jamk.fi

Preventing email forgery in Finland

Research on the current SPF and DMARC implementations

Ville Kontinen

Master's thesis
November 2020
School of Technology
Master's Degree Programme in Information Technology, Cyber Security

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

jamk.fi

Description

Author(s)	Type of publication	Date
Kontinen, Ville	Master's thesis	November 2020
		Language of publication: English
	Number of pages 78	Permission for web publication: x

Title of publication

Preventing email forgery in Finland

Research on the current SPF and DMARC implementations

Degree programme

Master's Degree Programme in Information Technology, Cyber Security

Supervisor(s)

Saharinen Karo, Kokkonen Tero

Assigned by

Finnish Transport and Communications Agency, National Cyber Security Centre

Abstract

The key objectives for Finnish Transport and Communications Agency's National Cyber Security Centre are to form situational awareness regarding the cyber security situation in Finland and to guide organizations towards secure behavior. During fall of 2019 a lack of knowledge regarding the implementation of email forgery preventing technologies arose. A goal of forming situational awareness regarding the usage of these in the whole .fi-zone and in the public sector in Finland. Additionally, draft-level guidelines were needed for a national guideline set.

The study was executed by mapping out the SPF and DMARC implementation rates and options using multiple DNS query sets to the whole .fi-zone. The gained results were sorted to numeric representations based on the criteria defined in the technology specifications. The public sector was divided into four target groups: municipalities, cities, agencies and ministries. The domain names in .fi-zones owned by these organizations were acquired from open sources. The target group data was extracted from the whole .fi-zone based on these domain lists. International guidelines were synthesized for gaining insight into the general situation. Comparable statistical data was gathered from various international sources.

A general consensus was found regarding safe implementation of the mentioned technologies, and this is used as base for a national level guideline. The technologies were found to be in use within the .fi-zone in similar rates when looking at comparable foreign data. The public sector was found to have significantly lower implementation rates when compared to the .fi-zone and the foreign counterparts.

Keywords/tags (subjects)

DNS, SPF, DKIM, DMARC, statistic, national guideline, implementation rate, Finland

Miscellaneous (Confidential information)

jamk.fi

Kuvailulehti

Tekijä(t)	Julkaisun laji	Päivämäärä
Ville Kontinen	Opinnäytetyö, ylempi AMK	marraskuu 2020
		Julkaisun kieli: englanti
	Sivumäärä: 78	Verkkojulkaisulupa myönnetty: x

Työn nimi

Preventing email forgery in Finland

Research on the current SPF and DMARC implementations

Tutkinto-ohjelma

Master's Degree Programme in Information Technology, Cyber Security

Työn ohjaaja(t)

Karo Saharinen, Tero Kokkonen

Toimeksiantaja(t)

Liikenne- ja viestintävirasto, Kyberturvallisuuskeskus

Tiivistelmä

Liikenne- ja viestintäviraston Kyberturvallisuuskeskuksen perustehtäviä ovat muodostaa tilannekuvaa kyberturvallisuudesta Suomessa ja ohjata ja opastaa organisaatioita turvalliseen toimintaan. Syksyllä 2019 havaittiin puute väärennettyjen sähköpostien estämiseen tarkoitettujen teknologioiden käytöstä ja toteutustavoista. Tavoitteeksi asetettiin muodostaa selkeä kuva yleistilanteesta näiden osalta koko .fi-osoiteavaruudessa sekä tarkennetusti julkishallinnossa Suomessa. Tämän lisäksi tavoitteeksi asetettiin alustava suosituslista ilmiötä vastaan taistelemiseksi kansallista ohjetta varten.

Tutkimus toteutettiin kartoittamalla SPF:n ja DMARC:in käytössä olleita asetuksia suorittaen useita nimipalvelutietuekyselykierroksia koko .fi-osoiteavaruuteen. Saatujen kyselyiden luokittelumallina käytettiin kyseisten tekniikoiden avainmuuttujia toimintalogiikan suhteen. Saadut tilastot kerättiin yhteen numeerisessa muodossa näiden muuttujien perusteella luokiteltuina. Julkishallinnon osalta tutkimusryhmät jaettiin neljään osaan: kunnat, kaupungit, virastot ja ministeriöt. Organisaatioiden hallussa olevat .fi-osoitteet selvitettiin avoimista palveluista. Kollektiivisesta datamassasta eroteltiin jokaisen tarkemman tutkimusryhmän tulokset omaksi datakseen. Pohjustavaksi osaksi useista kansainvälisistä julkaisuista kerättiin suositus- ja määräysmateriaalia kyseessä olevien tekniikoiden turvallisesta käytöstä. Tekniikoiden käyttöasteista kerättiin vertailudataa eri lähteiden julkaisemien tilastojen pohjalta.

Kyseisten tekniikoiden osalta löytyi selkeä konsensus kansainvälisistä ohjeista, joka toimii pohjana kansalliselle ohjeelle. Tekniikoiden käyttöasteen .fi-osoiteavaruudessa havaittiin vastaavan keskimäärin vertailukelpoisia otosryhmiä. Julkishallinnon osalta tekniikoiden käyttöaste oli merkittävästi heikompaa kuin .fi-osoiteavaruudessa tai vertailuryhmissä.

Avainsanat (asiasanat)

DNS, SPF, DKIM, DMARC, tilasto, kansallinen suositus, käyttöaste, Suomi

Muut tiedot (salassa pidettävät liitteet)

Contents

Figures

Tables

Acronyms

ACSC	Australian Cyber Security Centre
BEC	Business Email Compromise
BOD	Binding Operational Directive, issued by DHS
CIS	Center for Internet Security
DHS	The Department of Homeland Security (USA)
DKIM	DomainKeys Identified Mail
DMARC	Domain Message Authentication Reporting
DNS	Domain Name System
DNSSEC	Domain Name System Security Extensions
FBI	Federal Bureau of Investigation (USA)
IC3	Internet Crime Complaint Center, subdivision of FBI
IETF	Internet Engineering Task Force
M3AAWG	The Messaging, Malware and Mobile Anti-Abuse Working Group
MTA	Mail transfer agent
NCSC	National Cyber Security Centre
RFC	Request for Comments
SMTP	Simple Mail Transfer Protocol
SPF	Sender Policy Framework
(cc)TLD	(country code) Top Level Domain
TLS	Transport Layer Security

1 Introduction

Email is one of the cornerstones in modern communication. As a technology the standard for Simple Mail Transfer Protocol (SMTP) was originally drafted in August 1982 (Postel 1982). At this era the face of the Internet was significantly different, and the original technology related to mail transfer did not take into account malicious activities performed by hostile operators – the model was based on mutual trust between all communicating parties. As years passed by, it became clear that further development was needed in order to combat email forgery, and technologies such as Sender Policy Framework (SPF), DomainKeys Identified Mail (DKIM) and Domain Message Authentication Reporting (DMARC) were developed. Each of these technologies add upon the existing SMTP specification via records published in domain name services (DNS), and partially relying on cryptography. (IETF 2014 and 2015; Network Working Group 2009).

The National Cyber Security Centre of Finland (NCSC-FI) has a legislative mandate of supporting, guiding and monitoring information security in electronic communications. NCSC-FI also gathers information in the cyber field in Finland and forms a situational awareness in relation to that (Laki Liikenne- ja viestintävirastosta 2018). The largest role of targeted audience are the governmental parties, such as agencies and ministries, and organizations operating in critical fields.

During 2019 NCSC-FI established an internal project where various aspects of network presence of certain parties is mapped out. The found technical implementations would be assessed and best practices and guidelines would be formed and presented to the general audience. As email is probably one of the most common electronical communication methods within in the Finnish society, the technologies related to securing email communications and fighting forgeries were chosen as the primary research target.

Email is currently the de facto communication method between different organizations operating in the Western world. Due to this position email is also used for a significant amount of critical communication regarding business relations and other critical communication. As such, criminals have taken email-based attacks as one of the preferred methods in various types of attacks. Impersonation leading to

financial fraud or credential compromise, malicious document delivery are prime examples of attack types. According to Verizon report detailing cyber-attacks on 2019, email was the delivery type of malware in 94% of the cases where the initial vector was known (Verizon, 2020 p. 13). FBI Internet Crime Complaint Center (IC3) estimates Business Email Compromise related cases accounting to total losses globally of USD26 billion in the time period between June 2016 and July 2019. (FBI, 2019). Specifically to the topic of this work, the researchers at Agari detail in their report that Russian originating criminals target specifically organizations lacking proper DMARC implementation (Hassold, 2020).

In order to combat these phenomena, several sources define proper implementation of domain protection to be one of the basic aspects for building and maintaining efficient cybersecurity defenses. The Department of Homeland Security (DHS) issued Binding Operational Directive 18-01 regarding the implementation of email security measures on 2017 (DHS, 2017). M3AAWG released a statement on June 2020 that all organizations should implement proper SPF, DKIM and DMARC configurations regarding all the domains in their operations, including domains not used for mailing purposes (M3AAWG, 2020). Center for Internet Security (CIS) lists implementation of DMARC and other receiver-side verification methods in their 20 critical controls for any medium to advanced-level organization (CIS, 2019, p. 29).

2 Research framework

2.1 Technical operation

The key technologies related to this study are Sender Policy Framework (SPF), DomainKeys Identified Mail (DKIM) and Domain Message Authentication Reporting (DMARC). These email security protocols are inherently tied to DNS records published by the domain owners. Logically when an email message is sent between two mail transfer agents (MTA), the receiving host performs checks on the SMTP message data and performs DNS queries based on the domain names and subdomains found in the message. This operational logic allows the domain owners to modify and fine-tune their security stance on a domain/subdomain level, and mail

receivers can accordingly tune their mail filtering systems based on the filtering decisions made by the system. The operational logic of SPF is shown in Figure 1.

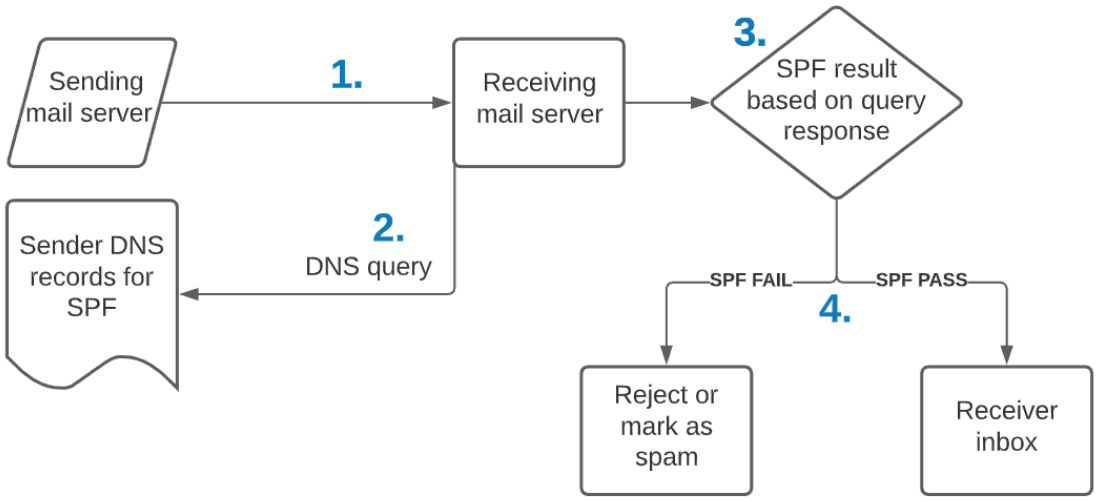

Figure 1. SPF operational logic

The numbers in the figure are logical steps in the process.

1. Mail is sent to the receiving server according to SMTP protocol.
2. The receiving mail server performs DNS checks according to the data given in the SMTP session. DNS queries may be performed to external domains depending on the SPF settings and the given data.
3. SPF evaluation is performed with the data gained from the SMTP session versus the data retrieved via DNS queries
4. Depending on the SPF result and configured mail filtering settings the mail is either rejected, marked as spam or delivered to the receiver address inbox.

DMARC-based filtering adds layers on top of this. The operational logic of DMARC is shown in Figure 2.

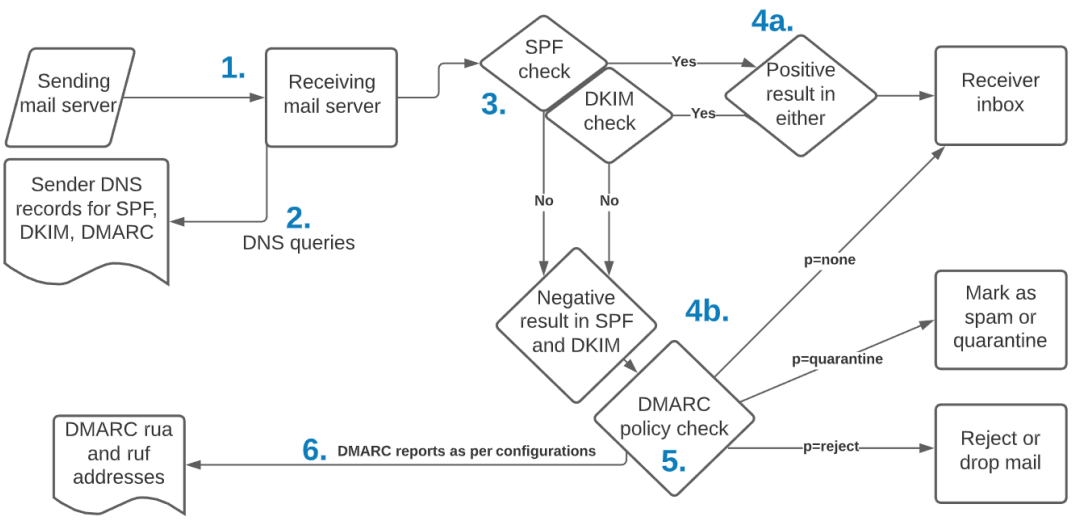

Figure 2. DMARC operational logic

The numeric steps represent logical operations:

1. Mail is sent to the receiving server according to SMTP protocol.
2. The receiving mail server performs DNS checks.
3. SPF policy is evaluated as depicted before, and DKIM signatures are verified against the data found in DNS queries.
4. Filtering decisions are made
    a. If SPF or DKIM checks succeed the mail is delivered to the recipient inbox and the corresponding protocol and DMARC are marked as passed.
    b. If SPF and DKIM checks both result in failure, DMARC policy stance is checked.
5. Depending on the DMARC policy or lack of DMARC in overall the mail is either delivered to the recipient inbox, marked as spam or quarantined or rejected/dropped.
6. Possible aggregated and forensic reports are delivered to the given mail addresses based on the DMARC settings.

2.1.1 Interpreting relevant DNS records

From a researching perspective the usage of public DNS records allows exploring and measuring certain aspects without having access to either end of the mail transfer chain, and the research results project the actual status of the operational systems. This results into having highly accurate data resembling the current state of the technologies. However, the gained data depicts only a snapshot of that particular moment as the domain owners can change the DNS records practically constantly. Therefore, in order to gain a better picture of how what the actual situation in long-

term is, it is advisable to collect and compare the DNS records from a longer timeframe in order to draw better conclusions of the situation in overall.

From a technological standpoint SPF and DMARC are directly measurable by querying a domain name, as their operational logic is identical: a DNS record is published at a specific location defined by the relevant RFC, and the data in the record contains the settings. DKIM poses a problem: a cryptographic signature is added to a message by the outbound MTA, and the corresponding public key for that signature is published in a specific DNS record. The DNS records should be stored in a specific location, which the receiving mail filter queries based on the signature found in the message. Two critical questions arise from this operational logic:

- Which DKIM keys are in use, as there can be multiple in use simultaneously?
- Are all of the used keys in the assumed locations?
- Is DKIM signing in use at all?

Answers to these questions cannot be known without having access to a legitimate mail sent from the examined domain, and therefore from an external perspective only rough estimates can be made on whether DKIM is at use and in what depth.

Because of these uncertainties DKIM implementation rates or models are not measured in this study at all.

2.1.2 About SPF

SPF specifications allow for various settings to be defined quite liberally by the domain administrators (IETF, 2014). SPF records can also include data from other domains or the query can be redirected to other domains via statements of “include:otherdomain.fi” and “redirect=otherdomain.fi”. A simplified but valid SPF record might look like the following:

```
domain.fi IN TXT “v=spf1 +mx ?a a:mail.domain.fi ~ipv4:1.2.3.4
-all”
```

The record above is interpreted as follows:

- The record is published as a TXT record in the domain root
- A valid record begins with declaration “v=spf1” and is case-insensitive
- Specific signs are qualifiers for the data in that field

    - Plus-sign in front of any field indicates a “Pass” for that data – mail matching this should pass
    - Minus-sign indicates a “Fail” – colloquially HardFail – mail matching to this data must be blocked
    - Tilde indicates “SoftFail” – mail matching to this should be blocked
    - Question mark indicates “Neutral” stance – the domain owner has no stance for matches in this category
    - A field without a specific qualifier (in the example “a:www.domain.fi”) defaults to “Pass”
- There cannot be more than ten subqueries for additional DNS queries
    - Assuming that all of the stated records in the example resolve directly without CNAME redirections there are three queries in this SPF record
- If this count is exceeded or if the record does not meet the specified syntax, the SPF evaluation results in a “PermError”
- If there are e.g. DNS errors during the process the SPF evaluation process results in “TempError”

The last parameter “-all” is generally called as explicit HardFail: anything (IP addresses, network ranges, domain and host names, sending servers’ SMTP HELO/EHLO declared hostname) not declared specifically in the SPF record matches to the “all”-field.

2.1.3 About DMARC

DMARC DNS records are published in a TXT record in _dmarc.domain.fi (IETF, 2015). A valid DMARC record looks like the following:

```
_dmarc.domain.fi IN TXT “v=DMARC1; pct=100; p=none;
sp=quarantine; ruf=mailto:forensics@domain.fi;
rua=mailto:aggregated@domain.fi”
```

The fields and their corresponding states are the following:

- Valid declaration of DMARC version 1 is stated in **v=DMARC1**
- **pct=[0-100]** states the percentage of mails which are to be filtered
- **p=none/reject/quarantine** defines the selected policy for the domain
- **sp=none/reject/quarantine** defines the policy for subdomains
- **ruf=mailto:** defines the recipient email address for forensic reports
- **rua=mailto:** defines the destination of aggregated reports

Stating as policy “p=none” is commonly known as DMARC in monitoring mode – the recipient filter is not supposed to act based on the DMARC evaluation results and the

sender domain is often requesting reports from filtering decisions (possibly both forensic and aggregated) in order to adjust their SPF and DKIM settings accordingly.

2.2 The study goals

The goals for this study and reasoning for each of them are elaborated in this sub-chapter. In short, the goals for the study are:

1. What are the usage rates regarding SPF and DMARC implementations in the .fi-zone?
2. What is the situation regarding these technologies in the public sector, divided into four groups (municipalities, cities, agencies, ministries)?
3. Is there a consensus regarding secure implementations of these technologies?
4. Can the observed results in the previous goals be influenced in order to see a statistical difference?

2.2.1 Situational awareness regarding .fi-zone

As it stands there are no current statistics available from the whole .fi-zone in the context of email forgery prevention technologies. Certain statistics regarding these technologies have been collected earlier to some extent, but these have gone stale and a consistent trend of development cannot be gathered from them.

Therefore, the first goal is to form an overall statistic over SPF and DMARC implementation rates and methods in top level domain .fi-zone (referred as the zone from here on). A projected trend is to be established from the gathered statistics.

Individual organizations and researchers have gathered statistics regarding various, very limited parts of the zone. While these may be of interest in certain context, the availability of public overall statistics is clearly needed. Some foreign TLDs publish their total statistics, and ccTLD comparisons can be made with these regarding technology implementation levels.

DKIM is inherently important in the process of deploying DMARC policies to the strictest state. However, as DKIM implementation status and key usage cannot be measured or tested externally with sufficient confidence levels, analyzing DKIM related DNS records is left out of scope for this work. The reasoning behind this decision was explored in depth in chapter 2.1.2.

2.2.2 Implementation rates in the public sector

Due to NCSC-FI's position the public sector is a particularly interesting target regarding any situational awareness or statistics. The second goal for this work is to gather situational awareness regarding the implementation of the defined protocols in the public sector. The public sector in general is divided into four parts in this study:

- Ministries
- Governmental agencies
- Cities
- Municipalities

The group division is self-explanatory for ministries and agencies. During the time of the study there were 12 ministries (Laki valtioneuvostosta, 2003) and a total of 59 governmental agencies (Valtiokonttori, 2019).

Finland is divided into 310 municipalities as of 2020, with certain municipalities merging together on January 1st 2020. Each of the municipalities can declare themselves as cities, and such a declaration has been made by 107 cities. That leaves the total amount of municipalities to 203 (Kuntaliitto, 2019), but on the survey the number of municipalities from 2019 was used – leaving the total count for individual municipalities to 205. Generally speaking, the main difference in between cities and municipalities is the number of inhabitants within the land area of said entity, but there are variations to this as there are larger municipalities in both senses than certain cities are. The decision on dividing cities and municipalities as their own interest groups was made by NCSC-FI.

The complete lists of the inspected target groups and their corresponding business IDs are listed in Appendix 1.

2.2.3 Best practice guidelines

A qualitative data gathering and analysis is performed by the author of the thesis from available reference material regarding current best practices for implementing SPF, DKIM and DMARC policies. This material is gathered from other national level

guidelines along with available commercial or non-profit material, and a summary of these is then distilled into most recommended actions.

NCSC-FI will use this resulting material in drafting a national level voluntary guideline, which is to be published and distributed freely. The process of making this document and publishing is out of scope for this study, as it is a purely internal task for the NCSC-FI. Deriving from this, the third goal is to gain an in-depth understanding on the current situation of guidelines in other nations and generate a synthesized shortlist of practices listed in them.

2.2.4 Projected development versus influence results

As stated before, one of the responsibilities for NCSC-FI is to guide and monitor the status of information security in Finland. NCSC-FI publishes both voluntary and mandatory guides and directives on selected areas and topics in order to improve the cyber security capabilities in overall in Finland. Mandatory directives are targeted on specific entities defined in the legislation (such as internet service providers and other telecommunication providers). Voluntary guides are aimed at the general public: any organization can use them in their own operations as they see fit, and there is no binding legislation or sanctions related to enforcing the guidelines.

This leads to a theoretical research question: what is the measurable impact of these guides? Many of the guides are touching areas very hard to measure by external means (such as distributed denial of service attack guideline), but as email security technologies rely heavily on domain name records, one can gather statistics and monitor what impact is achieved, if any. The presence and structure of said records is an objective definition of the current status of these technologies, as they are used directly by operational systems as-is.

The fourth goal of the study is to assess the overall impact of the guideline stated in the previous sub-chapter and any communication and promotion related to it against projected development trends gathered from the .fi-zone and the interest groups.

2.3 Research methods

The study is an action research in practice, consisting of three steps. The logical process for the whole study is shown in Figure 3.

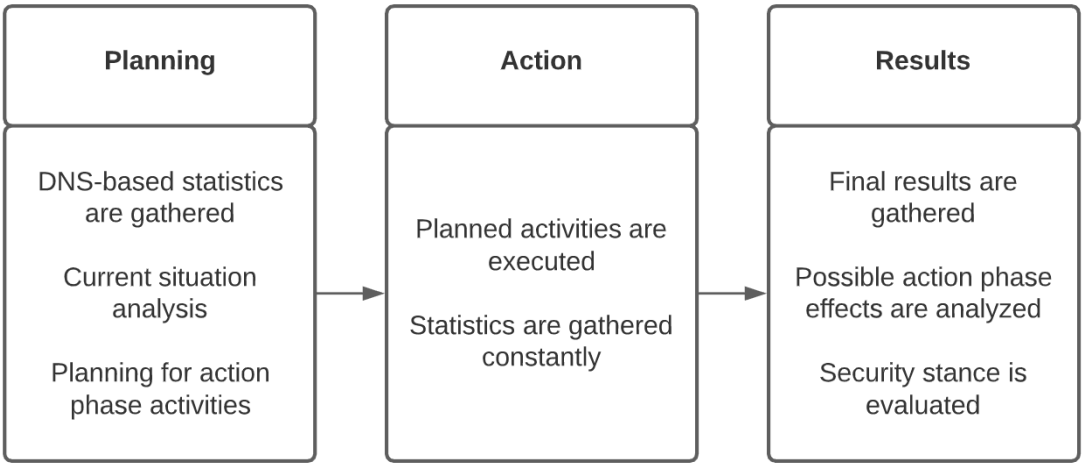

Figure 3. Logical research phases

The logical phases follow closely the model represented for a typical action research presented by Willis and Edwards (Willis, J. W., Edwards, C. L.,2014, p. 59). The initial reflection phase is missing, as it was already performed within NCSC-FI before this study had officially begun. There is one difference in comparison to the theoretical model: there is no flowback from latter phases to the prior, as the study follows a single iteration loop.

In the planning phase a qualitative statistical analysis is performed on data collected from DNS queries performed to the .fi zone, forming an overview of the implemented technologies and their configuration status. The process of data gathering and filing criteria are described in chapter 2.4. After this a plan of action is devised and that is implemented in the action phase in an attempt to affect the measured situation. Data points are gathered throughout the action phase. In the results phase the gathered statistics are analyzed for both possible impacts from the action phase and for forming a security awareness regarding the final measured situation regarding secure SPF and DMARC utilization.

Going into further detail the target research group subset results for the public sector are extracted from the results gained from the whole .fi-zone in order to form their own specific statistics. The process of forming the target groups is described in

chapter 2.5. The formed statistics from both the complete zone and the target subsets are then analyzed and compared to available foreign statistics where comparable data is present – ccTLD zones to the .fi-zone and public sector data against the fitting research group data. The subsets are also compared to each other.

The gathered data on itself may be considered as quantitative research, but keeping in mind that the main research questions relate to situational awareness, the data must be analyzed using qualitative methods: what can be considered secure, and does the situation in the measured data match that? Because of this the study in overall is best described as a qualitative action research.

2.4 DNS statistic gathering methods

NCSC-FI operates as it's on division in the Finnish Transport and Communications Agency (Traficom). Traficom has a separate unit responsible for providing and maintaining the .fi zone, and these responsibilities do not fall under the jurisdiction defined for NCSC-FI. However, as the NCSC-FI has the duty of monitoring and guiding all Finnish entities regarding cyber security and the domain authority has interest in upkeeping the security stance for the .fi in general at a high level, there are mutual interests present. Because of this, NCSC-FI has access to the .fi zone as a whole on the level of registered and operational domain names and their authoritative domain name servers as a daily dump given by the domain authority. NCSC-FI has no access to domain owner data besides public WHOIS-record data.

In this work the zone data was used in the following logic:

- Individual domains from the daily zone were singled out along with their authoritative name servers
- The authoritative name servers were queried for specified DNS records
- If the records pointed out references to external zones (such as _dmarc.domain.fi IN CNAME _dmarc.domainb.fi) recursive resolving was used
- The received responses were analyzed for specific keyword responses
- The full DNS record responses were not stored, but only their interpreted value
- The gained data was then formed as a CSV file
- The CSV was analyzed by gathering all possible value combinations and their corresponding numeric counts along with the individual value counts
- If a nameserver did not respond to a query due to any reason, the query was repeated a total of three times with timeout set to five seconds

- If there was no response or the response did not match to the expected values, the evaluated value was set to “false”

The DNS responses and their corresponding assignment values are presented in Table 1.

Table 1. DNS responses and their assigned values

Assigned value	DNS response
mx-true	IN MX response (not checking response syntax)
mx-false	No response to IN MX
spf-false	No SPF-syntax valid response to IN TXT domain.fi
spf-hardfail	IN TXT response containing "v=spf1" and "-all"
spf-softfail	IN TXT response containing "v=spf1" and "~all"
spf-nostance	IN TXT response containing "v=spf1" and none of the above
dmarc-false	No response to IN ANY _dmarc.domain.fi
dmarc-test	IN ANY response to _dmarc.domain.fi containing "v=DMARC1" and "p=none"
dmarc-quarantine	IN ANY response to _dmarc.domain.fi containing "v=DMARC1" and "p=quarantine"
dmarc-reject	IN ANY response to _dmarc.domain.fi containing "v=DMARC1" and "p=reject"
dmarc-unknown	DMARC syntax invalid response to IN ANY _dmarc.domain.fi

A couple of examples for possible response lines stored for analysis look like the following:

```
domaina.fi,mx-true,spf-false,dmarc-false
domainb.fi,mx-false,spf-nostance,dmarc-test
domainc.fi,mx-true,spf-hardfail,dmarc-quarantine
```

This statistic gathering process was repeated throughout the study period, beginning on week 36 (Sep 9th-15th) 2019 and ending on week 36 (Aug 31st-Sep 6th) 2020, spanning approximately one year. The complete results of the gathered statistics are shown in Appendix 2.

MX record presence was collected also, even though it does not directly implicate whether a domain has operational mail services or not, as the RFC for mailing allows the usage of A record as a fallback whenever MX is not found. However, the lack of a valid MX record is often considered to be a criterion in labeling incoming mail as

possible spam, which is relevant to the overall topic of this work of combatting forgeries and spam.

2.5 Establishing the interest groups

As NCSC-FI does not have access to domain owner data on the .fi-zone dump, other methods had to be devised in order to establish the interest groups for this study (ministries, agencies, cities and municipalities). Each legal entity in Finland possesses at least one business ID code (Y-tunnus), but there are scenarios where one logical entity has multiple, such as a large corporation with sub-branches with each having their own. Public sector entities such as the ministries and agencies only have one, and these are available from the State Treasure.

There seems to be no public listing for each of the municipalities and their corresponding business ID codes. This was circumvented by using the current listing of municipalities in Finland, replacing ä's and ö's with a-o and adding a .fi for mocking up an expected domain for that municipality. This domain was then WHOIS-queried from the public service, and the owner business ID code was extracted and manually verified to correspond to that exact municipality. This worked for approximately 90% of the municipalities in overall, with manual seeking needed only for the smallest Swedish-speaking municipalities. As a result of this full lists containing names and business ID codes for all of the target groups were acquired.

Domain ownership for organizations in the .fi-zone is bound to their corresponding business ID codes. The domain authority does provide a public OpenData source which can be queried for domains based on given seek parameters. One of these seek parameters is the business ID code. Using this OpenData source and the gathered business ID code list a combined list for each of the domains owned by that entity was gained as target material. The individual domain ownership results were then joined together according to the study group division.

2.6 Study ethics

The study consists primarily of DNS queries performed to technical systems. The queries are performed over the Internet to the name servers designated by the

domain owners. No human interaction is performed in the process nor any people are interviewed or examined in the process. The DNS query set sent to each of the domains corresponds to the same queries an email filter would send out when it is inspecting an incoming message. All of the polled data is public and acquirable by practically anyone with access to the Internet.

The complete operational domain listing for the .fi-zone is used in the query source material. Theoretically this zone data could be of value to someone performing open source intelligence against certain parties, but it has to be kept in mind that the same data can be acquired from the mentioned OpenData sources by polling individual business IDs.

Considering that the study results will very likely reveal technical security stances within the defined study groups, a malicious actor could use this information in planning hostile operations. However, as stated the polled information is public and readily available, so it can be assumed that this data has already been examined to certain extent by such possible actors.

In overall, the study represents a current situation regarding the technologies and their implementation models and does not pose any problematic questions regarding ethicality in the process or the gained results.

3 Comparison points

3.1 Overview of policies developed abroad

In this subchapter certain selected policies and guidelines are briefly iterated and finally collected for comparison. The different sources were selected by the following criteria:

- Source of the guideline
    - Governmental authority or
    - Generally acknowledged and known party
- Applicability of the policy target group in relation to the research goals (public sector preference)
- Guideline timeliness – is the guideline published recently, within three years?

Australian Cyber Security Centre (ACSC) has published an instructional set regarding the implementation of SPF, DKIM and DMARC (ACSC, 2020a). The set goes in-depth in technicality with examples regarding different use cases and possible caveats when deploying these technologies in both self-owned domains and in mail filtering cases. In addition to this, ACSC has published a guideline regarding these technologies (ACSC, 2020b). In the guideline the following recommendations are made:

- SPF related DNS records are to be published for all domains
- A HardFail SPF record is to be used with email sending servers
- DMARC is to be used for all domains in “p=reject” mode for SPF or DKIM failures

Center for Cybersikkerhed in Denmark (CFCS-DK) has issued a guideline for fighting forged emails (CFCS, 2017). Sikkerdigital.dk is a joint operation between central security organizations in Denmark, and they have published a mandatory baseline requirement set targeted for state authorities regarding the implementation of several security hardening sets (Sikkerdigital, 2020). The DMARC recommendation is simply to enforce reject policy for all domains owned by a governmental entity.

The Department of Homeland Security issued Binding Operational Directive 18-01: Enhance Email and Web Security for governmental bodies of the United States of America (DHS, 2017). In this BOD the DHS defined a timeline with different stages of required implementation regarding various aspects of email security. The goals were the following:

- After 15 days of establishing a reporting location, DHS is to be included in DMARC aggregate reports
- After 90 days of issuing all agency domains are to have valid SPF and DMARC records, with DMARC policy set to “p=none” and an address defined for receiving reports
- After a year of issuing a DMARC policy is to be set for all domains and mail hosts with policy status “p=reject”

M3AAWG has published a “best practices” guide regarding parked domains. The guide has been updated on December 2015 (M3AAWG, 2015). Parked domain in this context are domains registered and operated by an organization, but not intended for any email traffic. M3AAWG recommends the following for parked domains:

- An explicit hard fail in SPF should be published for both the main domain and possible subdomains
- DMARC should be used in “p=reject” policy with reporting used at least for aggregate reports
- A null MX record is to be used to indicate that there are no operational mail servers for the domain or subdomains

National Cyber Security Centre UK (NCSC-UK in this document) has published an extensive guideline set regarding email security. The set in total goes beyond the scope of this work, but the key aspects regarding SPF, DKIM and DMARC are the following (NCSC-UK 2019a and 2019b):

- Implementing a monitoring state DMARC policy for all domains
- Creating and refining a SPF record, starting out with SoftFail
- Adopting DKIM into outbound mail usage
- Changing DMARC policy to reject once the organization is ready for it

National Institute of Standards and Technology (NIST) has published an in-depth document related to email security regarding both the in and outbound mail (NIST, 2019). The document recommends various key aspects regarding multiple email security related technologies, and the SPF related recommendations are summarized below (NIST, 2019, chapter 4.8). DMARC-related recommendation is specified in chapter 4.6.3 in the same document.

- All domains should have a SPF record stated
- Domains not designated for outbound mail traffic should have SPF records stating no valid email sending servers exist
- If SPF and/or DKIM is in usage, DMARC usage is recommended and the policy should be defined according to the senders’ own choice

The key takeaways from the recommendations and guidelines explored above are summarized in Table 2.

Table 2. SPF and DMARC recommendation summary

Source	SPF policy, parked domains	SPF policy, mailing domains	DMARC policy
ACSC	HardFail	HardFail	reject
CFCS	in use, not defined	in use, not defined	reject
DHS	in use, not defined	in use, not defined	reject
M3AAWG	HardFail	not defined	reject
NCSC-UK	SoftFail	SoftFail	at least none
NIST	HardFail	in use, not defined	at least none

As average it can be said that SPF should be in use with preferably HardFail set for parked domains, and a suitable setting level defined for mailing domains, preferring to explicitly state a selected Fail-option. DMARC must be in use and the target is to have policy set to reject. Some of the guidelines mention a developing path leading to these options, so implementing the strictest settings from day 1 is not necessary.

3.2 Statistics of various domains

Various sources publish statistics related to email security technologies and their adaptation rates. Some of the available statistics are explored in this chapter. It has to be noted that statistics are not directly comparable in numeric count, as there are significant differences in sample sizes between surveys. Some surveys analyze only the primary mail sending domains, while others attempt to analyze a larger sample size of the designated target group. When considering the accuracy of an analysis one has to also bear in mind that an organization might own domains through 3rd party registrations (e.g. marketing company owns certain domains related to an individual campaign), which then cannot be tied to that particular organization at all and therefore are left out of scope. Additionally, there are not that many publicly available zone-wide statistics for a complete top-level domain.

The sources included in this work are selected based on the following criteria:

- Statistic source correlation to the guidelines used before
- Statistic source credibility
    - Publishing party in relation to the surveyed domains
    - Surveyed size
    - Organization sector and focus
- Statistic source timeliness

3.2.1 Statistics from the Commonwealth

ACSC has published a report regarding the cyber security posture of the Commonwealth in 2019 (ACSC, 2020c). The report states that statistics are gathered from a total of 18 000 Australian Government domain. In the report SPF and DMARC implementation rates in the subdomains of .gov.au are shown in Table 3.

Table 3. SPF and DMARC statistics for .gov.au

Setting	Percentage
SPF protected	40,5 %
SPF unprotected	59,5 %
DMARC protected	55,5 %
DMARC unprotected	44,5 %

The report does not iterate what the technical qualifiers and policies are for these settings. Deriving from the recommendations made by the ACSC, it can be assumed that SPF “all” qualifier must be set in either SoftFail or HardFail state. Following the same logic DMARC should be set to either “p=quarantine” or “p=reject” state, as “p=none” does not enforce mail filtering.

3.2.2 Statistics for country code top-level domain .nl

SIDN is the administrative organ for the ccTLD .nl, and their laboratory section (SIDN Labs) provide extensive public statistics regarding various DNS-measurable technologies, including email (SIDN Labs, 2020a and 2020b). The statistics are updated approximately once or twice per month and cover the whole of .nl zone. MX record presence statistics are shown in Figure 4.

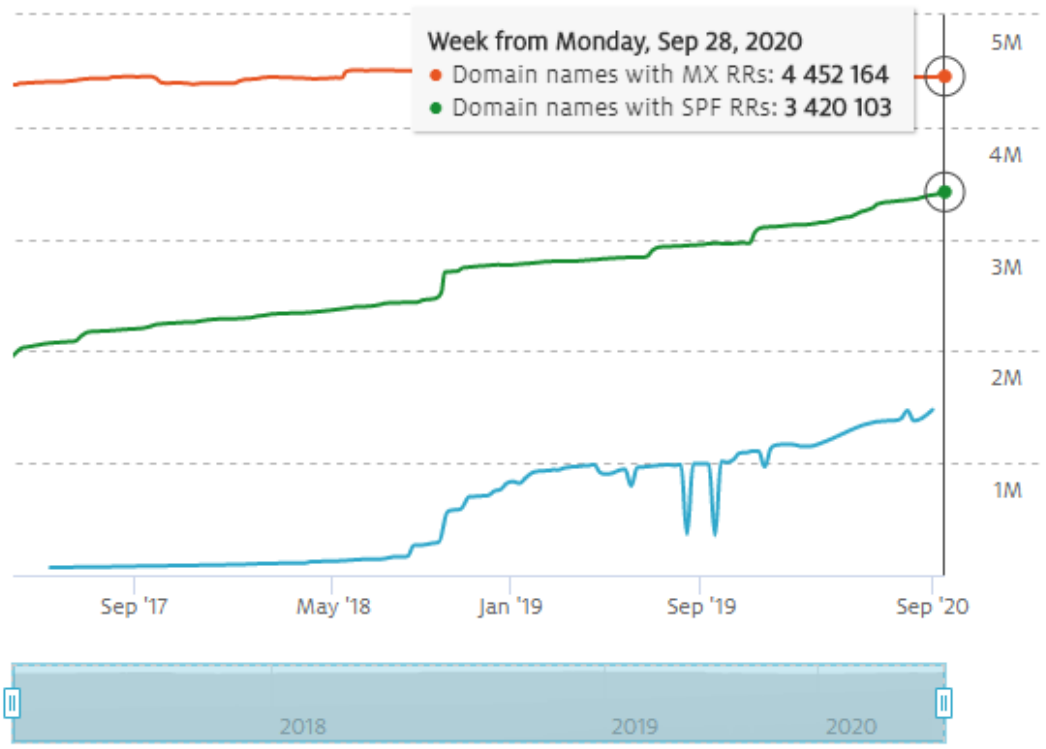

Figure 4. .nl MX record presence (as of Sep 28, 2020)

Corresponding SPF qualifier statistics are shown in Figure 5.

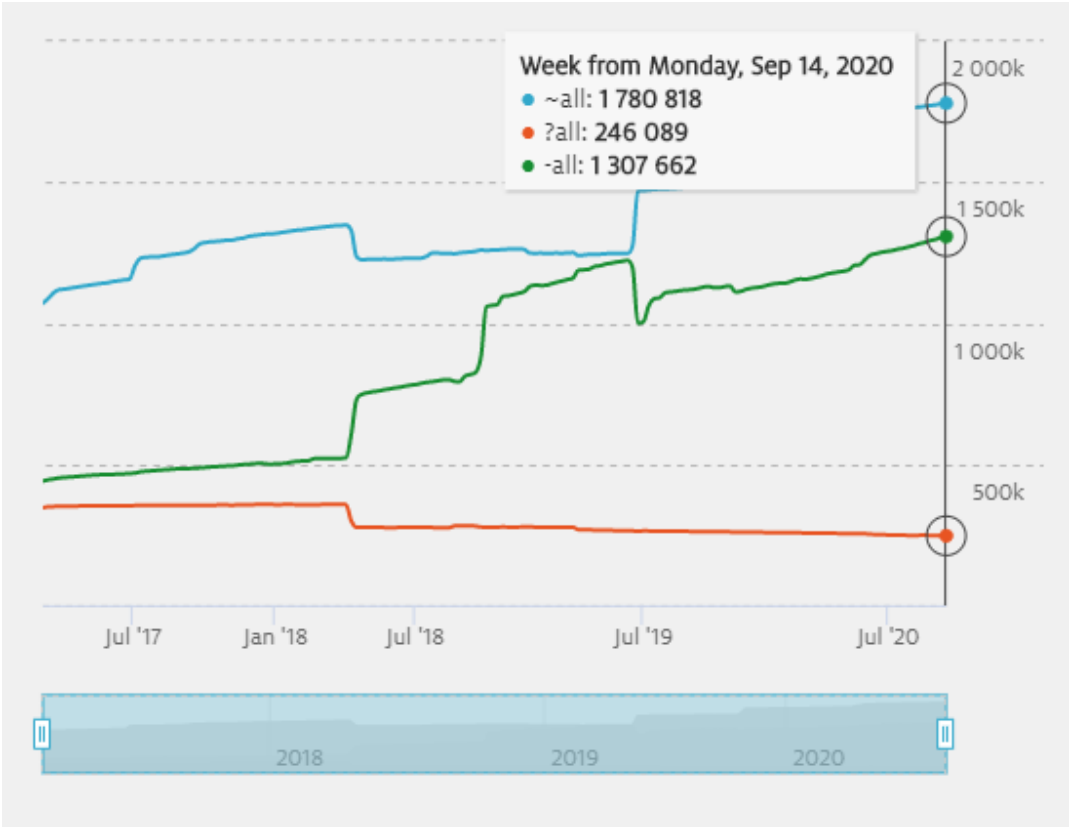

Figure 5. .nl SPF qualifiers statistics graph (as of Sep 14, 2020)

In the data there are a couple of significant changes occurring in the past. The first significant change has occurred approximately in the beginning of April 2018: in that step approximately 200 000 domains switched from using SPF qualifier Neutral and SoftFail to HardFail. The next major change occurred around late September 2018, when approximately 150 000 domains started using the HardFail qualifier. In the beginning of July 2019 about 198 000 domains switched their qualifier from HardFail to SoftFail. Corresponding DMARC policy statistics are shown in Figure 6.

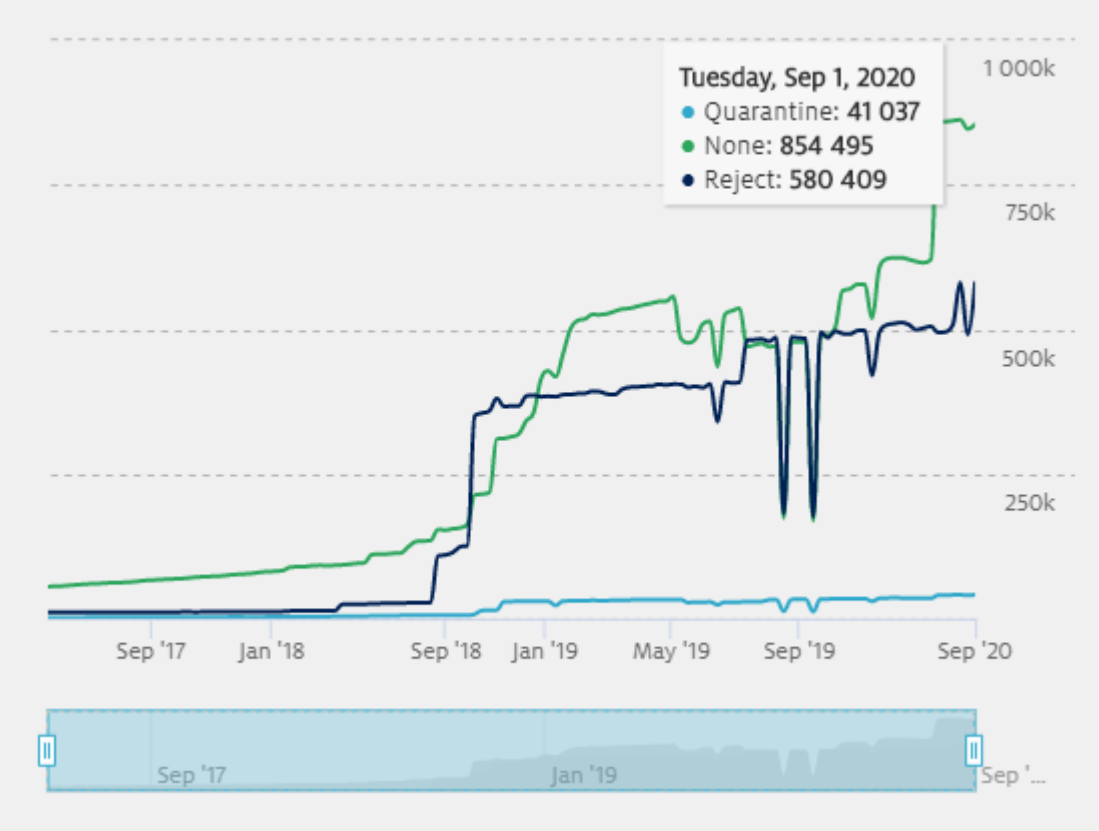

Figure 6. .nl DMARC policy statistics graph (as of Sep 1, 2020)

It has to be noted that the stepping interval in the statistics graph is not linear: the last major increase in DMARC none-policy implementation has occurred between January and July 2020. Considering the cumulative increase, the change has still been significant in that time period. Two of the major changes regarding SPF and DMARC implementations have occurred approximately at the same time: early September 2018. SIDN provides a Registrar Scorecard ranking to the domain name registrars as a dashboard view, in which they detail how that particular registrar performs in comparison to the other registrars. The logic for this is to boost technological capabilities and adaptation rates in overall. On June 2018, SIDN added the SPF and

DMARC implementation statistics to the scorecard system (SIDN, 2018). It can be assumed that the increases in the technology adaptation rates are connected to this.

To calculate the usage percentages in the whole .nl zone, the amount of operational domain names must be known. SIDN Labs provides this data also (SIDN Labs, 2020c), and the according numbers are shown in Figure 7.

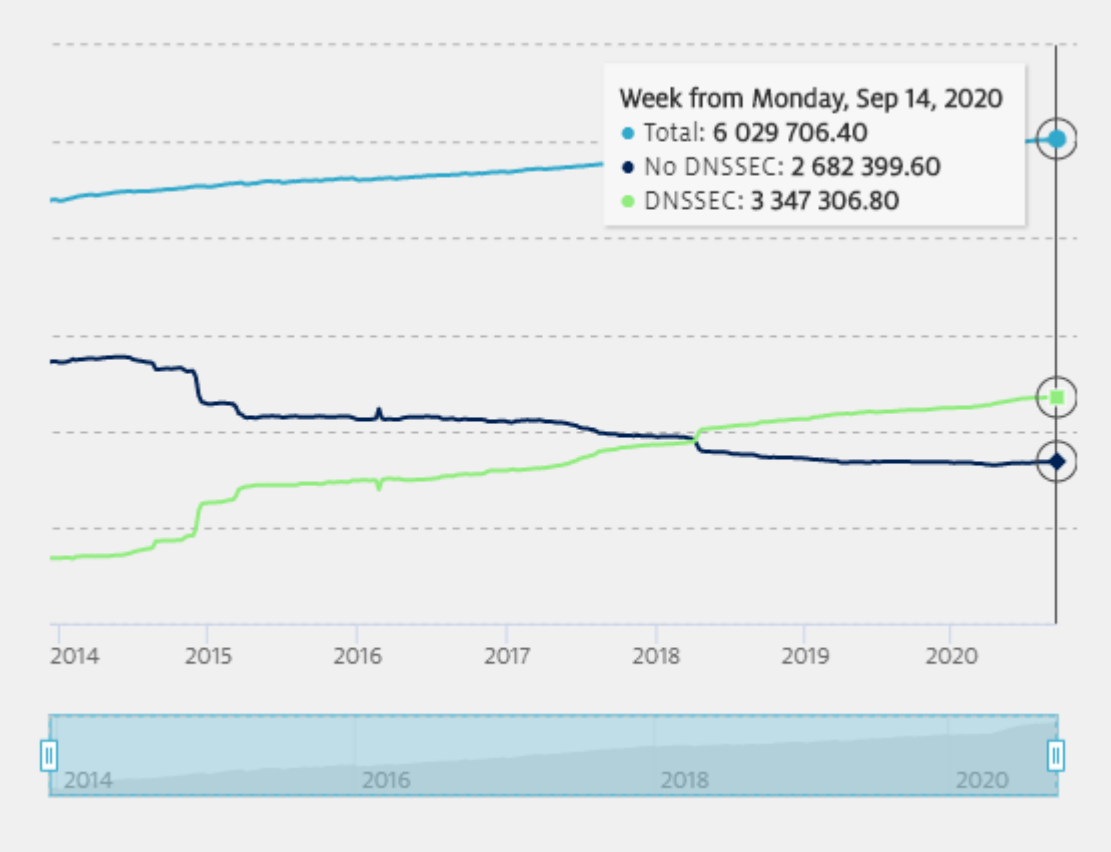

Figure 7. .nl registered domains statistics graph (as of Sep 14, 2020)

It is unclear why there are portions of domains registered, and the number is rounded to the closest integer. From these the total usage rates for the technologies can be calculated. SPF statistics derived from this are presented in Table 4.

Table 4. MX and SPF statistics for .nl

Setting	Count	Percentage
MX record present	4 452 164	73,75 %
SPF not defined	2 695 137	44,70 %
SPF Neutral	246 089	4,08 %
SPF SoftFail	1 780 818	29,53 %
SPF HardFail	1 307 662	21,69 %
SPF total usage	3 334 569	55,30 %

Corresponding DMARC adoption statistics are presented in Table 5.

Table 5. DMARC statistics for .nl

Policy	Count	Percentage
Not in use	4 553 765	75,52 %
Policy none	854 495	14,17 %
Policy quarantine	41 037	0,68 %
Policy reject	580 409	9,63 %
Total usage	1 475 941	24,48 %

3.2.3 Statistics for governmental organisations in the USA

Agari (Agari Inc and Agari Data Inc) is an organization focusing in email security related projects and services. Agari has been working together with the DHS after the issuance of BOD 18-01 mentioned before (DHS, 2017), and a progress report was published in September 2018 on the execution status of the directive (Agari Inc, 2018). The report was gathered one month before the BOD 18-01 final deadlines. The statistics of that report are iterated in Table 6.

Table 6. DMARC statistics for .gov.us as of Sep 2018 according to Agari (n=1144)

Policy	Count	Percentage
Not in use	192	17 %
Policy none	Not stated	18 %
Policy quarantine	Not stated	1 %
Policy reject	727	64 %
Total usage	952	83 %

250ok is an organization providing similar services as Agari does. The annual report from 2019 published by 250ok details statistics specifically for the subdomain .gov.us with references to the BOD 18-01 (Vernhout, M., 2019, p. 24). The statistics from that report are detailed in Table 7.

Table 7. DMARC statistics for .gov.us as of 2019 according to 250ok (n=1262)

Policy	Count	Percentage
Not in use	259	20,52 %
Policy none	64	5,07 %
Policy quarantine	8	0,63 %
Policy reject	931	73,77 %
Total usage	1003	79,48 %

3.2.4 Global statistics

Valimail provides a dataset regarding Fortune 500 dated to Q2 2020 (Valimail, 2020, p. 6). The total sample size for the survey is not stated, but the report states that 70% of the primary domains for Fortune 500 companies use DMARC in any of the configuration states. Only 21% of the primary domains use DMARC with either reject or quarantine policy status. Similar statistics are available from Agari for H2/2020, although also without declared sample size: reject policy is in action for 20% of the domains and 8% use quarantine. 45% of the surveyed domains use DMARC in monitoring mode (Agari Data Inc, 2020, p. 18).

In the same report Agari states the overall statistics for a massive 477 million surveyed domains to be as stated in Table 8 (Agari Data Inc, 2020, p. 16).

Table 8. DMARC statistics according to Agari Data Inc, H2/2020 (n=477 million)

Policy	Count	Percentage
Not in use	468 925 623	98,31 %
Policy none	n/a	n/a %
Policy quarantine	n/a	n/a %
Policy reject	2 041 442	0,44 %
Total usage	8 074 377	1,69 %

Gathering statistics regarding the implementation of SPF and DMARC proved problematic. There are not many parties publishing data regarding ccTLDs in the same manner as .nl has, which then presents problems in comparing implementation rates on a national level. Certain sources monitor and gather statistics from a set of domains and follow their trends in a constant cycle, but the quality of these also vary: some follow only the primary domains used by the interest groups, while some

follow a larger portion of domains. The sample size and exact content of the surveys is not always defined precisely. Additionally multiple SPF related statistics were outdated, with dates going back several years – as the situation can change considerably within a few weeks (as seen from the changes occurring in the .nl zone in 2018), these statistics have only historical value but provide no current insights to the situation. Considering that public sector operations are quite generally open to the public audience, the statistics in this sector seem to have slightly better quality in the research set from an academic perspective.

In summary the variation between technology implementation rates is quite vast. When looking at any large size samples the usage rates for enforced DMARC (reject and quarantine) seem to drop to a few dozen percent units, usually below 30%. However, there is a significant change to this if there are mandatory guidelines set for that target group, as the public sector statistics show. A comparison of the combined statistics for DMARC policy settings reject and quarantine are shown in Figure 8.

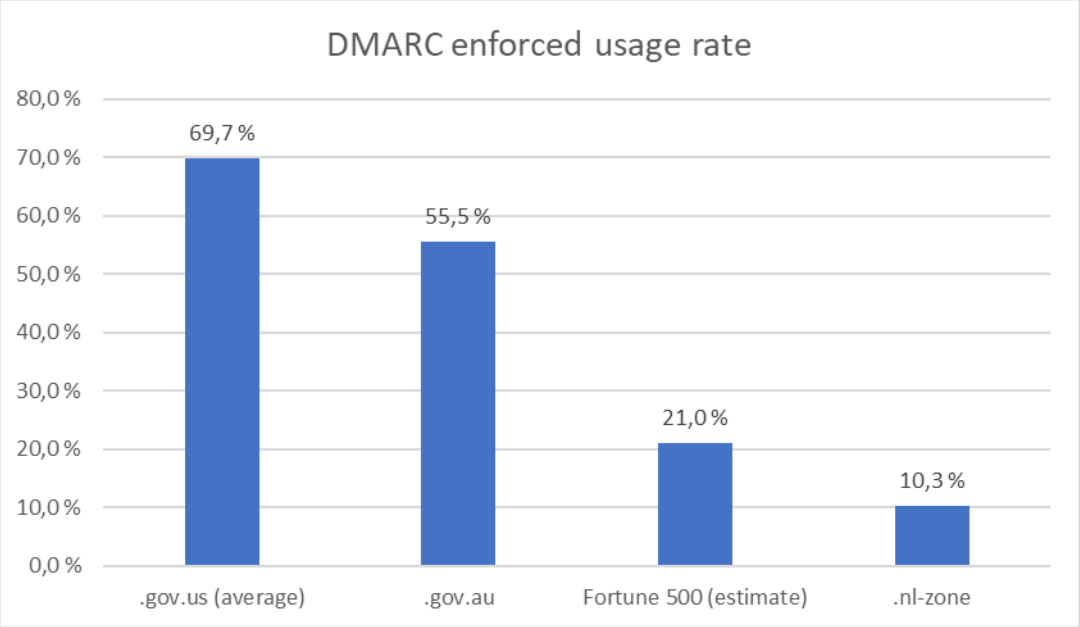

Figure 8. DMARC enforced usage rate

4 Current status and trend forecast

4.1 Statistics and forecast for the .fi-zone

Using the methods described in chapter 2.4 statistics based on the DNS data were collected using the most current .fi-zone for that particular sample date. This data was then formatted based on the criteria specified in chapter 2.4, and visual graphs were formed. The full gathered dataset is in Appendix 2, and this chapter contains extracts and graphs derived from that data for visualization purposes.

Trend forecast was locked on week 7/2020, based on the results gathered from week 36/2019 onwards. After this communication from NCSC-FI to external parties begun. Further iteration of the used communication methods, target audiences and the dates are in chapter 5.2.

Initial statistics for the used technologies were collected during August 2019. On the preliminary rounds the statistics were gathered several times within short timeframes. This was performed as a sort of agile scrap method: as there were no tools specifically designed for this, they were developed from ground up with multiple iterations. Gained results were then compared to runs performed on the same data within close time vicinity in order to establish reliability. Anomalies and errors were manually checked upon and verified whether it was due to errors in the tool coding, DNS functionality or the results themselves. After honing the process for a couple of weeks the final tooling script was finished and left as such for the rest of the project, in order to avoid errors in results due to configuration changes. The statistics gathered on week 36/2019 are shown in Table 9.

Table 9. .fi-zone statistics for week 36/2019 (n=474 032)

Assigned value	Count	Percentage
mx-true	359 504	75,84 %
mx-false	111 783	23,58 %
spf-false	318 541	67,20 %
spf-hardfail	55 475	11,70 %
spf-softfail	81 068	17,10 %
spf-nostance	16 203	3,42 %
dmarc-false	455 214	96,03 %
dmarc-test	13 325	2,81 %
dmarc-quarantine	564	0,12 %
dmarc-reject	2184	0,46 %
dmarc-unknown	0	0,00 %

Summing up from the statistics three key findings are found:

1. 75,84 % of the domains in the whole zone state valid MX records.
2. SPF in either SoftFail or HardFail state is in use for 28,80 % of the domains.
3. DMARC implementation with enforced policies in total is at 0,58 %, with additionally 2,81 % of domains in testing/reporting phase.

These values are directly comparable to the statistics found regarding the .nl-zone. The MX record statement rate is nearly equal between the zones at approximately 74 % versus 76 %, which at itself is a relatively interesting fact. Initial assumption was that the operational amount of mail related domains would be significantly lower but based on this statistic the vast majority of the registered domains do have valid mail servers present, whether or not this is intended by the domain owners.

SPF implementation differs heavily between the zones. The total usage rate in .nl-zone for either SoftFail or HardFail state is at 55,30 % versus the 28,80 % in the .fi-zone. Looking back at the historical statistics from .nl-zone, the situation for .fi was roughly 2,5 years behind the situation in the .nl-zone. The earliest statistics for .nl-zone from around July 2017 (before the campaigns launched by SDIN) show a total SPF usage rate at around similar percentages as the .fi-zone had in September 2019, and the difference between the SPF qualifiers has evolved from that point on.

As per DMARC usage rates the situation was also very different: the .nl-zone had similar enforced usage rates before August 2018 (below 2 %), with testing/reporting phase usage rates at between 2-3 %. As the .nl-zone had significant leaps in

implementation rates on September-October 2018, a major difference in the statistics is present. When looking at the implementation rates of DMARC in overall in the .fi-zone in comparison to the largest available sample size provided by Agari similar results are found for the enforced policies – the general usage rates are around 2 %.

Based on these comparison points the initial situation for the .fi-zone regarding the implementation rates of secure SPF and DMARC policies can be considered as average in comparison to other available larger sample sets.

Upon continuing the collection of the statistics, a general development trend was established. Keeping in mind that the sample size average for the time period is approximately 478 800 domains, one percentage point of change corresponds to changes in about 4 800 domains. The gathered baseline regarding the observed trend for SPF implementation is shown in Figure 9.

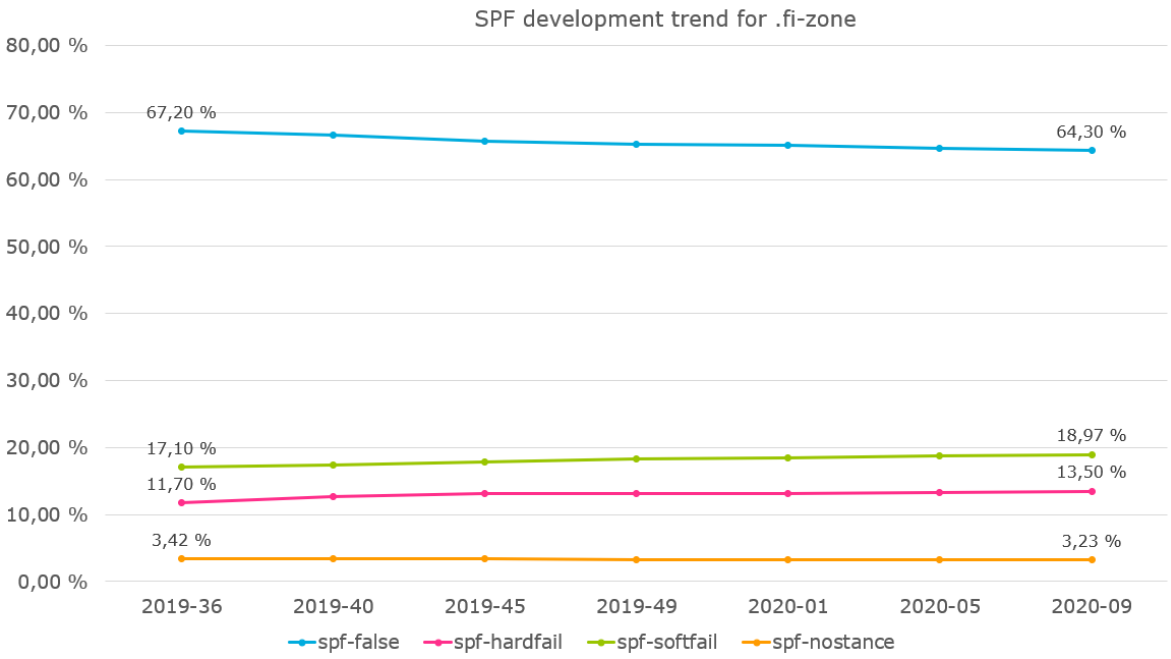

Figure 9. SPF development trend for .fi-zone (weeks 36/2019-09/2020)

Based on these statistics the development trends follow a slight but constant change. The total development in percentage unit changes over a period of approximately half a year (25 weeks) are the following:

- Lack of SPF implementation dropped from 67,20 % to 64,30 %; net decrease of 2,90 percentage points

- The total implementation of enforced SPF increased from 28,80 % to 32,47 %; net increase of 3,67 percentage points

The corresponding DMARC development graph is shown in Figure 10. The Y-axis is split in order to fit the values logically into the same graph with the development changes still visible.

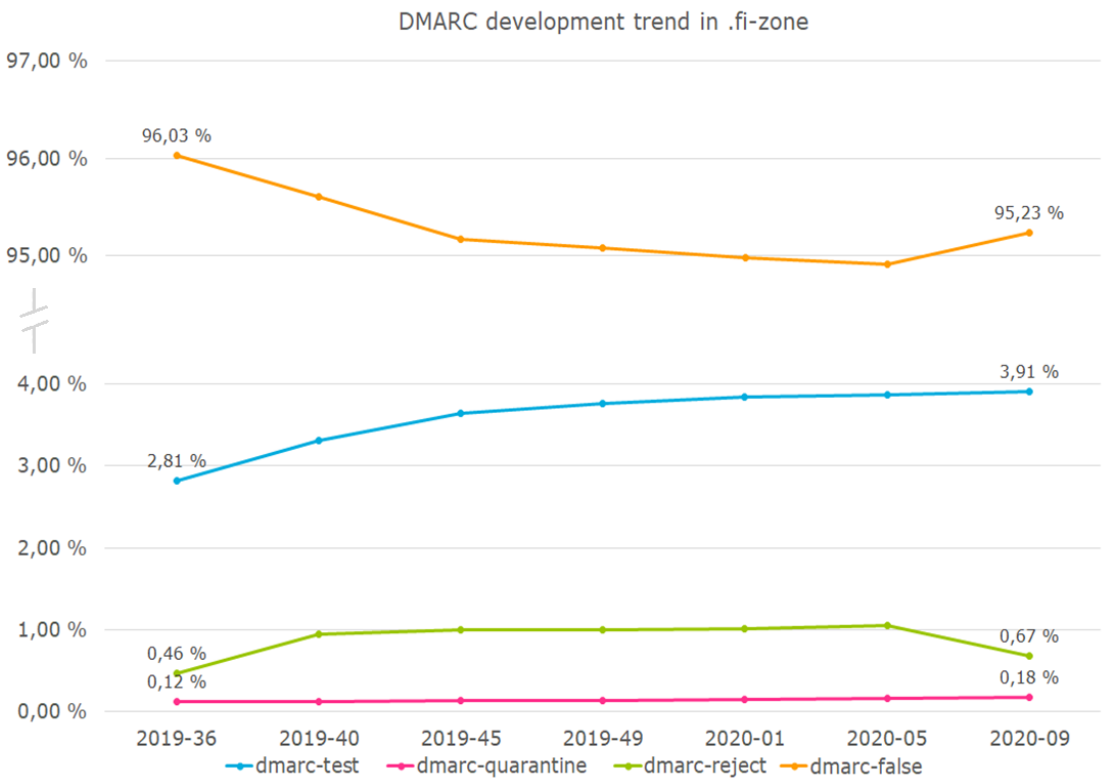

Figure 10. DMARC development trend for .fi-zone (weeks 36/2019-09/2020)

The overall development trend for DMARC implementation seems to vary slightly more than for SPF. Even though there was initially an observable trend towards DMARC reject policy being implemented to a larger portion of the domains, in the sample of week 9 on 2020 the trends reversed for domains not implementing DMARC at all and the domains implementing reject-policy. Interestingly the portion of domains in testing phase did not change at all at that point. Net changes in the policy deployment portions are as follows:

- Lack of DMARC deployment rate decreased by 0,8 percentage points, from 96,03 % to 95,23 %
- Enforced DMARC deployment rate increased from 0,46 % + 0,12 % to 0,67 % + 0,18 %, representing a cumulative increase of 0,27 percentage points

Based on this it can be said that the deployment average of DMARC enforced policies is relatively low – approximately 220 domains per month adopt either quarantine or reject as their policy. SPF implementation trend is slightly better, as approximately 2900 domains implement SPF qualifier of HardFail or SoftFail per month. Comparing these values to the average change in the total count of domains in the zone, SPF implementation rate is surpassing the increase in the zone size, although with a minor gap. This indicates that in a very long run the implementation trend of SPF would cover the whole zone if this pace keeps up. For secure DMARC implementations this does not seem to be naturally achievable.

4.2 Interest group statistics and forecast

As stated before, the interest groups are divided into four groups: municipalities, cities, governmental agencies, and ministries. In this chapter the initial status and forecast for development changes regarding these groups is explored. In each of the figures and charts “n” represents the amount of operational domains for that target group found at the point of data gathering. There may be more .fi-domains registered for each of the organizations, but if no valid name server records are found the domain is not included in the survey nor statistics for not being operational. Detailed breakdowns of the statistics are found in Appendix 2: Target group data. The initial detected rate for SPF deployment is shown in Figure 11.

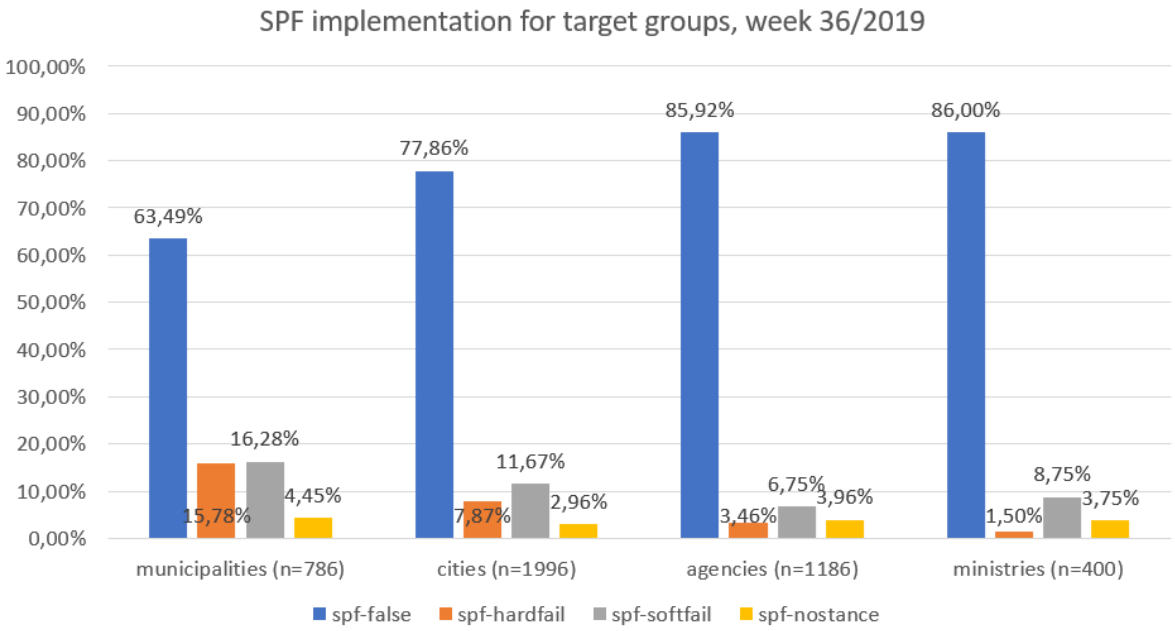

Figure 11. SPF implementation status for target groups, week 36/2019

Breaking down the sample sizes in relation to the target group sizes is essential in understanding the statistics.

- There are 205 municipalities, and one municipality owns 4 .fi-domains in average
- Correspondingly one city owns an average of 18 .fi-domains
- Agencies have approximately 20 domains on average
- And finally, ministries own 33 domains each

The first thing arising from the statistics is the following fact: as the number of domains owned by an organization (in average) increases, the secure deployment of SPF falls.

Second interesting note concerns the deployment rates regarding domains with declared MX records – these represent a direct attack surface due to the presence of the MX record and possible “quick wins” regarding SPF deployment ease. Figure 12 represents this in detail.

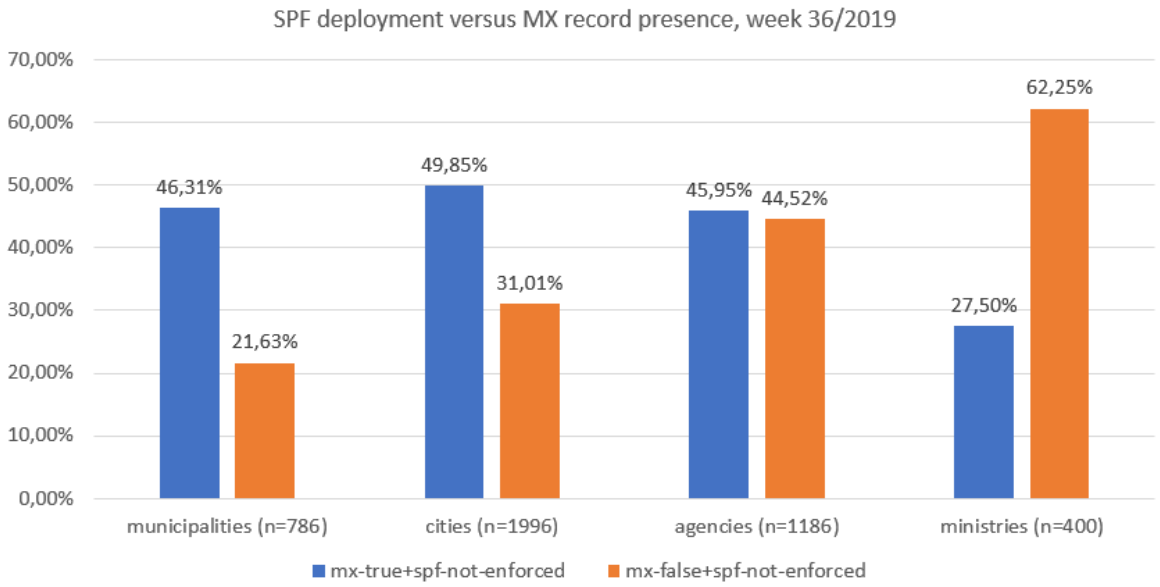

Figure 12. SPF deployment versus MX record presence, week 36/2019

In the figure the blue bar represents the potential attack surface: domains with declared MX records (incoming mail is likely to be processed) with SPF either not in use at all (record not present) or an explicit “all” qualifier is missing in either HardFail or SoftFail state. From these statistics it can be said that an average of 42,40 % of the domains owned by the public sector in Finland can be easily forged without causing SPF failures.

On the other hand, the orange bar represents the low hanging fruits: domains which are not designed to receive mail and therefore very likely being used for only other purposes, such as brand presence on websites only, but which are lacking proper SPF protection. It should be assumed that a valid and security-enforced SPF implementation would be very easy to deploy to these domains.

The situation for DMARC implementation status for the same time period of week 36/2019 is shown in Figure 13. It has to be noted that in order to see the values regarding anything else than DMARC not being deployed at all, the Y-axis in the figure cuts off at 3,00 % from the higher end.

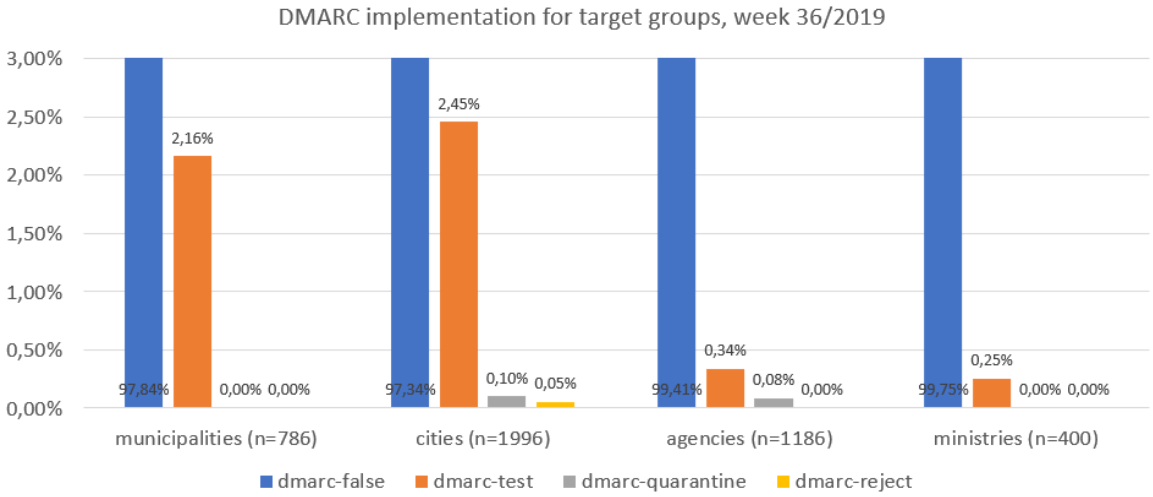

Figure 13. DMARC implementation for target groups, week 36/2019

The first key finding is that the total count of domains implementing a DMARC policy of either reject or quarantine is four (two for each policy state) out of the total surveyed count of 4 368. This leads to secure implementation rate of 0,09 % in the whole of public sector owned domains in the .fi-zone. Municipalities and cities do have a substantially higher percentage of domains with the policy status of “none”, which leads to the conclusion that some organizations are possibly attempting to implement the technology to some of their domains. In overall, the initial status of deployment status and settings can be considered suboptimal.

In order to present coherent figures for SPF implementation trends, two data points are joined together: SPF qualifiers SoftFail and HardFail. The observed development trend regarding this over the weeks 36/2019-05/2020 is shown in Figure 14.

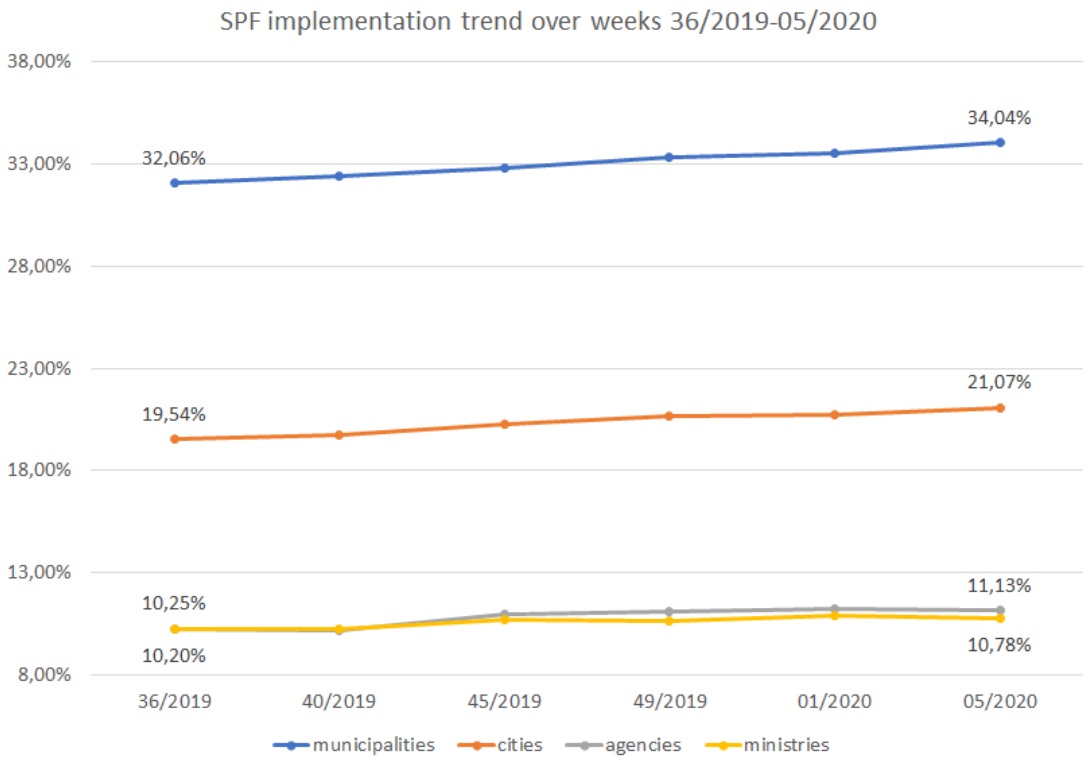

Figure 14. SPF implementation trend target groups, weeks 36/2019-05/2020

The SPF implementation trend for the municipalities and cities seems to follow a slow and steady path towards better implementation rates. The overall change over a five-month period is 1,98 and 1,53 percent points, respectively. This is slightly below the .fi-zone trend (3,67 percent points). The situation regarding agencies and ministries is different: considering the sample sizes there is barely a noticeable difference: 0,88 percent points for the agencies and 0,58 percent points for the ministries. Considering that the agencies and ministries represent a significantly smaller amount of total organizations in comparison to the municipalities and cities, even one organization performing an improvement project would be clearly visible on the statistics due to the higher count of domains owned per organization.

DMARC implementation trend for the same period is shown in Figure 15. As it is with SPF, all applicable DMARC policy settings (none, quarantine, reject) are joined to present the status of any deployment models regarding this technology.

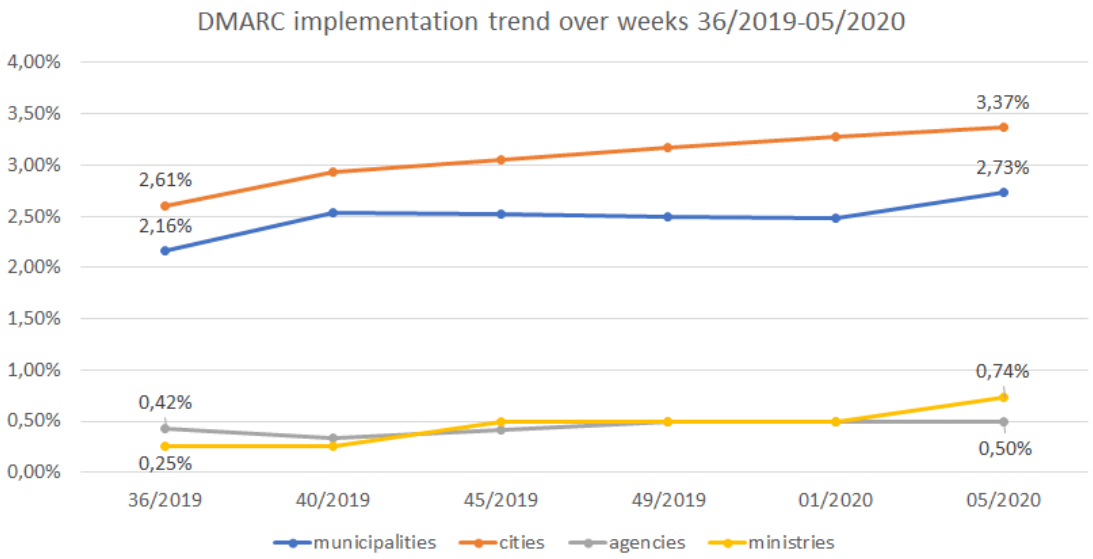

Figure 15. DMARC implementation trend target groups, weeks 36/2019-05/2020

The total implementation rates for any DMARC adaptation policy were increasing in each of the target groups. Considering the change in percentages, ministries approximately tripled their implementation rate – but to put into an actual scale, the increase was only 0,49 percentage points. It is worthwhile noting that the implementation rate for the agencies did not increase practically at all, and the ministries' implementation rate surpassed them. Cities increased the adaptation rate by 0,76 percentage and are leading within the target groups on a semi-stable development path. Municipalities present slight steps in the adaptation rate: there is an initial change of approximately 0,4 percentage points, then practically nothing happens, and in the last four weeks another bump of 0,25 percentage points occurs.

5 How to improve

The action phase of the research is explored in this chapter. The process of forming the best practices and guidelines is visited upon, along with a comparison analysis of the resulting work in comparison to the foreign counterparts which were visited in chapter three. After this the timeline and methods how these were communicated to the selected interest groups is explored.

Original plans for this part of the project were developed in January-February 2020. A significant portion of this part was originally devised to be performed within the National Cyber Security Centre Finland (NCSC-FI) along with other current projects.

The original plan was accepted and set to motion, but the global onset of Covid-19 changed the operational situation within the organization significantly – resources (both manpower and financial) were re-allocated in some extent, and this had a direct effect on this part of the project. Therefore, the original and intended plan is elaborated here, along with the now-postponed timeframe and plan of actions.

5.1 Best practices and guidelines

NCSC-FI is currently utilizing three main methods in the process of instructing organizations and individuals in Finland: Information Security Now! (Tietoturva Nyt!, abbreviated as TTN from here on)-posts, topic-specific guideline and instruction sets and finally mandatory guidelines (Määräys in Finnish) for constituents under legal requirements. Mandatory guidelines were ruled out immediately from the selection set: there is currently no legal basis or existing mandatory guidelines under which NCSC-FI could issue out anything regarding the data published in a DNS record by the domain owners. Therefore, the first two options are the only choices: a guideline document and a supporting publishment under the TTN! “brand”. The guideline document was decided to be published in electronical form only (as a PDF document), as it was seen that physical material distribution would not serve a purpose in this context.

From the beginning of the project it was already clear that foreign counterparts and partners to the NCSC-FI had already developed and published material directly regarding this subject. It was seen foolish not to begin the process by first collecting any available material and inspecting whether a consensus from them could be found. As already explored earlier in chapter 2.2.3, there is a relatively clear consensus on the matter. The guideline main topics and the rationale for them are the following.

Deploy SPF, DKIM and DMARC for all owned domains, both with parked and mail traffic related and for all used subdomains. Deploying the technologies in a secure manner should be considered as a basic security hardening option. Parked domains appear in public WHOIS records to be owned as the organization, and they should

not be ignored – the recipient organization may not have any indicators which of the domains are designed to handle mail traffic.

Remove unnecessary MX records for parked domains or replace them with a null MX option (“.” as the record data). While this is not strictly necessary if proper SPF and DMARC settings have been made, it does further enforce the stance that mail concerning these domains should not be handled at all.

Deploy SPF with explicit “all” setting along with primarily HardFail qualifier. If DMARC is not yet deployed, SoftFail is also valid qualifier. SPF default stance without a specific match to a stated record is “Neutral”, so therefore the explicit “all” should be published. While deploying the technologies the SoftFail qualifier “~all” is adequate, but efforts should be made towards adopting the HardFail qualifier. For parked domains a SPF record of “v=spf1 -all” should be used.

Deploy DKIM signing on internet-facing mail relays where applicable. Deploy empty DKIM records for the rest of the domains and subdomains where mail services are not used. DKIM signing should be used only on the last owned mail relay before the message is sent out from the organization. There can be scenarios where a message is modified in internal transit (adding disclaimers, for example) which will cause a DKIM signature to fail. It is advisable to sign a message only once. If a (sub)domain does not use DKIM signing at all, it should be indicated with a “v=DKIM1; p=” record.

Deploy DMARC with policy option “reject” as a target for both mail-related domains and parked domains. Aim for deploying DMARC with percentage of mails to be filtered at 100% (the setting “pct=N”). Monitor for DMARC forensic and aggregated reports. As a goal this is self-evident, but the process may take a while. Deploying DMARC initially in the reporting mode will nevertheless give immediately a better visibility in mail delivery and forgery status and utilizing this will be helpful when adjusting SPF and DKIM records and settings. Deploying DMARC with the policy none without stating any recipient addresses for aggregated reports should not be used under any circumstances, as this approach undermines the whole purpose of the policy option.

Considering that deploying these technologies from a domain owner’s perspective is practically only half of the way in utilizing these in practice, as this requires effort

from mail filter services also. Each organization should be in control regarding the mail flow inbound to their organization even if they do not manage the mail services themselves, so therefore the following best practices are to be followed:

- Filter incoming mail based on SPF, DKIM and DMARC checks (along with other filtering options such as content checking)
- Respect SPF qualifier settings and send out Non-Delivery Reports (NDRs) according to the SMTP RFC
- If possible, respect the DMARC policy setting of “quarantine”
- If the previous is not possible, treat quarantine-policy as like policy none (do not reject), but apply a spam marking on the message
- Send out at least aggregated reports based on DMARC settings
- If possible, send also forensic DMARC reports

5.2 Communicating guidelines to interest parties

In this portion different communication methods about the project, the instructions and their realized or planned execution times are iterated.

Initial briefing of the project and current statistical situation of the public sector was presented in Disobey 2020 event closed-doors GovTrack-portion on Friday, February 14th (week 7). An undisclosed amount of governmental security organizations took part in this session. At this stage the initial results from a six-month observation period were already available, and this material was presented. Some of the organizations had discovered the issue concerning the low deployment rates on their own but did not have the larger data available. The participants all agreed that promoting the secure use of these technologies would be beneficial in overall for the public sector organizations. As the participants did not represent the IT administration or manager level from their own organizations, it could be assumed that no direct impact to the statistics would be gained from this session.

On March 21st, 2020 (week 12) a virtual meeting was organized for governmental organizations. The meeting was a part of series of meetings in which different information security matters and phenomena are discussed in-depth with the representatives of governmental organizations. Participants generally represent their organization from a chief of information security role or similar and participation is voluntary for the organizations. Notes from the meetings and possible additional material was distributed to an email list operated by NCSC-FI. This list contains solely

governmental staff. In this meeting the gathered statistical data regarding the situation in both the whole .fi-zone and the public sector was shown, and the desired course of action was presented. The course of action plan matches the best practices list defined in chapter 5.1, added with some practical examples and steps to keep in mind when beginning the implementation. The displayed material was distributed to the organizations on the mailing list on week 13 for future reference. Considering that the target audience in the meeting and the mailing list are generally the deciding persons in each organization when it comes to IT or information security related issues, this forum may be the best available option in influencing the situation in the governmental sector. While the participation rate of organizations does not cover the governmental sector fully, a significant portion is present and therefore a visible impact could be achieved.

Municipalities (incl. cities) are self-organized and governed units when the subject is on how they organize their own operations. As for information technology related matters they can use their own discretion on how and what they decide to implement. Municipalities do have a certain level of information sharing between each other and the Association of Finnish Municipalities (Kuntaliitto), and the latter organizes training events and provides material to the municipalities on different matters. In discussions between NCSC-FI and Kuntaliitto in late 2019 it was found that there is no actual contact method regarding information security staff for reaching out to the municipalities (besides individual names from seemingly random sources), and a project was set afoot in fixing this. The outcome of this project in the context of this study was planned to be used as another medium in influencing the email security settings level in the municipalities. The planned course of action would have been to organize a similar meeting to the municipalities as was held to the governmental units and distribute the same material to them. This portion was not fulfilled, as the project of collecting the contact points did not finish in time before the project regarding this study was postponed. Nevertheless, the project will highly likely deliver the desired recipient audience considering from the perspective of this study and the target study groups, and this plan of action is therefore sound in the future also.

Having an influence on the large audience of generic .fi domain owners is a bigger challenge. NCSC-FI has mailing lists and a certain amount of reach when it comes to targeted media presence, but this is usually limited to security-aware organizations. These mailing lists will be utilized “automatically”, but they will not provide as large of an audience as desired, as domain owners may vary greatly from individual persons to governmental organizations. Keeping this in mind it was initially planned that an official media release and a small campaign will very likely result in a larger audience regarding the subject. This may result eventually into a measurable impact – the key issue here are budget limitations regarding financing the campaign. Because of this uncertainty larger media campaigns have to be considered separately depending on the current situation.

As an alternative to this the domain administration division within Traficom hosts a portal for domain registrars and has contact details for each of the registrars. Utilizing these would distribute the message to the parties responsible for registering the domains, who might also be in a service provider role to the actual domain owners. Publishing the guideline set and instructions in the portal and distributing them to the registrars would be a project with significantly lower implementation costs and would also reach a relatively large audience. The only minor downside for this is that NCSC-FI cannot perform this on its own – this must be done in collaboration with the domain administration. Based on initial discussions there is mutual interest in the subject, so therefore this option will very likely be explored in the future.

Considering all of the methods above it was seen that significant portion of the intended target audiences could be reached with relatively low financial costs. The biggest obstacle in executing the individual steps in the plan relies on available human resources on planning the campaigns in depth and then actually executing them. Executing the steps with time gaps in between could also indicate which methods provide the best results, but a separate survey to a randomly selected audience could be organized after the whole campaign to verify the objective DNS-surveyed results.

6 Monitoring change and evaluating results

As stated before, the forecast trend for the implementation rates was locked on week 7/2020 when external communication from NCSC-FI begun. This chapter summarizes the possibly observed changes in the statistics in comparison to the established forecast, and an assessment of the latest available data is performed using the previously defined criteria as a basis for the evaluation. Additionally, the latest available data is compared to the foreign statistics explored earlier.

It must be noted that the statistics were not collected between weeks 24 through 31 (June and July) 2020 due to task prioritization caused by the annual holiday season.

6.1 Observed results for the .fi-zone

All the measured data along with the graphs presented here are available in Appendix 2: Full dataset. Key statistic points and observed trends are pinpointed in this chapter, with the starting point being the last timestamp before any external communication on the matter begun.

The final measured results from week 36/2020 in comparison to the data from week 9/2020 for all collected data points are shown in Table 10.

Table 10. .fi-zone statistics differences between weeks 9 and 36/2020

Assigned value	Percentage on week 9/2020	Percentage on week 36/2020	Difference
mx-true	75,50 %	74,67 %	-0,83 %-pts
mx-false	24,50 %	25,33 %	+0,83 %-pts
spf-false	64,30 %	62,37 %	-1,93 %-pts
spf-hardfail	13,50 %	14,81 %	+1,31 %-pts
spf-softfail	18,97 %	19,78 %	+0,81 %-pts
spf-nostance	3,23 %	3,03 %	-0,20 %-pts
dmarc-false	95,23 %	94,57 %	-0,66 %-pts
dmarc-test	3,91 %	3,83 %	-0,08 %-pts
dmarc-quarantine	0,18 %	0,25 %	+0,07 %-pts
dmarc-reject	0,67 %	1,34 %	+0,67 %-pts
dmarc-unknown	0,01 %	0,01 %	0,00 %-pts

In comparison to the key findings explored in chapter 4.1 the following was found:

1. A total of 74,67 % of the domains in the whole zone state valid MX records – this declined by 0,83 percentage points in six months and by 1,17 percentage points in a year.
2. SPF in either SoftFail or HardFail state is in use for 34,59 % of the domains – in total there was an increase of 2,12 percentage points in six months and 5,79 percentage points over the one-year period.
3. Enforced DMARC implementation resulted to 1,59 %, which increased by 0,74 percentage points in six months and by 1,01 percentage points over the year.

The observed SPF implementation trend derived from the data from the data points collected in that same timeframe on the whole .fi-zone is shown in Figure 16.

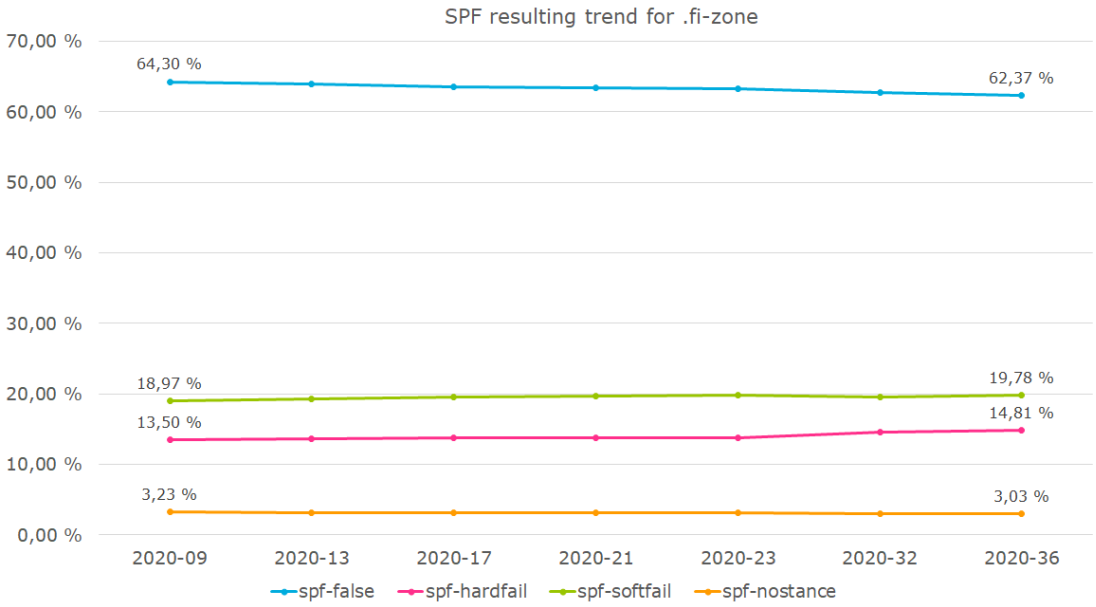

Figure 16. SPF resulting trend, weeks 09/2020-36/2020

The corresponding statistics regarding observed DMARC implementation trends from the same time period are shown in Figure 17.

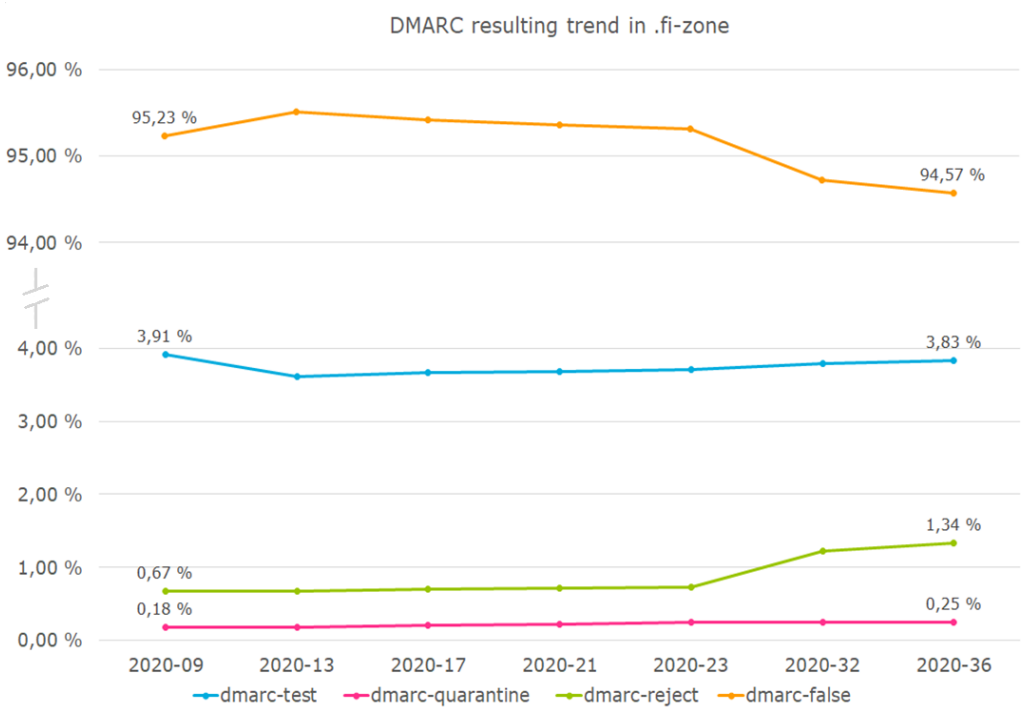

Figure 17. DMARC resulting trend, weeks 09/2020-36/2020

An interesting note is that between weeks 9 and 13 DMARC implementation from "p=none" switched apparently over on approximately 0,3 % of the sample size (appr. 1500 domains) to having no DMARC DNS records at all. Then between weeks 23 and 32 approximately 0,5 % of the sample size (appr. 2500 domains) switched over directly to DMARC "p=reject".

Complete trendlines regarding all the measured data points from the whole statistics collecting period (week 36/2019 through week 36/2020) are shown in Appendix 3.

6.2 Observed results for the target groups

Following the same logic from chapter 4.2, first the latest SPF and DMARC statistics are explored, and then the observed trendline leading to the final result is shown. The fine-grained SPF results from week 36 in 2020 is shown in Figure 18.

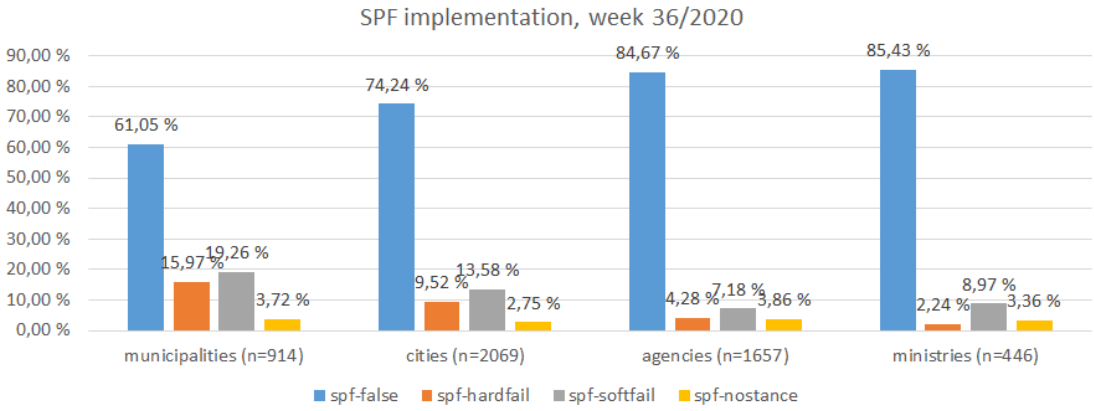

Figure 18. SPF implementation for the target groups, week 36/2020

Following up from the statistics gathered on week 5 in 2020, a development chart is shown in Figure 19. The figure contains both SPF SoftFail and HardFail qualifiers in order to show the total enforced policy deployment.

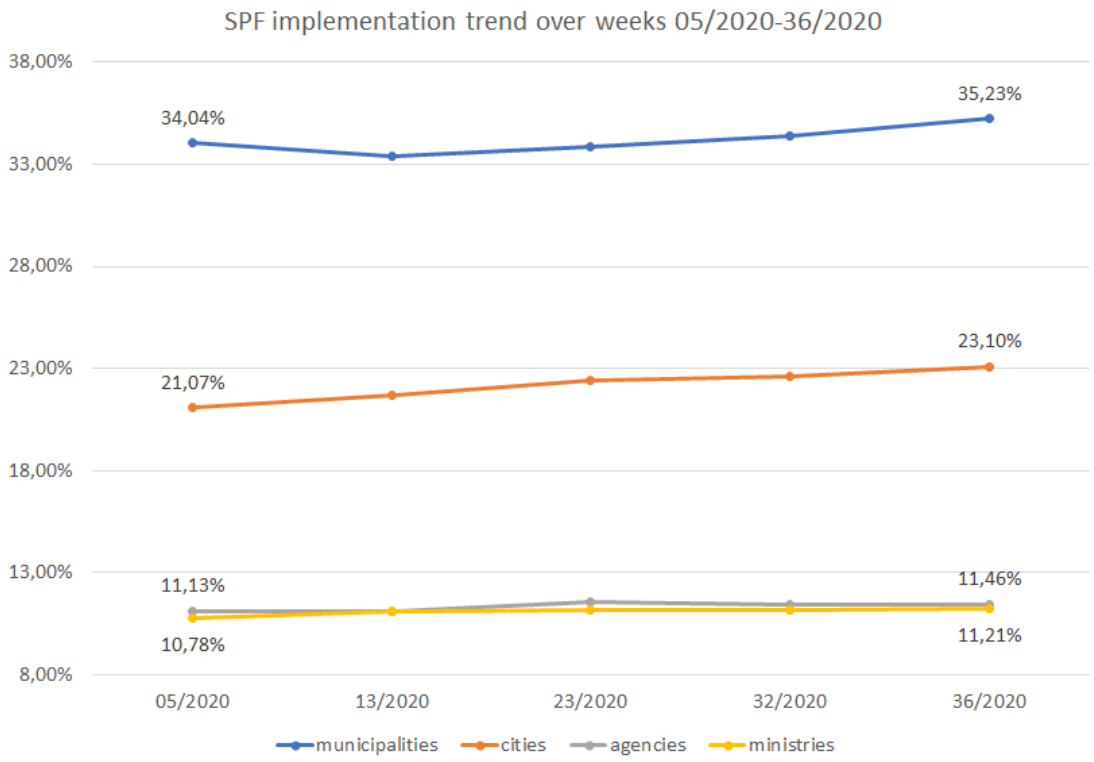

Figure 19. SPF resulting trend for target groups, weeks 05/2020-36/2020

It is interesting to note that the municipalities' development dipped approximately by one percentage unit, until it recovered and then kept a steady development pace. The cities followed a relatively steady development path, gaining 2,03 percentage points – on the other hand the ministries gained only 0,43 percentage points, and the agencies advanced a mere 0,33 percentage points. Combining the key data

points observed together a cumulative trend from one year was formed, and it is shown in Figure 20.

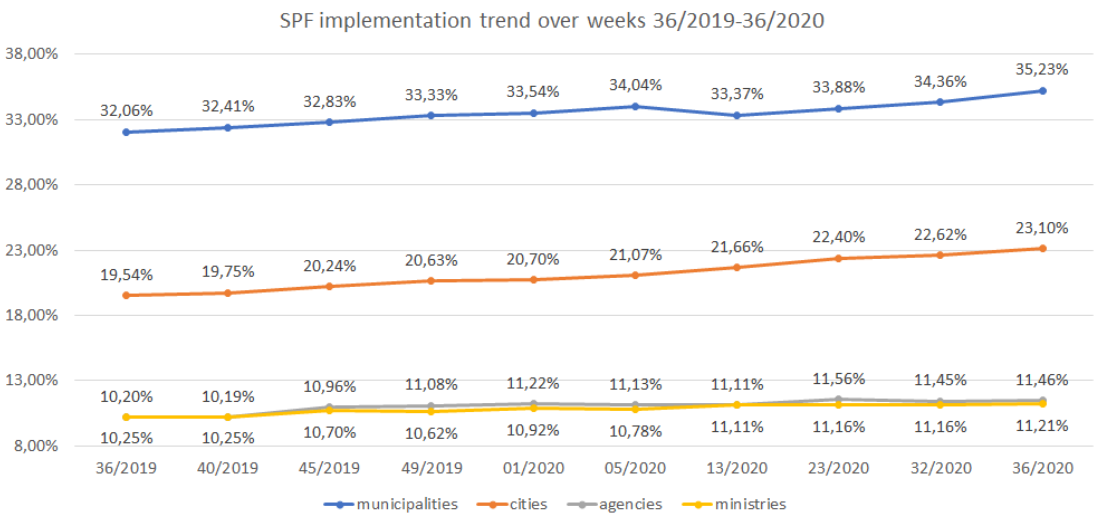

Figure 20. Complete SPF trend for target groups, weeks 36/2019-36/2020

The dip observed for the municipalities implementation rate shown in Figure 19 is clearly present the whole year statistics. The trends are quite consistent in the other groups: the cities follow a relatively steady development path with consistent increases in adaptation rates, but the governmental agencies and ministries are practically in a stale position in comparison to the aforementioned groups.

The total breakdown of measured DMARC adaptation policies on week 36 of 2020 is shown in Figure 21. Again, the Y-axis in the figure is capped off at 5 % in order to show anything besides the DMARC “false” status, which is the predominant result for all of the target groups.

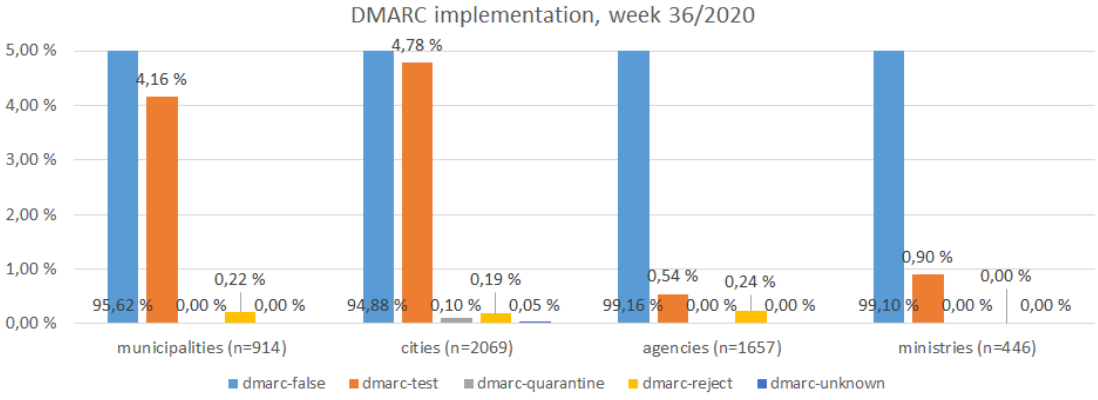

Figure 21. DMARC implementation for the target groups, week 36/2020

It is interesting to note that there are 0,05 % (approximately 103 domains) in the status “DMARC unknown” for the cities: this in practice means that the initial “v=DMARC1;” declaration is present at _domain.city.fi, but the syntax is not technically correct in the DNS record.

Following up with the data collected from week 5 through week 36 in 2020 a graph was formed, and it is shown in Figure 22. The graph contains combined data for any applicable and functional DMARC policy (none, quarantine, reject) to indicate any usage of the technology in practice.

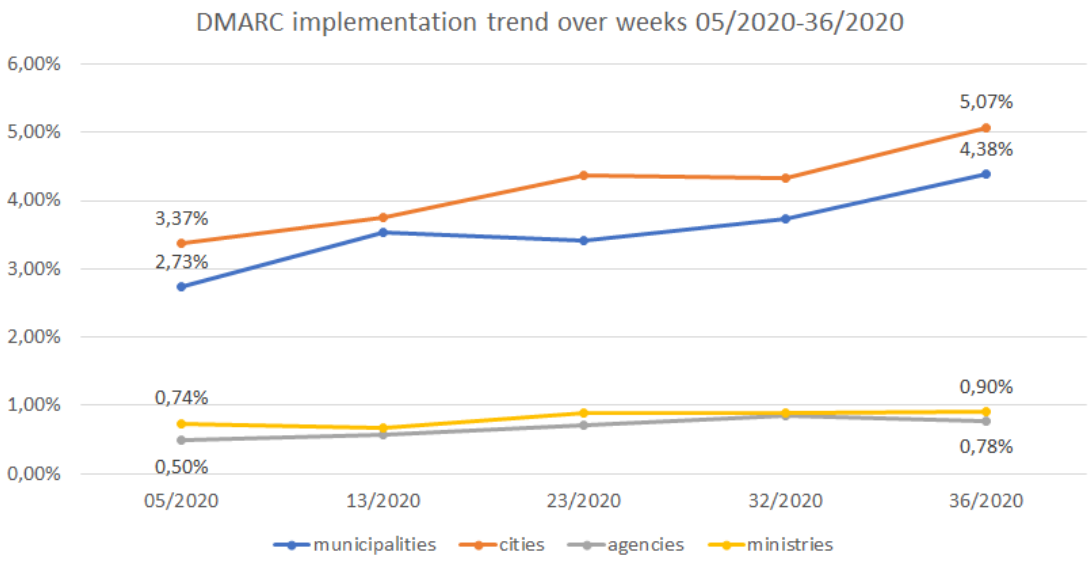

Figure 22. DMARC resulting trend for target groups, weeks 05/2020-36/2020

A similar dip in DMARC development curve for the municipalities can be seen as it was with the SPF development – although it happens one cycle later, between weeks 13 and 23. It is unclear if these two events are connected somehow to each other. Combining the data accumulated over the year, the total observed DMARC implementation trend is shown in Figure 23.

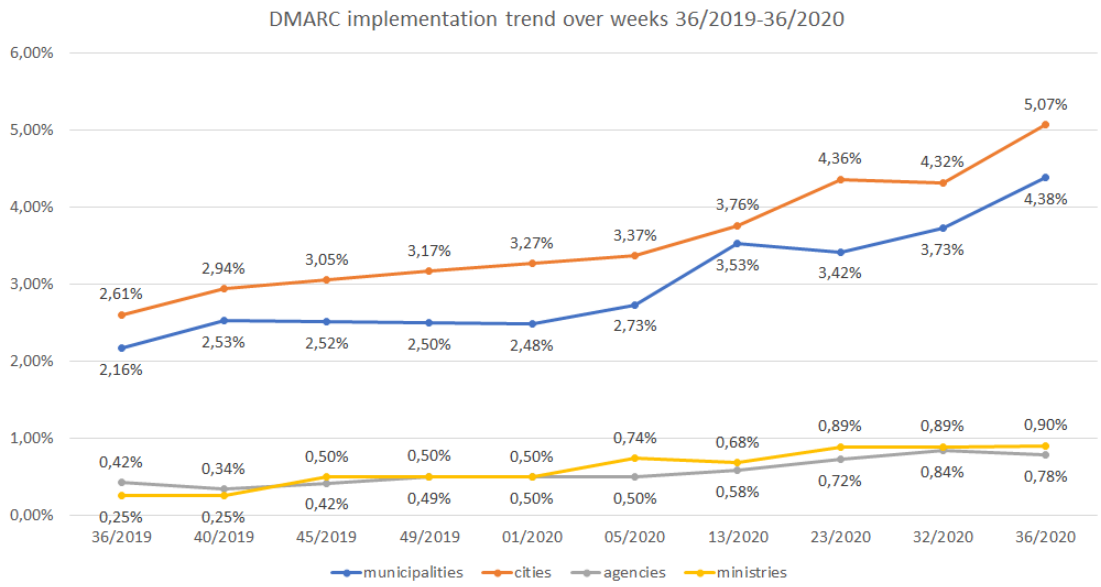

Figure 23. Complete DMARC trend for target groups, weeks 36/2019-36/2020

The resulting development path for the municipalities and cities is interesting: while there were some minor changes in the initial six-month observation period, changes occurred which do not follow the expected trendline, in both advancements and declines. Observing the trend from the perspective over a whole year the trends are consistent and upward-heading – the technology is being adapted in more domains somewhat consistently. The situation for the governmental bodies follows the SPF trends: slight changes are occurring, but not in the pace and scale of the cities and municipalities.

6.3 Impact of actions taken

In this chapter the impact of the actions and possible external influences on the observed trends are explored, starting out on an analysis regarding the whole .fi-zone and then separately the target groups.

6.3.1 Changes in the .fi-zone trends

As the study focuses on technology implementation trends there is a possibility that external factors to this study would have an influence on the measured results. These would include factors such as

- Domestic service provider, e.g. domain registrar, implements new policies by default to their customers

- Major global service providers, such as Microsoft, would implement instructions in their guidelines or default policies
- External campaigns or media attention on the subject of email forgery causes individual organizations to take action

The two first factors were not observed. Microsoft is in the process of implementing further support to separate protocols related to the subject in Office 365 environment, but the default policies regarding SPF, DKIM and DMARC have not changed. No major domestic service providers had an active campaign on the matter which would be known of.

Email forgery related issues, such as different types of frauds, did make the national headlines in IT related news outlets (Halonen, 2020; Kantomaa, 2020; Mikrobitti, 2020), but the topic of forgery prevention via the studied technologies was not heavily present in the news.

Considering that the intended plans for reaching out to the majority of domain owners and operators in the .fi-zone were not executed, as it was stated in chapter 5 it was expected that no major changes would occur due to the actions (or lack thereof) performed by the NCSC-FI.

As discussed in chapter 4.1, the observed initial increase in enforced SPF implementation rate was 3,67 percentage points for the initial 25 weeks' trend establishment period. For the observation period an increase of 2,12 percentage points was observed. Looking at the total development graph found in Appendix 2, no major changes in the trendline can be seen.

Enforced deployment of a DMARC policy had a small fluctuation regarding the policy status "p=reject". As it was seen in chapter 4.1, Figure 10, DMARC implementation peaked from the initial status on week 36/2019 only to drop lower on week 09/2020. Looking at the total development graph, found in Appendix 2, this drop evened out in between weeks 23 and 32 on 2020. In overall the development trend follows a relatively steady path without major fluctuations aside from this deviation.

Summarizing from all of the factors above no major changes regarding the implementation trends of SPF and DMARC were observed during the whole study. The technology implementation trend is directed to more domains using the

technologies securely, but the adaptation trend is relatively slow and did not peak in any direction in a significant manner.

6.3.2 Target group change

The breakdown of the changes in combined SPF implementation rates for the target groups in relative percentage points for the initial observation period, the following action phase and the resulting change over on year are shown in Table 11.

Table 11. Combined SPF statistics changes in the target groups (units percentage points)

Target group	Change between 36/2019-05/2020	Change between 05/2020-36/2020	Total difference 36/2019-36/2020
Municipalities	1,98	1,19	3,17
Cities	1,53	2,03	3,56
Agencies	0,93	0,33	1,26
Ministries	0,53	0,43	0,96

The comparable combined DMARC implementation rates are shown in Table 12.

Table 12. Combined DMARC statistics changes in the target groups (units percentage points)

Target group	Change between 36/2019-05/2020	Change between 05/2020-36/2020	Total difference 36/2019-36/2020
Municipalities	0,57	1,65	2,22
Cities	0,76	1,70	2,46
Agencies	0,08	0,28	0,36
Ministries	0,49	0,16	0,65

Considering the actions performed by NCSC-FI as explored in chapter 5.2, or more of the lack of actions, no major changes in the trends can be attributed to NCSC-FI. Looking at the gathered statistics and the cumulative trend paths for SPF and DMARC implementations (Figure 20, Table 11; and Figure 23, Table 12 respectively), the governmental agencies and ministries show no practical change development. The cities and municipalities had already a consistent development path set and kept that pace through some variations in the trends, even surpassing the expectations regarding DMARC implementation trends.

6.4 Evaluating technology implementation level

Based on the guidelines and instruction sets visited in chapter 3.1 and the planned content for the guidelines to be published by NCSC-FI discussed in chapter 5.1 the following criteria must be met for a secure deployment of SPF and DMARC:

1. A synthetically valid SPF record has been published with explicit “all” parameter for all possible domains.
2. The “all” parameter must be set to either SoftFail or HardFail state.
3. A synthetically valid DMARC record has been published for all possible domains.
4. For domains without valid MX records DMARC policy must be set to reject.
5. For domains with MX records the policy must be set to either quarantine or reject.

Using these parameters as baseline the data collected in this study with final labels in any of the following combinations satisfy the requirements:

```
mx-true+spf-hardfail+dmarc-quarantine
mx-true+spf-hardfail+dmarc-reject
mx-true+spf-softfail+dmarc-quarantine
mx-true+spf-softfail+dmarc-reject
mx-false+spf-hardfail+dmarc-quarantine
mx-false+spf-hardfail+dmarc-reject
mx-false+spf-softfail+dmarc-quarantine
mx-false+spf-softfail+dmarc-reject
```

Using these strings as search criteria from the statistics gathered on the final week 36/2020, the statistics found from the .fi-zone are shown in Table 13.

Table 13. Total domains meeting the secure deployment criteria within the .fi-zone

Criteria	Percentage	Count
mx-true+spf-hardfail+dmarc-quarantine	0,11 %	542
mx-true+spf-hardfail+dmarc-reject	0,40 %	1984
mx-true+spf-softfail+dmarc-quarantine	0,10 %	488
mx-true+spf-softfail+dmarc-reject	0,05 %	250
mx-false+spf-hardfail+dmarc-quarantine	0,01 %	55
mx-false+spf-hardfail+dmarc-reject	0,76 %	3750
mx-false+spf-softfail+dmarc-quarantine	0,00 %	4
mx-false+spf-softfail+dmarc-reject	0,00 %	24
Total	**1,43 %**	**7097**

Comparing this to the found larger sample sets, the largest sample measured by Agari, shown in Table 8 and the .nl-zone statistics shown in Figure 8, the situation can be considered somewhat mediocre within the .fi-zone in comparison to the larger sample set usage of 1,69 %. The situation in .fi-zone falls short when comparing to the situation in .nl-zone, where the secure implementation rate is at 10,3 % through the gains in the recent years.

By using the same criteria as search parameters on the public sector target groups a comparable picture can be drawn. These statistics are shown in Table 14.

Table 14. Total domains meeting the secure deployment criteria within the target groups

Criteria	Municipalities	Cities	Agencies	Ministries
mx-true+spf-hardfail+dmarc-quarantine	0.00%	0.00%	0.00%	0.00%
mx-true+spf-hardfail+dmarc-reject	0.00%	0.10%	0.00%	0.00%
mx-true+spf-softfail+dmarc-quarantine	0.00%	0.10%	0.00%	0.00%
mx-true+spf-softfail+dmarc-reject	0.11%	0.05%	0.00%	0.00%
mx-false+spf-hardfail+dmarc-quarantine	0.00%	0.00%	0.00%	0.00%
mx-false+spf-hardfail+dmarc-reject	0.11%	0.05%	0.24%	0.00%
mx-false+spf-softfail+dmarc-quarantine	0.00%	0.00%	0.00%	0.00%
mx-false+spf-softfail+dmarc-reject	0.00%	0.00%	0.00%	0.00%
Total	**0.22%**	**0.30%**	**0.24%**	**0.00%**

The situation in all of the target groups regarding secure implementation of the inspected technologies is significantly below all of the available reference groups: the averages in .fi-zone, global domains, foreign public sector statistics exceed the measured situation in the Finnish public sector on a logarithmic scale.

Considering that SPF and DMARC implementation is done from a technical perspective quickest when a domain is not designed to send any mail, there are some key takeaways. Any domain not having valid MX records could be very easily turned to having secure SPF and DMARC policies in place, as there is no need for monitoring the mail activity and adjusting the records accordingly. Keeping this in mind, the following statistics can be seen from week 36/2020:

- Municipalities have 22,98 % of the domains in this state
- Cities have 34,17 %
- Agencies have 54,13 %
- Ministries have 66,59 %
- And the whole .fi-zone has 25,33 % of domains in this state

In practice this means that the previously listed findings could be radically changed within a week or less with very small (if any) impact to mail flow. While this does not improve the security stance on the mail-related domains, it does prevent forgery using other domains owned by the organizations, which is a worthy goal as itself.

Another point to bear in mind is that DMARC in the monitoring state (p=none) does not influence mail flow at all – it is designed by default to be just that: observing whether SPF and DKIM implementations are functioning as expected before implementing stricter policies. Deploying a simple DMARC record and setting up a service monitoring the received aggregated reports could be easily done within a few days at most. Even if the policy state would never be "upgraded" to quarantine or reject in order to prevent forgeries, this would at least provide situational awareness into knowing whether someone is attempting to forge mail using the domains owned by that entity.

7 Conclusions

7.1 Answers and reliability

To recap, the research questions for the study were:

1. What are the usage rates regarding SPF and DMARC implementations in the .fi-zone?
2. What is the situation regarding these technologies in the public sector, divided into four groups (municipalities, cities, agencies, ministries)?
3. Is there a consensus regarding secure implementations of these technologies?
4. Can the observed results in the previous goals be influenced in order to see a statistical difference?

Three of the questions can be considered as answered and met. The general situation in the .fi-zone corresponds to what the situation is in comparable top-level domains, but there is still room to improve towards more secure status. As seen from the foreign examples significant strides can be made within short periods of time, and possible steps for improvement were presented.

The public sector owned domains fall short in secure deployment level when measured against practically any comparison group. Some degree of improvement does occur, but it does not result into implementing the technologies in a secured state. There is significant room for improvement in multiple ways when it comes to the deployment of DMARC.

Available benchmark points in the form of mandatory and voluntary guidelines generally agree on certain key points: SPF must be deployed with explicit qualifiers on all domains, and DMARC must be deployed correspondingly. There are no major contradictions between the available guidelines, and they have a level of internal cross-references to each other, which is not surprising.

No measurable change to the statistics was achieved. While this can be a perfectly normal result in a study, it should be considered as a failure due to the lack of practical effort and communication attempts. It cannot be assumed that major changes would occur without focusing on the matter for a set period of time. However, an action plan has been drafted and as seen above, a baseline to compare to has been achieved, which pave way to further activity on the matter.

When assessing the accuracy of the measured results in the study two key things must be kept in mind: the whole .fi-zone was tested on live DNS servers and the target group accuracy depends on the detected domain ownership, as explained in chapter 2.5. As the study method was to test the whole zone thoroughly and inspect the current status in general, any real-life issues in the implementations are shown in the statistics and this does not pose an error in the study results. Looking at the target groups there is a possibility of error regarding the sample sizes: in manual data validity inspections missing data was found. There were domains missing from their expected owners. This boiled down to two main reasons:

1. Domains were registered to a third party instead of the actual owner. While this is perfectly normal for marketing domains etc., it does have an effect the accuracy of the target group profiling to an unknown extent.
2. The OpenData service provided by Traficom does not iterate all of the domains registered to that particular business ID code. Certain domains are withdrawn from the public dataset due to unknown reasons, and therefore this behavior has an unknown amount of effect to the survey.

Considering the drawbacks stated above, the resulting domain sample sizes per organization and target group were considered of satisfactory size in order to have a general overview of the current situation. There are some domains missing, but this is estimated to be under 5% from the total count of domains. All of the sample sizes are n>10 per organization, so a reliable and average situational awareness picture can be acquired from the data.

7.2 Discussion and setbacks

The chosen research method was an action study. When considering this in hindsight the method was not executed as it was supposed to: as stated before, the action phase did not have as much effort as it should have had due to external factors, namely COVID-19 and the severely increased workload within NCSC-FI. From an academic research method perspective, the study is probably close to a quantitative research with brief touches on literature reviews. This outcome does not affect the plan, however: one cannot foresee possible hurdles altering the ultimate path a study undertakes, and satisfactory results were still delivered to the subscriber of the work.

During the process of the study some critical remarks regarding the subject of email usage, threats and statistics were found. First of all, there are apparently no academically valid studies regarding email usage in Finland. The Statistics Finland (Tilastokeskus) does collect data on internet and online service usage, but there are no specific questions nor data regarding email usage. Different academic sources have studies on both bachelor's and master's levels, but the studies focus on specific organizations and their usage patterns – there seem to be no larger studies regarding the overall usage of email on either corporate or private level.

Second problematic issue was finding comparable data regarding the usage of SPF, DKIM or DMARC. Again, smaller studies focusing on the usage within single entities (usually companies) can be found within Finland. When looking abroad for the situation in different TLDs only rough estimates seem to be the closest matches, with the statistics from SIDN Labs regarding .nl-zone being the exception. Companies operating in the email security field generally publish some statistics, but their

sample sizes vary and are not always even given out with the usage percentages. In some cases, even exact timestamps for when the data is gathered are not found, even on the level of naming the month of the year.

Third problem lies in the subject itself: there is no objective way of measuring en masse how many service providers use DMARC-based filtering in their inbound mail and to what extent (portion of spam scoring or direct qualifier for rejection). Google and Microsoft M365 services represent probably a relatively large portion of organizational mail providers and they both state that SPF, DKIM and DMARC are used in filtering criterion, but there is a significant amount of other mail filtering companies and self-run instances with unknown settings. Even if all the domains in .fi-zone would use these technologies on the mail sending side, it does not make any difference in combating cybercrime if the inbound filters do not use these technologies as filtering criteria.

7.3 Different approaches and further studies

When looking at the original problem of email spoofing and forgery being used as a tool in cybercrime, there are also other options not covered in this work for tackling this issue: Domain Name System Security Extensions (DNSSEC) combined with the usage of DNS-based Authentication of Named Entities (DANE) could also be one approach of validating incoming mail with trusted partners: if a mail relay sending mail regarding a specific DANE-protected domain does not present the correct certificate detailed within given DNS records, the mail session must be terminated and thus forgery is extremely hard. Downside to this approach is the mandatory need for a DNSSEC implementation and verification in both ends of the mail chain.

Another approach is Mail Transfer Agent Strict Transport Security (MTA-STS): the sending party publishes legitimate mail server names in a specific location over HTTPS and a DNS TXT record indicating that the MTA-STS is in action. DANE and MTA-STS are not mutually exclusive, and both can be deployed at the same time. Both DANE and MTA-STS are relatively fresh technologies: RFCs have been published in August 2012 and September 2018, respectively (IETF, 2012; IETF 2018). It may be

that only a small fraction of domains use these technologies – there is no data or research regarding these.

All of the stated problems and alternative solutions pose possible topics for future research. Given that this study represents only one calendar year regarding the status of the .fi-zone and the public sector, repeating this study with the same target groups would produce a longer trendline and situational awareness regarding them. In this study the public sector was chosen, but by using similar methods other target groups can easily be defined and inspected, including domains from other TLDs, different sectors etc. The only issue is to establish the domain ownership for that study group in order to correctly attribute the results.

References

Agari Inc. 2018. DHS Mandates DMARC for Email Security – September 2018 BOD 18-01 Progress Report. Accessed on 20.9.2020. Retrieved from https://www.agari.com/wp-content/uploads/2018/09/Agari_DMARC_Adoption_Report_Federal_2018_v2_online-1.pdf

Agari Data Inc. 2020. H2 2020 Report – Email Fraud & Identity Deception Trends. Accessed on 21.9.2020. Retrieved from https://www.agari.com/cyber-intelligence-research/e-books/agari-h2-2020-email-fraud-report.pdf

Australian Cyber Security Centre. 2020a. How to Combat Fake Emails. Accessed on 15.9.2020. Retrieved from https://www.cyber.gov.au/publications/how-to-combat-fake-emails

Australian Cyber Security Centre. 2020b. Email gateways and servers. Accessed on 15.9.2020. Retrieved from https://www.cyber.gov.au/acsc/view-all-content/guidance/email-gateways-and-servers

Australian Cyber Security Centre. 2020c. The Commonwealth Cyber Security Posture in 2019. Accessed on 21.9.2020. Retrieved from https://www.cyber.gov.au/acsc/view-all-content/reports-and-statistics/commonwealth-cyber-security-posture-2019

Center for Cybersikkerhed (CFCS). 2017. Reducér risikoen for falske mails. Accessed on 20.9.2020. Retrieved from https://fe-ddis.dk/cfcs/publikationer/Documents/Vejledning%20Reducer%20risikoen%20for%20falske%20mails.pdf

Center for Internet Security. 2019. The 20 CIS Controls v7.1. Accessed on 13.9.2020. Retrieved from https://www.cisecurity.org/controls/cis-controls-list/

The Department of Homeland Security. 2017. Binding Operational Directive 18-01: Enhance Email and Web Security. Accessed 15.9.2020. Retrieved from https://cyber.dhs.gov/bod/18-01/

Willis, J. W., Edwards, C. L. & Casamassa, M. 2014. Action research: Models, methods, and examples. Charlotte, North Carolina: Information Age Publishing Inc.

Federal Bureau of Investigation. 2019. I-091019-PSA: Business Email Compromise The $26 Billion Scam. Accessed on 13.9.2020. Retrieved from https://www.ic3.gov/media/2019/190910.aspx

Hassold, C. 2020. Cosmic Lynx: A Russian Threat Hits the BEC Scene. Accessed on 13.9.2020. Retrieved from https://www.agari.com/email-security-blog/cosmic-lynx-russian-bec/

National Cyber Security Centre (UK). 2019a. Email security and anti-spoofing. Accessed 19.9.2020. Retrieved from https://www.ncsc.gov.uk/collection/email-security-and-anti-spoofing/configure-anti-spoofing-controls-

National Cyber Security Centre (UK). 2019b. Email security and anti-spoofing. Accessed on 19.9.2020. Retrieved from https://www.ncsc.gov.uk/collection/email-security-and-anti-spoofing/continuous-improvement

National Institute of Standards and Technology. 2019. Trustworthy Email. Accessed on 19.9.2020. Retrieved from https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-177r1.pdf

Halonen, Antti. 2020. Kalliiksi käyvä ”toimitusjohtajahuijaus” leviää aina kesällä – kohteena erityisesti kesätyöntekijät ja tuuraajat. Accessed on 15.10.2020. Retrieved from https://www.iltalehti.fi/digiuutiset/a/27a057a0-69fb-490b-8a97-80c57ff70014

Internet Engineering Task Force (IETF). 2012 The DNS-Based Authentication of Named Entities (DANE). Accessed on 25.10.2020. Retrieved from https://tools.ietf.org/html/rfc6698

Internet Engineering Task Force (IETF). 2014. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. Accessed on 3.2.2020. Retrieved from https://tools.ietf.org/html/rfc7208

Internet Engineering Task Force (IETF). 2015. Domain-based Message Authentication, Reporting, and Conformance (DMARC). Accessed on 3.2.2020. Retrieved from https://tools.ietf.org/html/rfc7489

Internet Engineering Task Force (IETF). 2015. A "Null MX" No Service Resource Record for Domains That Accept No Mail. Accessed on 3.2.2020. Retrieved from https://tools.ietf.org/html/rfc7505

Internet Engineering Task Force (IETF). 2018 SMTP MTA Strict Transport Security (MTA-STS). Accessed on 25.10.2020. Retrieved from https://tools.ietf.org/html/rfc8461

Kantomaa, Raija. Hakkereiden huijausviestien määrä kasvanut Suomessa koronakeväällä – myynnissä olemattomia maskeja ja rokotteita. Accessed on 15.10.2020. Retrieved from https://www.mtvuutiset.fi/artikkeli/hakkereiden-huijausviestien-maara-kasvanut-suomessa-koronakevaalla-myynnissa-olemattomia-maskeja-ja-rokotteita/7826080#gs.709e4u

Kuntaliitto. 2019. Kaupunkien ja kuntien lukumäärät ja väestötiedot (Amount of cities and municipalities with population details). Accessed on 14.9.2020. Retrieved from https://www.kuntaliitto.fi/tilastot-ja-julkaisut/kaupunkien-ja-kuntien-lukumaarat-ja-vaestotiedot

Laki Liikenne- ja viestintävirastosta (Law on Finnish Transport and Communications Agency) L935/2018. 2018. Accessed on 3.2.2020. Retrieved from https://finlex.fi/fi/laki/ajantasa/2018/20180935

Laki valtioneuvostosta (Law on Finnish Government) 28.2.2003/175. 2003. Accessed on 14.9.2020. Retrieved from https://www.finlex.fi/fi/laki/ajantasa/2003/20030175

Messaging,Malware and Mobile Anti-Abuse Working Group. 2015. M3AAWG Protecting Parked Domains Best Common Practices. Accessed 3.2.2020. Retrieved from https://www.m3aawg.org/sites/default/files/m3aawg_parked_domains_bp-2015-12.pdf

Messaging, Malware and Mobile Anti-Abuse Working Group. 2020. M3AAWG statement on email authentication for COVID-19 mailings. Accessed on 13.9.2020. Retrieved from https://www.m3aawg.org/blog/m3aawg-statement-on-email-authentication-for-covid-19-mailings

Mikrobitti, 2020. Älä lankea juoneen: Netflix-huijaus liikkeellä, näin tunnistat sen. Accessed on 15.10.2020. Retrieved from https://www.mikrobitti.fi/uutiset/ala-lankea-juoneen-netflix-huijaus-liikkeella-nain-tunnistat-sen/44143b82-d3fb-44b3-9e58-2df4b2e7fa93

Network Working Group. 2009. DomainKeys Identified Mail (DKIM) Service Overview. Accessed on 3.2.2020. Retrieved from https://tools.ietf.org/html/rfc5585

Postel, J. 1982. Simple Mail Transfer Protocol. Accessed on 3.2.2020. Retrieved from https://tools.ietf.org/html/rfc821

SIDN. 2018. People should be able to trust e-mail from their local authority. Accessed on 20.9.2020. Retrieved from https://www.sidn.nl/en/news-and-blogs/people-should-be-able-to-trust-e-mail-from-their-local-authority

SIDN Labs. 2020a. Mail – SPF qualifiers. Accessed on 19.9.2020. Retrieved from https://stats.sidnlabs.nl/en/mail.html#spf%20qualifiers

SIDN Labs. 2020b. Mail – DMARC policies. Accessed on 19.9.2020. Retrieved from https://stats.sidnlabs.nl/en/mail.html#dmarc%20policies

SIDN Labs. 2020c. Registration – Registered domain names. Accessed on 19.9.2020. Retrieved from https://stats.sidnlabs.nl/en/registration.html#registered%20domain%20names

Sikkerdigital. 2020. Tekniske minimumskrav for statslige myndigheder. Accessed on 20.9.2020. Retrieved from https://sikkerdigital.dk/myndighed/tekniske-tiltag/tekniske-minimumskrav

Valtiokonttori. 2019. Valtion kirjanpitoyksiköt, virastot ja laitokset sekä talousarvion ulkopuolella olevat valtion rahastot 1.1.2019. Accessed on 14.9.2020. Retrieved from https://www.valtiokonttori.fi/maaraykset-ja-ohjeet/valtion-kirjanpitoyksikot-virastot-ja-laitokset-seka-talousarvion-ulkopuolella-olevat-valtion-rahastot-1-1-2019/

Verizon. 2020. 2019 Data Breach Investigations Report. Accessed on 13.9.2020. Retrieved from https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf

Vernhout, M. 2019. Global DMARC Adoption 2019. Accessed on 3.2.2020. Retrieved from https://s3.amazonaws.com/250ok-wordpress/wp-content/uploads/2019/07/09140509/Global-DMARC-Adoption-2019.pdf

Valimail. 2020. Sender identity movement continues, with over 1 million DMARC-enabled domains. Accessed on 13.9.2020. Retrieved from https://www.valimail.com/resources/email-fraud-landscape-summer-2020/

Appendices

Appendix 1. Interest groups and their business IDs

Business ID	Organization
0244632-1	Liikenne- ja viestintäministeriö
0913655-3	Maa- ja metsätalousministeriö
0245974-7	Oikeusministeriö
0245872-8	Opetus- ja kulttuuriministeriö
0146010-5	Puolustusministeriö
0245992-3	Sisäministeriö
0244685-8	Sosiaali- ja terveysministeriö
2160307-0	Työ- ja elinkeinoministeriö
0245973-9	Ulkoministeriö
0245975-5	Valtioneuvoston kanslia
0245439-9	Valtiovarainministeriö
0519456-1	Ympäristöministeriö
0282400-3	Ahvenanmaan valtionvirasto
0948320-5	Asumisen rahoitus- ja kehittämiskeskus
0245977-1	Eduskunta
2296962-1	ELY-keskusten ja TE-toimistojen kehittämis- ja hallintokeskus
1738354-6	Energiavirasto
1094544-6	Etelä-Suomen aluehallintovirasto
0244680-7	Geologian tutkimuskeskus
2409452-3	Hätäkeskuslaitos
0202419-6	Huoltovarmuusrahasto
0244664-7	Ilmatieteen laitos
0512696-4	Innovaatiorahoituskeskus Business Finland
0245885-9	Kansallisarkisto
2502067-3	Kilpailu- ja kuluttajavirasto
0921536-6	Lääkealan turvallisuus- ja kehittämiskeskus
2924753-3	Liikenne- ja viestintävirasto
0244629-2	Luonnonvarakeskus
1019953-5	Maahanmuuttovirasto
0245954-4	Maanmittauslaitos
0994911-6	Maatalouden interventiorahasto (MIRA)
1106498-2	Maatilatalouden kehittämisrahasto (MAKERA)
0292559-2	Museovirasto
2722042-5	Oikeusrekisterikeskus
2769790-1	Opetushallitus
0244683-1	Patentti- ja rekisterihallitus
0909306-3	Pelastusopisto
2288666-6	Poliisihallitus
0988874-7	Puolustushallinnon rakennuslaitos
0952029-9	Puolustusvoimat
2683902-3	Rahoitusvakausrahasto

Business ID	Organization
2683902-3	Rahoitusvakausvirasto
0246003-5	Rajavartiolaitos
0312081-7	Rikosseuraamuslaitos
2911686-7	Ruokavirasto
0245869-9	Säteilyturvakeskus
1567057-6	Sosiaali- ja terveysalan lupa- ja valvontavirasto
2841366-7	Sosiaaliturva-asioiden muutoksenhakulautakunta
0245435-6	Suojelupoliisi
0245893-9	Suomen Akatemia
0996189-5	Suomen ympäristökeskus
0245880-8	Suomenlinnan hoitokunta
2722043-3	Syyttäjälaitos
0245979-8	Tasavallan presidentin kanslia
2229500-6	Terveyden ja hyvinvoinnin laitos
0245491-1	Tilastokeskus
0245442-8	Tulli
1021277-9	Turvallisuus- ja kemikaalivirasto
2062721-4	Ulkopoliittinen instituutti
2302414-4	Ulosottolaitos
0245437-2	Väestörekisterikeskus
0245440-1	Valtiokonttori
0948320-5	Valtion asuntorahasto (VAR)
1583293-4	Valtion eläkerahasto (VER)
0809880-7	Valtion taloudellinen tutkimuskeskus
2272612-8	Valtion talous- ja henkilöstöhallinnon palvelukeskus
2574261-7	Valtion tieto- ja viestintätekniikkakeskus Valtori
0702479-3	Valtion ydinjätehuoltorahasto
0245456-7	Valtiontalouden tarkastusvirasto
1010547-1	Väylävirasto
0245458-3	Verohallinto
0208591-7	Äänekoski
0177736-4	Ähtäri
0194529-2	Akaa
0177619-3	Alajärvi
0177736-4	Alavus
0101263-6	Espoo
0145626-1	Forssa
0209756-3	Haapajärvi
0184872-4	Haapavesi
0146921-4	Hämeenlinna
0242496-6	Hamina
0103166-9	Hanko
0132585-1	Harjavalta
1068892-9	Heinola
0201256-6	Helsinki
0203762-4	Huittinen

Business ID	Organization
0125866-0	Hyvinkää
9086071-6	Iisalmi
0203797-4	Ikaalinen
0159216-7	Imatra
0175622-1	Jämsä
0126541-4	Järvenpää
0242746-2	Joensuu
0174666-4	Jyväskylä
0133226-9	Kaarina
0214958-9	Kajaani
0185924-7	Kalajoki
0147907-6	Kangasala
0133596-1	Kankaanpää
0178455-6	Kannus
0127046-7	Karkkila
0208787-5	Kaskinen
0178718-3	Kauhajoki
0208852-8	Kauhava
0203026-2	Kauniainen
0210427-6	Kemi
0191717-9	Kemijärvi
0127485-5	Kerava
0208388-2	Keuruu
0168900-6	Kitee
0170843-0	Kiuruvesi
0203925-9	Kokemäki
0179377-8	Kokkola
0160225-7	Kotka
0161075-9	Kouvola
0216509-5	Kristiinankaupunki
0186204-0	Kuhmo
0171450-7	Kuopio
0209046-8	Kurikka
0186418-5	Kuusamo
0149669-3	Lahti
0134480-9	Laitila
0162193-3	Lappeenranta
0209113-7	Lapua
0169321-6	Lieksa
1068322-0	Lohja
1927453-8	Loimaa
0203263-9	Loviisa
0205071-4	Maarianhamina
0157867-2	Mänttä-Vilppula
0165116-3	Mikkeli
0135457-2	Naantali
0181367-9	Närpiö

Business ID	Organization
0186757-0	Nivala
0112038-9	Nokia
0207669-0	Nurmes
0129920-0	Orimattila
0151789-6	Orivesi
0186852-2	Oulainen
0245895-5	Oulu
0215254-2	Outokumpu
0136169-2	Paimio
0136082-5	Parainen
0136311-0	Parkano
2048903-4	Pieksämäki
0209242-0	Pietarsaari
0137323-9	Pori
1061512-1	Porvoo
0188962-2	Pudasjärvi
0210261-7	Pyhäjärvi
1791817-6	Raahe
0131297-0	Raasepori
0204428-5	Raisio
0138780-9	Rauma
0152563-4	Riihimäki
1978283-1	Rovaniemi
0176975-1	Saarijärvi
0139533-1	Salo
0144411-3	Sastamala
0166906-4	Savonlinna
0182381-8	Seinäjoki
0153082-0	Somero
0208061-4	Suonenjoki
0211675-2	Tampere
0193524-6	Tornio
0204819-8	Turku
0204910-7	Ulvila
0183077-8	Uusikaarlepyy
0144036-6	Uusikaupunki
0209602-6	Vaasa
0157568-2	Valkeakoski
0124610-9	Vantaa
0173416-1	Varkaus
0208573-0	Viitasaari
0206333-9	Virrat
0190557-3	Ylivieska
0158221-7	Ylöjärvi
0184674-5	Alavieska
0145208-4	Asikkala

Business ID	Organization
9000162-0	Askola
0132103-3	Aura
0967337-5	Brandö
0280703-5	Eckerö
0163687-9	Enonkoski
0190662-1	Enontekiö
0132239-4	Eura
0132322-3	Eurajoki
0177804-1	Evijärvi
0205003-6	Finström
0282394-0	Föglö
0205012-4	Geta
0184918-8	Hailuoto
0177826-0	Halsua
0132947-3	Hämeenkyrö
0205014-0	Hammarland
0174035-0	Hankasalmi
0163734-5	Hartola
0145801-3	Hattula
0145997-2	Hausjärvi
0164308-3	Heinävesi
0164384-1	Hirvensalmi
0146248-5	Hollola
0132697-7	Honkajoki
0146556-0	Humppila
0185075-2	Hyrynsalmi
2054621-1	Ii
0158766-7	Iitti
0178008-8	Ilmajoki
0167589-4	Ilomantsi
0190758-7	Inari
0126293-4	Inkoo
0178071-5	Isojoki
0178131-2	Isokyrö
1872300-4	Ivalo
0133127-4	Jämijärvi
0147510-4	Janakkala
0147645-7	Jokioinen
0205023-9	Jomala
0207112-8	Joroinen
0174108-9	Joutsa
0168900-6	Juuka
0147705-4	Juupajoki
0164551-3	Juva
0170664-6	Kaavi
0164690-3	Kangasniemi
0175798-8	Kannonkoski

Business ID	Organization
0178498-6	Karijoki
0148268-9	Kärkölä
0186511-0	Kärsämäki
9094917-1	Karstula
0133735-0	Karvia
0178981-6	Kaustinen
0170773-7	Keitele
0210469-8	Keminmaa
0133833-7	Kemiönsaari
0186002-9	Kempele
0133862-8	Kihniö
0242816-6	Kinnula
0203107-0	Kirkkonummi
0191406-6	Kittilä
0176150-6	Kivijärvi
0205032-7	Kökar
0191528-8	Kolari
0176227-7	Konnevesi
0169048-8	Kontiolahti
0179699-5	Korsnäs
0213007-9	Koski
0180065-9	Kronoby
0176357-9	Kuhmoinen
0205030-0	Kumlinge
0180117-6	Kuortane
0134349-4	Kustavi
0176410-9	Kyyjärvi
0180451-0	Laihia
0203135-3	Lapinjärvi
0172127-2	Lapinlahti
0180516-9	Lappajärvi
0180857-0	Larsmo
0176478-2	Laukaa
0162576-6	Lemi
0205034-3	Lemland
0150783-1	Lempäälä
0172231-2	Leppävirta
0180774-6	Lestijärvi
0134698-6	Lieto
0186553-2	Liminka
0169583-6	Liperi
0150919-1	Loppi
0176592-9	Luhanka
0186580-7	Lumijoki
0205038-6	Lumparland
0162631-2	Luumäki
0180948-5	Maalahti

Business ID	Organization
0129261-5	Mäntsälä
0165761-0	Mäntyharju
0135086-2	Marttila
0204064-7	Masku
0186588-2	Merijärvi
0135202-4	Merikarvia
0162675-0	Miehikkälä
0186646-3	Muhos
0208471-1	Multia
0191824-3	Muonio
0181101-6	Mustasaari
0176699-9	Muurame
2048364-4	Mynämäki
0203282-3	Myrskylä
0135662-3	Nakkila
0135821-5	Nousiainen
9014643-2	Nurmijärvi
0135869-6	Oripää
0151924-2	Padasjoki
2050961-3	Pälkäne
0188808-0	Paltamo
1913642-6	Parikkala
0198517-1	Pedersöre
0191866-5	Pelkosenniemi
0193729-2	Pello
0116845-4	Perho
0165867-2	Pertunmaa
0176769-2	Petäjävesi
0172446-5	Pielavesi
0243027-4	Pihtipudas
0152084-1	Pirkkala
0169823-6	Polvijärvi
0136610-0	Pomarkku
0130095-3	Pornainen
0191908-6	Posio
1929519-5	Pöytyä
0130729-0	Pukkila
0138037-5	Punkalaidun
0189081-8	Puolanka
0166400-1	Puumala
0189127-1	Pyhäjoki
0189226-6	Pyhäntä
0204403-1	Pyhäranta
0162798-0	Pyhtää
0169967-7	Rääkkylä
0166507-1	Rantasalmi
0191974-8	Ranua

Business ID	Organization
0172586-3	Rautalampi
0172646-0	Rautavaara
0206951-1	Rautjärvi
0189548-3	Reisjärvi
0189576-6	Ristijärvi
0163013-5	Ruokolahti
0152842-1	Ruovesi
0204524-5	Rusko
0139937-5	Säkylä
0192936-4	Salla
0205119-4	Saltvik
9038213-6	Sauvo
0163109-0	Savitaipale
0210704-7	Savukoski
0189615-2	Sievi
0139842-8	Siikainen
2047359-3	Siikajoki
0189019-9	Siikalatva
0172718-0	Siilinjärvi
0193015-4	Simo
0203533-8	Sipoo
0131156-4	Siuntio
0193169-1	Sodankylä
0177736-4	Soini
9090160-2	Sonkajärvi
0189766-5	Sotkamo
0205121-5	Sottunga
0167265-0	Sulkava
0921395-0	Sund
0189925-7	Suomussalmi
0167352-2	Sysmä
0163320-5	Taipalsaari
0190100-3	Taivalkoski
0139991-4	Taivassalo
0153179-4	Tammela
0173081-4	Tervo
0193249-1	Tervola
0182734-1	Teuva
0170026-0	Tohmajärvi
0182779-8	Toholampi
0177201-0	Toivakka
0173128-6	Tuusniemi
0131661-3	Tuusula
0190140-9	Tyrnävä
0157323-0	Urjala
0190224-1	Utajärvi
9129466-4	Utsjoki

Business ID	Organization
0177224-8	Uurainen
0190027-0	Vaala
0207709-5	Valtimo
0205126-6	Vårdö
0144561-8	Vehmaa
0173787-2	Vesanto
0157711-9	Vesilahti
0184278-7	Veteli
0173835-7	Vieremä
0131905-6	Vihti
0184318-1	Vimpeli
0207033-6	Virolahti
2050514-5	Vöyri
0210826-9	Ylitornio
0158301-7	Ypäjä

Appendix 2. Full dataset and target group dataset

The full gathered dataset file can be found at https://github.com/vkontinen/vkontinen.github.io/blob/main/Attachment_1_Full_dataset.xlsx

The dataset gathered regarding the target groups can be found at https://github.com/vkontinen/vkontinen.github.io/blob/main/Attachment_2_Target_group_data.xlsx

Appendix 3. Complete collected trendlines for .fi-zone