National College of Ireland

Email spoofing defence techniques: a comprehensive review and development of a novel measurement tool

MSc Research Project
Cybersecurity

Marcello D'Angelone
Student ID: x21113777

School of Computing
National College of Ireland

Supervisor: Mark Monaghan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Marcello D'Angelone
Student ID:	X21113777
Programme:	MSc Cybersecurity **Year:** 2022/23
Module:	Research Project
Supervisor:	Mark Monaghan
Submission Due Date:	Aug. 14 2022
Project Title:	Email spoofing defence techniques: a comprehensive review and development of a novel measurement tool
Word Count:	12193 **Page Count** 36

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:

Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Email spoofing defence techniques: a comprehensive review and development of a novel measurement tool

Marcello D'Angelone
Student ID: x21113777

Abstract

In the past year, organisations have faced tenfold losses compared to the previous five years due to Business Email Compromise (BEC). This phishing attack leverages social engineering techniques by spoofing the email sender's address to deceive the victims into making fraudulent financial payments. Since early 2000, the Internet Engineering Task Force proposed email anti-spoofing protocols, such as SPF and DMARC, to mitigate this cyber threat. Many researchers have tried to understand their limitations and adoption rate. Although it has been three years since the last survey, they all reported an overall low adoption rate; However, there are inconsistencies in the methodology and datasets used to perform such measurements, which may misrepresent the real adoption rate. This research proposes a novel domain crawler tool which provides detailed statistics about SPF and DMARC deployment. The tool has been tested with over 1.4 Million unique domains collected from seven different datasets. This large-scale empirical analysis demonstrated that only 29.58% of the 20,349 US governmental domains comply with the Department of Homeland Security directive, which mandates a more restrictive DMARC policy. By performing a statistical hypothesis, it has also been demonstrated that there is a significant increase in the SPF and DMARC adoption rate compared to previous measurements of the Alexa Top 1 Million dataset, with a 59.6% and 25.6% respectively, and a dramatic reduction in misconfigured domains. Furthermore, the Tranco dataset is included through its Python package to provide security researchers with a more research-oriented domain crawler tool. The objective is to lay the foundations to understand trends better and provide recommendations based on more scientific and reproducible measurements.

1 Introduction

Email spoofing is one of the main phishing techniques attackers leverage to induce the victims to open or click a malicious email. These attacks are constantly changing and increasingly sophisticated, with devastating consequences for organisations of all sizes. Despite the advancement in phishing detection technology and user awareness training, email phishing remains one of the most common attack vectors to steal credentials, deliver ransomware, compromise data and ultimately for illicit financial gains. According to the Anti-Phishing Working Group (APWG), phishing attacks have increased by 209% over the past ten years (Dalvi et al., 2020). Most high-profile data breaches originated from phishing emails; however, the **Business Email Compromise** (BEC) represents one of the most lucrative scams. In this case, the victim receives **a spoofed email** which appears to be sent by a legitimate or trusted

sender who requests payment for an invoice or transfers a certain amount of money to an account. The email typically contains phishing links, malicious attachments or details for unlawful payment to the attacker's bank account. Even companies like Google and Facebook were hit by this scam. However, one of the most severe cases occurred in 2019, when Toyota Corporation disclosed that they were the victim of fraudulent payment under the direction of a malicious third party which caused a loss of $37.3 Million. The attackers compromised and monitored a Toyota subsidiary's email activities and learned about internal processes and procedures until, on August 14th, they impersonated a legitimate business partner submitting a payment invoice via email. The request created a sense of urgency, claiming that any further delay in transferring the money could have resulted in a slowdown of Toyota production. The accounting department employee fell victim to a typical BEC attack, which induced him to bypass the required approvals to transfer a large amount of money and wire it into the attacker's bank account. When Toyota's upper management became aware of the fraud, they reported the monetary loss to local authorities, investigated the incident and hired a legal professional team to recover the sum. In addition to the substantial financial loss and reputational damage, Toyota had to adjust the earnings forecast for 2020, which negatively affected its stock price (Lindsey, 2019). The FBI's Internet Crime Complaint Center (IC3) received 21,832 BEC complaints amounting to over 2,7 billion in losses in 2022 (FBI Internet Crime Report Center, 2022). This is an increase of more than ten times compared to the losses sustained by businesses over the past five years. Furthermore, recent Email Security reports (ProofPoint, 2023; Armorblox, 2023) reported that BEC-related attacks increased by 22% over 2022, almost half bypassing traditional email security filters. This represents one of the top security concerns for 41% of Board Members, especially with new tools such as ChatGPT, the number of BEC emails is expected to increase dramatically from 2023 onward.

1.1 Research background and objectives

Since early 2000, many efforts have been made toward implementing email anti-spoof protocols such as **SPF, DKIM** and **DMARC** (Hu et al., 2018). However, after more than 20 years, their adoption appears to be still low. According to a recent large-scale analysis, 50.3% of domains have an SPF record, and only 11.5% of them have a DMARC record. Some qualitative surveys reported that email administrators believe they are ineffective and the cost overweight the benefits (Maroofi et al., 2021), while others highlight the challenges in their implementation, which lead to misconfigurations (Hu et al., 2018). Their common concern is to see spoofed emails bypassing security controls but also legitimate emails being blocked. The first step to understanding the reasons behind such mixed responses is to measure their adoption regularly to understand how they can be improved by rectifying misconfigurations. To the best of my knowledge, there is no consistent way to measure the deployment of such protocols, and no automation tool specifically addresses these challenges. This research provides a comprehensive review of email anti-spoofing protocols and their adoption by developing a novel tool which gathers SPF and DMARC details. The results are compared and contrasted with previous measurements to determine future trends. The severity of cybercrime's impact originated from a spoofed email, and the limitations of the technology standard developed to mitigate them require a comprehensive analysis of the phenomenon and the proposal of new tools.

2 Related Work

The Simple Mail Transfer Protocol (SMTP) was developed in 1982 without any built-in security features to prevent attackers from impersonating an email sender. It has been estimated that 6.4 billion spoofed emails are circulating daily (Tatang et al., 2021). According to Shen et al. (2021), four different email communication stages must be secured in order to validate its authenticity. These are sender authentication, receiver verification, forwarder verification and UI rendering. The following paragraphs focus on receiver verification and some aspects of forwarder verification and UI rendering.

2.1 Email communication weak points and mitigations

Figure 1 represents the main sections of email communication, their weak points and how Email filtering solutions mitigate those weaknesses.

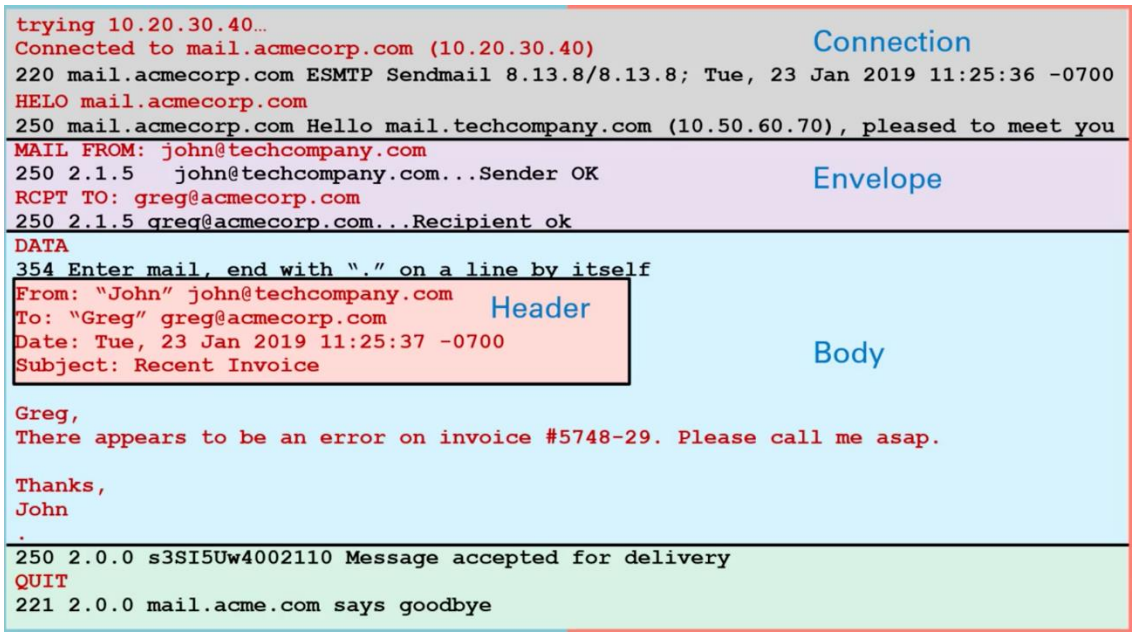

Figure 1: Email sections

1. Connection

During the first part of the SMTP connection, information that is typically used for filtering messages or even rejecting the connection is exchanged. Some examples include:

- Sender HELO domain
- Sender IP address
- Sender Hostname

Online reputational tools such as Spamhaus[1] build their block list from this data type providing the first line of defence.

[1] https://www.spamhaus.org/

2. Envelope

During the envelope part of the SMTP conversation, information that can be used to configure conditions for filtering messages is exchanged. Also, envelope information can be used to block the conversation and reject the connection. A couple of examples of envelope information include:

- Envelope Sender (email address)
- Envelope Recipient (email address)

3. Body

After the DATA command is received and accepted, the sending Mail Transfer Agent (MTA) sends the body of the message, which includes the header and any attachments. Attributes of the message body can be used to filter messages, but the connection cannot be rejected after the Protection Server (receiving MTA) accepts the DATA command. Examples of body information include:

- Message text contents
- Message size (in bytes)
- Attachment attributes

4. Header

The message header is a subset of the message body. Attributes of the message header can also be used to filter messages. Examples of message header information include:

- From: (visible to the recipient as the sender of the email)
- To: (visible to the recipient as the recipient of the email)
- Date:
- Subject: (visible to the recipient of the email)
- Reply-To: (visible to the recipient when he selects “reply”)

Threat actors usually spoof the header information, such as the "From:" and "Reply-To:" information, so it does not match the envelope or connection information. They are doing so because these are the most recognisable parts of the email message visible to the recipients and can be used to make them believe the email is coming from a trustworthy source.

2.2 Phishing mitigations evolution

Most research papers initially focused on developing algorithms that effectively detected the Phishing link's characteristics in the message's body. They did not combine features of the medium delivering the link, in other words, the email itself. As noted by Smadi et al. (2015), however, the average lifespan of a phishing website is only 2.25 days. According to Muneer et al. (2021), only a few papers published between 2015 and 2019 focused on investigating email features, while the rest still focus on Web and URL detection techniques typically present in the body of the message. More recent studies (Smadi et al., 2015; Li et al., 2022) propose filtering suspicious emails, leveraging data mining techniques, Bayesian Filters and Natural Language processing classifications. As Smadi et al. (2015) pointed out, more features can be extracted from the email alone compared to those that can be extracted from the URLs only, increasing the chance of detecting a suspicious attribute. Their algorithm extracts features from email headers and content, increasing the accuracy rate to 98.87%. However, because of the high computational costs of such filters, blocking suspicious emails by inspecting their origin

IP addresses, domains, or email servers is more efficient than processing their content. While significant progress has been made in detecting traditional phishing emails, the security industry still faces significant challenges in mitigating a particular type of email threat called Business Email Compromise.

2.3 Business Email Compromise challenges

Cidon et al. (2019) highlight that most email security systems do not block Business Email Compromise threats because they inspect two main attributes: malicious origin/content or volumetric. Since BEC emails are usually tailored to deceive targeted recipients in an organisation, they evade volumetric defences. Furthermore, it has been reported that 60% of them do not even contain a link or attachment, which a reputational rule set can easily block. The nature of BEC emails relies on plain text communication, which leverages social engineering techniques to transmit a sense of urgency or authority to deceive the victim. Because of the human nature of this phenomenon, it is essential to understand what makes some people more inclined to phishing attacks than others. As Alkhalil et al. (2021) reported, many studies are devoted to identifying more vulnerable people's personal and contextual attributes. Some point out the lack of technical knowledge, inexperience, or age group as the main reason some victims are more likely to fall than others. However, attackers prefer to exploit human "psychological triggers" since they surpass conscious decisions. Triggers such as a sense of urgency, reward/recognition, and authority are the most common and effective (T N et al., 2021), especially when they impersonate familiar senders by spoofing high-profile domains. In summary, BEC emails represent one of the more complex technical and social engineering threats organisations face. The following sections cover the solutions proposed by the Internet Engineering Task Force to hamper attackers' ability to send spoofed emails and their limitations.

2.4 Sender Policy Framework

The Sender Policy Framework (SPF), which was proposed in early 2000 and standardised in 2014 by the Internet Engineering Task Force (IETF) (Hu et al., 2018), is considered the first attempt to prevent attackers from sending emails from a domain they don't own. The SPF requires a DNS TXT record which specifies what IP address is authorised to send emails from the email sender's domain. As defined by RFC7208 (Kitterman, 2014), an SPF record may contain the following tags:

- **v** indicates the version. (This is mandatory).
- **ipv4** or **ipv6** indicates version 4 or version 6 of the IP addresses authorised to send emails.
- A tag which indicates the behaviour of the policy:
    - **+all** pass or accept all messages from any email server (*Pass*)
    - **-all** drop the message if the IP is not authorised (*Hard Fail*)
    - **~all** accept and tag the email (*Soft Fail*)
    - **?all** neither pass nor fails, likely to be accepted (*Neutral*)
- **a** is used to perform a DNS lookup for an A record
- **mx** is used to perform a DNS lookup for an MX record

- **include** tag is used to allow third-party to send emails on behalf of the domain in question
- **redirect** is used to delegate the SPF policy to another domain which has been configured with the relevant SPF tags

Typical SPF records may look like the following ones:

- v=spf1 ip4:192.0.2.0/24 +all
- v=spf1 redirect=_spf.example.com

As per RFC7208, an SPF record must end with the all or redirect tag to be valid. The SPF specification also permits using macros, or variables, which the MTA evaluates. This overcomes the limitations of the number of DNS lookups allowed and enables a more dynamic policy processing. In 2021 two critical vulnerabilities were discovered in the SPF library libspf2, which allowed an attacker to perform a Denial-of-Service attack and remote code execution to an email server by crafting a message that leads the DNS to execute the payload. According to Bennett et al. (2022), 80% of MTAs are still vulnerable even after private notifications and public disclosure.

2.5 Domain Keys Identified Mail

The Domain Keys Identified Mail (DKIM), drafted in 2004 and standardised in 2011, tried to address the same threats by leveraging public key cryptography to authenticate and validate the integrity of an email. Similarly to SPF, a DNS TXT record (_domainkey) is used to publish the DKIM public key. The corresponding private key is required on the sender's email server to sign each email's header and body digitally. The former signature is optional, whereas the latter is mandatory. The recipient email server is then querying the DKIM signature to obtain the signer's public key from the domain "d" field and validate the integrity of the message.
As specified by the RFC6376 (Crocker et al., 2011), a DKIM record may contain the following tags:

- **v** indicates the protocol version
- **a** indicates the algorithm used for the signature
- **h** lists colon-separated headers fields
- **b** the hash value of the message headers listed in the **h** tag
- **bh** the hash value of the message body
- **d** represents the domain of the signing entity
- **s** specifies the selector being used

The selector is necessary as DKIM supports multiple key pairs to handle large-scale email operations and quickly identify which system sent an email. Durumeric et al. (2015) pointed out that this protocol is a step forward compared to SPF since it validates the message's integrity and authenticity. However, there is no automated enforcement for emails with invalid or missing cryptographic signatures. Hence there must be an agreement between the sender and receiver in advance. Furthermore, measuring how many domains have deployed DKIM correctly requires obtaining the selector tag in the email header. For this reason, only a few studies attempted to measure its adoption rate with approximate results (Tatang et al., 2021).

2.6 Domain-based Message Authentication, Reporting and Conformance

Finally, the Domain-based Message Authentication, Reporting and Conformance (DMARC) was drafted in 2011 and standardised in 2015 to overcome the limitations of both SPF and DKIM by verifying that the displayed header "from" is from the domain validated by those two protocols. It is not a stand-alone protocol, as it requires both SPF and DKIM to work; however, it completes them by enforcing an email delivery policy to apply in case of a mismatch between the "from" and "replies to" addresses (Hu et al., 2018). Furthermore, as the name implies provides extensive aggregate and forensic reporting capabilities to obtain visibility of the email flow and to monitor delivery issues. Once again, another DNS TXT record is used to establish the policy and alignment. As defined by RFC7489 (Kucherawy and Zwicky, 2015), a DMARC record contains at least two mandatory tags:

- **v** indicates the version of the protocol
- **p** tag indicates the policy action, which can be equal to *none*, *quarantine*, or *reject*.
    - **none:** the email is delivered regardless of SPF/DKIM failure or validation.
    - **quarantine:** the email is flagged as suspicious and delivered to the spam folder if either SPF or DKIM fail.
    - **reject:** the email is not delivered and discarded if SPF or DKIM fails.

Non-mandatory tags include:

- **sp** has the same syntax as p but applies to subdomains
- **aspf** indicates the alignment mode for SPF, which can be relaxed (**r**) or strict (**s**)
- **fo** specifies the behaviour of the failure reports
- **rua** specifies the email address to which the aggregate report must be sent
- **ruf** specifies the email address to which failure reports must be sent

A valid DMARC record may look like the example below (RFC7489):

- v=DMARC1; p=reject; aspf=r; rua=mailto:dmarc-report@example.com

Many commercial and open-source tools, such as DMARCBox, assist organisations in generating analytical reports and statistics about emails sent from their domain or subdomain (Nanaware et al., 2019). This is a critical aspect of ensuring legitimate Email Servers have been authorised to send Emails on behalf of the domain in question and reduce the number of False Positives. It is common practice to deploy DMARC with a policy of *p=none* to gather intel about the email servers sending on behalf of the domain, and once all email legitimate email servers have been onboarded, move the domain to a more restrictive policy of *quarantine* or *reject*.

The consensus is that spoofing a trustworthy website or domain remains the most common technique to lower the defences and lead the victim to the trap. This is where the implementation of anti-spoofing technology plays a crucial role by flagging suspicious emails and alerting the user or rejecting them together so they don't even reach the user's inbox. The vast literature on detecting spoofed emails can be grouped into two main branches. One focuses on the analysis of the implementation and challenges with the adoption of anti-spoofing protocols. The other focuses on the limitations and technical gaps of such protocols proposing improved detection techniques.

2.7 Anti-spoofing protocols adoption and challenges

Although many recent studies surveyed the adoption of email anti-spoofing protocols, they all differ in the methodology and the number of domains surveyed. It is, therefore, difficult to consistently track their trends and how they are implemented (Hu et al., 2018; Maroofi et al., 2021; Deccio et al., 2021; Shen et al., 2021). Durumeric et al. (2015) propose one of the first studies based on analysing the Alexa Top 1 Million domains list. They performed an MX record lookup to validate if the domain contains an email server, followed by a DNS query to identify SPF and DMARC records. They reported 792,494 domains with an operational email server (80% of the Alexa Top 1 Million), 47% with an SPF policy but only 1% with a DMARC record. 21.7% of the SPF policies implemented *Hard Fail* (-all) 58% *Soft Fail* (~all), whereas the rest with a *Pass* (+all). The DMARC standard was recently updated at the time of the measurement, and almost the entire number of domains with a DMARC record did not have a policy published 98.9%, followed by a 0.82% with a *none* DMARC Policy, 0.22% with *reject* and 0.09% *quarantine*. Contemporary research from Foster et al. (2015) reported fewer SPF records from the same list of Alexa domains (40%). This could be explained because their DNS lookup included domains without an MX record. The measurement was repeated three years later by Hu et al. (2018), and they reported a mild increase in the adoption rate, with 44.9% and 5.1% of the Alexa Top 1 Million domains presenting an SPF and a DMARC record, respectively. They also found a slightly higher percentage, 54.3% SPF and 6.0% DMARC, when performing a domain lookup with an MX record (79% among the Alexa Top 1 Million). A smaller pool of the Top 500 domains from 139 countries was selected by Maroofi et al. (2020) from the Alexa Top 1 Million list. Their findings show 65.9% and 34.3% SPF and DMARC adoption rates, respectively. Comparable results were discovered for 7,022 domains of banking and financial domains. In the following research paper from the same authors (Maroofi et al., 2021), a much larger dataset is considered with a scan of approximately 236 Million domains and a scan of the same datasets investigated the previous year. With such a large-scale analysis, the domains with SPF records drop to 31% with an even lower DMARC adoption rate of 0.13%. However, after one year from their previous findings, they don't report much difference between the Top 500 and the banking and financial domains either. Instead of measuring SPF and DMARC diffusion from DNS lookups, Deccio et al. (2021) propose another large-scale analysis by reviewing the validation behaviour of the receiving MTAs. This experiment occurred during a worldwide mass email notification initiated to disclose a vulnerability on their network. This means they conducted an email notification campaign without sending spam or illegitimate emails to users' inboxes. A total of 26,695 domains were extracted from a set of email addresses whose email messages were received by their respective MTAs. They observed up to 85% of MTAs configured to validate SPF, but only half of them checked SPF, DKIM and DMARC combined. Shen et al. (2021) adopted a similar approach by using only dedicated email accounts owned by themselves, sending a low rate with over 100 minutes of interval between emails to minimise the impact. Their test revealed that only 23 out of 30 major email service providers are checking SPF, DKIM and DMARC combined.

Furthermore, all of them are vulnerable to at least one spoofing attack that bypasses the weakest link in the email authentication chain. Tatang et al. (2021) provide one of the latest measurements of the SPF and DMARC adoption implementing a DNS crawler. According to

them, the SPF protocol has been deployed to 50.7% of the domains, with an increase of 25% since 2015. They also report an increase of Hard Fail policies (-all) compared to the previous statistics and a higher than expected DMARC deployment on 11.5% of domains. However, the vast majority, 75%, is configured with a *none* Policy. Their survey of .gov top-level domains is also worth mentioning as they are restricted to US-based governmental agencies and public sector organisations. The scan took place four years after the Binding Operational Directive 18-01 from the Department of Homeland Security, which mandates SPF and DMARC for all governmental domains and a transition to a DMARC policy of *reject* by October 2018. According to them, this legislation has significantly contributed to increasing their adoption rate to 92% and 88% of all .gov top-level domains. Finally, they also attempt to provide an estimated global number of DKIM deployments: 113,866 (13%). Although they recognise that the measurement is a partial lower bound obtained by analysing DKIM key selector strings in email dumps, they contribute to discovering multiple domains sharing duplicate keys (2,302) and 4,312 domains using cryptographically weak keys with less than 1024 bits. However, their pool of domains is based on a subset of high-profile domains from the Alexa Top 1 Million list, and percentages may decrease with a broader scope of them.

Regarding measuring anti-spoofing protocol adoption, it should be noted that the Alexa list is no longer maintained, and other static lists are compiled with different proprietary domain ranking methods, which are not disclosed and may skew the results. Furthermore, most measurements do not indicate the total number of domains surveyed or the date when the list was retrieved, which hinders the reproducibility of the experiments (Le Pochat et al., 2019). A more controlled domain survey via API or open ranking systems would improve the reliability and consistency of such measurements. Understanding whether Anti-spoof protocols reached the so-called “critical mass” or “network externalities” is crucial. In other words, the tipping point whereby the value of the protocols increases exponentially as more people adopt the same technology. In fact, according to some surveys (Hu et al., 2018), there is a misconception among email administrators that publishing SPF and DMARC records on their DNS would only help other email services to identify spoofed emails sent from their own domain and, therefore, the cost overweight the benefits. It is still imperative to consistently measure their implementation, understand how this can be improved, and publish the results to persuade organizations to adopt them or remediate misconfigurations.

2.8 Technical limitations of anti-spoofing protocols and new approaches

Nowadays, companies are so reliant on email communication that the efficiency of Anti-Phishing systems is not only measured by their ability to detect malicious emails. The equilibrium between the False Positive Rate (FPR) and the False Negative Rate (FNR) is decisive. Both are equally important, as a high FPR or FNR would lead users to lose confidence in the system's accuracy. Less risk-prone organisations would tolerate a higher number of False Positives, whereas organisations more tolerant toward the risk would not accept such inconvenience, and users would be more frustrated and try to circumvent the security control. According to Dalvi et al. (2020), the Company's risk appetite should be used to tip the scale in one direction or another. However, when it comes to detecting spoofed emails, False Positives are a more severe problem than False negatives, as Konno et al. (2020) pointed out. The main reason is that the sender authentication and validation are performed before processing any

content or attachment due to the email communication flow. While other email security filters may block False Negatives, False Positive may be rejected entirely before reaching the user's inbox. This concern is also shared among email administrators in recent surveys (Hu et al., 2018). However, they overlook that DMARC could be implemented with a less restrictive policy of *none* to allow an assessment of the impact before moving to the enforcement and *reject* mode. Finally, Gupta et al. (2014) propose a novel method that does not rely on established anti-spoofing protocols. Their approach combines memory forensic techniques by capturing the browser's processes with DNS lookups to fetch the MX DNS record and match it with the email header extracted from the memory. This approach cannot block spoofed emails before they reach the user's inbox, as the emails must be loaded in memory and subsequently analysed leveraging UI rendering techniques. Although their experiment yielded lower overhead and resource consumption with a minimum false positives rate, they limited the investigation to web-based email clients. Similarly, Opazo et al. (2017) recognise the need for a layered defence approach. They suggest the idea of a sentinel software concept that could assist the user in automatically parsing email headers and the body of the message to highlight suspicious attributes, including SPF soft fail. Therefore the end user should also be provided with a client-side self-defence tool to detect spoofed emails reaching their inboxes. BEC is not only a technical challenge but also a social engineering one. For this reason, the user should be part of the solution as the first line of defence. In other words, any email security solution should be viewed as a holistic system where each component plays a crucial role.

2.9 SPF limitations

With the advent of cloud services, many organisations are outsourcing their email infrastructure, undermining the assumptions on which the SPF was designed. For instance, when an email service is sending emails on behalf of another service, also known as email forwarding, it is another scenario where the SPF protocol fails. As reported by Hu et al. (2018), Chauhan and Shah (2023), and Liu et al. (2023), the IP address of the original sender does not correspond with the IP address of the email server forwarding the message, furthermore the variety of methodologies on how to implement email forwarding, present significant challenges with the compatibility of anti-spoofing protocols in general and to find a solution that accommodates every scenario. The Authenticated Received Chain (ARC) was recently introduced in 2019 to overcome such limitations and maintain authentication in transit. Although email providers like Gmail and Outlook have already adopted this new protocol, little is known about its diffusion (Wang and Wang, 2022). However, it should be further investigated in separate studies.

2.10 DKIM limitations

On the other hand, Chen et al. (2020) focused on DKIM vulnerabilities, particularly signature replay attacks. These are more likely to occur when only the body is signed while the header is not so that the attacker appends additional email headers or body contents. The mitigation proposed consists of signing all the 19 headers available using the "h=" tag and avoiding using the optional "l=" tag, which indicates the length of the email body and allows it to append malicious content in the body of the message without breaking the digital signature. Also, Yu et al. (2022) reproduced this vulnerability in a lab environment by appending multiple display

names in the “From” header. Additionally, they performed two new types of attacks: the “negligible character attack” and the “special character hidden attack”. These consist of passing a special character such as “#” or “;” in the header “From” and bypassing the email verification process. Their proposed solution combines anti-spoofing protocols, such as DMARC, with email content filtering, inspecting the header for anomalies and alerting the user in the email client.

2.11 DMARC Limitations

As the DMARC protocol specification acknowledges, one of its main limitations consists in solving only for exact-domain spoofing. None of the authentication protocols currently available are solving for the proliferation of “cousin domain”. For instance, an attacker can register a similar well-known domain publishing SPF, DKIM and DMARC records, and the email will always be delivered as legitimate (Chauhan and Shah, 2023). This attack exploits the human mind's ability to fill the gaps when reading something with what is expected. An example of a cousin domain was used to compromise the insurance company Wellpoint creating a domain which replaced the “lls” with ones obtaining we11point.com (Zager, 2017). Tools such as dnstwist[2] may assist organisations in identifying all possible cousin domain permutations and finding which one is already registered. This tool could be used to detect phishing campaigns in advance and “defensively register domains”. Similarly to the cousin domain, attackers may also leverage subdomains to deceive the victims. Maroofi, Korczynski and Duda (2020) highlight the weaknesses of DMARC in subdomains, revealing how it is possible to send spoofed emails from existing subdomains without a specific configuration of the sp record (Subdomain Policy) within the DMARC record. The same result is likely for non-existing subdomains when the DMARC record does not contain a wildcard to cover non-existing subdomains. To evaluate such risks, they propose a new method that generates a list of subdomains and measures their implementation of email anti-spoofing protocols. As expected, their finding confirms that the situation for subdomains is even worse, with 70% of them without any SPF record or DMARC policy. To remediate this weakness, they propose a “defensive domain registration”, which consists of enforcing SPF and DMARC policies on subdomains. In the subsequent paper, Maroofi et al. (2021) expand their original findings by reviewing misconfigured SPF records to notify the domain owners. Furthermore, they present novel contributions such as large-scale measurement of SPF and DMARC adoption, a methodology to analyse DNS logs to prevent domain spoofing and perform a proof-of-concept subdomain spoofing attack of a high-profile domain.

3 Research Methodology

The research methodology consists of collecting the datasets, processing the data, extracting the relevant features from the SPF and DMARC records and automatically measuring the results.

[2] Dnstwist: https://github.com/elceef/dnstwist

3.1 Dataset collection

The proper dataset selection is crucial for this research type and comprises multiple domain lists. This ensures maximum coverage of domain classification methods and expands the scope of the analysis. The Alexa Top 1 Million domain list[3] is the most referenced in similar studies (Durumeric et al., 2015; Foster et al., 2015; Hu et al., 2018; Maroofi et al., 2020; Maroofi et al., 2021; Tatang et al., 2021; Deccio et al., 2021; Liu et al., 2023). According to Le Pochat et al. (2019), this list is used by 133 top-tier papers about internet security measurements. For this reason, it has been used as the benchmark to compare the SPF and DMARC protocol adoption rates. The Alexa list contains the most popular domains ranked by the number of visitors per day. The service, which is no longer updated since February 2023, was initially intended for marketing purposes until Amazon acquired it. The second dataset is Moz's list of the 500 most popular domains based on Domain Authority (DA) which is part of their search optimization service (SEO). The DA is based on a machine learning algorithm, which calculates how often a domain is referenced in a Google search[4]. The fourth dataset has been compiled by combining all the .gov and .mil top-level domains from the DomCop dataset with all non-.gov and non-.mil domains registered by governmental agencies in the US[5]. DomCop list contains 10 Million domains compiled using web crawled data and it is available through the Open PageRank initiative[6]. This combined dataset has been selected to understand the impact of the binding Operational Directive 18-01 on a broader range of governmental domains managed by US federal agencies to expand and validate the research carried out by Tatang et al. (2021) and Zager (2017), which focused only on .gov domains. Finally, the dataset provided via API by Similarweb ensured a more up-to-date list of the top 5000 most popular domains[7]. This data is collected from various sources such as Google Analytics and other anonymous traffic data. Le Pochat et al. (2019) argue that although several recent studies base their experiment on commercial domain ranking systems, they lack transparency and, sometimes, contain malicious domains, resulting in biased and skewed research results. They provide a list of 1 Million domains[8] aiming at a more scientific and transparent approach. Their list is compiled by combining different sources and applying the Dowdall rule to score each domain. According to them, this is a more suitable and reliable source to conduct such studies and encourage the scientific community to reproduce the results. For this reason, their dataset is included in this research and implemented through their Python library in the domain crawler tool. Table 1 summarises the datasets analysed in this survey.

[3] Alexa Internet: http://s3.amazonaws.com/alexa-static/top-1m.csv.zip
[4] Moz: https://moz.com/top500
[5] DomCop: https://www.domcop.com/top-10-million-domains
[6] U.S. General Services Administration: https://github.com/GSA/govt-urls/blob/main/1_govt_urls_full.csv
[7] Similarweb https://developers.similarweb.com/docs/digital-rank-api
[8] Tranco: https://tranco-list.eu/

Table 1: Domain Datasets

Dataset	Date collected	Number of records
Alexa Top 1 Million	February 2023	864,552
Moz	July 2023	500
DomCop .gov domains	June 2023	9,914
DomCop .mil domains	June 2023	1,210
U.S. General Services Administration	January 2023	9,225
Similarweb	June 2023	5,000
Tranco	June 2023	1,000,000

The total number of unique domains in all the datasets combined is 1,433,982. The format of the datasets containing the domains is CSV file, while the Similarweb API output is in JSON format.

3.2 Data processing

Data processing consists of two steps. First, the domain crawler tool imports the domain list either from the supplied CSV files, the Similarweb API or the Tranco library. Second, a DNS lookup retrieves each domain's relevant SPF and DMARC records. Specifically, it validates if the SPF and DMARC records are present in the domain in question if there are no misconfigurations, and what are the respective policies.

3.3 Data extraction

The output generated by the domain query is filtered, and only the SPF and DMARC policies for each domain are extracted and compiled in a separate CSV file. The only relevant information about the SPF record is the policy tag (-all, +all, ~all, ?all), whether there is a misconfiguration which is flagged with the keyword “Error” or the SPF record is missing and flagged with a Null value. The relevant information extracted about the DMARC record is the policy tag (none, quarantine, reject), whether there is a misconfiguration, also flagged with “Error”, or if the DMARC record is missing and also flagged with a Null value. The output is saved into a separate CSV file containing the domain name and related policies.

3.4 Data measurement and hypothesis

The new CSV file can be imported by a separate tool, which groups the SPF and DMARC results by their respective policies and creates a pie chart representing each policy's total number and percentage. The data measured are compared with previous studies to understand the trends in adopting anti-spoofing protocols on a global scale. This analysis compares and contrasts the absolute number of domains with SPF or DMARC records and how many have adopted a more restrictive policy in relation to previous measurements. The Null hypothesis assumes no significant increases in adopting anti-spoofing protocols compared to the last Alexa Top 1 Million domains measurement. The alternative hypothesis assumes an increase with a confidence level interval of 95% or an alpha value of 0.05.

3.5 Ethical considerations

The datasets were collected from the indicated sources and are available for public consumption to allow the experiment's reproducibility. Conscious of the potential impact when conducting such large-scale measurements of Internet systems, the scanning tool was designed to have no effect and to be minimally intrusive. Each measurement has been taken separately.

4 Design Specification

The domain crawler has been designed following three main guidelines: open-source, platform-independent and implementing existing libraries. The tool is written in Python 3.9 and can be executed on both Windows and Linux platforms.

4.1 Tool's Python libraries

The tool leverages an extensive array of built-in libraries and modules available with this programming language, such as:

- **Checkdmarc** is a Python library that parses and validates the SPF and DMARC records, performs DNS queries, and tests multiple domains[9].
- **Easygui** allows launching an Explorer Windows to browse and select the CSV file containing the output of the domain scan[10].
- **Matplotlib** allows the creation of static and interactive data visualisation in Python as well as publication quality plots[11].
- **Numpy** is the standard for working with numerical data in Python[12].
- **Pandas** is a data analysis and parsing module that allows the creation of data frames from CSV files and graphical representations[13].
- **Requests** is an HTTP library which allows API queries and data retrieval[14].
- **Tranco** is a library allowing access to a domain list maintained by security researchers and is more suitable for research purposes[15].

The libraries and their version are listed in a pre-compiled file called "requirements.txt", displayed in Figure 2.

```
PS C:\Users\marcello.WOODLINK\OneDrive\NCI\Thesis\pythonProject> cat .\requirements.txt
checkdmarc==4.5.2
easygui==0.98.2
matplotlib==3.4.3
numpy==1.21.2
pandas==2.0.2
requests==2.25.1
tranco==0.6
```

Figure 2: Python libraries and their versions

[9] checkdmarc: https://pypi.org/project/checkdmarc/
[10] easygui: https://pypi.org/project/easygui/
[11] matplotlib: https://pypi.org/project/matplotlib/
[12] numpy: https://pypi.org/project/numpy/
[13] pandas: https://pypi.org/project/pandas/
[14] requests: https://pypi.org/project/requests/
[15] tranco: https://pypi.org/project/tranco/

4.2 Data input specifications

The tool is designed to parse extremely large domain lists from different sources, including:

- API from third-party domain ranking services
- Standard CSV files containing domain lists
- Tranco Python library

Conscious of the potential bias and lack of transparency when using third-party and commercial domain ranking algorithms, the tool has been designed to leverage domain lists provided by a more "research-oriented" ranking system such as Tranco.

4.3 Program execution requirements

Another design goal was to develop a lightweight tool which does not require a graphical user interface, and that can be executed from a Windows shell, reducing the processing power required. The domain crawler software can be executed from a GUI directly on an IDE or from a command line by typing: *python3 domain_crawler_MainMenu.py*.

In either case, the output of the tool is printed on the screen to allow monitoring of execution. The output of the DNS query can be easily analysed and imported into a spreadsheet as a CSV file. This design choice facilitates interoperability with different tools and improves data analysis. Finally, a separate stand-alone function has been developed to represent the CSV output, including the total number of domains and a breakdown by number and percentage of each SPF and DMARC policy.

4.4 Algorithm diagram

The diagram in Figure 3 represents the algorithm's workflow.

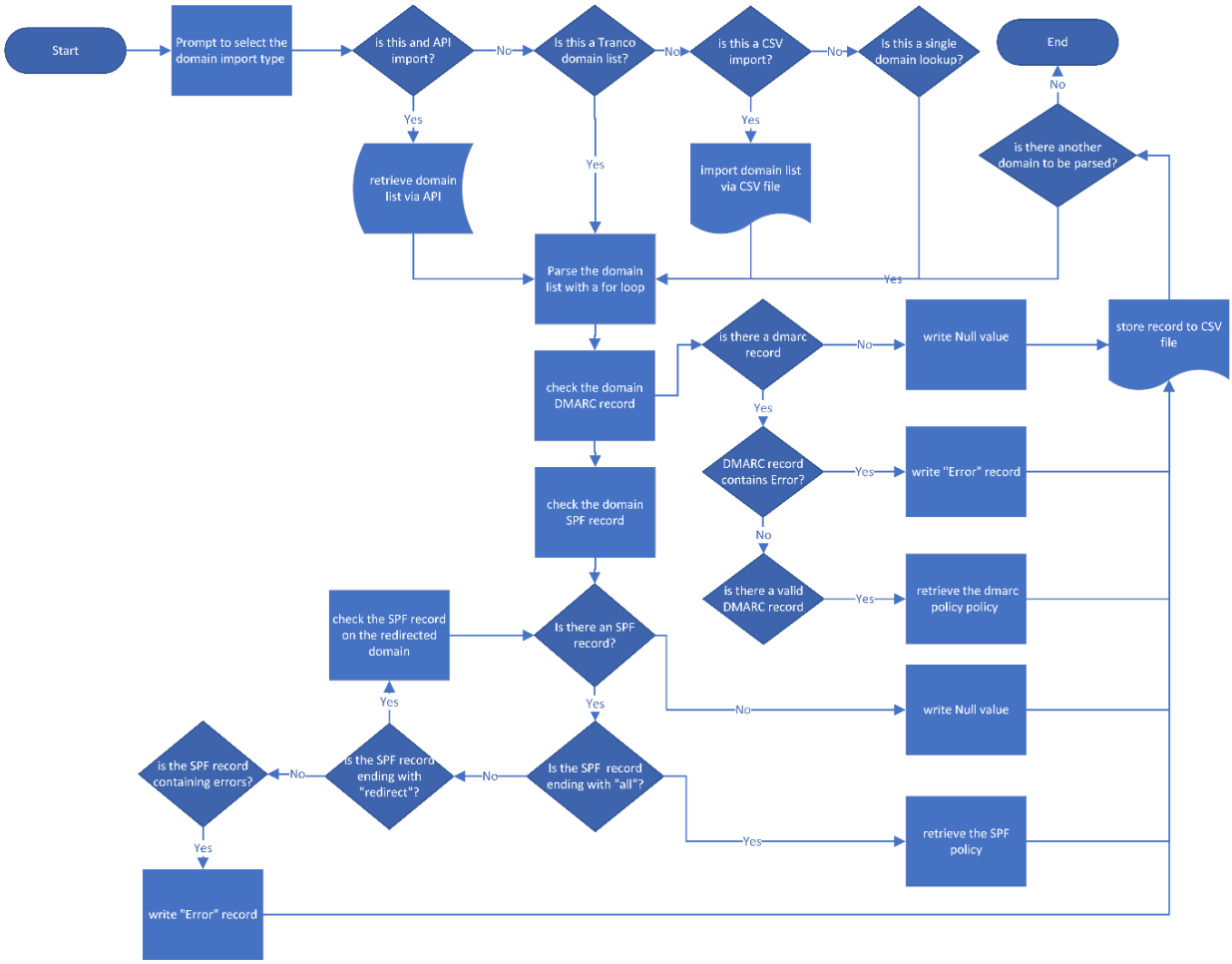

Figure 3: Domain crawler algorithm diagram

5 Implementation

The domain crawler has been developed in Python 3.9 using the PyCharm 2023.1 Community Edition IDE. To reliably gather information about SPF and DMARC and measure their deployment, the tool contains three user-defined functions: *domain_crawler_MainMenu.py*, the main interface, *get_domain_list.py*, *get_spf_dmarc.py*, and *parse_dmarc_spf_csv_result.py*, a stand-alone program which imports the output created by the tool and calculates the percentage and statistics about the SPF and DMARC policies end displays them with a pie chart.

5.1 Main Menu function

The *domain_crawler_MainMenu.py* user-defined function represents the interface which prompts the user to select the domain import type (API, Tranco library or CSV file) or complete DMARC/SPF record for a single domain. If option one is chosen, the user is prompted to decide how many domains must be retrieved by the Similarweb API, as represented by Figure 4. The free API is capped with a 5,000 monthly query limit.

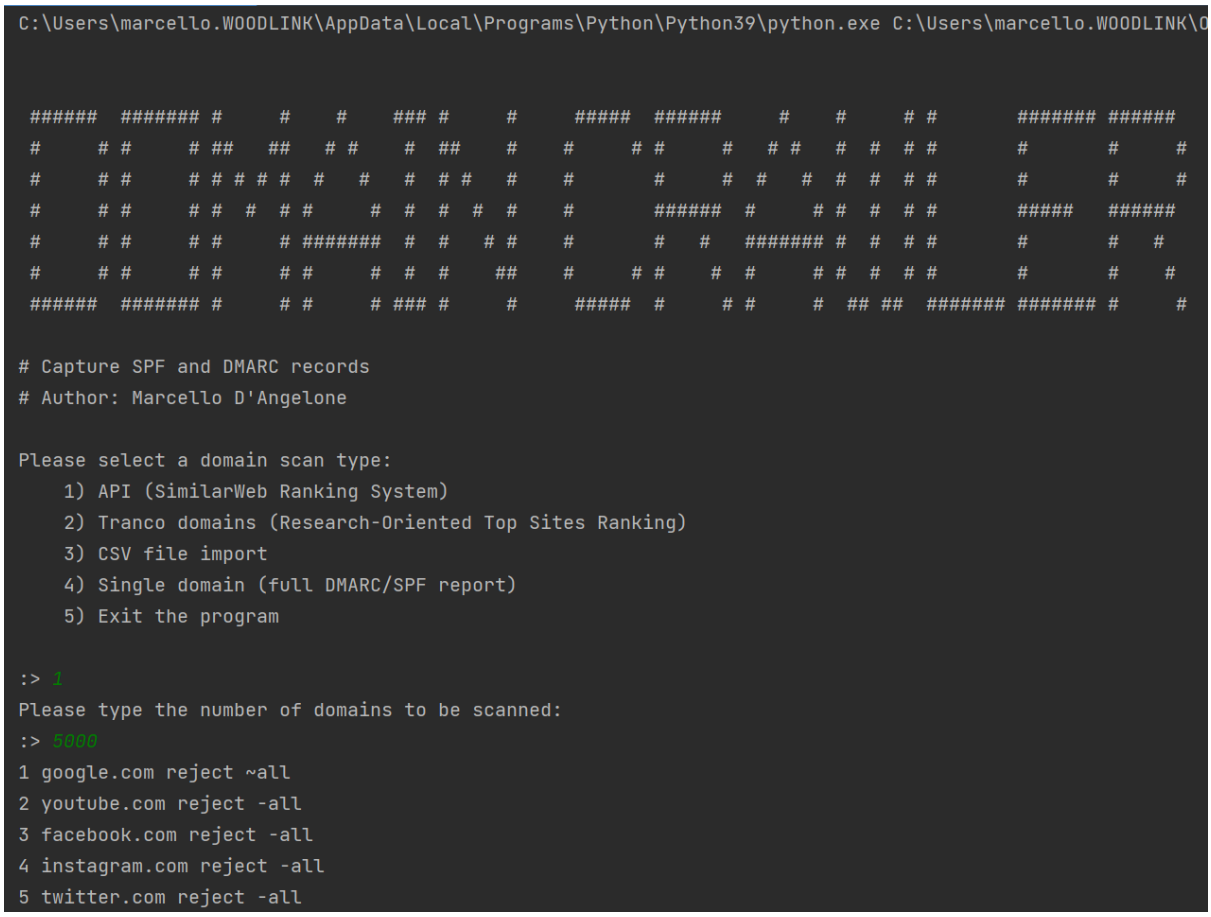

Figure 4: API domain scan

If option two is chosen, the user is prompted to type the number of domains that must be retrieved by the Tranco Python library, as represented by Figure 5. The library provides a maximum number of 1 Million domains.

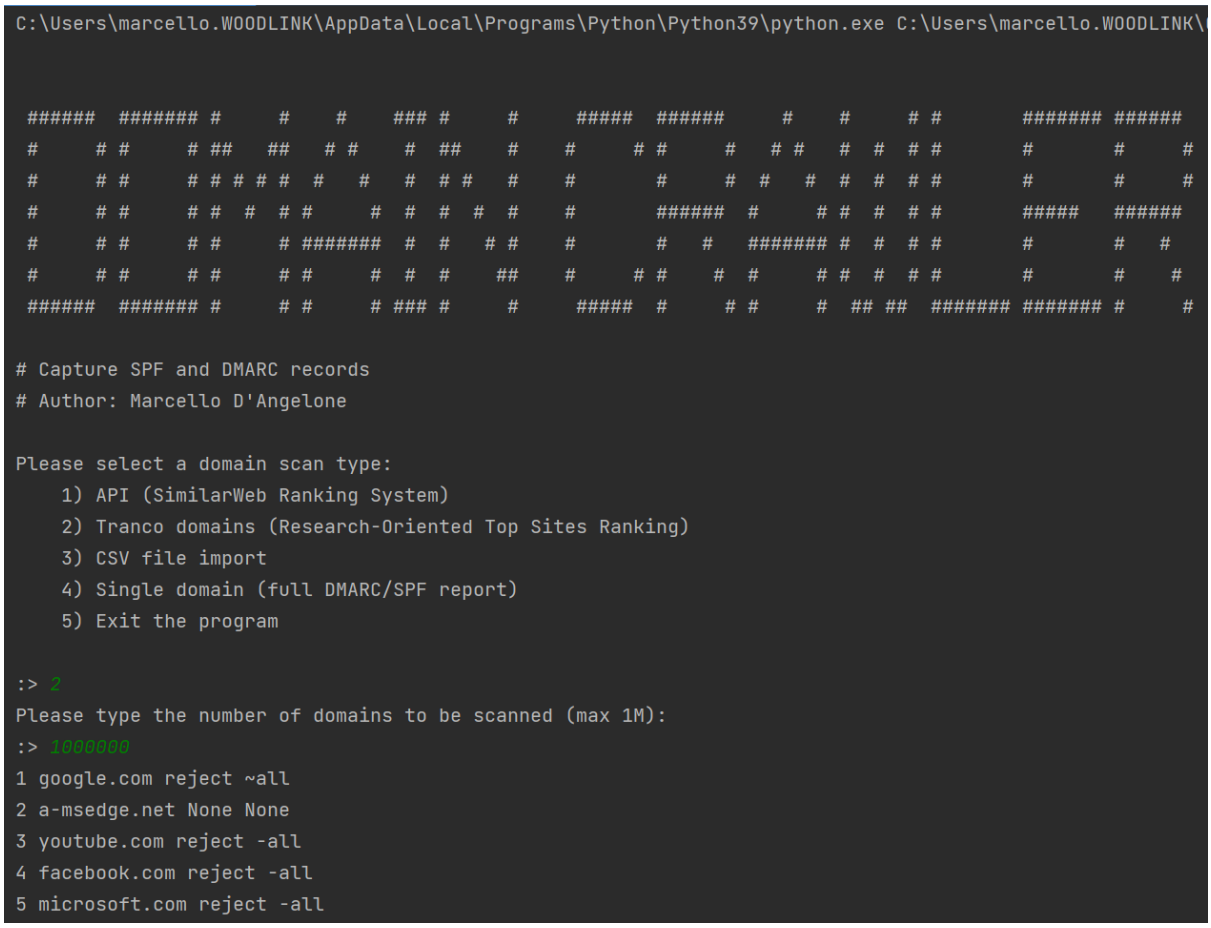

Figure 5: Tranco domain scan

If option three is chosen, the user is prompted to type the path and the name of the CSV file containing the list of domains, as represented in Figure 6.

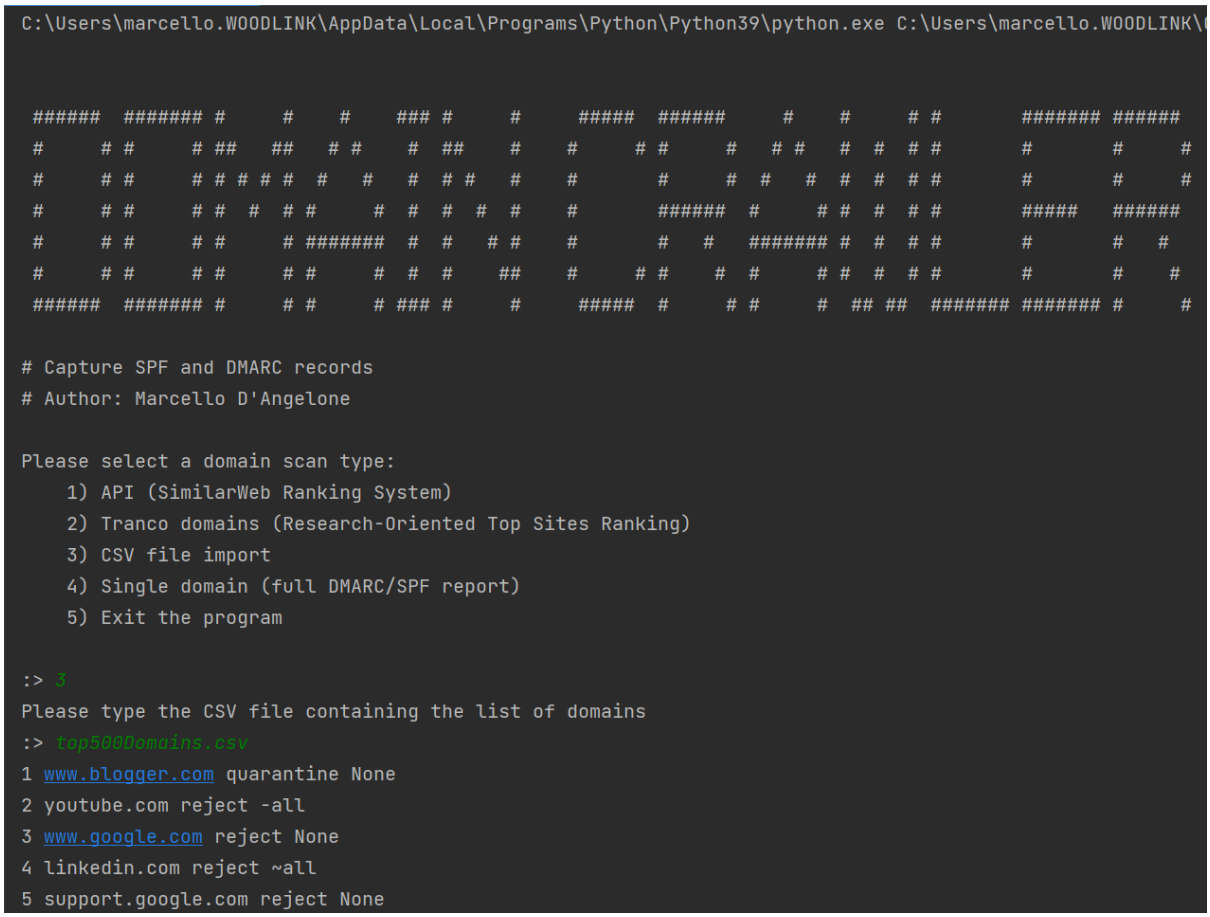

Figure 6: CSV import domain scan

Although there is no limit to the number of domains that can be imported via CSV, the tool has been tested with 10 Million entries. The Python *os* module ensures operating system portability to read file path location on Windows and Linux Operating Systems. The while condition will keep prompting the user until the CSV file exists and the path is correct with the *os.path.exists* method. If option four is chosen, the user is prompted to type the domain name for which full DMARC and SPF records must be extracted, as represented in Figure 7.

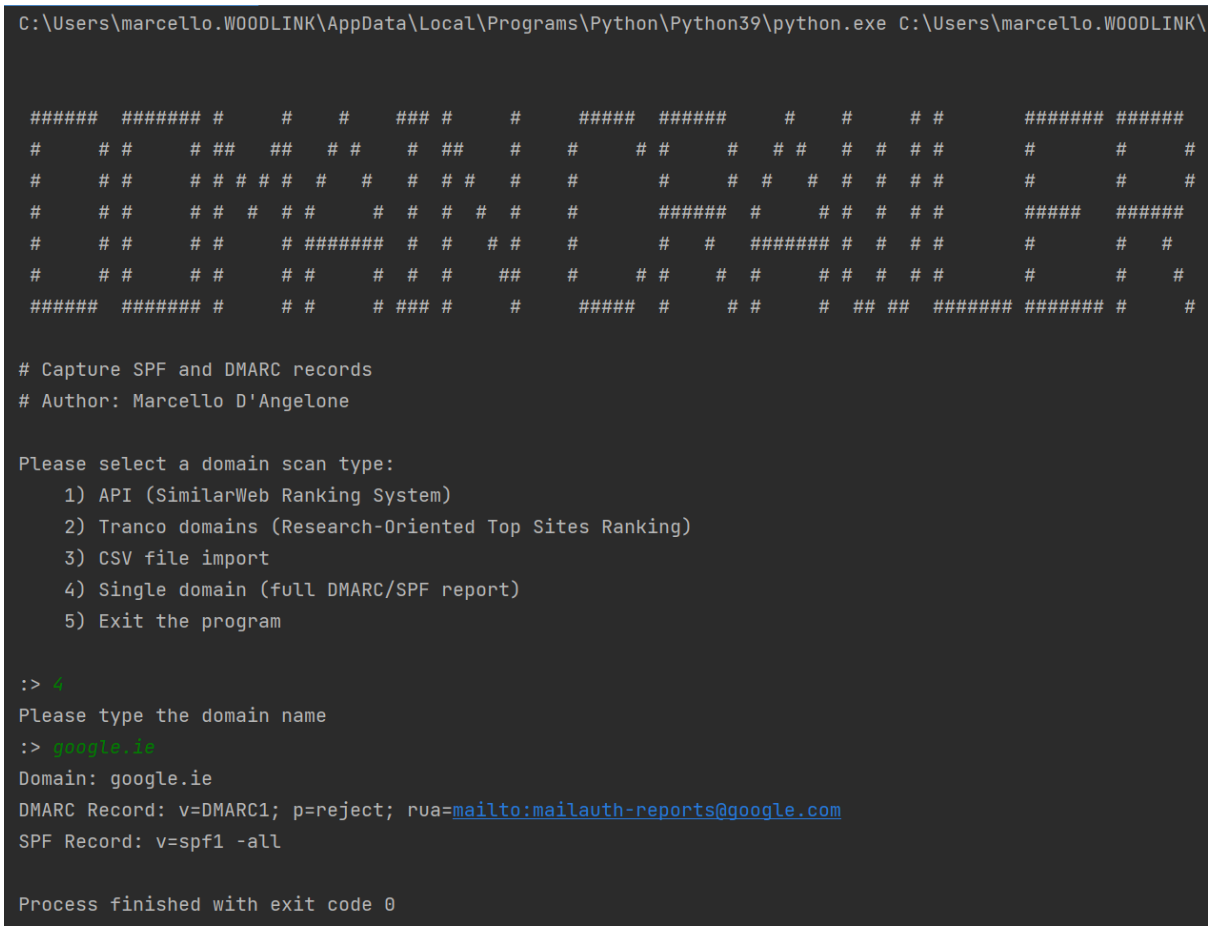

Figure 7: Single domain DMARC and SPF full records

Finally, by pressing option Five, the user can exit the program.

5.2 Import domain list function

The *get_domain_list.py* function reads the option selected from the Main Menu function as a parameter. If option one is selected, a query with the requests library is executed against the Similarweb APIs. The JSON output is parsed with a for loop, and only the "top_sites" key is extracted from the output and appended to the final list. The "try" block performs an API query with the "GET" method. The "except" block handles any error if the request returns an error and prints it to the screen. For instance, "Error: 429 Client Error: Too Many Requests" if the number of domains exceeds the API limit or "Error: 401 Client Error: Unauthorized for url" if the API key expired. Finally, the "else" block executes the next code; if there is no error, convert the response to JSON and extract the "top_sites" key from the response. A for loop parses the top_site variable and creates a list of domains extracting the value of each "domain" key. If option two is selected, the "Tranco" library caches the latest domain list available at the https://tranco-list.eu/ website and returns a tuple of domains which is passed as a parameter to the *get_spf_dmarc.py* function. If option three is selected, the "pandas" library imports the CSV file and parses the "Domain" column, only discarding any other information. In both cases, this returns a tuple as output containing the domains passed to the *get_spf_dmarc.py* function as a parameter.

5.3 SPF and DMARC query function

The *get_spf_dmarc.py* module is the primary tool's engine. It contains two functions, one receives the domain tuple parameter in a for loop and creates a CSV file report, and the second receives a single domain parameter and prints the result to the screen. In both cases, an SPF and DMARC query is performed using the "checkdmarc" Python library. The "try" block executes the get_dmarc_record function against the domain. The "except" block handles exceptions such as DMARCRecordNotfound if there is no DMARC record and assigns a None keyword to the DMARC policy in Python corresponding to the null value. All the other exceptions are captured with the "Error" keyword. If there are no exceptions, then the DMARC policy is extracted from the JSON output recording the value of the ['parsed']['tags']['p']['value'] key. The "for loop" then continues with the domain, using the get_spf_record function in a nested "try" block. The "except" block handles exceptions such as SPFRecordNotFound and assigns the None Keyword and the remaining exceptions such as SPFSyntaxError, SPFToomanyDNSLookups, SPFRedirectLoop, with the "Error" keyword. If there are no exceptions, the first condition checked is whether the SPF record ends with the "all" tag. If this is true, a regular expression pattern validates the policy with a matching expression: "[-+?~]?(all)$". Whereas "[-+?~]" matches a single character in the list, "?" matches the previous list between zero and one time, (all) literally matches "all", and "$" matches at the end of the string. The second condition checked is whether the last SPF record starts with "redirect=". If the SPF record ends with the latter, a new while loop recursively queries the SPF record of the target domain until another exception is met or there is a valid SPF policy. As per RFC7208, a valid SPF record must end with either the "all" "mechanism" or the "redirect" "modifier". The "qualifiers" "+", "-", "~", and "?" are optional and default to "+" (pass) if not present. An "if" statement checks this condition and prepends "+" to the "all" qualifier to normalize the data. Furthermore, any "redirect" modifier is ignored if an "all" tag is in the same record. The new target domain specified in the "redirect" modifier can also contain another "redirect" statement. However, a limit of ten recursive lookups during an SPF evaluation is enforced by design (Kitterman, 2014). A "SPFTooManyDNSLookups" exception is thrown and recorded with the "Error" keyword if the ten lookups threshold is exceeded. The for loop appends the result to a list, and each row is saved as a CSV file, represented in Figure 8. This file contains four columns, one with the ordinal number, the parsed domain, and its DMARC and SPF policies, respectively.

A	B	C	D
#	domain	DMARC Policy	SPF Policy
1	google.com	reject	~all
2	youtube.com	reject	-all
3	baidu.com		-all
4	bilibili.com	reject	-all
5	facebook.com	reject	-all
6	qq.com	quarantine	-all
7	twitter.com	reject	-all
8	zhihu.com	quarantine	-all
9	wikipedia.org	none	~all
10	amazon.com	quarantine	-all
11	instagram.com	reject	-all
12	linkedin.com	reject	~all
13	reddit.com	reject	~all
14	whatsapp.com	reject	-all
15	openai.com	reject	-all
16	yahoo.com	reject	?all
17	bing.com	none	-all

Figure 8 SPF and DMARC CSV output

The CSV file is called "spf-dmarc_result_" appended with the timestamp of the year, month, day, hour and minute when the file was created, for instance, "spf-dmarc_result_2023_06_03-11-06.csv". If a single domain option is selected from the main menu, the full DMARC and SPF records are extracted and printed on the screen instead, as shown in Figure 5.

5.4 Plot SPF and DMARC CSV file

The *parse_dmarc_spf_csv_result.py* function is a stand-alone script which imports the "spf-dmarc_result" CSV file and provides a graphical representation of the percentage and absolute number of both SPF and DMARC policies. Upon execution, an Explorer window is launched to allow the user to browse and select the CSV file created by the domain crawler program. The "while True" loop validates the user interaction with the *easygui* library. If the user press "Cancel", the program will exit. If the user does not select a CSV file, he will be prompted with an "Invalid path or filename" error message until a valid CSV file has been selected. The file is then parsed with the "pandas" library, and both the "SPF Policy" and "DMARC Policy" columns are grouped with the "df.groupby" function. The parameters "sort=True" are used to sort the groups by size, as_index=True returns the objects with a group label, and the dropna=False includes the "Null" values in the grouping. The "fig.add_subplot" function plots two pie charts, one next to the other, and a lambda function is called to represent the numerical and percentage value for each DMARC and SPF grouped data frame. Finally, a title is displayed using the data frame index, representing the total number of values, as represented in Figure 9.

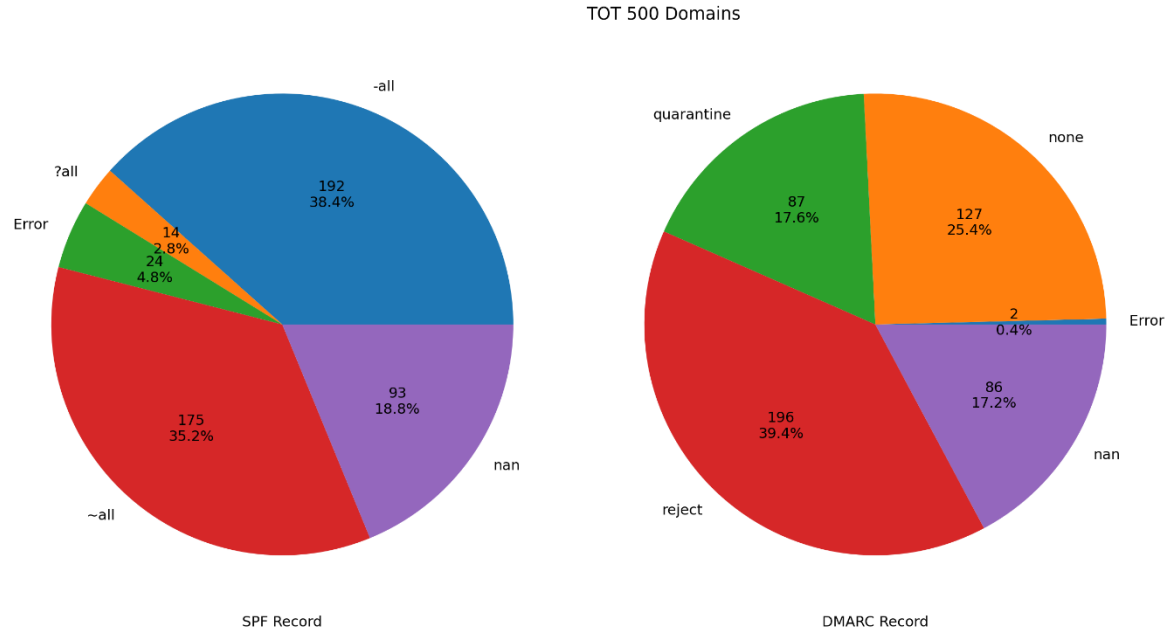

Figure 9 SPF and DMARC pie chart plots

The fig.savefig() function automatically creates an image in Scalable Vector Graphics format (SVG) which can be directly used in publications.

6 Evaluation

The CSV output created by the different domain list scans has been imported into an Excel spreadsheet for further analysis and comparison.

6.1 US government domains evaluation

The measurements related to the US .gov domains did not return the expected results compared to previous studies. As mentioned, Tatang et al. (2021) report an 88% DMARC and a 92% SPF adoption rate for all .gov top-level domains. Two years later, and five years after the Binding Operational Directive 18-01, the scan of 9,914 .gov domains shows 75.47% with a valid DMARC policy, less than half of them, 45.95%, with a *reject* policy. This is a step forward compared to the 20% measured before the Department of Homeland Security (DHS) issued the Directive. Nonetheless, only 54.05% of the .gov domains comply with the mandate. Table 2 summarises the DMARC policy implementation.

Table 2: DMARC results on .gov domains

DMARC policy	Count by policy type	Rate by policy type	Count by valid/invalid - N/A	Rate by valid/invalid - N/A
reject	4555	45.95%	7482	75.47%
quarantine	377	3.8%		
none	2550	25.72%		
Error	74	0.75%	2432	24.53%
N/A	2358	23.78%		

Regarding the SPF protocol, even lower rates were detected, with only 31.75% of the domains with a valid SPF record and 20.62% with a *Hard Fail* policy type (-all). Tatang et al. (2021)

do not provide the total number of .gov domains scanned or a breakdown for each policy. They highlight a misconfiguration with the "cia.gov" domain, which appears now to be configured correctly but with a *quarantine* DMARC policy instead of a *reject* and a *Soft Fail* SPF policy (~all), as reported in Figure 10.

```
Please type the domain name
:> cia.gov
Domain: cia.gov
DMARC Record: v=DMARC1; p=quarantine; pct=100; rua=mailto:demarcreports@uce.cia.gov; ruf=mailto:demarcfailures@uce.cia.gov
SPF Record: v=spf1 mx a:mail1a.cia.gov a:mail1b.cia.gov a:mail2a.cia.gov a:mail2b.cia.gov mx:cia.gov mx:ucia.gov ~all

Process finished with exit code 0
```

Figure 10 cia.gov full DMARC and SPF records

Table 3 summarises the findings of the SPF protocol's implementation.

Table 3: SPF results on .gov domains

SPF policy	Count by policy type	Rate by policy type	Count by valid/invalid - N/A	Rate by valid/invalid - N/A
-all	2044	20.62%	3148	31.75%
~all	1058	10.62%		
?all	42	0.42%		
+all	4	0.04%	6766	68.25%
Error	91	0.92%		
N/A	6675	67.33%		

Figure 11 represents both protocols' implementation on a pie chart created with the *parse_dmarc_spf_csv_result.py* function.

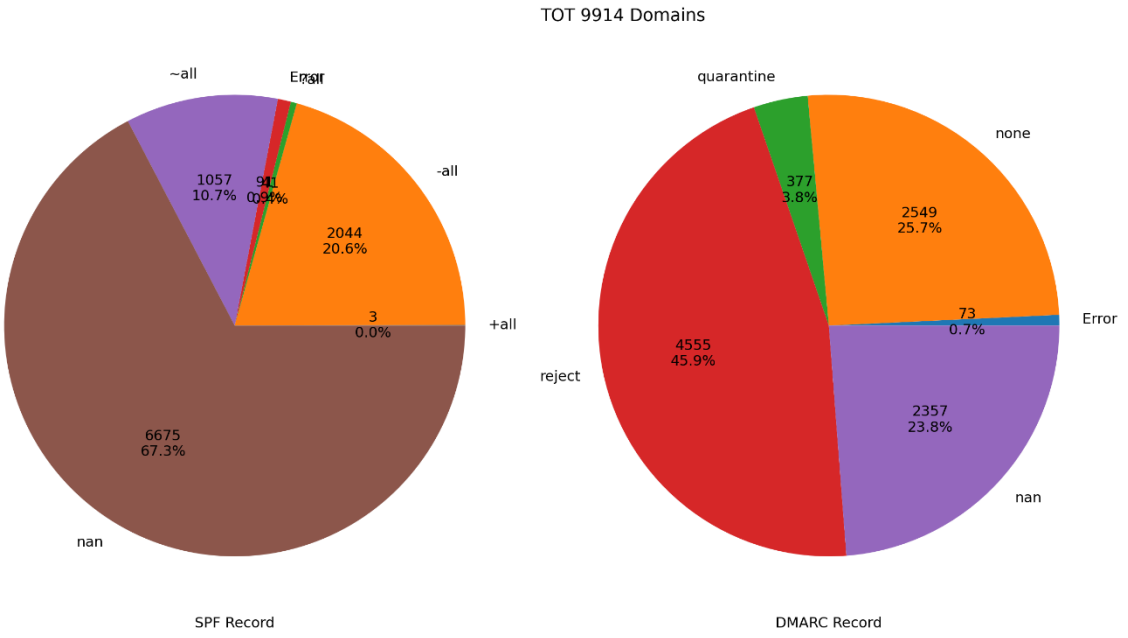

Figure 11: SPF and DMARC .gov implementation pie chart

It should be noted, however, that the 18-01 directive applies to all federal agencies, not only the ones using .gov top-level domain. Therefore, to expand the scope of the previous research

papers, a more significant number of domains has been scanned: 20,349. This dataset includes .gov and .mil domains collected from the DomCop dataset and the domain list maintained by the U.S. General Services Administration. The result shows that the DMARC policy is deployed to only 51.74% of all governmental domains, with only 29.58% of them complying with a *reject* policy. A further inspection of the 10,529 domains with a valid DMARC policy reveals that 57.18% of them are set to *reject*, only 9.22% to *quarantine*, and 33.6% to *none*. This means that on third is configured with a less restrictive policy in monitoring mode.

Table 4: DMARC results on US governmental domains

DMARC policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
reject	6020	29.58%	10529	51.74%
quarantine	971	4.77%		
none	3538	25.72%		
Error	101	0.50%	9820	48.26%
N/A	9719	47.76%		

Regarding the SPF policy, slightly better results are recorded compared to the previous survey on .gov domains. Less than half have a valid SPF policy (45.83%), and little over a quarter are implemented with a *Hard Fail* (28.11%). Overall the number of misconfigured domains is still very low (2.19%); however, half of them do not present any SPF record (52.02%). A drill-down of the 9,326 domains with a valid SPF policy shows that 61.34% of them are restricted with a *Hard Fail*, more than one-third (36.55%) with a *Soft Fail* and a negligible number with a *Neutral* and *Pass* policy.

Table 5: SPF results on US governmental domains

SPF policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
-all	5721	28.11%	9326	45.83%
~all	3409	16.75%		
?all	182	0.89%		
+all	14	0.07%	11023	54.17%
Error	437	2.19%		
N/A	10586	52.02%		

Although Tatang et al. (2021) do not report the dataset sources and the number of .gov domains surveyed for their measurements, they quote another research carried out by a leading email security vendor, which examined a total of only 1,311 Federal civilian domains (ProofPoint, 2018). With a limited dataset of the most popular .gov Top-level domains, it is more likely to obtain higher deployment rates. However, it should be noted that the Binding Operational Directive 18-01 scope is much broader. As highlighted by the Cybersecurity & Infrastructure Security Agency (2017), the Directive applies not only to government-owned systems but also

to all federal agencies and those operating on their behalf. Furthermore, as correctly pointed out, even domains not used to send emails should be protected with the DMARC policy of *reject* to thwart attackers from spoofing governmental services. Both surveys of 9,914 and 20,349 datasets demonstrated that US agencies are still halfway through their journey to comply with the Directive and secure their infrastructure against spoofing attacks. Limiting the survey to the top one thousand domains may skew the results and misrepresent the actual figures.

6.2 Top 500 domains evaluation

The survey of the top 500 most popular domains presented a more expected result. The organisations behind those domains are some of the most popular and recognisable. Therefore, they are more willing to protect themselves and their customers against spoofing attacks, which may represent reputational damage. Notably, 82.4% have a DMARC policy, with almost half, 39.4%, having the most restrictive *reject* policy. Only two domains are misconfigured: clarin.com, as reported by the single domain scan DMARC record in Figure 12, the mandatory v=DMARC1, and the p= policy tags are missing.

```
Domain: clarin.com
DMARC Record: dmarc@planisys.com is not a valid DMARC report URI
```

Figure 12 DMARC misconfiguration

Whereas, as reported in Figure 13, the failure reporting option "fo=" is misconfigured on the ig.com.br domain.

```
Domain: ig.com.br
DMARC Record: Error: Expected tag_value at position 27 in: v=DMARC1; p=none; sp=none; fo=; ri=3600
```

Figure 13 DMARC misconfiguration

However, the remaining 88, 17.6%, do not have any DMARC records. Table 6 reports the DMARC statistics about the TOP 500 domains.

Table 6: DMARC results on TOP 500 domains

DMARC policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
reject	197	39.4%	412	82.4%
quarantine	88	17.6%		
none	127	25.4%		
Error	2	0.4%	88	17.6%
N/A	88	17.2%		

Out of the 412 domains with a valid DMARC configuration, almost half of them (47.82%) are secured with a *reject* policy, less than one-third (30.83%) are in monitoring mode with a *none* policy, and finally, more than one-fourth (21.36%) present a less restrictive policy of *quarantine*.

The SPF policy of the top 500 domains presents a symmetrical configuration, with 38.4% configured with a more restrictive *Hard Fail* and 18.8% without an SPF record. A slightly

higher number of domains present a misconfiguration, mainly because of exceeding the number of lookups allowed. Table 7 represents the findings about the SPF record.

Table 7: SPF results on TOP 500 domains

SPF policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
-all	192	38.4%	382	76.4%
~all	176	35.2%		
?all	14	2.8%		
+all	0	0%	118	23.6%
Error	24	4.8%		
N/A	94	18.8%		

Out of the 382 domains with a valid SPF configuration, half of them are secured with a *Hard Fail* (50%), slightly less than half with a *Soft Fail* (46.07%) and the remaining 3.66% with a *Neutral* configuration. None of them presents a *Pass* policy.
Additionally, 143 of the Top 500 domains with a DMARC policy of *reject* also have an SPF policy of either *Hard Fail* (-all) or *Soft Fail* (~all), which ensures a more secure configuration. Among them, it is possible to find popular online platforms such as youtube.com, linkedin.com, paypal.com, governmental agencies such as nasa.gov, nih.gov, and usda.gov and also online newspapers such as theguardian.com, nytimes.com and cnn.com

6.3 Alexa top 1 Million domains evaluation

The total runtime to parse the Alexa Top 1 Million domains was eight days, two hours and four minutes. Its measurement confirmed a trend that has been reported over the past eight years with three years intervals. Table 8 shows the percentages of the SPF and DMARC deployments published by previous studies and the latest rate captured by the domain crawler tool.

Table 8: Alexa Top 1 Million SPF and DMARC published measurements

	Durumeric et al.; Foster et al. (2015)	Hu et al., (2018)	Tatang et al., (2021)	Domain Crawler 2023
DMARC	1.1%	5.1%	11.5%	25.6%
SPF	40.1%	44.9%	50.3%	59.6%

Although, the percentages are much lower compared with the top 500 domains. The overall coverage of the DMARC protocol is as high as 25.63%. Also, the most restrictive policy is not as widely implemented, with only 6.87% and the majority of domains configured in monitoring mode with the policy of *none*. The percentage of misconfigured domains is equivalent to the top 500, with only 0.41%, as represented in Table 9.

Table 9: DMARC results on Alexa Top 1 Million domains

DMARC policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A

reject	59,417	6.87%	221,607	25.63%
quarantine	45,247	5.23%		
none	116,943	13.53%		
Error	3514	0.41%	642,945	74.37%
N/A	639,431	73.96%		

Out of the 221,607 domains with a valid DMARC record, a little over half of them are configured with monitoring mode with a *none* policy (52.77%), where one-fourth and one-fifth are configured with a *reject* and *quarantine* policies, respectively (26.81% and 20.42%). In regards to the SPF measurements, the *Soft Fail* policy is the most prevalent, with 35% of the total domains, followed by the *Hard Fail* with 21.53%. There is also a negligible number of *Neutral* and *Pass* policies, with 2.83% and 0.06%, respectively. Similarly with the DMARC records also, the percentage of misconfigured SPF domains is comparable with the TOP 500 (4.19%); however, the percentage of Null values is twice as much as the Top 500 (36.22% versus 18.8%). Interestingly, the rate of misconfigured SPF records is less than one-third compared to the same dataset measured by Tatang et al. (2021) three years ago. With a total of 75,403 invalid SPF records (13.0%) in 2020 versus the 36,219 (4.19%) reported by the domain crawler tool in 2023. Table 10 summarises the SPF results on the Alexa list.

Table 10: SPF results on Alexa Top 1 Million domains

SPF policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
-all	186,167	21.53%	515,164	59.59%
~all	303,965	35.16%		
?all	24,480	2.83%		
+all	552	0.06%		
Error	36,219	4.19%	349,388	40.41%
N/A	313,169	36.22%		

Out of the 515,164 domains with a valid SPF record, more than half present a *Soft Fail* policy (59%) and more than one-third a *Hard Fail* policy (36%). This leaves with less than 5% shared with the *Neutral* and *Pass* policies. The Alexa Top 1 Million measurements confirm that similarly to the governmental domains, the larger the pool, the lower the adoption rate of the DMARC and SPF protocols.

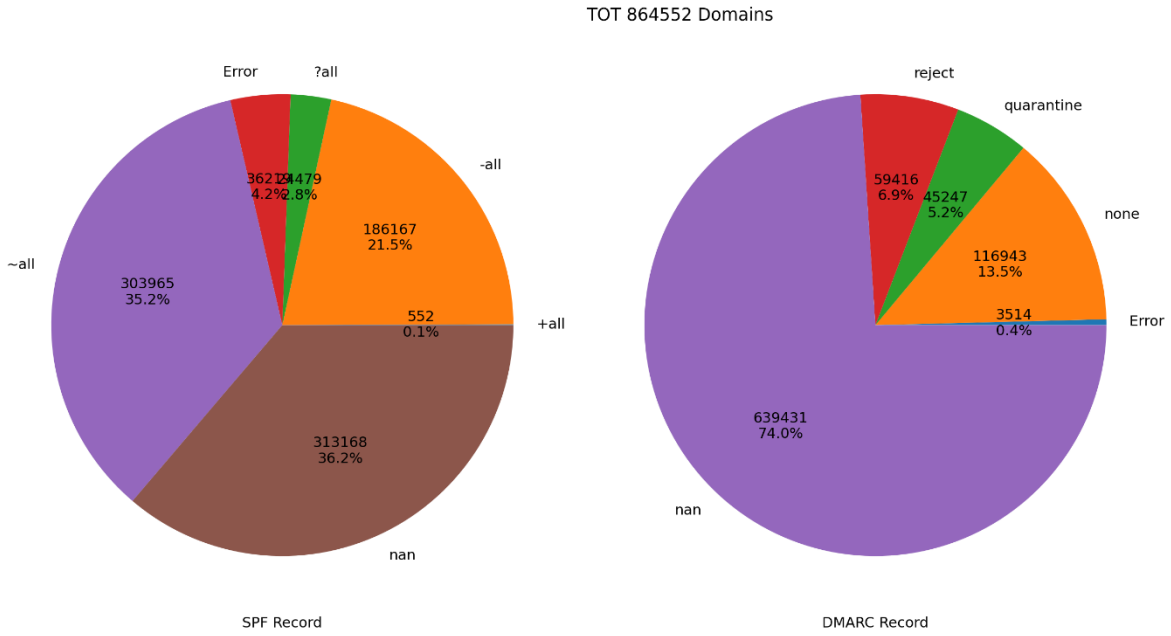

Figure 14: Alexa Top 1 Million SPF and DMARC pie chart

To properly compare the results with previous studies, it should be noted that one of the latest surveys of DMARC and SPF was based on the complete 1 Million Alexa list. Since then, the number of domains provided by Amazon has fluctuated from 839,000 to 864,552 in the last release. Table 11 provides a snapshot of the data to be compared.

Table 11: Alexa Top 1 Million measurements comparison

	Total valid DMARC		Total valid SPF		Total
Tatang et al. (2021)	114,706	11.47%	503,310	50.33%	1,000,000
domain crawler tool	221,607	25.63%	515,164	59.59%	864,552

To measure two variables with different sample sizes and confirm an increase in the adoption of the DMARC and SPF protocols, a 2-sample test for equality of proportions with continuity correction was used with R (prep.test function). Because the p-value of 2.2e-16 is much smaller than the alpha value of 0.05, with both DMARC and SPF data, the Null Hypothesis has been rejected, which leads to the conclusion that the Top 1 Million Alexa domain list presents a significant increase in DMARC and SPF adoption between 2021 to 2023

6.4 SimilarWeb API domain evaluation

Also, the survey of the Top 5000 SimilarWeb domains shows how the rate of domains with a DMARC implementation is drastically lower as the number of domains in scope increases, with only 55.56% compared to the 82.40% of the Top 500 domains list. The distribution of the policy is similar, with a majority of *reject* 22.74% followed by *none* 19.16% and 13.66% with *quarantine*. The overall number of misconfigured domains is low, with high-profile domains such as gov.uk among them, slightly better regarding the SPF deployment, with a total of 69.6% of domains compared with 76.04% of Moz's Top 500.

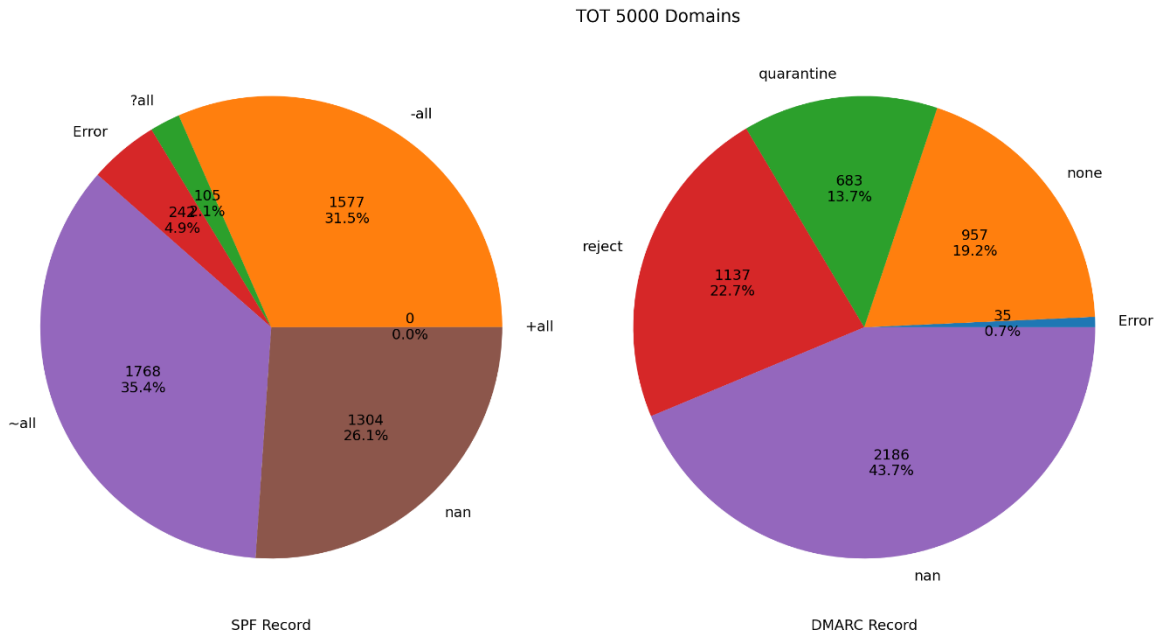

Figure 15: Similarweb TOP 5000 domains

6.5 Tranco Top 1 Million evaluation

The last measurement leveraged the Tranco Python package, which pulled an aggregate list of the Top 1,000,000 domains provided by Alexa, Umbrella, Majestic and Farsight from the 26th of May to the 24th of June 2023[16]. As mentioned, the researchers who created this list supply a permanent link which can always be referenced for future comparative studies. Furthermore, they remove any malicious domain which could skew the results. The results are in line with the Alexa Top 1 Million measurements. A few points percentage lower in the DMARC adoption rate with an overall 23.78% versus 25.63% of the Alexa list and a few points percentages higher number of domains without a DMARC policy, 76.22% versus 74.37%, almost the same percentage of misconfigured domains, 0.49% versus 0.41% Table 12 represent a complete break down of the numbers and percentages.

Table 12: DMARC results on Tranco list

DMARC policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
reject	58,712	5.87%	237,811	23.78%
quarantine	54,006	5.40%		
none	125,093	12.51%		
Error	4,949	0.49%	762,189	76.22%
N/A	757,240	75.72%		

[16] Tranco: https://tranco-list.eu/list/VXVVN

Also, the SPF measurements of the Tranco dataset show a slightly lower adoption rate than the Alexa one. The total percentage of SPF adoption is 55.36% versus 59.59% of the Alexa list, and the percentage of domains without an SPF record or with an error is slightly higher at 44.64% versus 40.41%. The complete breakdown is represented in Table 13.

Table 13: SPF results on Tranco list

SPF policy	Count by policy type	Rate by policy type	Count by valid / invalid - N/A	Rate by valid / invalid - N/A
-all	207206	20.72%	553,584	55.36%
~all	320149	32.01%		
?all	25528	2.55%		
+all	701	0.07%		
Error	38105	3.81%	446,416	44.64%
N/A	408311	40.83%		

Although the Tranco measurement shows a slightly lower adoption of the Alexa Top 1 Million, it is still possible to confirm an increase compared to the previous measurements and an overall upward trend.

7 Conclusion and Future Work

This thesis's contribution is twofold. One aspect provided an in-depth review of the current state of email anti-spoofing protocols, their limitations and how recent studies tried to measure their adoption rate with different techniques. The second aspect provided a novel domain crawler, which aims to be the standard in measuring SPF and DMARC adoption rates, thus, facilitating a consistent large-scale trend analysis. Spoofed emails to perform BEC attacks remain one of the primary vectors attackers leverage for financial gains or data breaches. Due to the fundamental authentication weaknesses in the SMTP protocol, security extensions have been designed as an afterthought to mitigate spoofing attacks.

Nonetheless, SPF and DMARC adoption rate is not as widespread as expected. It is crucial to understand their adoption rate to rectify misconfigurations whenever detected and promote awareness of email security best practices. Previous studies proposed different measuring techniques and did not report the total number of domains surveyed or used. This can lead to misinterpretation of their deployment status. As demonstrated for the US governmental domains, previous studies claimed a DMARC and SPF adoption rate of 88% and 92%, respectively. This research reported 75% and 31% on a broader range of .gov top-level domains, highlighting how the dataset selection is as important as the reliability of the measurement itself.

Furthermore, the measurement has been carried out on different datasets and compared with previous studies. Comparing previous measurements of the Alexa Top 1 Million domain list, it has been statistically demonstrated that the DMARC and the SPF adoption rate keep growing, with the DMARC protocol doubling from 11.5% to 25.6% and the SPF increasing from 50.3% to 59.6%. The crawler tool has also been equipped with the Tranco domain list for future measurements. This research-oriented Python library provides a more consistent and suitable

domain ranking system which can be referenced at different points in time. Future work may leverage such functionality to represent the increase or variation in the SPF and DMARC adoption rate against the same dataset. For instance, how many domains went from a *none* to a *reject* DMARC policy or went from a *Soft Fail* to a *Hard Fail* SPF policy. Which domain with a misconfiguration fixed their anti-spoofing policy. Finally, it should be noted that the total duration time for the Alexa Top 1 Million domain measurement was over a week. This was deliberate for ethical reasons and to avoid an excessive burden on the DNS infrastructure. However, future work could be dedicated to building a multi-thread distributed system that could increase the efficiency in parsing the domain list and dramatically reduce the run time.

References

Alkhalil, Z., Hewage, C., Nawaf, L. and Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, [online] 3, p.563060. doi:https://doi.org/10.3389/fcomp.2021.563060.

Armorblox (2023). *2023 Email Security Threat Report*. [online] Armorblox. Available at: https://assets.armorblox.com/f/52352/x/d0929bd19f/armorblox_2023-email-security-threat-report.pdf?cv=1683760508039 [Accessed Apr. 2023].

Bennett, N., Sowards, R. and Deccio, C. (2022). SPFail: Discovering, Measuring, and Remediating Vulnerabilities in Email Sender Validation. *Proceedings of the 22nd ACM Internet Measurement Conference*, pp.633–646. doi:https://doi.org/10.1145/3517745.3561468.

Chauhan, P.D. and Shah, A.M. (2023). Effectiveness of Anti-Spoofing Protocols for Email Authentication. *IEEE Xplore*. [online] doi:https://doi.org/10.1109/ICCT56969.2023.10076098.

Chen, J., Paxson, V. and Jiang, J. (2020). *Composition Kills: A Case Study of Email Sender Authentication*.

Cidon, A., Korshun, N., Schweighauser, M., Tsitkin, A., Gavish, L. and Bleier, I. (2019). *High Precision Detection of Business Email Compromise High Precision Detection of Business Email Compromise*. [online] Available at: https://www.usenix.org/system/files/sec19-cidon.pdf.

Crocker, D., Hansen, T. and Kucherawy, M. (2011). DomainKeys Identified Mail (DKIM) Signatures. *www.rfc-editor.org*. [online] doi:https://doi.org/10.17487/RFC6376.

Cybersecurity & Infrastructure Security Agency (2017). *Binding Operational Directive 18-01 | CISA*. [online] Binding Operational Directive 18-01. Available at: https://www.cisa.gov/news-events/directives/binding-operational-directive-18-01 [Accessed Jun. 2023].

Dalvi, S., Gressel, G. and Achuthan, K. (2020). Tuning the False Positive Rate / False Negative Rate with Phishing Detection Models. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9, pp.2249–8958. doi:https://doi.org/10.35940/ijeat.A1002.1291S52019.

Deccio, C., Yadav, T., Bennett, N., Hilton, A., Howe, M., Norton, T., Rohde, J., Tan, E. and Taylor, B. (2021). Measuring email sender validation in the wild. *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*. [online] doi:https://doi.org/10.1145/3485983.3494868.

Durumeric, Z., Adrian, D., Mirian, A., Kasten, J., Bursztein, E., Lidzborski, N., Thomas, K., Eranti, V., Bailey, M. and Alex, H.J. (2015). Neither Snow Nor Rain Nor MITM... An Empirical Analysis of Email Delivery Security. *Proceedings of the 2015 Internet Measurement Conference*. doi:https://doi.org/10.1145/2815675.2815695.

Federal Bureau of Investigation (2022). *Internet Crime report*. [online] https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf. Available at: https://www.documentcloud.org/documents/23707016-2022_ic3report [Accessed 3 Feb. 2023].

Foster, I.D., Larson, J., Masich, M., Snoeren, A.C., Savage, S. and Levchenko, K. (2015). Security by Any Other Name: On the Effectiveness of Provider Based Email Security. [online] ACM, pp.450–464. doi:https://doi.org/10.1145/2810103.2813607.

Gupta, S., Pilli, E.S., Mishra, P., Pundir, S. and C, Joshi R (2014). Forensic analysis of E-mail address spoofing. pp.898–904. doi:https://doi.org/10.1109/CONFLUENCE.2014.6949302.

Hu, H., Peng, P. and Wang, G. (2018). Towards Understanding the Adoption of Anti-Spoofing Protocols in Email Systems. In: *2018 IEEE Cybersecurity Development (SecDev)*. 2018 IEEE Cybersecurity Development (SecDev). pp.94–101. doi:https://doi.org/10.1109/SecDev.2018.00020.

Kitterman, S. (2014). *Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1*. [online] IETF. Available at: https://www.rfc-editor.org/rfc/rfc7208.html [Accessed 16 May 2023].

Konno, K., Kitagawa, N. and Yamai, N. (2020). False Positive Detection in Sender Domain Authentication by DMARC Report Analysis. *Proceedings of the 2020 The 3rd International Conference on Information Science and System*. doi:https://doi.org/10.1145/3388176.3388217.

Kucherawy, M. and Zwicky, E. (2015). *Domain-based Message Authentication, Reporting, and Conformance (DMARC)*. [online] RFC Editor, p.RFC7489. doi:https://doi.org/10.17487/rfc7489.

Lindsey, N. (2019). *Toyota Subsidiary Loses $37 Million Due to BEC Scam*. [online] CPO Magazine. Available at: https://www.cpomagazine.com/cyber-security/toyota-subsidiary-loses-37-million-due-to-bec-scam/ [Accessed Mar. 2023].

Liu, E., Akiwate, G., Jonker, M., Mirian, A., Ho, G., Voelker, G. and Savage, S. (2023). *Forward Pass: On the Security Implications of Email Forwarding Mechanism and Policy*. [online] Available at: https://arxiv.org/pdf/2302.07287.pdf [Accessed 2 Apr. 2023].

Maroofi, S., Korczynski, M. and Duda, A. (2020). *From Defensive Registration to Subdomain Protection: Evaluation of Email Anti-Spoofing Schemes for High-Profile Domains*. [online] Available at: https://dl.ifip.org/db/conf/tma/tma2020/tma2020-camera-paper54.pdf.

Maroofi, S., Korczynski, M., Holzel, A. and Duda, A. (2021). Adoption of Email Anti-Spoofing Schemes: A Large Scale Analysis. *IEEE Transactions on Network and Service Management*, [online] 18(3), pp.3184–3196. doi:https://doi.org/10.1109/TNSM.2021.3065422.

Nanaware, T., Mohite, P. and Patil, R. (2019). DMARCBox – Corporate Email Security and Analytics using DMARC. In: *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. pp.1–5. doi:https://doi.org/10.1109/I2CT45611.2019.9033552.

Opazo, B., Whitteker, D. and Shing, C.-C. (2017). Email trouble: Secrets of spoofing, the dangers of social engineering, and how we can help. [online] IEEE, pp.2812–2817. doi:https://doi.org/10.1109/FSKD.2017.8393226.

Pochat, L., Goethem, V., Tajalizadehkhoob, S., Korczynski, M. and Joosen, W. (2019). Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. [online] Internet Society. doi:https://doi.org/10.14722/ndss.2019.23386.

ProofPoint (2018). *Federal DMARC Adoption Rates Increase Significantly to Address BOD 18-01 Deadline*. [online] Proofpoint. Available at: https://www.proofpoint.com/us/corporate-blog/post/federal-dmarc-adoption-rates-increase-significantly-address-bod-18-01-deadline [Accessed Jun. 2023].

ProofPoint (2023). *2023 State of the Phish*. [online] Proofpoint. Available at: https://www.proofpoint.com/us/resources/threat-reports/state-of-phish. [Accessed Apr. 2023].

Shen, K., Wang, C., Guo, M., Zheng, X., Lu, C., Liu, B., Zhao, Y., Hao, S., Duan, H., Pan, Q. and Yang, M. (2021). *Weak Links in Authentication Chains: A Large-scale Analysis of Email Sender Spoofing Attacks*. [online] USENIX Association, pp.3201--3217. Available at: https://www.usenix.org/system/files/sec21summer_shen-kaiwen.pdf [Accessed 1 Apr. 2023].

T N, N., Bakari, D. and Shukla, C. (2021). Business E-mail Compromise — Techniques and Countermeasures. In: *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. International Conference on AdvanceComputing and Innovative Technologies in Engineering (ICACITE). pp.217–222. doi:https://doi.org/10.1109/ICACITE51222.2021.9404587.

Tatang, D., Zettl, F. and Holz, T. (2021a). The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws. *24th International Symposium on Research in Attacks, Intrusions and Defenses*. doi:https://doi.org/10.1145/3471621.3471842.

Tatang, D., Zettl, F. and Holz, T. (2021b). The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws. *24th International Symposium on Research in Attacks, Intrusions and Defenses*. doi:https://doi.org/10.1145/3471621.3471842.

Wang, C. and Wang, G. (2022). Revisiting Email Forwarding Security under the Authenticated Received Chain Protocol. *Proceedings of the ACM Web Conference 2022*. doi:https://doi.org/10.1145/3485447.3512228.

Yu, B., Li, P., Liu, J., Zhou, Z., Han, Y. and Li, Z. (2022). Advanced Analysis of Email Sender Spoofing Attack and Related Security Problems. In: *2022 IEEE 9th International Conference on Cyber Security and Cloud Computing (CSCloud)*. pp.80–85. doi:https://doi.org/10.1109/CSCloud-EdgeCom54986.2022.00023.

Zager, R. (2017). *A Maginot Line in Cyberspace: The Binding Operational Directive BOD-18-01 DMARC Mandate*.