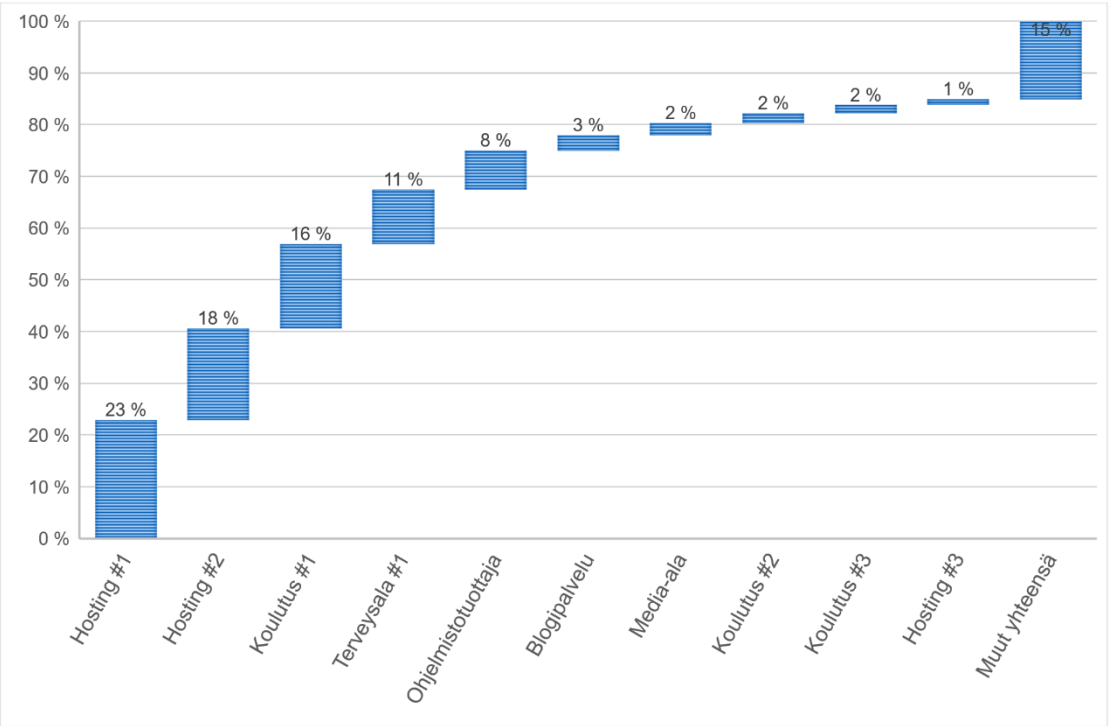
Julkaistujen security.txt-tiedostojen analyysin keskeinen havainto oli, että muutama toimija oli julkaissut tiedostoja lukuisissa eri verkkotunnuksissa. Lokakuussa 2022 useampaan verkkotunnukseen yhdistettäviä toimijoita löytyi 30 ja helmikuussa 2023 yksi enemmän eli 31. Tämän vuoksi samankaltaiset tiedostot julkaissut toimija nostettiin havaintoyksiköksi yksittäisen verkkotunnuksen rinnalle. Yhdistetyt verkkotunnukset eivät aina olleet saman organisaation hallussa, sillä yhdistämiseen käytettiin tiedostojen sisällön samankaltaisuutta, kuten yhteisiä "Contact"-kentän yhteystietoja. Organisaatioiden Y-tunnukset olisivat mahdollistaneet myös saman organisaation hallitsemien verkkotunnusten yhdistämisen yhteiseen tarkasteluun, mutta ilmiön harvinaisuuden vuoksi tämä ei ollut vielä ajankohtaista.

Kuvasta 7 ilmenee, miten lokakuussa 2022 kymmenen suurimman toimijan julkaisemat tiedostot kattoivat 86 % verkkotunnuksista (663), joissa tiedostoja esiintyi, ja loppu 14 % (107) jäi kuudellekymmenelle muulle toimijalle. Kuvassa

8 sama asetelma toistuu helmikuussa 2023, jolloin kymmenen suurinta toimijaa julkaisi tiedoston 85 %:ssa verkkotunnuksista (686), ja 68 toimijaa julkaisi tiedoston 15 %:ssa (122) verkkotunnuksista. Toimijoiden keskinäiset osuudet vaihtelevat hieman, ja kaksi toimijaa (hosting-palvelu #3 ja media-alan organisaatio) ovat vaihtaneet paikkoja keskenään.



Kuva 7. Samaan toimijaan yhdistyvät verkkotunnukset 10/2022.



Kuva 8. Samaan toimijaan yhdistyvät verkkotunnukset 2/2023.

Kymmenen suurimman toimijan joukossa oli kolme hosting-palvelua, jotka julkaisivat tiedoston asiakkaidensa verkkotunnuksissa omilla yhteystiedoillaan. Eräs blogijulkaisualusta käytti tiedostoa asiakkaiden verkkotunnuksissa, jolloin yhteystiedot oli julkaistu väärinkäytöksistä ilmoittamista varten. Näiden lisäksi verkkotunnuksen haltijasta riippumaton julkaisu liittyi sovellukseen, joka julkaisi tiedoston ohjelmistotuottajan yhteystiedoilla. Lopun kärkikymmeniköstä muodostivat kolme koulutusalan organisaatiota, yksi terveysalan organisaatio sekä yksi media-alan organisaatio.

5.2.2 Tiedostojen laatu suhteessa RFC 9116:n määritelmiin

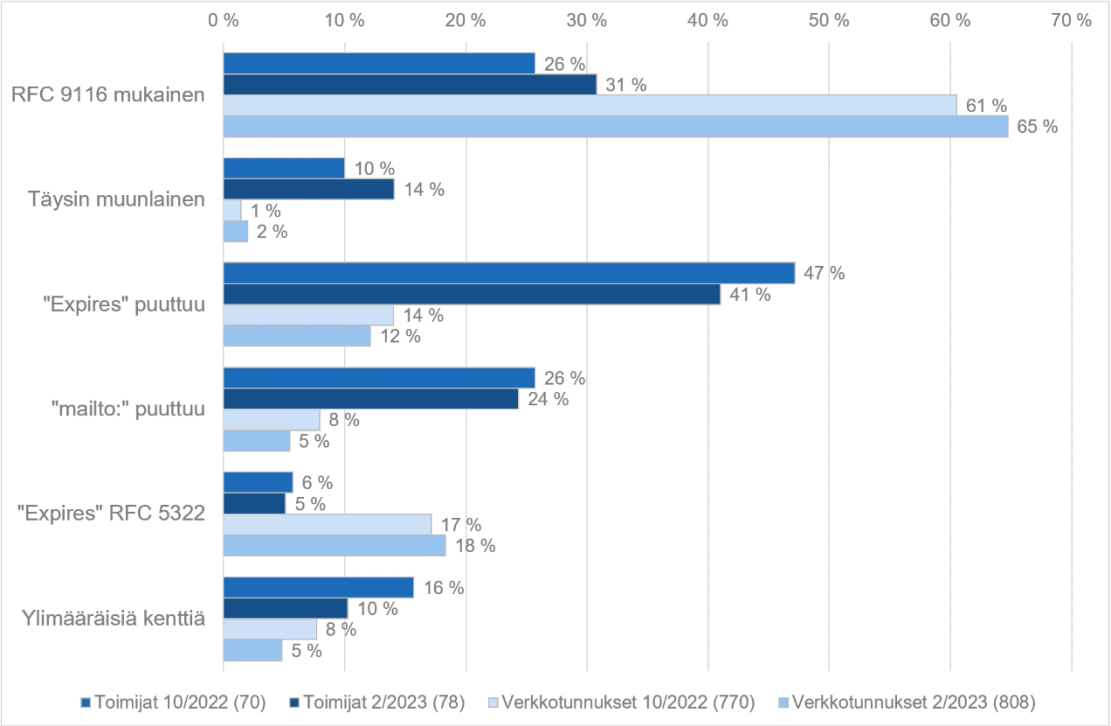
Jotta tiedosto läpäisee RFC 9116:n formaatin mukaisen tarkistuksen, täytyy vain muutamien ehtojen täyttyä. Turvallisuusnäkökohtien lisäsuosituksia ei tässä yhteydessä ole huomioitu, vaan niitä käsitellään myöhemmin. Kuten luvussa 3.4.3 tarkemmin eriteltiin, tällöin tiedosto läpäisee tarkistuksen, jos:

1. Tiedosto sisältää pakolliset kentät "Contact" ja "Expires".
2. Kenttien sisällöissä ei ole muotovirheitä.
3. Tiedosto ei sisällä ylimääräisiä kenttiä tai muita rivejä. OpenPGP-allekirjoitus, kommentit ja tyhjät rivit ovat kuitenkin sallittuja.

Lokakuussa nämä ehdot täytti 61 % (466) verkkotunnuksista, mutta tämä selittyy osittain sillä, että viiden suurimman toimijan joukosta neljä läpäisi validoinnin. Toimijoista puolestaan 26 % (18) oli julkaissut muodoltaan oikeanlaisen tiedoston. Täysin RFC 9116:n määrittelystä poikkeavia tiedostoja oli 10 %:lla (17) toimijoista ja 1 %:lla (11) verkkotunnuksista. Nämä jätettiin pois laskennasta tarkempien virheiden osalta, vaikka niissä toki oli sekä ylimääräisiä kenttiä että niistä puuttui pakolliset kentät. Pääasiassa RFC 9116:n mukaisten tiedostojen osalta on laskettu virheet tyypeittäin. Samassa tiedostossa saattoi silti olla useampi virhe.

Helmikuussa RFC 9116:n mukaisia tiedostoja oli 65 %:ssa (523) verkkotunnuksista 31 %:lla (24) toimijoista eli osuudet olivat hieman korkeampia. Samalla kuitenkin myös täysin muunlaisten tiedostojen osuus oli noussut 14 %:iin (11) toimijoista ja 2 %:iin (16) verkkotunnuksista. On tärkeä huomata, että tässä luvussa esiintyvistä muutoksesta ei voida kahden mittauspisteen perusteella päätellä trendiä kumpaankaan suuntaan, vaan vaihtelu on pikemminkin satunnaista, ja yksittäisten suurempien toimijoiden vaikutus siihen on

merkittävää. Kuvasta 9 ilmenee näiden havaintojen prosenttiosuudet toimijoittain ja verkkotunnuksittain sekä lokakuussa 2022 että helmikuussa 2023.



Kuva 9. Tiedostojen RFC 9116:n mukaisuus ja tyypilliset formaattivirheet.

Pakollinen "Contact"-kenttä puuttui kokonaan vain niiltä, joiden tiedosto ei vastannut lainkaan RFC 9116:n formaattia. Tyypillinen virhe "Contact"-kentässä oli sähköpostiosoitteen esittäminen ilman "mailto:" URI-muotoa 26 %:lla toimijoista (18) ja 8 %:lla verkkotunnuksista (61) lokakuussa 2022 sekä 24 %:lla toimijoista (19) ja 5 %:lla verkkotunnuksista (44) helmikuussa 2023.

Pakollinen "Expires"-kenttä puuttui 47 %:lla toimijoista (33) ja 14 %:lla verkkotunnuksista (108) lokakuussa 2022 sekä 41 %:lla toimijoista (32) ja 12 %:lla verkkotunnuksista (98) helmikuussa 2023. "Expires"-kentässä oli luonnosversiossa tuettu RFC 5322:n mukainen aikaleima 6 %:lla toimijoista (4) lokakuussa 2022, mikä muodosti 17 % verkkotunnuksista (132) sekä 5 %:lla (4) helmikuussa 2023 kattaen 18 % verkkotunnuksista (148). RFC 5322:n mukaisen aikaleiman käytön kasvu selittyy sillä, että tässä muodossa aikaleimaa käyttänyt yksittäinen toimija julkaisi tiedostoa useammassa verkkotunnuksessa kuin aikaisemmin.

Ylimääräisiä kenttiä oli 16 %:lla toimijoista (11) ja 8 %:ssa verkkotunnuksista (59) lokakuussa 2022 sekä 10 %:lla toimijoista (8) ja 5 %:ssa verkkotunnuksista (39) helmikuussa 2023. Seuraavassa alaluvussa tarkastellaan yksittäisten kenttien esiintyvyyttä lähemmin. Saadut tulokset poikkeavat tässä luvussa esitetystä muotovirheiden luokittelusta siksi, että kenttien esiintyvyyttä tarkasteltaessa myös täysin muunlaisiksi luokitellut tiedostot ovat mukana. Lisäksi samassa tiedostossa voi olla useampi ylimääräinen kenttä, kun aiemmin oli laskettu tällaisia kenttiä sisältäneiden verkkotunnusten lukumäärää.

5.2.3 Pakollisten, vapaaehtoisten ja ylimääräisten kenttien esiintyvyys

RFC 9116:n määrittelemien pakollisten ja valinnaisten kenttien lisäksi julkaistuissa tiedostoissa esiintyi jonkin verran aiemmista luonnoksista jääneitä kenttiä sekä kenttiä, jotka ovat kirjoitusvirheiden seurauksia tai itse keksittyjä. Nämä tulokset ovat luokiteltuna lokakuun 2022 osalta taulukossa 8 ja helmikuun 2023 osalta taulukossa 9, joissa kenttien yleisyys on esitetty sekä toimijoittain että verkkotunnuksittain. Taulukoissa on mukana myös prosenttiosuudet kaikista toimijoista ja verkkotunnuksista.

Taulukko 8. Kenttien yleisyys toimijoittain ja verkkotunnuksittain 10/2022.

	Kenttä	Toimijat		Verkkotunnukset		Yht.
Pakolliset	Contact	66	94 %	758	98 %	911
	Expires*	28	40 %	610	79 %	610
Valinnaiset	Acknowledgments	8	11 %	216	28 %	223
	Canonical	8	11 %	192	25 %	192
	Encryption	15	21 %	54	7 %	54
	Hiring	11	16 %	41	5 %	41
	Policy	25	36 %	276	36 %	280
	Preferred-Languages*	33	47 %	335	44 %	335
Aiemmat	Disclosure	1	1 %	5	1 %	5
	Permission	1	1 %	1	0 %	1
	Signature	4	6 %	12	2 %	12
Muut	Acknowledgements	8	11 %	30	4 %	30
	Expiration	4	6 %	26	3 %	26
	Out-of-scope	1	1 %	15	2 %	15
	OpenBugBounty	9	13 %	13	2 %	13
Yhteensä		*70*	*(100 %)*	*770*	*(100 %)*	*2748*

* Kenttä sallittu vain kerran samassa tiedostossa.

Taulukko 9. Kenttien yleisyys toimijoittain ja verkkotunnuksittain 2/2023.

	Kenttä	Toimijat		Verkkotunnukset		Yht.
Pakolliset	Contact	73	94 %	795	98 %	963
	Expires*	35	45 %	662	82 %	662
Valinnaiset	Acknowledgments	5	6 %	221	27 %	224
	Canonical	10	13 %	206	25 %	212
	Encryption	18	23 %	47	6 %	51
	Hiring	12	15 %	44	5 %	44
	Policy	24	31 %	278	34 %	278
	Preferred-Languages*	38	49 %	371	46 %	371
Aiemmat	Disclosure	1	1 %	4	0 %	4
	Permission	1	1 %	1	0 %	1
	Signature	4	5 %	12	1 %	12
Muut	Acknowledgements	8	10 %	14	2 %	14
	Expiration	4	5 %	10	1 %	10
	Out-of-scope	1	1 %	15	2 %	15
	OpenBugBounty	9	12 %	13	2 %	13
Yhteensä		*78*	*(100 %)*	*808*	*(100 %)*	*2874*

* Kenttä sallittu vain kerran samassa tiedostossa.

Taulukoiden 8 ja 9 viimeisissä sarakkeissa on kenttien kokonaismäärä, sillä valtaosa kentistä saa esiintyä samassa tiedostossa useamman kuin yhden kerran. Niistä ilmenee, että vain kerran sallittuja kenttiä "Expires" ja "Preferred-Languages" ei esiintynyt tiedostoissa useampia kertoja. Ainoastaan kenttiä "Contact", "Acknowledgments" ja "Policy" oli käytetty useamman kerran.

Pakollinen kenttä "Contact" esiintyi 98 %:ssa (758 lokakuussa 2022 ja 795 helmikuussa 2023) verkkotunnuksista, ja sitä oli käyttänyt 94 % tiedoston julkaisseista toimijoista (66 lokakuussa 2022 ja 73 helmikuussa 2023). Kenttä puuttui käytännössä vain niistä tiedostoista, joissa oli RFC 9116:n ulkopuolinen kenttä "OpenBugBounty", ja vain kahdessa verkkotunnuksessa nämä esiintyivät yhdessä.

Toinen pakollinen kenttä "Expires" esiintyi 79 %:ssa verkkotunnuksista lokakuussa 2022 ja 82 %:ssa helmikuussa 2023, mutta vain 39 % toimijoista käytti sitä lokakuussa 2022 ja 45 % helmikuussa 2023. Huolimatta siitä, että kenttä on Suomessakin heikommin tunnettu, on tämä silti huomattavasti parempi tulos kuin Poteatin & Lin (2021, 529–530) 1,7 % tiedoston julkaisseista verkkotunnuksista.

"Contact"-kentässä voidaan käyttää useampaa eri URI-skeemaa, joista kolme on erikseen mainittuna RFC 9116:n esimerkeissä. Sähköpostiosoitteen sisältävä "mailto:" oli selkeästi yleisin 28 toimijalla 609 verkkotunnuksessa lokakuussa 2022 ja 46 toimijalla 657 verkkotunnuksessa helmikuussa 2023. Verkkosivuille viittaavia "https:"-osoitteita oli 11 toimijalla 222 verkkotunnuksessa lokakuussa 2022 ja 9 toimijalla 241 verkkotunnuksessa helmikuussa 2023. Sen sijaan puhelinnumeron ilmaisevaa "tel:"-skeemaa ei havaittu lokakuussa 2022 lainkaan, ja helmikuussa 2023 se löytyi ainoastaan Kyberturvallisuuskeskuksen omasta security.txt-tiedostosta (ks. liite 2).

Valinnaisista kentistä sekä toimijoittain että verkkotunnuksittain suosituimmat olivat "Preferred-Languages" 47 %:lla toimijoista lokakuussa 2022 ja 49 %:lla helmikuussa 2023 sekä "Policy" 36 %:lla toimijoista lokakuussa 2022 ja 31 %:lla helmikuussa 2023. Kentät "Acknowledgments" ja "Canonical" olivat käytössä vai muutamalla toimijalla, mutta silti 27–28 %:ssa ja 25 %:ssa verkkotunnuksista. Tämän selittää se, että kumpikin oli käytössä samoilla suurilla toimijoilla, joka olivat julkaisseet tiedoston 130:ssä ja 54:ssä eri verkkotunnuksessa lokakuussa 2022 ja 132:ssa ja 61:ssä verkkotunnuksessa helmikuussa 2023. Lisäksi on huomioitava, että "Canonical"-kentällä on todellista merkitystä vain yhdessä OpenPGP-allekirjoituksen kanssa tiedoston luotettavuutta arvioitaessa. Toisella näistä toimijoista "Canonical"-kenttä kohdistui RFC 9116:n tarkoittamalla tavalla siihen verkkotunnukseen, jossa tiedosto oli julkaistu, mutta toinen oli ohjelmistotuottaja, joten kaikissa verkkotunnuksissa käytettiin samaa, ulkoista osoitetta.

Luonnosversioiden kenttiä "Disclosure", "Permission" ja "Signature" esiintyi vain satunnaisesti. Vähäisiä olivat myös määritelmästä poikkeavassa muodossa kirjoitetut "Acknowledgements" (brittienglannin ylimääräisellä e-kirjaimella) ja "Expiration". Kokonaan itse keksitty "Out-of-scope" oli vain yhdellä toimijalla ilmaisemassa sitä, että heidän pääasiallisen verkkotunnuksensa aliverkkotunnukset eivät ole tiedoston piirissä, vaan yhteystieto koskee niitä viittätoista apuverkkotunnusta, joissa tiedosto on julkaistu. Tämän ilmaiseminen erikseen on kuitenkin tarpeetonta, sillä RFC 9116:n mukaan tiedosto koskee vain sitä verkkotunnusta, josta se on noudettu, eikä sen aliverkkotunnuksia tai ylemmän tason verkkotunnuksia (Foudil & Shafranovich 2022, 10).

5.2.4 Yhteydenotoissa toivotut kielet

Lokakuussa 2022 kentässä "Preferred-Languages" (335 verkkotunnusta) esiintyneet kielet olivat pääasiassa englanti (en) ja suomi (fi), joiden järjestys vaihteli siten, että kahdeksan toimijaa (155 verkkotunnusta) toivoi englantia ensisijaisena kielenä ja kahdeksan toimijaa (88 verkkotunnusta) suomea. Kymmenen toimijaa (75 verkkotunnusta) toivoi yhteydenottoja pelkästään englanniksi ja kolme toimijaa (5 verkkotunnusta) englannin lisäksi saksaksi. Englanti – joka olisi oletuskieli myös kentän puuttuessa (Foudil & Shafranovich 2022, 8) – oli mukana kaikissa havainnoissa.

Helmikuussa 2023 pelkästään englanniksi yhteydenottoja toivovien toimijoiden määrä oli kasvanut kahdella (12 toimijaa ja 89 verkkotunnusta), mikä oli siirtymää aiemmin myös suomeksi yhteydenottoja toivoneista toimijoista, joista englantia ensisijaisena piti nyt seitsemän toimijaa (164 verkkotunnusta) ja suomea seitsemän toimijaa (91 verkkotunnusta). Tässä toimija oli merkittävämpi havaintoyksikkö, sillä suurempien toimijoiden lisääntynyt verkkotunnusten määrä vääristi trendejä, kun kasvua esiintyi niiden vuoksi kaikissa vaihtoehdoissa.

Yhden toimijan (4 verkkotunnusta) kielet oli ilmoitettu muulla tavalla kuin RFC 5646:ssa (Phillips & Davis 2009) tarkoitetuilla IETF-kielikoodeilla; määritelmästä poikkeavilla lyhenteillä "fin, eng". Muut kielet esiintyivät kaikki vaihtoehtoisina kielinä ja olivat yksittäistapauksia: ruotsi (sv), hollanti (nl), saame (se), puola (pl) ja espanja (es). Samat kielet esiintyivät sekä lokakuussa 2022 että helmikuussa 2023. Lisäksi ei voida varmuudella sanoa, oliko "se"-tunnisteella tarkoitettu IETF-kielikoodin mukaista saamen kieltä vai oliko se sekoitettu Ruotsin maakohtaiseen ylätason verkkotunnukseen ".se".

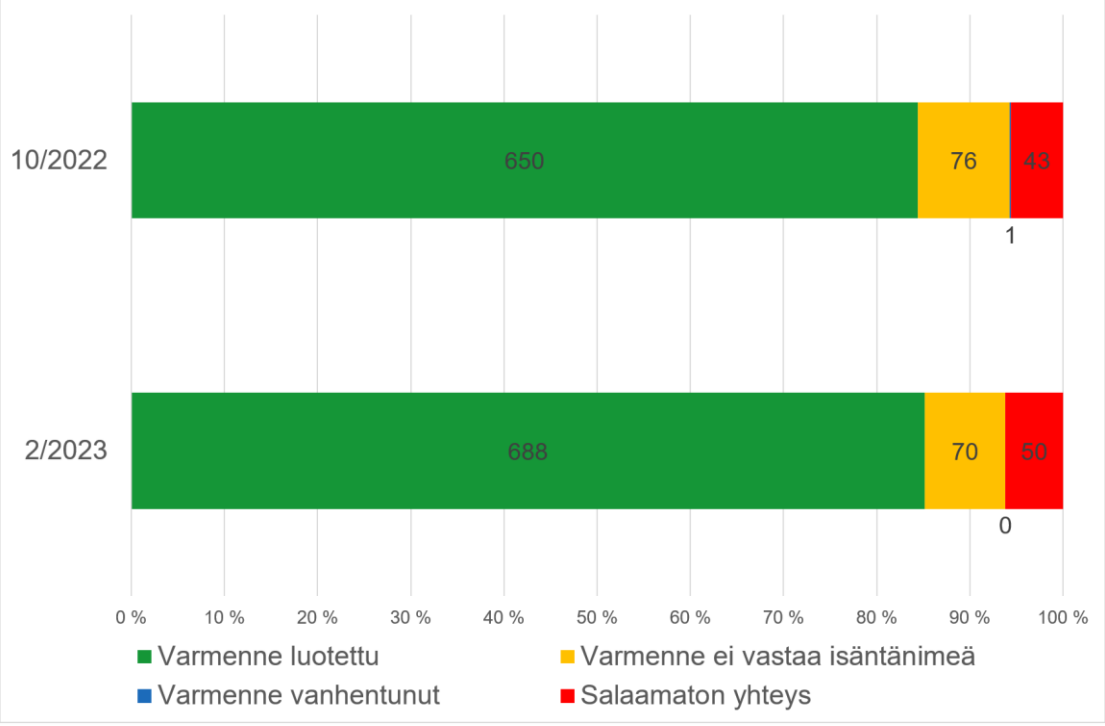
5.3 Turvallisuusnäkökohtien huomioiminen

Neljäntenä tutkimuskysymyksenä oli selvittää, miten hyvin julkaistut tiedostot ottavat huomioon RFC 9116:n nostamia turvallisuusnäkökohtia. Tähän saa vastauksen tarkastelemalla käytettyjen yhteyksien salausta ja luotettavuutta, OpenPGP:n käyttöä tiedostojen allekirjoittamisessa ja haavoittuvuuksiin liittyvän kommunikaation salaamisessa sekä tiedostojen voimassaoloaikoja.

5.3.1 Tiedoston lataamiseen käytetyn yhteyden luotettavuus

TLS-salattujen yhteyksien (Rescorla 2018) luotettavuutta arvioitaessa verkkotunnusten yhdistäminen toimijoihin ei ollut enää tarkoituksenmukaista, sillä muiden verkkotunnuksissa security.txt:n julkaisseet hosting-palvelut ja ohjelmistotuottajat eivät kontrolloi käytettävää varmennetta eivätkä sitä, valitseeko asiakas tai käyttäjä salatun vai salaamattoman yhteyden. Tämä näkyi tuloksissa siitä, että samat toimijat saattoivat julkaista tiedoston verkkotunnuksissa, joiden verkkosivuilla käytettiin sekä luotettuja varmenteita, virheellisiä varmenteita että kokonaan salaamattomia yhteyksiä. Siksi TLS-yhteyksiin liittyvät tilastot on esitetty ainoastaan verkkotunnuksittain.

Luotettavuuden arvioinnin kannalta merkittäväksi katsotaan se, vastaako yhteydessä käytetty varmenne verkkotunnusta ja onko se varmentajan allekirjoittama (Foudil & Shafranovich 2022, 15). Näiden lisäksi vain salaamattomilla yhteyksillä tiedoston saataville asettaneiden verkkotunnusten (6 % kummallakin kerralla) sekä vanhentuneiden varmenteiden osuus ilmenee kuvasta 10.



Kuva 10. Yhteyden salaus ja varmenteen luotettavuus; osuus verkkotunnuksista.

Lokakuussa 2022 varmenteista vain yksi ei ollut luotettu: tämäkin varmenne oli varmentajan allekirjoittama, mutta se oli vanhentunut noin viikon ennen ha-

vainnointihetkeä. Tällaista virhettä voidaan pitää tilapäisenä, ja se olikin korjattu helmikuussa 2023. Sen sijaan varmenteista 10 % lokakuussa 2022 ja 9 % helmikuussa 2023 kuuluivat eri isäntänimelle. Tyypillinen tapaus oli verkkotunnus, josta oli edelleenohjaus toiseen verkkotunnukseen, kuten haltijan pääverkkotunnukseen.

TLS-yhteyksistä (727 verkkotunnusta lokakuussa 2022 ja 758 helmikuussa 2023) kaikki oli kätelty protokollan nykyisellä versiolla TLS 1.2. Koska ZGrab 2.0 ei tue TLS 1.3 -versiota, tarkoittaa tämä sitä, että kaikkien security.txt:n julkaisseiden verkkotunnusten HTTPS-protokollaa käyttävät verkkopalvelut kykenivät kättelemään vähintään versiolla TLS 1.2. Tutkimuksen ulkopuolelle jäi kuitenkin sen kartoittaminen, tukivatko nämä palvelut myös vanhempia protokollaversioita, joita RFC 8996:n (Moriarty & Farrell 2021) mukaan ei pitäisi enää tukea, sillä RFC 9116 ei ota kantaa käytettävään versioon.

5.3.2 OpenPGP-allekirjoitukset

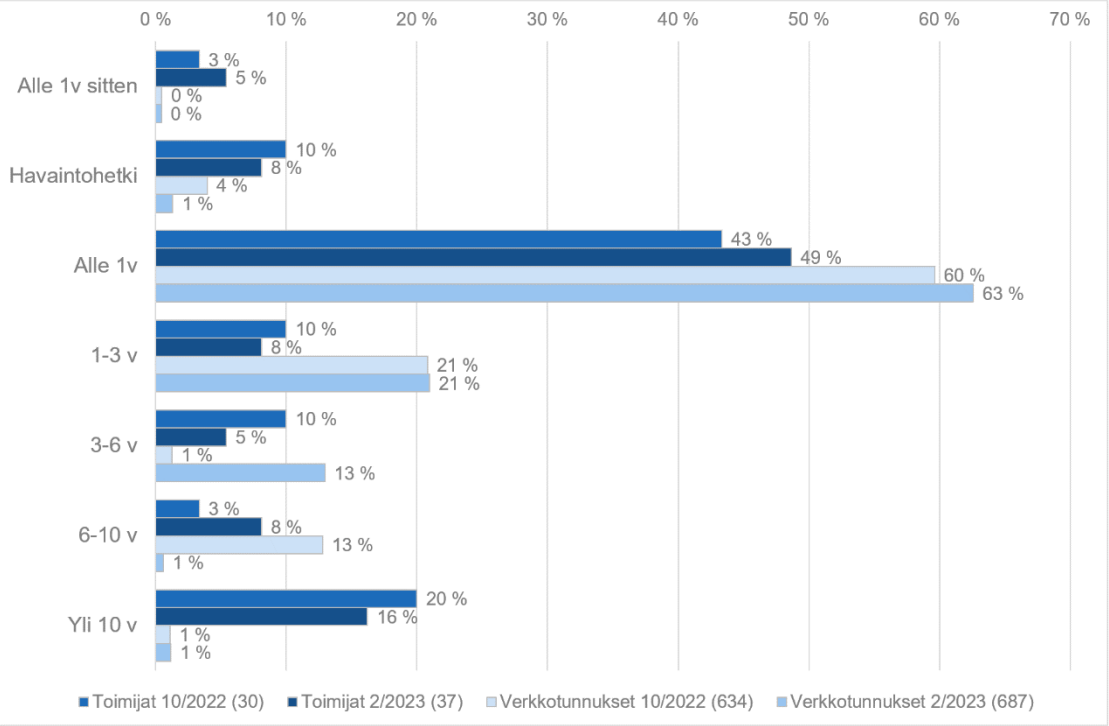
Vaikka toimijoista 20 % (14) tarjosi salausmenetelmien käyttöä "Encryption"-kentässä 7 %:ssa verkkotunnuksista (57) lokakuussa 2022 ja 23 % (18) 6 %:ssa verkkotunnuksista (51), oli itse tiedoston allekirjoittaminen OpenPGP:llä hyvin harvinaista. Vain kaksi toimijaa allekirjoitti tiedoston RFC 9116:n tukemalla tavalla seitsemässä verkkotunnuksista lokakuussa 2022. Helmikuussa 2023 uusia toimijoita oli kolme, mutta yksi niistä oli Kyberturvallisuuskeskuksen oma tiedosto (ks. liite 2), jota ei näkynyt lokakuun 2022 havainnoissa. Yhden toimijan allekirjoituksen tiiviste ei vastannut tiedoston sisältöä, ja kahdella allekirjoituksiin käytetyt OpenPGP-avaimet olivat vanhentuneita. Näiden lisäksi yksi toimija tarjosi kummallakin kerralla allekirjoituksen yhdeksässä verkkotunnuksessa "Signature"-kentässä viitatulla sig-tiedostolla, mikä oli varhaisissa luonnosversioissa tuettu tapa tiedoston allekirjoittamiseen.

"Encryption"-kentässä OpenPGP-avaimiin voidaan viitata myös esimerkiksi URI-skeemojen "dns:" ja "openpgp4fpr:" mukaisilla osoitteilla (Foudil & Shafranovich 2022, 7), mutta käytännössä vain HTTPS-osoitteita oli käytössä. Helmikuussa 2023 kaksi toimijaa käytti "openpgp4fpr:"-osoitetta neljässä verkkotunnuksessa, mutta toinen näistä oli Kyberturvallisuuskeskuksen oma tie-

dosto. Mikäli OpenPGP-avain ladataan samasta verkkotunnuksesta kuin security.txt, voi molempien luotettavuus vaarantua yhtä aikaa, mikäli sivustolle murtaudutaan. Toisaalta myös ulkopuolisesta verkkotunnuksesta ladatun avaimen luotettavuutta on yhtä vaikea arvioida, jos tiedoston on päässyt julkaisemaan pahantahtoinen taho. Vain OPENPGPKEY-tietueisiin (Wouters 2016) viittaavat "dns:"-osoitteet voisivat käytännössä tarjota riippumattoman ja DNSSEC PKI:llä allekirjoitettuna myös luotettavan kanavan avaimen lataamiseen.

5.3.3 Voimassaoloaikojen jakauma

RFC 9116 suosittelee, että "Expires"-kentässä julkaistu aika olisi korkeintaan vuoden tulevaisuudessa, jotta tiedoston sisällön ajantasaisuus tulisi tarkastettua säännöllisesti (Foudil & Shafranovich 2022, 8). Voimassaoloaikojen jakaumaa arvioitaessa on otettu huomioon myös luonnosversion mukaiset aikaleimat sekä väärin kirjoitettu "Expiration"-kenttä, sillä kumpikin voidaan tulkita aikomukseksi julkaista voimassaolon loppumisajankohta. Täten ne kertovat aivan vastaavalla tavalla siitä, millaisia aikaleimoja on käytetty ja miten hyvin niitä on päivitetty. Kuva 11 esittää aikaleimojen jakaumaa suhteessa havaintohetkeen sekä toimijoiden että verkkotunnusten osuuksina.



Kuva 11. "Expires"-kentän aikojen jakauma etäisyytenä havaintohetkestä.

Suositeltuun aikahaarukkaan mahtui 43 % toimijoista ja 60 % verkkotunnuksista lokakuussa 2022. Helmikuussa 2023 osuudet olivat hieman kasvaneet, ja suosituksen täytti 49 % toimijoista ja 63 % verkkotunnuksista. Kasvuun vaikuttivat neljä uutta toimijaa, joiden aikaleimat olivat alle vuoden tulevaisuudessa, yksi toimija, joka oli päivittänyt vanhentuneet aikaleimansa sekä kaksi toimijaa, joiden muuttumattomana pysynyt aikaleima osui nyt uuteen haarukkaan. Tämän vuoksi kasvusta ei voida suoraan päätellä nousevaa trendiä. Positiivista kuitenkin oli, että helmikuussa 2023 suositusten mukaiseen aikaväliin osuvista molemmissa havainnoissa mukana olleista 14 toimijasta 11 oli päivittänyt aikaleimojaan näiden neljän kuukauden sisällä, mikä kattoi peräti 97 % vastaavista verkkotunnuksista (415 verkkotunnusta 430:stä).

Vanhentuneita, pitkään päivittämättä olleita tiedostoja löytyi lokakuussa 2022 vain yhden toimijan kolmesta verkkotunnuksesta ja helmikuussa 2023 kahden toimijan kolmesta verkkotunnuksesta. Verkkotunnuksista 26 sisälsivät aikaleiman samalta päivältä kuin tiedosto oli ladattu, joista osa oli ennen ja osa jälkeen lataushetken. Näitä oli neljällä eri toimijalla, joiden tiedostot olivat muodoltaan hyvin samankaltaisia. Vaikutti siltä, että tiedosto generoitaisiin päivittäin automaattisesti uudelleen saman päivän aikaleimalla.

Erityisesti yli 10 vuoden päässä tulevaisuudessa olevat aikaleimat antoivat olettaa, ettei niitä ollut tarkoitettu päivitettäviksi lainkaan, sillä niissä käytettiin aikaleimoja vuosilta 2038, 2050, 2099 ja 2100, eikä niihin tullut lainkaan muutoksia neljän kuukauden aikana. Nämä olivat kuitenkin harvinaisia, sillä niitä löytyi vain kuudelta toimijalta, joiden hallussa oli 7 verkkotunnusta lokakuussa 2022 ja 6 verkkotunnusta helmikuussa 2023.

6 POHDINTA

6.1 Johtopäätökset

Aiempien tutkimusten tavoin RFC 9116:n käyttöönottoa voidaan myös suomalaisten organisaatioiden hallitsemien fi-verkkotunnusten osalta pitää vielä marginaalisena ilmiönä, sillä security.txt oli julkaistu helmikuussa 2023 vain noin 2,2 ‰:ssa kaikista verkkotunnuksista. Lisäksi ilmiössä painottui se, että muutama toimija oli vaikuttanut valtaosaan tiedostoista useammassa verkkotun-

nuksessa julkaistujen tiedostojen samankaltaisuuksien perusteella. Tällöin yksittäisen suuren toimijan panostus julkaisemiensa tiedostojen sisällön laatuun voi vaikuttaa merkittävästi kaikkiin tiedostojen sisällön laatuun liittyviin tuloksiin, mikäli tarkastelua tehdään vain yksittäisten verkkotunnusten tasolla. Tämän vuoksi oli mielekästä kääntää tarkastelu verkkotunnustasolta myös toimijatasolle. Myös käyttötapaukset, joissa tiedoston oli julkaissut jokin muu taho kuin se organisaatio, jonka hallussa verkkotunnus oli, kuten hosting-palvelun-tarjoaja tai ohjelmistotuottaja, olivat yleisiä.

Kuten RFC 9116:n luonnosversioiden lukuisten muutosten, varhaisista luonnosversioista tehtyjen uutisartikkeleiden päivittämättömyyden sekä myös aiempien tutkimusten perusteella oli arvioitavissa, tämäkin tutkimus vahvisti, että tiedostojen sisällön laadussa on suurta vaihtelevuutta. Tämä vaihtelevuus vaikuttaa suuresti koneluettavuuteen, joka on RFC 9116:n keskeisimpiä tavoitteita verrattuna siihen, että tiedot etsittäisiin manuaalisesti organisaation kotisivuilta. Tällä hetkellä ei ole takeita siitä, että haavoittuvuuden löytyessä voitaisiin yksittäiseltä verkkosivulta koneellisesti löytää security.txt:n avulla yhteystiedot ja siten esimerkiksi automatisoida raportointia. Vielä vaikeampaa on koneellisesti arvioida tätä kautta saadun tiedon luotettavuutta, mikä jäänee yksittäisen tietoturvatutkijan tehtäväksi myös siinä tapauksessa, että käytäntö yleistyisi ja tiedostojen sisällön laatu yhtenäistyisi.

TLS-yhteyksissä käytettyjen versioiden ajantasaisuudesta ja varmenteiden luotettavuuksista saadut tulokset olivat positiivisia. Lähestulkoon kaikki sivustot tukevatkin jo vähintään TLS:n versiota 1.2, ja TLS 1.1 ja 1.0 ovat käytössä vain taaksepäin yhteensopivuuden vuoksi, kuten ilmenee esimerkiksi F5 Labs:n raportista (Warburton 2021). Varmenteet olivat pääasiassa luotettuja ja ajantasaisia, vaikka osa varmenteista saattoikin kuulua esimerkiksi saman organisaation toiselle verkkotunnukselle. Sen sijaan OpenPGP:n käyttö tiedostojen allekirjoittamiseen oli hyvin harvinaista, vaikkakin OpenPGP-avaimia tarjottiin jonkin verran "Encryption"-kentän kautta kommunikaation salaamista varten.

Näissä johtopäätöksissä tiivistettiin keskeiset tulokset, jotka vastasivat neljään tutkimuskysymykseen, sekä mitä niistä voidaan suoraan kiistatta päätellä.

Seuraavassa alaluvussa arvioidaan saatujen tulosten merkitystä tutkimusongelman näkökulmasta, kuten miten hyvin tarkasteltu käytäntö kykenee käytännössä ratkaisemaan tutkimusongelmaa. Lisäksi pohditaan tekijöitä, jotka saattavat vaikuttaa käytännön yleistymiseen. Pohdinnan päätteeksi tarkastellaan vielä kriittisesti tutkimuksen luotettavuutta sekä sen aikana nousseita avoimeksi jääneitä kysymyksiä, jotka tuottavat jatkotutkimusaiheita.

6.2 Tulosten merkitys, hyödynnettävyys ja tulevaisuuden näkymät

Ilmiön marginaalisuudesta seuraa, että toistaiseksi tiedoston etsiminen verkkosivuston polusta "/.well-known/security.txt" pysyy vain yhtenä – joskin helpoutensa vuoksi mahdollisesti ensisijaisena – keinona, kun etsitään yhteystietoja haavoittuvuuksista ilmoittamiseen. Tämä rajoittaa myös security.txt:n hyödyllisyyttä Kyberturvallisuuskeskuksen haavoittuvuuskoordinaatiossa, mutta tilanteen kehittymistä kannattaa edelleen seurata. Tällaisen tiedoston julkaisemisen mahdollisuudesta ja sen hyödyistä tiedottamisesta ei ole haittaakaan, kunhan tiedostetaan, ettei voida vielä puhua yleisestä käytännöstä.

Hieman yllättävämpää oli niiden käyttötapausten yleisyys, joissa tiedoston oli julkaissut jokin muu taho kuin se organisaatio, jonka hallussa verkkotunnus oli, kuten hosting-palveluntarjoaja tai ohjelmistotuottaja. RFC 9116 (Foudil & Shafranovich 2022, 3–4) ei kuitenkaan selkeästi määrittele, mikä suhde organisaatiolla, joka haluaa käyttää menetelmää yhteystietojensa ja ilmoittamiskäytäntöjensä julkaisemiseen, tulisi olla verkkotunnukseen, jossa tiedosto on julkaistu. Erityisesti silloin, kun kyseessä on kaikkien asiakkaiden verkkotunnuksiin samalla sovelluksella tuotetusta palvelusta tai sovelluksesta, jonka asiakas tai käyttäjä asentaa itse esimerkiksi aliverkkotunnukseen, saattaa palvelun tuottajan tai ohjelmiston kehittäjän yhteystiedot olla jopa parempi kanava haavoittuvuuksista ilmoittamiseen. Mikäli näissä tapauksissa tiedon haavoittuvuudesta saisi vain verkkotunnuksen haltija eli käytännössä palvelun tai sovelluksen käyttäjä, tulisi tästä ylimääräinen välikäsi, joka hidastaisi tiedon kulkua niille tahoille, jotka oikeasti pystyvät haavoittuvuuden paikkaamaan. Pahimmassa tapauksessa tieto haavoittuvuudesta saattaisi tällöin välittyä jopa sellaisiin käsiin, joissa sitä käytettäisiin pikemminkin väärin kuin ongelman korjaamiseen.

Varmenteiden luotettavuus, ajantasaisuus ja jopa niiden melko hyvä vastaavuus verkkotunnuksiin ylitti odotukseni. Tätä voisi selittää esimerkiksi se, että security.txt:n käyttöönottaneet saattavat olla yleisesti ottaen keskimääräistä tietoturvatietoisempia, jolloin myös muista tietoturvallisuuteen vaikuttavista teknologioista pidetään hyvää huolta. Sen sijaan TLS-versioiden ajantasaisuudesta ei voida päätellä käyttöjärjestelmien ja sovellusten ajantasaisuutta, vaikka kääntäen TLS 1.2:n tuen puute viittaisi päivittämättömään ja jopa vanhentuneeseen ympäristöön. Security.txt:n julkaisemisen ja muun tietoturvatietoisuuden välisen yhteyden selvittäminen vaatisi kuitenkin lukuisten muiden tietoturvakäytänteiden kartoittamista sekä niiden välisten korrelaatioiden havaitsemista. Tämän lisäksi tulisi vielä pystyä osoittamaan korrelaatioiden johtuvan tietoturvatietoisuudesta eli näiden välillä pitäisi osoittaa olevan riippuvuutta tai kausaalisuutta (Töttö 2012, 179).

Koska OpenPGP-avaimia esiintyi "Encryption"-kentässä, ei niiden harvinaisuus tiedoston allekirjoittamisessa ehkä suoraan kerro niinkään OpenPGP:n käytön harvinaisuudesta kuin siitä, ettei tiedoston allekirjoittamisen mahdollisesti nähdä tuottavan riittävästi lisäarvoa suhteessa nähtyyn vaivaan tilanteessa, jossa yhteys on kuitenkin salattu ja luotettu. Luottamusverkon puuttumisen (Ulrich ym. 2011, 490, 504) vuoksi joudutaan avaimen luotettavuutta arvioimaan usein muilla perusteilla, jotka ovat osittain samat kuin security.txt:n luotettavuuden arvioinnin perusteet.

Jotta security.txt:n OpenPGP-allekirjoitukset yleistyisivät, täytyisi OpenPGP-avainten varmentamisen luottamusverkosta riippumattomien menetelmien, kuten OPENPGPKEY DNS-tietueiden (Wouters 2016) tai luonnosvaiheessa (*Internet-Draft*) olevan *OpenPGP Web Key Directory*n (Koch 2022), yleistyä ensin. Koska GnuPG:n kehittäjät ovat lähteneet kehittämään tätä jälkimmäistä omaa toteutustaan avainten jakamiseen ja tuomiseen GnuPG-sovellukseen, pidän epätodennäköisenä sitä, että he alkaisivat tukea myös OPENPGPKEY-tietueita. Keskenään kilpailevilla standardeilla ja muilla käytänteillä tuntuu olevan tapana hidastaa toistensa yleistymistä, koska organisaatioilla ei ole aikaa tutustua, saati omaksua niitä kaikkia. Myös RFC 9116 näyttää jo saaneen vastaavalla tavalla kilpailevan DNS-pohjaisen toteutuksen "*DNS Security TXT*", joka pyrkii lisäämään TXT-tietueina yhteystietoja ja linkkejä turvallisuuskäytäntöihin (Carroll & Ellis s.a.). Vaikka kehittäjien mukaan tavoitteena on laajentaa

security.txt nimipalveluihin, epäilen tämän aiheuttavan pikemminkin kilpailua ja sekaannuksen vaaraa.

6.3 Tulosten luotettavuuden arviointi

6.3.1 Validiteetti

Tutkimusongelmasta johdettuina tutkimuskysymyksinä oli selvittää, miten yleistä security.txt:n julkaiseminen on suomalaisissa organisaatioissa, millaisia tietoja niissä on julkaistu, sekä miten hyvin toteutuneet käytännöt noudattavat RFC 9116:n asettamia määrittelyjä ja ottavat huomioon sen nostamia turvallisuusnäkökohtia. Validiteetti mittaa sitä, miten hyvin käytetty menetelmä soveltuu tällaisten vastausten saamiseen. Käsitevaliditeetin näkökulmasta validiteetin argumentointi korostuu, jolloin pelkästään käytetyn mittarin ja mitatun ilmiön vastaavuus ei riitä, vaan myös mittaustuloksille annettu tulkinta pitää pystyä perustelemaan (Nummenmaa ym. 1997, 206). Jotta tämä vaatimus täyttyisi, valitut muuttujat johdettiin tutkimuskysymyksistä perustellen valintoja teoreettisella viitekehyksellä, käytettyjen teknisten menetelmien ja valittujen muuttujien looginen yhteys osoitettiin tutkimuksen toteutuksen kuvauksessa, ristiintarkastuksia käytettiin varmistamaan havaintojen luotettavuutta sekä niiden oikeaa tulkintaa, ja johtopäätökset johdettiin tuloksista loogisella päättelyllä. Tutkimuksen tulokset vastasivat kaikkiin esitettyihin tutkimuskysymyksiin, ja niistä voitiin tehdä tutkimusongelman kannalta merkityksellisiä johtopäätöksiä.

Validiteettiin vaikutti oleellisesti tutkimusjoukon valinta. Rajaus organisaatioihin oli perusteltua, sillä RFC 9116 on tarkoitettu organisaatioiden käyttöön. Lisäksi oli tärkeä arvioida, miten tutkimus on rajattavissa luotettavasti vain suomalaisiin organisaatioihin, ja tämän jälkeen saada kattavaa tietoa valitun joukon käytännöistä. Maakohtainen ylätason verkkotunnus sekä siitä helposti avoimesta rajapinnasta saadut tiedot, jotka sisälsivät vain organisaatioiden hallinnassa olevat verkkotunnukset sekä tiedon haltijan kotimaasta, tarjosivat määriteltyjä rajauksia vastaavan tutkimusjoukon.

Suppeampi otantatutkimus ei olisi tarjonnut kattavaa kuvaa ilmiöstä – varsinkin kun otetaan huomioon, miten marginaaliseksi ilmiö osoittautui. Koska tutkimuksen aineistona käytettiin kaikista rajauksen mukaisista verkkotunnuksista

kaikilla mahdollisilla polkuyhdistelmillä kerättyjä tietoja, oli sen muodostama aineisto kattava poikkileikkaus kummaltakin havainnointihetkeltä. Näin kattavan aineiston pohjalta pystyi analysoimaan kaikkia niitä yksityiskohtia, joita tutkimuskysymyksiin vastaamiseksi oli tarkasteltava.

Kvantitatiiviset tulokset olivat yksiselitteisiä ja saatu tarkastelemalla tutkimuskysymyksistä johdettuja, hyvin määriteltyjä muuttujia. Validiteetin kriittisessä tarkastelussa on kuitenkin tiedostettava, että mitä harvinaisemmista yksittäisistä valinnoista oli kyse, sitä heikommin tulokset ovat yleistettävissä. Koska kyseessä on kokonaistutkimus, ei tässä tapauksessa ole kyse otannan tulosten yleistämisestä perusjoukkoon vaan siitä, ettei niiden perusteella voida ennustaa samanlaisten valintojen tulevan olemaan yhtä yleisiä käytännön yleistyessä. Tutkimuksen ennustevaliditeetti eli kyky ennustaa myöhemmin tapahtuvaa ilmiötä on siis heikko (Nummenmaa ym. 1997, 205). Tämä näkyi hyvin jo lokakuun 2022 ja helmikuun 2023 tulosten eroista, joissa selkeiden trendien asemesta esiintyi paljon satunnaisuutta ja yksittäisten toimijoiden vaikutusta. Kyseessä ei ole varsinaisesti käytetyistä menetelmistä johtuva puute, vaan olosuhteilla on vaikutusta ennustevaliditeettiin: mitä enemmän ilmiö mahdollisesti yleistyy ja mitä pidemmälle seurantaa jatketaan, sitä paremmin samalla menetelmällä toteutettu tutkimus kykenee ennustamaan myös tulevaa kehitystä. Mitä useammat toimijat ottavat menetelmän käyttöönsä, sitä pienemmäksi myös yksittäisen toimijan vaikutus tuloksiin laimenee.

Syyt tiedoston julkaisseiden tahojen tekemien valintojen takana eivät ole suoraan luettavissa julkaistuista tiedostoista, mutta tätä tutkimuskysymykset eivät myöskään edellytä. Valitusta menetelmästä koituu tutkimukselle samalla kuitenkin rajoite, josta olisi voitu päästä yli käyttämällä menetelmätriangulaatiota, jossa ilmiötä lähestytään monelta suunnalta monimenetelmäisesti (Kananen 2011, 124). Kvantitatiivisella menetelmällä saatujen tulosten pohjalta olisi voitu tehdä kvalitatiivista, täsmentävää tutkimusta, jossa tiedoston julkaisseilta toimijoilta olisi kyselty avoimen kyselyn tai haastattelun muodossa, miten he ovat päätyneet valintoihinsa. Tutkimusjoukon tavoittamisen olisi tehnyt helpoksi se, että nimenomaan tiedoston julkaisseiden toimijoiden yhteystiedot olivat suoraan luettavissa tutkimusaineistosta. Tuloksia esittelevässä luvussa sekä johtopäätöksissä on tämän rajoituksen vuoksi pitäydytty niissä faktoissa, jotka

havainnoista voidaan suoraan päätellä. Johtopäätöksistä nousseet tulkinnat ja pohdinnat ovat selkeyden vuoksi eroteltu omaksi alaluvukseen.

6.3.2 Reliabiliteetti

Tutkimusaineiston kattavuus vaikuttaa merkittävästi myös tutkimuksen tulosten reliabiliteettiin eli luotettavuuteen. Koska aineisto on kerätty koko perusjoukosta, vältytään suoraan peitto- ja otantavirheiden mahdollisuudelta (Holopainen & Pulkkinen 2008, 29). Tällöin havaintojen yleistettävyyttä ei tarvitse arvioida, vaan näkyvyyttä yksittäiseen ajanhetkeen voidaan pitää sellaisenaan kattavana. Tuloksia tulkittaessa on kuitenkin tiedostettava, että ilmiön marginaalisuudesta ja muutaman suuremman toimijan vaikutuksesta lukuisiin verkkotunnuksiin seuraa vääristymiä, mikäli tuloksia tarkastellaan pelkästään yksittäisten verkkotunnusten näkökulmasta. Näitä vääristymiä on korjattu analysoimalla ilmiötä verkkotunnustason lisäksi toimijoiden tasolla.

Tutkimuksessa käytetyt aineistonkeruu- ja analyysimenetelmät sekä niissä käytetyt tekniset työkalut on kuvattu tarkasti, mikä tekee tutkimuksesta toistettavan. Vaikka tarkkoja komentosarjoja, joilla työkaluja on toisiinsa yhdistetty, ei olekaan julkaistu opinnäytetyön osana, pystyisi raportista ilmenevien kuvausten perusteella laatimaan samanlaisen tutkimusasetelman ja toistamaan tutkimuksen. Ilmiön muuttuvasta luonteesta johtuen tulokset saattaisivat poiketa tässä tutkimuksessa saaduista tuloksista, mutta tutkimus antaisi yhtä luotettavan kuvan havainnointihetken tilanteesta. Kananen (2012, 119–120) tähdentääkin, että tällä tavoin huomattu alhainen reliabiliteetti ei välttämättä johdu mittarin heikosta stabiiliudesta, ja että stabiiliutta voidaan parantaa uusintamittauksella. Tässä tutkimuksessa tehtiin uusintamittaus neljän kuukauden kuluttua ensimmäisestä aineistonkeruusta. Saadut tulokset olivat hyvinkin samankaltaisia, sillä niissä nousi uudestaan esiin samoja verkkotunnuksia ja toimijoita, mikä viittaa mittarien stabiiliuteen. Kuitenkin eroavaisuuksiakin ilmeni, mikä puolestaan kertoo ilmiön hitaasta kehittymisestä.

Aineiston keräysvaiheeseen liittyy epätarkkuustekijöitä, koska yhdellä keräyskerralla samaa polkua on yritetty hakea vain kerran. Internetin toiminnan luonteesta johtuen näkyvyys joihinkin palveluihin voi hetkellisesti häiriintyä sekä havainnoitsijan että kohteen puolella samoin kuin missä tahansa kohtaa reittiä

näiden välillä. Ainakin Liikenne- ja viestintävirasto Traficomin omien verkkotunnusten voidaan varmuudella todeta puuttuneen kokonaan lokakuun 2022 ja osittain helmikuun 2023 aineistonkeruusta havainnointiin käytetyn verkon konfiguraatiosta johtuen. Erityisesti HTTPS-yhteyksissä aikakatkaisuja ja tuntemattomia virheitä esiintyi runsaasti. Tällaiset havainnot tulkittiin HTTPS-palveluiden puutteena kyseisessä osoitteessa. Reliabiliteetin kannalta on kuitenkin aiheellista tarkastella, olisiko niitä syytä pitää vastaamattomuutena eli katona, ja voisiko niiden vuoksi jokin ilmiön kannalta oleellinen osajoukko olla karsiutunut pois tai aliedustettu (Holopainen & Pulkkinen 2008, 41; Nummenmaa ym. 1997, 22). Vähintäänkin tulee raportoida, miten paljon tietoa puuttui, millaista puuttuvuus oli ja mitä sille tehtiin (Töttö 2012, 140).

Tulkintaa aikakatkaisuista ja tuntemattomista virheistä palvelun puutteena puoltaa se, että verkkotunnuksen juuren ja www-aliverkkotunnuksen sekä kahden haetun polun havainnointihetkien välillä oli aineistonkeruun hitaudesta johtuvaa viivettä, joka olisi antanut tilapäisille verkko-ongelmille aikaa ratketa. Tästä viiveestä huolimatta tulokset neljän samaa protokollaa käyttäneen polun välillä olivat saman verkkotunnuksen osalta pääasiassa johdonmukaisia, ja aikakatkaisuja esiintyi tasaisesti koko aineistonkeruun ajan. Verkkokatkokseen olisi viitannut se, että aikakatkaisuja olisi ollut tietyllä välillä enemmän kuin muulloin. Tulkintaan selkeästä mittausvirheestä olisi puolestaan johtanut se, mikäli saman verkkotunnuksen kahdesta eri polusta olisi saatu ristiriitaisia tuloksia aikakatkaisujen tai yhteysvirheiden suhteen. Aikakatkaisuihin liittyviä haasteita oli tutkittu jo työkalujen testauksen aikana, kuten tutkimuksen toteutuksesta kertovassa luvussa tarkemmin kuvattiin, ja niistä mahdollisesti aiheutuvat vääristymät oli täten minimoitu.

Mikäli kaikkia aikakatkaisuja kuitenkin ajateltaisiin katona, voidaan niiden todeta kohdistuvan samalla todennäköisyydellä sekä security.txt:n julkaisseihin että julkaisemattomiin verkkotunnuksiin. Katoa voidaan toisin sanoen pitää havaittujen julkaistujen tiedostojen näkökulmasta satunnaisena puuttuvuutena, joka ei ole riippuvainen tarkasteltavasta muuttujasta, ja joka siksi on luotettavuuden kannalta harmiton (Töttö 2012, 126). Vaikka kyseessä on kokonaistutkimus, kadon aiheuttaman virheen suuruutta voitaisiin vertailun vuoksi arvioida samoilla menetelmillä kuin pienemmästä otannasta aiheutuvaa virhemarginaalia. Prosenttiyksikön piste-estimaatin ($\hat{p}$) enimmäisvirheen määrittelyyn

desimaalilukuna voidaan käyttää yhtälön 1 virhemarginaalin (*e*) kaavaa, johon on sijoitettu 99 %:n luottamustasoon liittyvä kriittinen arvo 2,58 (Holopainen & Pulkkinen 2008, 168–174).

$$e = \pm 2{,}58 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \quad (1)$$

Yhtälöstä nähdään, että enimmäisvirhe on suurimmillaan, kun $\hat{p}$ = 0,5 eli piste-estimaatin ollessa 50 %, josta se pienenee kumpaankin suuntaan. Ratkaisevaa on kuitenkin otoskoon (*n*) vaikutus. Kato voi vaikuttaa vain niihin verkkotunnuksiin, joihin pyyntöjä on tehty, joten otoskooksi on laskettava onnistuneiden yhteyksien määrä. Kun helmikuussa 2023 saatiin yli 827 000 onnistunutta HTTPS-yhteyttä, niistä tehtävien laskelmien enimmäisvirhe korkeimmassa kohdassa on kaavan mukaan ±0,0014 eli 0,14 prosenttiyksikköä kumpaankin suuntaan, eikä prosenttiosuuksia ole tuloksissa esitetty desimaalien tarkkuudella. Täten on oletettavissa, ettei mahdollisella kadolla ole merkittävää vaikutusta prosenttiyksiköinä esitettyihin tuloksiin.

Kokonaan vaille havaintoja kaikista HTTPS-poluista aikakatkaisujen ja tuntemattomien virheiden vuoksi jäi helmikuussa 2023 noin 42 tuhatta verkkotunnusta. Mikäli aukon paikkaamiseksi tehdään tilastollinen arvaus (Töttö 2012, 140) ja oletetaan suoraviivaisesti, että näistä kaikista olisi ollut löydettävissä security.txt-tiedostoja samassa suhteessa kuin muista verkkotunnuksista, olisi havainnoimatta jäänyt arviolta 109 tiedoston julkaissutta verkkotunnusta. Tämä olisi nostanut HTTPS-poluista tuloksena saadun 2,1 ‰:n esiintyvyyden korkeintaan 2,4 ‰:een. Todellisuudessa virhe on todennäköisesti tätä pienempi, koska kaikki virheet eivät voineet olla havainnoinnista riippuvia, eikä mahdollisesta kadosta johtuvaa virhettä voida myöskään pitää merkittävänä johtopäätösten kannalta.

6.4 Jatkotutkimusaiheet

Opinnäytetyössä käytetyn menetelmän ja kehitettyjen työkalujen avulla ilmiön kehittymistä voidaan seurata Kyberturvallisuuskeskuksessa vaivattomasti, sillä sekä aineistonkeruu että analysointi ovat pitkälti automatisoituja. Tämä mahdollistaa sellaisen pitkittäisen tutkimuksen, johon opinnäytetyön aikaikkuna ei

riittänyt. Tutkimukseen saattaa olla mahdollista lisätä myös sen arviointia, miten Kyberturvallisuuskeskuksen omat julkaisut vaikuttavat ilmiön leviämiseen ja laadulliseen kehittymiseen, mikäli muutosten ja julkaisujen välillä havaitaan ajallisia yhteyksiä.

Edellä kuvattu jatkotutkimus on tyypiltään kehittämistutkimusta, jossa ei sitouduta objektiivisuuteen, vaan pyritään myös vaikuttamaan tarkasteltuun ilmiöön sekä mittaamaan näitä vaikutuksia. Tämän ansiosta siihen olisi luontevaa yhdistää validiteetin käsittelyssä menetelmätriangulaationa esiin nostettua kvalitatiivista haastattelututkimusta, joka varsinkin suurimpiin toimijoihin kohdistettuna voisi myös vaikuttaa julkaistavien tiedostojen sisällön laatuun. Tutkimusstrategisesti sekä aiheeseen liittyvät julkaisut että keskustelut toimijoiden kanssa voitaisiin nähdä esimerkiksi toimintatutkimukselle tyypillisinä käytännönläheisinä interventioina, joiden vaikutuksia pystytään reflektoimaan (Heikkinen 2010, 27–36). Tällaista strategiaa on Kyberturvallisuuskeskuksessa aikaisemmin sovellettu Kontisen (2020) opinnäytetyössä, jossa rajatut kohderyhmät saivat erityistä opastusta sähköpostiväärennösten torjuntaan, ja muutosta kohderyhmien käytännöissä verrattiin muutokseen vertailujoukossa.

Sekä tässä tutkimuksessa että aiemmissa tutkimuksissa verkkotunnuksia on pystytty yhdistämään sen perusteella, että ne ovat jakaneet samankaltaisen security.txt-tiedoston, mutta tämä ei ole suoraan viitannut siihen, että verkkotunnukset olisivat saman entiteetin hallinnassa, sillä tiedoston on saattanut lisätä myös esimerkiksi julkaisualusta. Jokaisen organisaation hallitseman fi-verkkotunnuksen tietoihin liittyy organisaation yksilöivä yritys- ja yhteisötunnus eli Y-tunnus, jonka perusteella saman organisaation verkkotunnuksia voidaan tarkastella yhtenä kokonaisuutena myös silloin, kun security.txt on julkaistu vain osassa. Tällöin voidaan saada tietoa myös siitä, miten monessa organisaatiossa menetelmä on otettu käyttöön vain joidenkin verkkotunnusten osalta; jotkin organisaatiot saattavat julkaista tiedoston esimerkiksi vain pääasiallisella verkkotunnuksellaan. Y-tunnuksen perusteella havaintoja on mahdollista vertailla myös toimialakohtaisesti. Tutkimusten laajentaminen ja täsmentäminen tähän suuntaan ei kuitenkaan ole vielä ajankohtaisia, sillä tällaista vertailua on mielekästä tehdä vasta, mikäli ilmiö myöhemmin yleistyy.

Tätä tutkimusta tehtäessä vastaavaa kokonaisen ylätason verkkotunnuksen kartoittamista oli tehty vasta Sveitsin maatunnuskohtaisen ch-verkkotunnuksen osalta (Foudil 2022), ja tämäkin oli ei-tieteellinen selvitys. Mielenkiintoisia jatkotutkimuskohteita olisivat vastaavat tutkimukset muista maista sekä niiden tulosten vertailu keskenään. Tutkimuksen laajentaminen tähän suuntaan ei kuitenkaan liity suoraan toimeksiantajan lakisääteisiin tehtäviin. Maakohtaisten ylätason verkkotunnusten osalta tutkimuksia tai selvityksiä voitaisiin tehdä esimerkiksi muiden maiden vastaavien viranomaisten toimesta. Näiden keskinäinen vertailu eli välillinen käyttäminen jatkotutkimusten aineistona taas mahdollistaisi laajempien tutkimusten tekemisen ilman tarvetta kerätä laajaa aineistoa koko Internetistä, jossa on satoja miljoonia verkkotunnuksia.

Yhden sovelluksen havaittiin julkaisevan tiedostoja kaikissa niissä isäntänimissä, joihin sovellus on asennettu. Tällaisia sovelluksia saattaisi löytyä muitakin, mikäli tutkimuksia laajennettaisiin muihin aliverkkotunnuksiin, sillä harvat sovellukset päätyvät julkaistuiksi suoraan verkkotunnuksen juureen tai www-aliverkkotunnukseen. Tällaisen tutkimuksen haasteena olisi päättää, mitä aliverkkotunnuksia tutkimukseen otettaisiin mukaan ja miten listalle löydettäisiin kandidaatteja.

Tutkimuksen johtopäätöksiin perustuvissa pohdinnoissa tunnistettiin liittyviä teknologioita ja ilmiöitä, joilla saattaa olla vaikutusta myös RFC 9116:n käytön yleistymiseen. Myös kilpailevia ja rinnakkaisia käytäntöjä sekä mahdollisia RFC 9116:n päivityksiä ja laajennoksia tulisi tunnistaa ja seurata. Lisäksi ainakin OpenPGP:n (Callas ym. 2007), *OpenPGP Web Key Directory*n (Koch 2022), OPENPGPKEY:n (Wouters 2016) ja sitä kautta myös DANE:n (Hoffman & Schlyter 2012) ja DNSSEC:in yleistymisen seuraaminen ja niiden korreloiminen RFC 9116:n yleistymisen kanssa saattaisivat avata tulevaisuudessa mielenkiintoisia näköaloja.

LÄHTEET

Berners-Lee, T., Fielding, R. & Masinter, L. 2005. Uniform Resource Identifier (URI): Generic Syntax. RFC 3986. Network Working Group.

Blechschmidt, B. s.a. MassDNS. A high-performance DNS stub resolver. GitHub-tietovarasto. Saatavissa: https://github.com/blechschmidt/massdns [viitattu 20.2.2023]

Bonderud, D. 2017. The Telltale Text File: Security Researcher Proposes Standardization for Reporting Vulnerabilities. Security Intelligence 19.9.2017. Verkkolehti. Saatavissa: https://securityintelligence.com/news/the-telltale-text-file-security-researcher-proposes-standardization-for-reporting-vulnerabilities/ [viitattu 28.9.2022]

Bradner, S. 1996. The Internet Standards Process -- Revision 3. RFC 2026. Network Working Group.

Callas, J., Donnerhacke, L., Finney, H., Shaw, D. & Thayer, R. 2007. OpenPGP Message Format. RFC 4880. Internet Engineering Task Force.

Carroll, J. & Ellis, C. s.a. DNS Security TXT. WWW-dokumentti. Saatavissa: https://dnssecuritytxt.org/ [viitattu 11.2.2023]

Cimpanu, C. 2017. Security.txt Standard Proposed, Similar to Robots.txt. Bleeping Computer 15.9.2017. Verkkolehti. Saatavissa: https://www.bleepingcomputer.com/news/security/security-txt-standard-proposed-similar-to-robots-txt/ [viitattu 27.9.2022]

CISA. 2020. Binding Operational Directive 20-01 - Develop and Publish a Vulnerability Disclosure Policy. Cybersecurity and Infrastructure Security Agency. WWW-dokumentti. Saatavissa: https://www.cisa.gov/binding-operational-directive-20-01 [viitattu 27.9.2022]

Day, J., Kearney, P., Moor, J., Marshall, R., Bott, A. & Poyner, I. 2021. Vulnerability Disclosure: Best Practice Guidelines. Release 2.0, September 2021. IoT Security Foundation. PDF-dokumentti. Saatavissa: https://www.iotsecurityfoundation.org/wp-content/uploads/2021/09/IoTSF-Vulnerability-Disclosure-Best-Practice-Guidelines-Release-2.0.pdf [viitattu 30.9.2022]

Dolan, S. s.a. Jq - Command-line JSON processor. GitHub-tietovarasto. Saatavissa: https://github.com/stedolan/jq [viitattu 30.11.2022]

Fielding, R., Nottingham, M. & Reschke, J. (toim.) 2022. HTTP Semantics. RFC 9110. Internet Engineering Task Force.

Findlay, P. & Abdou, A. R. 2022. Characterizing the Adoption of Security.txt Files and their Applications to Vulnerability Notification. Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb) 2022.

Flanagan, H. (toim.) 2019. Fifty Years of RFCs. RFC 8700. Internet Architecture Board.

Foudil, E. & Shafranovich, Y. 2022. A File Format to Aid in Security Vulnerability Disclosure. RFC 9116. Internet Engineering Task Force.

Foudil, E. 2022. Security.txt adoption in Switzerland. Blogi. Päivitetty 18.1.2022. Saatavissa: https://edoverflow.com/2022/swiss-security-txt/ [viitattu 2.10.2022]

HackerOne. 2017. Hacker Q&A with EdOverflow. Blogi. Päivitetty 28.12.2017. Saatavissa: https://www.hackerone.com/ethical-hacker/hacker-qa-edoverflow [viitattu 30.9.2022]

Handl, R., Pizzo, M. & Biamonte, M. (toim.) 2016. OData JSON Format Version 4.0 Plus Errata 03. OASIS. PDF-dokumentti. Saatavissa: https://docs.oasis-open.org/odata/odata-json-format/v4.0/odata-json-format-v4.0.pdf [viitattu 5.10.2022]

Heikkinen, H. 2010. Toimintatutkimuksen lähtökohdat. Teoksessa Heikkinen, H., Rovio, E. & Syrjälä, L. (toim.) Toiminnasta tietoon. Toimintatutkimuksen menetelmät ja lähestymistavat. 3. korjattu painos. Helsinki: Kansanvalistusseura, 16–38.

Hoffman, P. & Schlyter, J. 2012. The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. RFC 6698. Internet Engineering Task Force.

Holopainen, M. & Pulkkinen, P. 2008. Tilastolliset menetelmät. 5.–6. painos. Helsinki: WSOY.

Hunt, T. 2017. Fixing Data Breaches Part 3: The Ease of Disclosure. Blogi. Päivitetty 20.12.2017. Saatavissa: https://www.troyhunt.com/fixing-data-breaches-part-3-the-ease-of-disclosure/ [viitattu 28.9.2022]

Hunt, T. 2020. There is a Serious Lack of Corporate Responsibility during Breach Disclosures. Blogi. Päivitetty 19.3.2020. Saatavissa: https://www.troyhunt.com/there-is-a-serious-lack-of-corporate-responsibility-during-breach-disclosures/ [viitattu 28.9.2022]

IANA. 2022. Security.txt fields. Internet Assigned Numbers Authority. WWW-dokumentti. Saatavissa: https://www.iana.org/assignments/security-txt-fields/security-txt-fields.xhtml [viitattu 28.9.2022]

Jones, K. B. 2008. Search Engine Optimization: Your visual blueprint for effective Internet marketing (Vol. 22). John Wiley & Sons.

Josefsson, S. 2006. Domain Name System Uniform Resource Identifiers. RFC 4501. Network Working Group.

Kananen, J. 2011. Kvantti. Kvantitatiivisen opinnäytetyön kirjoittamisen käytännön opas. Jyväskylän ammattikorkeakoulun julkaisuja 118. Jyväskylä: Jyväskylän ammattikorkeakoulu.

Klyne, G. & Newman, C. 2002. Date and Time on the Internet: Timestamps. RFC 3339. Network Working Group.

Koch, W. 2022. OpenPGP Web Key Directory. Internet-Draft. Version 15. Network Working Group. Saatavissa: https://datatracker.ietf.org/doc/draft-koch-openpgp-webkey-service/15/ [viitattu 11.2.2023]

Kohonen, I., Kuula-Luumi, A. & Spoof, S. K. (toim.) 2019. Ihmiseen kohdistuvan tutkimuksen eettiset periaatteet ja ihmistieteiden eettinen ennakkoarviointi Suomessa. Tutkimuseettisen neuvottelukunnan julkaisuja 3/2019. Helsinki: Tutkimuseettinen neuvottelukunta.

Kontinen, V. 2020. Preventing email forgery in Finland. Research on the current SPF and DMARC implementations. JAMK University of Applied Sciences. School of Technology. Master's thesis. PDF-dokumentti. Saatavissa: https://urn.fi/URN:NBN:fi:amk-2020120826904 [viitattu 8.3.2023]

Krebs, B. 2021. Does your organization have a security.txt file? Blogi. Päivitetty 20.9.2021. Saatavissa: https://krebsonsecurity.com/2021/09/does-your-organization-have-a-security-txt-file/ [viitattu 25.2.2023]

Kuldell, H. 2019. The agency wants feedback on how the government should accept unsolicited bug reports. Nextgov 18.12.2019. Verkkolehti. Saatavissa: https://www.nextgov.com/cybersecurity/2019/12/cisa-still-wants-your-thoughts-its-vulnerability-disclosure-policy/161989/ [viitattu 28.9.2022]

Laki Liikenne- ja viestintävirastosta 23.11.2018/935.

Marsan, C. D. 2005. Yet another foolish network protocol. Network World 28.3.2005. Verkkolehti. Saatavissa: https://www.networkworld.com/article/2319638/yet-another-foolish-network-protocol.html [viitattu 10.3.2023]

McQuistin, S., Karan, M., Khare, P., Perkins, C., Tyson, G., Purver, M., Healey,P., Iqbal, W., Qadir, J. & Castro, I. 2021. Characterising the IETF Through the Lens of RFC Deployment. ACM Internet Measurement Conference (IMC), 2.-4.11.2021, 137–149.

Moriarty, K. & Farrell, S. 2021. Deprecating TLS 1.0 and TLS 1.1. RFC 8996. Internet Engineering Task Force.

NCSC UK. 2020. Vulnerability Disclosure Toolkit. National Cyber Security Centre. PDF-dokumentti. Saatavissa: https://www.ncsc.gov.uk/files/NCSC_Vulnerability_Toolkit.pdf [viitattu 30.9.2022]

Nottingham, M. 2019. Well-Known Uniform Resource Identifiers (URIs). RFC 8615. Internet Engineering Task Force.

Nummenmaa, T., Konttinen, R., Kuusinen, J. & Leskinen, E. 1997. Tutkimusaineiston analyysi. 1. painos. Porvoo: WSOY.

Phillips, A. & Davis, M. (toim.) 2009. Tags for Identifying Languages. RFC 5646. Network Working Group.

Pochat, V. L., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M. & Joosen, W. 2019. TRANCO: A research-oriented top sites ranking hardened against manipulation. Network and Distributed Systems Security (NDSS) Symposium 2019.

Poteat, T. & Li, F. 2021. Who you gonna call? An empirical evaluation of website security.txt deployment. Proceedings of the 21st ACM Internet Measurement Conference, 526–532.

ProjectDiscovery Inc. s.a. Nuclei FAQ. WWW-dokumentti. Saatavissa: https://nuclei.projectdiscovery.io/faq/nuclei/ [viitattu 2.10.2022]

Rashid, F. Y. 2020. CISA seeks comments on how government should handle vulnerability reports. Decipher 3.1.2020. Duo Security. Verkkolehti. Saatavissa: https://duo.com/decipher/cisa-seeks-comments-on-how-government-should-handle-vulnerability-reports [viitattu 28.9.2022]

Rescorla, R. 2018. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. Internet Engineering Task Force.

Resnick, P. (toim.) 2008. Internet Message Format. RFC 5322. Network Working Group.

Rousku, K. (toim.) 2017. Ohje riskienhallintaan. Valtiovarainministeriön julkaisuja 22/2017. Helsinki: Valtiovarainministeriö. PDF-dokumentti. Saatavissa: https://julkaisut.valtioneuvosto.fi/handle/10024/80013 [viitattu 4.3.2023]

Saint-Andre, P. & Hodges, J. 2011. Representation and verification of domain-based application service identity within internet public key infrastructure using X. 509 (PKIX) certificates in the context of transport layer security (TLS). RFC 6125. Internet Engineering Task Force.

Salz, R. 2022. Entities involved in the IETF standards process. RFC 9281. Internet Engineering Task Force.

SFS-EN 45020. 2007. Standardisointi ja siihen liittyvä toiminta. Yleissanasto.

The ZMap Project s.a. ZGrab 2.0 - Fast Go Application Scanner. GitHub-tietovarasto. Saatavissa: https://github.com/zmap/zgrab2 [viitattu 5.10.2022]

Tilastokeskus s.a. Tutkimus- ja kehittämistoiminta. Määritelmä 2. Saatavissa: https://www.stat.fi/meta/kas/t_ktoiminta.html [viitattu 3.10.2022]

Traficom. 2020. Haavoittuvuudet - miten niistä ilmoitetaan oikein. Tietoturva nyt! Kyberturvallisuuskeskus. WWW-dokumentti. Saatavissa: https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/haavoittuvuudet-miten-niista-ilmoitetaan-oikein [viitattu 27.9.2022]

Traficom. 2021a. Kyberturvallisuuskeskuksen haavoittuvuuskoordinaatio pähkinänkuoressa. Tietoturva nyt! Kyberturvallisuuskeskus. WWW-dokumentti. Saatavissa: https://www.kyberturvallisuuskeskus.fi/fi/ajankohtaista/kyberturvallisuuskeskuksen-haavoittuvuuskoordinaatio-pahkinankuoressa [viitattu 3.10.2022]

Traficom. 2021b. Verkkotunnusvälittäjän opas. Ohjeita ja vaatimuksia verkkotunnusvälittäjäksi aikovalle. Versio 1.2. PDF-dokumentti. Saatavissa: https://www.traficom.fi/sites/default/files/media/file/Verkkotunnusvalittajan-opas.pdf [viitattu 23.9.2022]

Töttö, P. 2012. Paljonko on paljon? Luvuilla argumentoinnista empiirisessä tutkimuksessa. Tampere: Vastapaino.

Ulrich, A., Holz, R., Hauck, P. & Carle, G. 2011. Investigating the OpenPGP Web of Trust. European Symposium on Research in Computer Security, 489–507.

Viestintävirasto. 2017. Fi-verkkotunnukset. OData-palvelukuvaus. Versio 2.1. 15.12.2017. PDF-dokumentti.

Warburton, D. 2021. The 2021 TLS telemetry report. F5 Labs. WWW-dokumentti. Saatavissa: https://www.f5.com/labs/articles/threat-intelligence/the-2021-tls-telemetry-report [viitattu 11.2.2023]

Wouters, P. 2016. DNS-based Authentication of Named Entities (DANE) bindings for OpenPGP. RFC 7929. Internet Engineering Task Force.

KUVALUETTELO

TAULUKKOLUETTELO

ZGRAB 2.0 -TYÖKALUN TUOTTAMA JSON-RAKENNE

Esimerkki ZGrab 2.0:n tuottamasta JSON-rakenteesta onnistuneessa HTTPS-yhteydessä pelkistettynä YAML-puuksi. Yksinkertaiset listat on korvattu "[]"-merkinnällä, kokonaislukuarvot N-kirjaimella sekä totuusarvot B-kirjaimella, ja sisennyksen hahmottamista on helpotettu pisteillä. Muut arvot ovat merkkijonoja. Rakenne pelkistyy, mikäli ominaisuuksia ei ole käytössä, ja esimerkiksi "headers" voi sisältää useampia avaimia riippuen vastauksen otsakkeista.

Esimerkiksi kohdan "response" alta löytyvään "body"-avaimeen voidaan viitata tekstissä ".data.http.result.response.body" ja kohdan "headers" alta löytyvään "content_type"-listaan ".data.http.result.response.headers.content_type[]". Lista "…server_certificates.chain[]" sisältää kättelyn varmenneketjun kaikki varmenteet samanlaisena rakenteena kuin "…server_certificates.certificate" sisältää palvelimen varmenteen, joten se on poistettu tarpeettomana toistona. Tässä mainitut kohdat on korostettu lihavoimalla.

ip:
domain:
data:
· http:
· · status:
· · protocol:
· · result:
· · · response:
· · · · status_line:
· · · · status_code: N
· · · · protocol:
· · · · · name:
· · · · · major: N
· · · · · minor: N
· · · · headers:
· · · · · accept_ranges:[]
· · · · · connection:·[]
· · · · · **content_type**:·[]
· · · · · date:·[]
· · · · · etag:·[]
· · · · · last_modified:·[]
· · · · · server:·[]
· · · · · strict_transport_security:·[]
· · · · · upgrade:·[]
· · · · · vary:·[]
· · · · · x_content_type_options:·[]
· · · · · x_frame_options:·[]
· · · · · x_xss_protection:·[]
· · · · **body**:
· · · · body_sha256:
· · · · content_length: N
· · · · request:
· · · · · url:
· · · · · · scheme:
· · · · · · host:
· · · · · · path:
· · · · · method:
· · · · · headers:
· · · · · · accept:·[]
· · · · · · user_agent:·[]
· · · · · host:
· · · · · tls_log:
· · · · · · handshake_log:
· · · · · · · server_hello:
· · · · · · · · version:
· · · · · · · · · name:
· · · · · · · · · value: N
· · · · · · · · random:
· · · · · · · · session_id:
· · · · · · · · cipher_suite:
· · · · · · · · · hex:
· · · · · · · · · name:
· · · · · · · · · value: N
· · · · · · · · compression_method: N
· · · · · · · · ocsp_stapling: B
· · · · · · · · ticket: B
· · · · · · · · secure_renegotiation: B
· · · · · · · · heartbeat: B
· · · · · · · · extended_master_secret: B
· · · · · · · server_certificates:
· · · · · · · · **certificate**:
· · · · · · · · · raw:
· · · · · · · · · parsed: N
· · · · · · · · · · version:
· · · · · · · · · · serial_number:
· · · · · · · · · · signature_algorithm:
· · · · · · · · · · · name:
· · · · · · · · · · · oid:
· · · · · · · · · · issuer:
· · · · · · · · · · · common_name:[]
· · · · · · · · · · · country:[]
· · · · · · · · · · · organization:·[]
· · · · · · · · · · issuer_dn:
· · · · · · · · · · validity:
· · · · · · · · · · · start:
· · · · · · · · · · · end:
· · · · · · · · · · · length: N
· · · · · · · · · · subject:
· · · · · · · · · · · common_name:[]
· · · · · · · · · · subject_dn:
· · · · · · · · · · subject_key_info:
· · · · · · · · · · · key_algorithm:

· · · · · · · · · · · · name:
· · · · · · · · · · · · rsa_public_key:
· · · · · · · · · · · · · exponent: N
· · · · · · · · · · · · · modulus:
· · · · · · · · · · · · · length: N
· · · · · · · · · · · · fingerprint_sha256:
· · · · · · · · · · · extensions:
· · · · · · · · · · · · key_usage:
· · · · · · · · · · · · · digital_signature: B
· · · · · · · · · · · · · key_encipherment: B
· · · · · · · · · · · · · value: N
· · · · · · · · · · · · basic_constraints:
· · · · · · · · · · · · · is_ca: B
· · · · · · · · · · · · subject_alt_name:
· · · · · · · · · · · · · dns_names:[]
· · · · · · · · · · · · authority_key_id:
· · · · · · · · · · · · subject_key_id:
· · · · · · · · · · · · extended_key_usage:
· · · · · · · · · · · · · server_auth: B
· · · · · · · · · · · · · client_auth: B
· · · · · · · · · · · · certificate_policies:[]
· · · · · · · · · · · · authority_info_access:
· · · · · · · · · · · · · ocsp_urls:[]
· · · · · · · · · · · · · issuer_urls:
· · · · · · · · · · · · signed_certificate_timestamps:
· · · · · · · · · · · · · -·version: N
· · · · · · · · · · · · · · log_id:
· · · · · · · · · · · · · · timestamp: N
· · · · · · · · · · · · · · signature:
· · · · · · · · · · · · signature:
· · · · · · · · · · · · signature_algorithm:
· · · · · · · · · · · · · name:
· · · · · · · · · · · · · oid:
· · · · · · · · · · · · value:
· · · · · · · · · · · · valid:B
· · · · · · · · · · · · self_signed:B
· · · · · · · · · · · fingerprint_md5:
· · · · · · · · · · · fingerprint_sha1:
· · · · · · · · · · · fingerprint_sha256:
· · · · · · · · · · · tbs_noct_fingerprint:
· · · · · · · · · · · spki_subject_fingerprint:
· · · · · · · · · · · tbs_fingerprint:
· · · · · · · · · · · validation_level:
· · · · · · · · · · · names:·[]
· · · · · · · · · · · redacted: B
· · · · · · · · · **chain:[]**
· · · · · · · · validation:
· · · · · · · · · · browser_trusted: B
· · · · · · · · · · matches_domain: B
· · · · · · · · server_key_exchange:
· · · · · · · · · ecdh_params:
· · · · · · · · · · curve_id:
· · · · · · · · · · · name:
· · · · · · · · · · · id: N
· · · · · · · · · · server_public:
· · · · · · · · · · · x:
· · · · · · · · · · · · value:
· · · · · · · · · · · · length: N
· · · · · · · · · · · y:
· · · · · · · · · · · · value:
· · · · · · · · · · · · length: N
· · · · · · · · · digest:
· · · · · · · · · signature:
· · · · · · · · · · raw:
· · · · · · · · · · type:
· · · · · · · · · · valid: B
· · · · · · · · · · signature_and_hash_type:
· · · · · · · · · · · signature_algorithm:
· · · · · · · · · · · hash_algorithm:
· · · · · · · · · · tls_version:
· · · · · · · · · · · name:
· · · · · · · · · · · value: N
· · · · · · · · client_key_exchange:
· · · · · · · · · ecdh_params:
· · · · · · · · · · curve_id:
· · · · · · · · · · · name:
· · · · · · · · · · · id: N
· · · · · · · · · · client_public:
· · · · · · · · · · · x:
· · · · · · · · · · · · value:
· · · · · · · · · · · · length: N
· · · · · · · · · · · y:
· · · · · · · · · · · · value:
· · · · · · · · · · · · length: N
· · · · · · · · · · client_private:
· · · · · · · · · · · value:
· · · · · · · · · · · length: N
· · · · · · · · client_finished:
· · · · · · · · · verify_data:
· · · · · · · · server_finished:
· · · · · · · · · verify_data:
· · · · · · · · key_material:
· · · · · · · · · master_secret:
· · · · · · · · · · value:
· · · · · · · · · · length: B
· · · · · · · · · pre_master_secret:
· · · · · · · · · · value:
· · · · · · · · · · length: B
· · timestamp:

Esimerkki security.txt-tiedostosta

Kyberturvallisuuskeskuksen security.txt-tiedosto 24.1.2023. Vasemmassa reunassa on selkeyden vuoksi rivinumerointi, sillä pitkille riveille 9 ja 12 on lisätty rivinvaihdot, joita tiedostossa itsessään ei ole. Rivit 1–3 ja 19–34 muodostavat OpenPGP-allekirjoituksen. Rivi 4 on kommentti. Rivin 18 tulisi päättyä isoon Z-kirjaimeen pienen sijaan; aikaleima vastaa tässä RFC 9116:n osiota 2.5.5, johon on jäänyt virhe RFC 3339:n osioon 2 ja ISO 8601 -standardiin nähden.

```
01  -----BEGIN PGP SIGNED MESSAGE-----
02  Hash: SHA512
03
04  # What's this file? RFC 9116; https://securitytxt.org/
05
06  Contact: mailto:cert@traficom.fi
07  Contact: tel:+358-29-534-5000
08
09  Encryption: https://www.kyberturvallisuuskeskus.fi/sites/default
    /files/media/file/NCSC-FI_Incident_Response_2023_pub.txt
10  Encryption: openpgp4fpr:68e6ae087d604e1bb7b077b1b08aeb4c0b702656
11
12  Canonical: https://www.kyberturvallisuuskeskus.fi/.well-known/se
    curity.txt
13  Canonical: https://www.cert.fi/.well-known/security.txt
14  Canonical: https://www.ncsc.fi/.well-known/security.txt
15
16  Preferred-Languages: fi, sv, en
17
18  Expires: 2024-01-23T12:28:53z
19  -----BEGIN PGP SIGNATURE-----
20
21  iQIzBAEBCgAdFiEEaOauCH1gThu3sHexsIrrTAtwJlYFAmPPzwkACgkQsIrrTAtw
22  JlaPHA//d1g2OIiFfBFd1af7XmzoqpolsodDPFuMpPcn6pGjLC12pVfPKypb8opv
23  GBAJt5BQgbOtUHGmNpVF2aFvHd7L0pGnpzPTWcp6IS/a527rf2NK/JCr9vRUcNbV
24  4nh6lJg2YSFmy7f5N/DvH3lBaDq7YzYT9nR4+8uxuMax9iLCJFZ1G/eBJILUJ9L9
25  65MWHumy1K2E5pHiqtgDSb2PIUUv9W2SoJJPZF17wVHZilGOhXZCphR2AIGIkax5
26  eRhxF5o+PvWIqm9Cx68PG7f/8k/Bvn++Sh7Idwjj2uymocmdb3VEEyePVME8UeWT
27  4Q07GdKyLG+84f+R/+lYP4N15QGBiTVu70a0JBIsBlf80JT1dX+73DT+Lv+2V0y2
28  d1VXCKH5fShj1EKn+P57olXDXfyHwk+/7DyJ8BrFqK+Dxiab4ratDGQvBSdLK2ps
29  aIl9sI+lBC0DADBbaNnugZMY3oxZVLGBvAFn8eK+hpf9tdp8bsoU6hPZd4kCJ7r5
30  AizAzIvDOPaKRL46nw1+TcWuLJav/Dgj0sCEYAUGKqFLqLmEfY7em5AA5oajikmy
31  a4owHRDW7wawOWMhZdUYRkpNC/Pu/742bYA78FbrnEF+B8k45ZhT3tOy3b+0Ce9s
32  XEDYaCAWUtwDze4x8ZxcXZm86hecOIaqfJPQvDWumfI8cf0Mt9c=
33  =ZhUV
34  -----END PGP SIGNATURE-----
35
```

Tiedostossa on käytetty Unix-tyyppisten järjestelmien rivinvaihtoa (LF / %x0A), ja se päättyy tällaiseen rivinvaihtoon rivin 34 lopussa. Käytetyt rivinvaihdot eivät ilmene suoraan tästä liitteestä. RFC 9116 sallii myös *carriage return & line feed* -rivinvaihdot (CRLF / %x0D %x0A). OpenPGP-allekirjoituksen varmentamiseksi on kuitenkin tiedettävä, kumpaa rivinvaihtoa on käytetty.