The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws

Dennis Tatang
Ruhr University Bochum
Bochum, Germany
dennis.tatang@rub.de

Florian Zettl
Ruhr University Bochum
Bochum, Germany
florian.zettl@rub.de

Thorsten Holz
Ruhr University Bochum
Bochum, Germany
thorsten.holz@rub.de

ABSTRACT

Email is still one of the most common ways of communication in our digital world, the underlying *Simple Mail Transport Protocol* (SMTP) is crucial for our information society. Back when SMTP was developed, security goals for the exchanged messages did not play a major role in the protocol design, resulting in many types of design limitations and vulnerabilities. Especially spear-phishing campaigns take advantage of the fact that it is easy to spoof the originating email address to appear more trustworthy. Furthermore, trusted brands can be abused in email spam or phishing campaigns. Thus, if no additional authentication mechanisms protect a given domain, attackers can misuse the domain. To enable proper authentication, various extensions for SMTP were developed in the past years.

In this paper, we analyze the three most common methods for originating DNS domain email authentication in a large-scale, longitudinal measurement study. Among other findings, we confirm that *Sender Policy Framework* (SPF) still constitutes the most widely used method for email authentication in practice. In general, we find that higher-ranked domains use more authentication mechanisms, but sometimes configuration errors emerge, e.g., we found that *amazon.co.jp* had an invalid SPF record. A trend analysis shows a (statistically significant) growing number of domains using SPF. Furthermore, we show that the *Domain-based Message Authentication, Reporting and Conformance* (DMARC) distribution evolved significantly as well by increasing tenfold over the last five years. However, is still far from being perfect with a total adoption rate of about 11%. The US and UK governmental domains are an exception, given that both have a high adoption rate due to binding legal directives. Finally, we study *DomainKeys Identified Mail* (DKIM) adoption in detail and find a lower bound of almost 13% for DKIM usage in practice. In addition, we reveal various flaws, such as weak or shared duplicate keys. As a whole, we find that about 3% of the domains use all three mechanisms in combination.

KEYWORDS

DNS, Measurement, SPF, DMARC, DKIM

ACM Reference Format:
Dennis Tatang, Florian Zettl, and Thorsten Holz. 2021. The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws. In *24th International Symposium on Research in Attacks, Intrusions and Defenses (RAID '21), October 6–8, 2021, San Sebastian, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3471621.3471842

1 INTRODUCTION

The exchange of email messages based on *Simple Mail Transport Protocol* (SMTP) is a standard communication channel on the Internet. SMTP does not provide any security features; thus, an attacker can easily abuse the protocol to launch spam, phishing, or other types of social-engineering campaigns. Even current malware like Emotet campaigns use spoofed E-mails for further spreading. In particular, sender identities cannot be verified, which enables different types of impersonation/spoofing attacks. This leads to fake emails representing a major Internet problem with an estimated 6.4 billion fake emails per day [42]. To combat this ever-increasing problem, several extensions for SMTP were developed to improve the security and especially authentication in the context of emails. For example, PGP and S/MIME provide a way to make email communication more secure for individual messages if both the sender and receiver of the message actively use these applications. Other techniques can be configured directly by the domain owner. First, encryption techniques, such as *Transport Layer Security* (TLS), encrypt the communication channel to guarantee the integrity of messages. In addition, several authentication mechanisms were developed to ensure authentic information who sent a specific message. More specifically, three DNS-based authentication methods were developed in the past years: *Sender Policy Framework* (SPF), *DomainKeys Identified Mail* (DKIM),

RAID '21, October 6–8, 2021, San Sebastian, Spain
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9058-3/21/10.
https://doi.org/10.1145/3471621.3471842

and *Domain-based Message Authentication, Reporting and Conformance* (DMARC).

Previous work already examined the distribution of SPF and DMARC adoption on the Internet in 2015 [11, 13] and most recently in January 2018 [19]. However, some time has passed since the last set of thorough measurements and email spoofing continues to be a major problem. Hence, we argue that we need to regularly measure, determine how much progress is being made, and make recommendations on how to move forward. In our work, we do not simply intend to reflect the current state of affairs again, but also to draw attention in particular to the trends in development. In addition, we think that this topic also needs to be revisited given the growing importance of this topic: a binding directive has been published by the Department of Homeland Security (DHS) which states that all official US domains should use DMARC with the *Reject* policy starting October 16, 2018 (BOD-18-01) [4]. There is also a British counterpart to the directive [40] and a recommendation for the EU to use DMARC for administrative domains [12]. Nevertheless, there are still examples where these basic methods are still not used, e.g., most of the 2020 US presidential candidates did not use DMARC for their campaign domains [41]. We believe that Internet-wide protocol measurements are necessary to understand the current state of adoption better and to provide guidance on how it can be improved. Furthermore, we understand this work as a wake-up call to further increase the support of these protocols.

In this paper, we analyze the three most common DNS-based authentication methods in a large-scale, one and a half year long measurement study motivated by a case study and discussion on design limitations. On the one hand, we implemented a DNS crawler that collects and stores the matching TXT records for SPF and DMARC over time. On the other hand, we analyzed email dumps to learn common patterns of selectors of DKIM keys to implement a second DNS crawler for measuring usage of DKIM in the wild. We performed our measurement study between December 2018 and May 2020 with the top 1 million domains listed in the three lists Alexa [2], Majestics [27], and Tranco [25]. Our results show that SPF and DMARC increase significantly in use, thus, we are making progress in this area. Furthermore, the results confirm that SPF is still the most widely used authentication protocol with an adoption rate of about 50%. We verify also that well-known, higher-ranked domains more often support mail authentication, but implementation errors can occur so that domains like *amazon.co.jp* or *imgur.com* have invalid SPF records. In case of the top 100 domains, we assume some kind of saturation (on average up to 89%), since we could not measure any changes over time. Additionally, we demonstrate that the used policies improved over the last few years and a graph-based analysis reveals that Google and Microsoft are the most trusted organizations on the Internet regarding SPF. We find that the DMARC adoption rate is developing well and shows high rates, especially for *.gov* and *.edu* domains. However, most DMARC records use the *None* policy (~70%), but not the *Reject* policy (~15%). Regarding DKIM our findings indicate that it is implemented by at least about 13% of all domains.

Our study provides enhanced insights into the use of the three protocols through an in-depth analysis, and we are the first to directly analyze DKIM besides SPF and DMARC with a new method to measure the adoption: for gathering DKIM keys, we need the so-called *DKIM key selector string*. Given that this selector is not public and chosen by the administrator, it can vary depending on the domain. By analyzing email dumps, we can learn how DKIM key selector strings are chosen in practice and this allows us to generate a list of potential selector strings that we can use in a measurement study to obtain a lower bound of DKIM adoption in the wild. Furthermore, this analysis enables us to collect DKIM keys and study their properties. Amongst other findings, we detected 4,312 weak keys with a key length of 384 bits (66 times) or 786 bits (4,246 times), and 2,302 duplicated keys which were used by 654,089 domains in total. Overall, we find that only 3% of the domains examined use all three mechanisms together.

In summary, we make the following key contributions:

1. We evaluate 25 popular email providers in a case study and discuss potential flaws that enable successful attacks against authentication methods for SMTP.
2. We study the current state of SPF, DKIM, and DMARC deployment in practice by conducting a large-scale, one and a half year long measurement study based on three different domain datasets and compare our findings to previous work on this topic. In particular, we focus on how the adoption state evolves over time.
3. We perform a comprehensive analysis of SPF usage, e.g., by creating two graphs based on the requested SPF records.
4. We contribute new insights on DKIM keys, in particular cryptographically weak keys and duplicate keys shared among multiple domains.

2 DNS-BASED EMAIL AUTHENTICATION

2.1 Sender Policy Framework (SPF)

SPF allows publishing hosts that are authorized to send emails on behalf of a given domain (RFC 7208 [20]). To use SPF, a domain owner specifies a range of hosts that are authorized to send emails on its behalf in a DNS TXT record. When receiving an email, the recipient can validate via a DNS request whether an authorized host sent the received email or not. Moreover, the recipient extracts the sender's SPF policy and decides whether to reject the message or not.

A SPF record contains the version and all IP addresses that are authorized to send messages on behalf of the domain. There are different mechanisms to define trusted servers. On the one hand, single IP addresses, as well as whole address ranges, can be configured directly. On the other hand, the *include* mechanism makes it possible to combine several SPF records. This allows a complete SPF record to be composed of several individual SPF records.

If an SPF verification is successful, the message is forwarded to the recipient. If it fails, the further processing of the message depends on the defined SPF qualifiers. Four different qualifiers exist: *pass* (default), *fail*, *softfail*, and *neutral*. The verification process can also lead to errors. If the syntax is not correct, this leads to a *permerror*, e.g., if a unresolved domain is *include*d that does not exist.

2.2 DomainKeys Identified Mail (DKIM)

DKIM offers the possibility to check received email messages for modifications or spoofing (see RFC 6376 [22]). To use DKIM, a sender must attach a DKIM signature to the message header. The following tags are required in every DKIM signature. `v=` indicates the version of DKIM, `a=` indicates the used algorithm for generating the signature, `s=` indicates the selector name that is necessary to gather the public key in DNS, `d=` indicates the used domain, `h=` is a list of header data which will be used during the signing algorithm to create the hash in `b=`, `b=` is the hash data (DKIM signature) Base64 encoded, and `bh=` is the hash of the message body.

The full message then includes the entire content of the email, the header signed with the private key of the sender's domain, and the selector string for the DKIM DNS request. Thus, the recipient can obtain the public key via a special DNS request and verify the signature. The DKIM DNS records are stored in the following format: `selector._domainkey.domain`. The domainkey is a fixed string, and the selector is a randomly chosen string by the domain owner. The receiving email server initiates signature verification. In the first step, the key is requested via the specified selector string and the corresponding domain via DNS. If no public key exists or if this key is revoked, the message is treated as if it had no DKIM signature. If a public key is retrieved, the signature can be verified.

2.3 Domain-based Message Authentication, Reporting and Conformance (DMARC)

DMARC is based on both SPF and DKIM. In DMARC, it is possible to propose policies for handling SPF and DKIM (see RFC 7489 [23]). Again, the DNS TXT record is used to publish that the domain email authentication is supported and which policies are defined, such as how the authentications are to be handled. For the DMARC verification, if available, a DKIM and a SPF verification are performed, and the results are stored for the DMARC verification. SPF and DKIM perform their validation on two different aspects of the message. These values might be different, but DMARC requires them to be equal to pass. If not, the specified DMARC policy will be applied. Three different policies exist: *none*, *quarantine*, and *reject*. The *none* policy states that the message will be delivered to the receiver and treated like any other email. The *quarantine* policy marks the message as suspicious, but it will be delivered. Depending on the receiver's client, the message will be placed in the spam folder. The *reject* policy states that the message will be rejected by the email server and not delivered to the receiver.

3 MOTIVATIONAL CASE STUDY AND DISCUSSION

3.1 Case Study: Free Email Providers

As part of the first case study, we analyzed 25 popular web-based email providers. For the selection of the providers, we combined different sources. We obtained all free email services (17 providers) analyzed by Hu et al. [19]. Second, we expanded this by adding domains from the same services, e.g., *hotmail.com* and *outlook.com*. Third, we searched on Google for further popular email providers. The most popular email providers of the US are Gmail, Yahoo, Hotmail, and AOL. Gmail alone has about 1.5 billion active users [15]. Thus, the results of this case study represent a significant fraction of people worldwide. Table 7 in Appendix A summarizes our findings. We found that only one email provider does not support any authentication (*t-online.de*). Otherwise, SPF is at least supported by all others and DKIM is supported by 18 out of the 25 studied email providers. The examined providers least support DMARC, 14 out of 25 implement it.

The policy for SPF was mostly configured as *softfail* or *hardfail*, and only in three cases as *neutral*. None of the providers uses the *pass* mechanism. The number of authorized subnets and IP addresses differs significantly: while Firemail, for example, only authorizes a single IP address via SPF, there are others like AOL or Hotmail that publish hundreds of thousands of IP addresses in their SPF records. Regarding DKIM, most providers with the top-level domain *.de* do not offer DKIM as *web.de*, *gmx.de*, or *freenet.de*. Interestingly, the premium version of *gmx.de* and *web.de* support DKIM. The observed key lengths are good (1024 or 2048 bits). All providers with DKIM use the *rsa-sha256* algorithm for signing and verification. In case of DMARC, eight providers use testing mode with the *None* policy (e.g., *gmail.com*), two use the *Quarantine* policy, and four use the strictest *Reject* policy.

The overall results indicate that most email providers are aware of email authentication and started to use it. We found that DMARC support has the biggest need to catch up, especially because it should not be applied with the *None* policy given that no security gain is obtained with this configuration. This leads to the question of what the overall adoption rate and configuration looks like on an Internet-wide scale.

3.2 Discussion: SPF and DKIM Design Limitations

We learned that most popular email providers implement at least SPF. Thus a simple spoofing attack is more difficult to perform. However, a design limitation of SPF is that it verifies the *MAIL FROM* header only. This header is inside the *FROM* header, which is not verified. Therefore, the verification does not check if both values contain the same domain, allowing to bypass SPF verification. This is possible for all domains that implement SPF only. So the research question arises: how widespread is the usage of SPF in practice?

Moreover, DKIM provides additional security compared to SPF because it attaches a signature of the header and body to the email message. A private key is used for signing and a public key for verification. When using DKIM, it is crucial to use secure keys. Keys under a length of 1024 bits are considered insecure because they can be factorized (i.e., an attacker can try to calculate the private key from the public key). Another typical problem in practice is key management; e.g., one should not use duplicate keys. Otherwise, other people can also sign messages successfully. The research question is: how many weak keys or duplicate keys are used in the wild?

A design issue with DKIM is that a client only knows if a domain supports DKIM when the DKIM signature header is included in a received email. An attacker who wants to spoof the email address of a domain that uses DKIM can send an email message without the DKIM signature header. The recipient checks the email headers, and if it does not find a DKIM signature header, it leads to the assumption that this domain does not support DKIM at all, and no DKIM verification can be performed. This attack vector does not target the DKIM verification itself, but the problem that a client cannot verify if the sending domain does not implement DKIM or an attacker just removed the DKIM headers. The described attack vector applies to all domains implementing DKIM without DMARC. This brings us to the third research question: what does the adoption of DMARC looks like overall?

4 DATA GATHERING AND MEASUREMENT SETUP

To obtain a general picture of the adoption states and their development, we performed a one and a half year long measurement study. More specifically, we measured the described DNS-based techniques from Section 2 on a large scale, pursuing two distinct data gathering approaches:

(1) We implemented a DNS crawler that collects and stores the matching TXT records for SPF and DMARC once a month.
(2) By analyzing email dumps, we learned common patterns of selectors of DKIM keys, and subsequently applied these to generate a list of selectors with which we can request as many DKIM keys as possible.

4.1 Data Sets

For our data collection, we need domain lists for extracting the DNS information regarding mail authentication methods that are as consistent and broad as possible. According to Scheitle et al. [34], individual top lists fluctuate and may bias results. Thus, we consider a fixed list of domains for each analyzed top list. Previous works usually use the Alexa top 1 million list as basis for further analysis [2, 17, 33, 34]. In addition, other papers use the Majestics top list [5, 27]. Moreover, since February 2019, a third list combining several lists called *Tranco* was published [25]. In our work, we use all three top lists to be as comparable as possible with other measurement studies and to provide an as complete survey as possible since different top lists include different domains.

There are only 214,015 common domains in all three top lists, thus, the intersection is comparatively small due to the significant differences in the methods used to create the different lists. Altogether, we consider 2,039,279 unique domains in the union of all lists. We analyzed monthly snapshots of the Majestics list from December 2018 to May 2020; for the Alexa list we collected snapshots from December 2018 to May 2020, and for the Tranco list, we have snapshots from March 2019 to May 2020. (Note: The scan from January 2020 is unfortunately unusable because we had memory problems. It is not included in all following analyses.) The data was collected periodically once a month and is available after publication, including the crawler implementation.

4.2 Data Crawlers

In case of SPF and DMARC, we implemented a crawler to gather DNS TXT records for a list of domains and save them for later analysis. Data collection for DKIM is not as easy since to collect the keys, we have to know the key selector string. For this reason, we analyzed several email dumps and subsequently generated a preferably complete list of valid selectors. We tested this list for all domains and collected corresponding DKIM keys and saved them for further analysis.

4.2.1 SPF and DMARC. The SPF and DMARC information is stored in DNS TXT records and must start with a valid version tag. In case of SPF, this is *v=spf1* and in case of DMARC, this is *v=DMARC1*. If no DNS TXT record starts with a valid version tag, this means that the method is not implemented for the given domain.

With SPF records, we have to check for the *include* or *redirect* mechanisms, as we have to collect all domains to get a complete SPF record.

4.2.2 DKIM. For gathering DKIM keys, we need the DKIM key selector string. This selector is not public and not a fixed string, it can vary depending on the domain. The only way to obtain a selector for a domain is to receive an email with a valid DKIM signature. To address this problem, we collect selectors from email dumps. For this, we examined the Linux kernel mail archive from the introduction of DKIM in 2011 until the end of 2018 (2018/12/22) with about one million emails. A potential limitation regarding this email dump is that these emails might not be representative for all emails. But this is not a big deal, as there might even be more technical or even security-savvy people sending emails, the number of identifiable selectors is maybe even higher. However, in addition, we included a private GMX and Gmail account from one of the authors with about 7,000 received emails from 2014/11/16 to 2018/12/06 in the analysis. From all analyzed emails, around 35% contained a DKIM signature header. Although a selector can be randomly selected, we noticed that most selectors are built based on certain patterns, such as a date or a fixed string like *selector1*, *default*, or *mail*.

After analyzing the dumps and having a sense for the structure of various selectors, we generated a list with 3,498 possible selector strings in a second step. To review our generated list, we analyzed a spam email archive, including more than one million emails. We use this selector list to find as many valid DKIM records as possible from the examined domain lists. Note that completely random selectors cannot be identified with this method. As a result, we can identify only a lower bound for the use of DKIM in the wild. Note that we provide the script for collecting the selectors of the Linux kernel mail archive, but unfortunately the private email accounts represent data we cannot share. However, the resulting list of selectors will be provided for reproducibility.

5 MEASUREMENT RESULTS

5.1 SPF Measurements

5.1.1 Adoption Rates. To obtain a long-term overview, we included the results from previous work as comparative values [11, 13, 19]. We find that the adoption rate of SPF is slowly, continuously rising on all domain lists; nowadays up to 50.7% of the examined domains implement SPF. Note that the comparison with previous work has the limitation that not the same set of domains were checked, but it is enough to obtain a broad overview and feeling for the overall adoption rates and their evolution. In total, we find that the general adoption rate increased by around 25% since 2015, but remains just around half of the examined domains.

Next, we are interested in the adoption rate for higher-ranked, more popular domains. We suspect that higher-ranked domains are potentially better maintained as they would have the economic incentive to defend against spoofing that hurts their brand, or phishing that targets their customers and therefore more likely to use security mechanisms like SPF. The top 100 domains of Majestics, Alexa, and Tranco show a higher rate of SPF usage (up to 90%) compared to the rest of the top 1 million domains, validating our assumption. Moreover, no further changes are visible in this cases, but it seems to remain stable for all lists. For the Alexa list, on average 88 of the top 100 domains implement SPF. The remaining domains that have not implemented SPF include, e.g., *cnblogs.com*, *office.com*, and *xinhuanet.com*. It might be surprising that the top-ranked domain *office.com* owned by Microsoft has not implemented SPF. One primary reason is likely that the domain does not operate an email server and has published a DMARC record, so an attacker cannot easily abuse this domain for spoofing.

We found that 12 out of 88 domains published an invalid SPF record (see Table 1 for details). It is surprising that *amazon.co.jp* and *imgur.com* use an SPF record where a non-existing DNS TXT record is used as the include mechanism. The inclusion of invalid or non-existing SPF records always leads to a *permerror*. Both domains have implemented DMARC, so the domains are protected, however, the incorrect SPF record can also lead to a rejection of an email if the DKIM signature is not present or also incorrect. The domains *espn.com* and *wiki.com* have another issue: they exceeded the number of total DNS lookups. According to RFC 7208 [20], the SPF implementations must limit the total number of DNS queries to 10, and if this value is exceeded, the validation must return a *permerror*.

For the top 1,000 domains, the adoption rate is still high with on average 74% for Alexa, 69% for Majestics, and 77% for Tranco. If we compare these values with the top 10,000 domains, we can see that the adoption rate is somewhat stable

Table 1: Domains from the top 100 Alexa list with no or invalid SPF records. There are three types of invalid SPF records, no use of SPF (NULL), too many DNS lookups in the record, and use of an illegal record.

Domain	SPF error	MX
amazon.co.jp	Perm Error: No valid SPF	yes
cnblogs.com	NULL	yes
detail.tmall.com	NULL	no
espn.com	Perm Error: Too many lookups	yes
gmw.cn	NULL	yes
imgur.com	Perm Error: No valid SPF	yes
login.tmall.com	NULL	no
office.com	NULL	no
pages.tmall.com	NULL	no
porn555.com	NULL	no
wiki.com	Perm Error: Too many lookups	yes
xinhuanet.com	NULL	yes

Table 2: The number of invalid or valid SPF records and their policies across related work, the first, and last Alexa scans. The use of the *hard fail* policy increased and the *neutral* policy use decreased.

Status	04/2015	12/2018	05/2020
SPF Total	n.a.	537,002 (100%)	578,713 (100%)
SPF Valid	401,356	459,423 (85.6%)	503,310 (87.0%)
SPF policy: Hard fail (-)	84,801	142,218 (26.5%)	173,005 (29.9%)
SPF policy: Soft fail (~)	226,117	265,609 (49.5%)	283,274 (48.9%)
SPF policy: Neutral (?)	80,394	50,824 (9.5%)	46,472 (8.0%)
SPF policy: Pass (+)	n.a.	772 (0.1%)	559 (0.1%)
SPF Invalid	n.a.	77,579 (14.4%)	75,403 (13.0%)
Too many DNS lookups	n.a.	38,750 (7.2%)	34,800 (6.0%)
Two or more type TXT spf record found	n.a.	18,659 (3.5%)	18,279 (3.2%)
No valid SPF record for included domain	n.a.	9,571 (1.8%)	10,347 (1.8%)

and decreased only to around 62% for Alexa, 65% for Majestics, and 67% for Tranco. This indicates that higher ranked domains, which are mostly operated by larger organizations, seem to take the security issue of spoofing more serious than smaller organizations with lower ranked domain names. It should be noted that known domains are probably more likely to be spoofed than unknown ones. However, this does not apply to all highly ranked domains, as multiple domains still do not implement SPF.

Further analyses, e.g. on the distribution of SPF on servers with or without MX records, or the detailed distributions for the higher-ranked domains, are excluded due to space reasons.

5.1.2 *Trend Analysis.*

In a first step, we performed the trend analysis by plotting the measured values as a linear trend function. For the Alexa top list, we calculated a rise of about $m = 1754.3$ per month (Tranco: $m = 653.4$, Majestics: $m = 280.74$).

To confirm that our measured values represent a significant increase over time, we performed a t-test (paired two samples for means). Our null hypothesis is that there was no relevant increase in use since the beginning and the end of our measurement period. We used the measurements from December 2018 to March 2019 (first measuring points) and the last measurements from February 2020 to May 2020 (last measuring points) as input for our t-test.

The p-value is < 0.05 and thus the null hypothesis is rejected. There is a significant difference between the average values at the beginning and at the end. Thus, the use of SPF has increased significantly in course of one year. If we look at the valid SPF records in the next step, we see the same situation, i.e., a significant increase in valid SPF records (p-value ≈ 0.002). In case of invalid SPF records, we also find a significant difference between the first and last measurement points (p-value ≈ 0.008), but at this point we observe a significant decrease (mean of start and end values: $[78977.5, 74583.5]$). We calculated all the t-tests also for the other two lists (Majestics and Tranco). For both lists, we obtained similar results, i.e., we observe a significant difference (increase for valid and decrease for invalid SPF records).

5.1.3 *Policies and Content.*

Next, we consider the actual content of the SPF records. We compare the number of valid SPF records (including their corresponding SPF policy) and the number of invalid SPF records over time from all datasets. (see Table 2 for details on numbers from related work [11], the first, and the last snapshots exemplary for the Alexa domain list). We find that most of the SPF records are valid. However, about 77,000 records contain invalid configurations across all measurement points in case of Alexa (67,000 in case of Majestics, 75,000 in case of Tranco). The numbers decrease slightly over the measurements, which is remarkable since there is an overall increase in total SPF records. The most common reason for invalidating a record was caused by exceeding the allowed number of DNS requests. As noted, this is limited according to RFC 7208 [20] to a maximum of ten lookups and it leads to a *permerror* SPF verification result. Almost half of the invalid domains belong to this category. Another widespread issue with about 24% of the invalid records was the implementation of multiple records. A domain that publishes multiple SPF records including an obsolete SPF record will cause that the records will be invalid and no SPF verification can be performed.

It is not surprising that valid SPF records are continuously increasing over time as the total amount of SPF records increases. The number of valid SPF records grew from 401,356 domains at the first snapshot to more than 500,000 in May 2020 (+25%). Almost half of the domains with a valid SPF record implemented the *soft fail* policy and the number is continuously rising. The *hard fail* policy, which is the strictest option of all the possible options, is adopted by maximum 33.1% of the valid SPF domains. The *neutral* policy is used by around 10% of all valid SPF record and is minimally decreasing over time. Interestingly, the adoption rate for the *pass* policy is first increasing between the snapshots 12/2018 and 01/2019 and afterwards decreasing. This behavior demonstrates that most of the domains that recently implemented SPF tend to use more strict policies like *hard fail* or *soft fail*. Both policy modes are recommended as they either completely block unauthorized IP addresses or mark them as untrustworthy. A positive aspect is a minimal decrease in the *pass* policy. This policy does not improve the security in any way, as every message will pass the validation process of SPF.

5.2 Graph-based SPF Analysis

To gain deeper insights into the use of SPF, we analyze the relationships between SPF entries. This analysis is based on the snapshot from April 2019 for the Alexa list. In particular, we examined two different aspects: (1) relation between domains and their include records and (2) relation between IP address and Autonomous System (AS). In total, these results demonstrate that there is a small number of SPF include records or AS that are trusted by a large number of domains. Especially the two companies Google and Microsoft are among the most commonly used SPF include records, which means that many parties use services in those domains to send as if they were sending from those parties: almost every fifth domain with a valid SPF record included an SPF record of these two companies. However, also the SPF records of other companies like Amazon, SoftLayer Technologies, and SendGrid were added by a large number of domains. All these domains would be the perfect target for an attacker as she would be able to pass the SPF validation of a large number of domains if she gets access to a single system that is listed in the SPF record. On the other hand, we observed that a small number of domains trusts a variety of IP addresses from up to 9,000 different ASes. This implies that some individual domains trust almost the whole Internet, indicating that some individual configurations are not reasonable at all.

5.2.1 Domains and their Includes. The data set analyzed for this purpose contains all domains with a valid SPF record (481,959 domains). Of these domains, about 66% (319,349 domains) use the include mechanism in their SPF records to add external SPF records. For graph creation, each domain represents one node, each domain in the include mechanism represents a node, and each use of the include mechanism is an edge between the domain and the SPF-included domain. The two largest nodes are Google with *_spf.google.com* and Microsoft with *spf.protection.outlook.com* which both are trusted by about 30%. Many nodes with a degree of zero lead to an average degree of 0.98. The average path length is 1.68, representing one or two include records. The value of the network diameter is 10, which represents the longest path of the graph. The majority of the SPF records has a path length of one or two and only a small number of nodes has a path length of nine or ten. Larger path lengths result in an invalid SPF record as the number of DNS lookups must be limited to ten and thus are not found in our graph.

5.2.2 IP Addresses and AS. In total, all domains with a valid SPF record contain almost nine million IP addresses. Nevertheless, the number of unique IP addresses is much smaller, showing that many IP addresses are more often than others trusted. In our analyzed data set, there are 579,690 unique IP addresses. We define for the graph creation that each node is a domain, as well as an AS derived from the IP addresses. The edges reflect the relationships between the domains and the included ASes.

In a SPF record, not only individual IP addresses can be added, but also entire subnets. Therefore, we checked which subnet size is often used. We found that /32 networks, i.e., single IP addresses, are most frequently used. However, also /24 subnets are often applied (1,747,501 times). Interestingly, /8 subnets are found in 1,181 cases and subnets larger than /8 in 41 cases. We even discovered /0 in 11 cases. To make the following analysis reasonable, we have to determine a subnet size that is as small as possible, but covers as many used subnets as possible. We calculated a cumulative distribution function. We cover about 60% of all addresses with /24 and apply this subnet size in the following for the analysis of the relations between IP address and AS. This means that networks larger than /24 are divided into multiple /24 networks. This procedure does not change the results about ASes, as these are divided into larger networks. This modification only minimizes the number of duplicate connections between trusted IP addresses of a domain and its AS.

The three largest ASes are AS8075 (Microsoft), AS15169 (Google LLS), and AS16509 (Amazon). The Amazon simple email service was only used by 2%, however, many companies host their email server on Amazon Web Services (AWS), which results in a large number of trusted IP addresses for the AS of Amazon. Google and Amazon even have both a second AS that is trusted by a smaller but still large number of domains. Besides Google, Microsoft and Amazon, cloud computing providers and ISPs are among the top connected nodes. We identified that the average degree of this graph is 3.3, which means that on average domains use IP addresses from three different ASes. Since we have no connections between the ASes and only from domains to AS, the average path length and also maximum path length is one. If we take a look at the out-degree, the majority of the domains uses IP addresses from one to six different ASes. However, some trust an enormous number of IP addresses from up to 9,000 different ASes.

5.3 DMARC Measurements

5.3.1 *Adoption Rates.*

According to our most recent scan, DMARC is only used by up to 11.5% of all examined domains. Nevertheless, the adoption rate of DMARC is rising continuously ; in comparison, it rises faster than SPF. The total adoption rates, however, are five times smaller compared to the SPF measurements. In particular, e.g., in the Alexa scans, we notice a rise of more than ten times between January 2015 and December 2019. Regarding the adoption rates for higher-ranked domains, again, we find that higher-ranked ones implement DMARC with a higher probability similar to the behaviour we already observed for the SPF measurements. However, the adoption rate drops rather quickly and especially low-ranked domains rarely adopt DMARC.

5.3.2 *Trend analysis.*

To show that our measured values represent a significant increase over time, we performed a t-test (paired two samples for means) for all our scans. We used our first four measuring points and the last four measuring points to perform the t-test and set an alpha value of 0.05 for each individual list. The result of the t-test is that the p-value is in each case smaller than 0.05. This allows us to reject the null hypothesis on all lists and conclude that we measured a significant difference, i.e., a significant increase in the use of DMARC in all measured domain lists between our first and last conducted scans.

5.3.3 *Policies and Content.*

Considering not only the adoption but also the content of the DMARC records, we observe that almost one percent invalid DMARC records exist. This implies if DMARC is in use, mostly it is implemented correctly. However, we notice that the *none* policy accounts for about 75% of all used policies. The most recommended *reject* policy is only applied by about 15%. The *quarantine* policy is the least used with about 14%. Table 3 summarizes the number of invalid and valid DMARC records, including their applied policies across related work [13], our first, and last Alexa measurements.

Table 3: Invalid or valid DMARC records and policies across related work, the first, and last Alexa snapshots.

Status	01/2015	12/2018	05/2020
DMARC Total	n.a.	72,303 (100%)	115,756 (100%)
DMARC Invalid	n.a.	705 (1.0%)	1,050 (0.9%)
DMARC Valid	9,700 (0.97%)	71,598 (99.0%)	114,706 (99.1%)
Policy: None	7,300 (0.73%)	54,503 (75.4%)	78,591 (67.9%)
Policy: Quarantine	800 (0.08%)	7,877 (10.9%)	18,263 (15.8%)
Policy: Reject	1,600 (0.16%)	9,218 (12.7%)	17,852 (15.4%)

5.4 Governmental Domains

Finally, we consider government domains and their use of DNS-based email authentication mechanisms. Especially the *.gov* top-level domain is exclusively used by authorities in the USA. A primary goal of this analysis is to find out if *.gov* domains comply with the Binding Operational Directive 18-01 [4] and implement all email authentication methods according to the directive. A blogpost from October 2018 showed that one year before the directive came into force (October 2017) only 20% of US federal domains had implemented DMARC, compared to 74% at the time the directive came into force [32]. Internationally there are comparable directives in the UK for *.gov.uk* domains [40] and a recommendation for the EU to use DMARC for administrative domains [12]. Other countries with specific top-level domains for their government institutions are, e.g., China with *.gov.cn* and France with *.gov.fr*. In the following, we compare the usage rates among government agencies. In addition, we include Japan and Germany, as these are two economically significant countries. There are significant differences between the countries. The USA has the highest adoption rate of DMARC (88%) and SPF (92%) regarding the newest scan. Especially from 2018 to 2019, both methods were implemented significantly more widely. However, some domains are still missing to meet BOD-18-01. A domain that is not implemented correctly is, e.g., *cia.gov*. The official domains of the UK are in second place with their adoption rate and are only slightly worse. There are major differences between other countries. Interestingly, SPF usage in Japan is very high, but DMARC usage is shallow. The results underline that regulatory directives in the US and Britain help to increase the deployment of these security mechanisms in a specific set of domains (government domains) significantly. Pure recommendations that also exist in the EU, however, only lead to small improvements due to their non-binding character.

5.5 DKIM Measurements

To complement our study, we performed two comprehensive analyses of the DKIM records and their corresponding DKIM keys collected in August 2019 and January 2020. This study is a so far new perspective of DKIM usage in the wild. The results of both analyses are comparable; the following results are based on our first collection in August 2019.

5.5.1 *Created Selector List.*

To perform DKIM measurements, we require a practical DKIM selector list to extract DKIM keys from the domains. From the analysis of the email dumps,

Table 4: Top 10 most seen DKIM selectors.

Selector	Total Number	Selector	Total Number
default	45,521 (4.48%)	mailjet	5,018 (0.50%)
s1	26,435 (2.60%)	selector1	3,821 (0.38%)
google	25,590 (2.51%)	20140924	1,882 (0.19%)
mail	13,618 (1.34%)	s1024	964 (0.10%)
dkim	7,190 (0.71%)	test	426 (0.04%)

we learned three types of selectors. First, selectors with syntax [year]-[month]-[day] with and without hyphen. Second, selectors consisting only of four numbers (year). Third, a group of 19 different strings. In our created list, there are 3,498 selectors in total, 3,468 entries belong to the first group starting with the year 2014. For the second group there are 11 entries from 2008 to 2018, and the strings for the last group are: `default`, `dkim`, `google`, `selektor`, `selector`, `selector1`, `s1024`, `s2048`, `s512`, `s1`, `postout`, `alpha`, `beta`, `gamma`, `test`, `mandilla`, `mailjet`, `mail`, and `mail2`. Table 4 shows the top 10 most seen DKIM selectors. Note that the distribution has a long tail, the top 10 only represent 12.9% of all selectors.

5.5.2 Spam email DKIM Usage. To verify our list, we retrieved 1,232,160 spam emails from a spam archive available at http://artinvoice.hu/spams/. The emails were collected over 10 years (from 15/5/2009 to 20/08/2019). Altogether, the files were over 7 GB zipped. We found 26,062 emails with DKIM-Signature headers in the whole archive. 23,489 (90%) include valid DKIM selectors. This represents about 2% of all emails and shows that a few individual cases also use DKIM for spam messages. About 89% (20,789) of the selectors are already in our list. Completely randomized selectors, selectors without patterns, or selectors not found in our generated list are at 11% (2,700). We argue that our created selector list is working properly and thus we use it in the following for collecting DKIM keys.

5.5.3 Gathered DKIM Keys. The DKIM keys were determined by analyzing email dumps and the selector list with 3,498 selector strings based on the results of the analysis presented above. In total, we gathered 998,336 DKIM records from 113,855 different domains. Altogether, we identified 757,470 valid keys and 259,720 invalid ones. As noted earlier, the adoption rate (11.4%) of DKIM may be higher as truly random selectors were not included in the created selector list. The same applies to the analysis of received emails, as the dataset did not contain emails from all the top 1 million domains. In total, we identified 105,683 unique DKIM keys.

Among the identified keys, we discovered a total of 4,312 weak keys. A key is weak if it is not RFC 8301 [21] compliant and therefore its key length is less than 1,024 bits. We also found 2,302 duplicate keys, i.e., keys used by more than one domain. We found that a total of 29,308 keys were invalid and 211,215 were withdrawn. Withdrawn keys contain an empty public key. As a result, they can no longer be used for signing. Invalid keys are DKIM keys that are flawed and thus do not provide a valid signature key.

5.5.4 Weak Keys. We first focus on the key length. We divided weak keys into keys with 384 bits (66 times) and 786 bits (4,246 times) key length. Weak keys are critical since they can be factorized in a reasonable amount of time and effort. Once an attacker is able to factorize a public key, she can obtain the private key and sign messages on behalf of a particular domain. The 384 bit keys use three different selector strings only. 56 domains use as selector *dkim*, nine domains *default*, and one domain *mail*. The most seen top-level domain is *nl* with more than half of all domains (36 domains). The 768 bit keys use 3,366 unique selector strings. A conspicuous domain here is *audiomicro.com*, which alone makes up 3,366 of the DKIM records with 768 bit keys. This domain also uses a date selector string which is apparently renewed every day. If we remove this domain, 880 domains remain. In 766 cases, these use *default* as selector string. Fortunately, the majority of more than 99% of the keys have either a key length of 1,024 (693,245 times) or 2,048 bits (59,910 times). These keys conform to the most recent RFC 8301 document that specified a recommendation for the used cryptographic algorithm for signing and the key length for DKIM [21].

Most domains with a weak key are ranked at lower positions and represent a negligible amount of the top 1,000 domains. However, individual cases that affect a large number of users exist. One case is Facebook with the domain *facebook.com*. This domain is using a weak key with 768 bits length and *default* as the DKIM selector. Other domains using weak keys as illustrated in Table 5, which represents the top 10 highly ranked domains with weak keys.

5.5.5 Shared Duplicate Keys. Next, we consider duplicated keys in detail. In total, we detected 2,302 duplicated keys which were used by 654,089 domains. In some cases, duplicated keys were used across different top-level domains, but with the same second-level domain. For example, Amazon uses the same DKIM key for the domains *amazon.com*, *amazon.de*, and *amazon.co.uk*. This behavior is not critical, as these domains belong to the same organization or company. A second scenario is a duplicated key for the same domain with different selectors. Obviously, this is not critical either. Therefore, we remove all domains that share keys only with the same second-level domain, but different top-level-domains and domains sharing keys among their alias domain names. This results in 13,373 remaining domains with 1,881 keys.

Table 5: Overview of the top 10 most popular domains contained in the Alexa list with a cryptographically weak DKIM key

Domain	Key Length (Bits)	Alexa Rank
facebook.com	768	2
surveymonkey.com	768	443
motherless.com	768	1145
asu.edu	768	1864
commentcamarche.com	768	2706
dailypakistan.com.pk	768	3057
androidauthority.com	768	3258
ubnt.com	768	3281
jpnn.com	768	3349
jobvite.com	768	3506

The duplicated keys are mostly shared across two different domains only (1161 groups). However, even larger groups of domains with the same key exist. Between three and 13 different domains there are still 630 groups. Between 14 and 30, the number of groups is reduced to 46. Over 30 domains are in 44 groups. Figure 1 in Appendix A features the nine largest groups with the number of domains and five higher-ranked domains, according to Alexa. By manual search, we tried to learn more about the individual groups. One specific key was used by 1,515 different second-level domains, which represent the biggest group of domains sharing the same key (DKIM Key Group 1). Interestingly, this key was displayed on multiple websites where the functionality of DKIM was explained, and this key was used as an exemplary DKIM key [36]. The reason why so many domain owners used this exemplary key is unclear, as the domains cannot test the functionality of DKIM if they do not possess the private key. Otherwise, if the private key is known, everyone can sign messages for those 1,515 domains.

At DKIM Key Group 2, four of five displayed domains have a CNAME record for `s1._domainkey.[domain]` which points to `s1acc903393.domainkey.freshdesk.com`. This again points to a domainkey subdomain of *sendgrid.net*. In this case, Freshdesk [14], a customer service software, is used, which in turn provides DKIM for its customers using Sendgrid. The same applies for DKIM Key Group 4: all domains have a CNAME record pointing to a domainkey subdomain of *sendgrid.net*. These are thus again Sendgrid customers. A description of how to configure DKIM at Sendgrid is available online [35].

For DKIM Key Group 8, all displayed domains have a domainkey CNAME subdomain pointing to a domainkey subdomain of *atlassian.net*, which further points to a subdomain of *sendgrid.net*. Atlassian provides software solutions for collaborative work and customer support. The domains in this group are Atlassian customers who have configured DKIM by using them. Atlassian, in turn, uses Sendgrid [3].

The domains in DKIM Key Group 3 all belong to *jimdo.com*. Jimdo is a provider of website development tools and Internet services. They all use the same nameservers, so it indicates that all domains in the zone of Jimdo use the same key, because *jimdo.com* also manages the email service for all domains in this zone. Similarly, all displayed domains from DKIM Key Group 5 contain nameservers of *mchost.ru* in their NS records. Mchost is a Russian provider of web services.

For DKIM Key Groups 6, 7, and 9, we have no additional information on why they share the same key.

To sum up, we have three groups (2, 4, and 8) where *sendgrid.net* was used to set up DKIM. In two cases with intermediary software (groups 2 and 8), thus there are three different duplicate DKIM keys used multiple times. Then we have one group (1) using a DKIM key from a tutorial and two groups (3 and 5) which are related to web service providers. For three groups (6, 7, and 9), we have no further information as to why the same key is used. In some cases, e.g., Sendgrid, Jimdo, or Mchost, sharing the DKIM key is not security critical as long as the domain owners trust these services/providers. The case of Group 1 is more critical as this is likely to have happened unintentionally. The groups without further information are also remarkable, as a shared duplicate key may not be intended as well.

As expected, most of the domains using a shared duplicate key are in general ranked low. However, some higher-ranked domains like *yahoo.com*, *adobe.com*, and also known domains like *symantec.com*, and *myspace.com* are included (see Table 6). However, with the somewhat more well-known domains, it should be noted that they share the same key across different groups in the same company:

- In the case of *yahoo.com*, this domain shares the same key with other Yahoo sites (*yahoo.com.tw*, *ymail.com*, *yahoo-inc.com*, *yahoomail.com*, and *yahoo.co.id*) and two other domains, *sky.com* and *ovi.com*. It is a 2048 bit key, and the selector is always *s2048*.
- The domain *live.com* shares its DKIM key with five other domains. In four cases, these are Microsoft services (*outlook.com*, *msn.com*, *onedrive.com*, and *hotmail.com*) each with the selector String *selector1*. In one case, these domains also share the key with *nuangel.net*. We did not find a connection to Microsoft for this domain.
- The domain *yandex.ru* also shares its key with other domains associated with Yandex (*ya.ru*, *yandex.by*, and *yandex.kz*). All use *mail* as selector string.
- Another interesting domain is *nytimes.com*, which shares the same key with 70 other domains. These

Table 6: Overview of most popular domains with duplicated DKIM key

Domain	Ranking	Domain	Ranking
yahoo.com	8	vimeo.com	132
live.com	13	chaturbate.com	141
yandex.ru	16	indeed.com	143
ebay.com	35	salesforce.com	147
yahoo.co.jp	39	instructure.com	152
msn.com	52	ebay.co.uk	160
xvideos.com	58	ebay.de	163
adobe.com	75	target.com	168
nytimes.com	116	txxx.com	171
bestbuy.com	128	mercadolivre.com.br	180

domains include *salesforce.com*, *hbo.com*, and *rollingstone.com*. All use the same selector string *s1*.

- A last finding is that *ebay.com* is using three selectors *dkim*, *google*, and *s1* and *adobe.com* is using two selectors *default*, and *s1*.

Overall, we observe that many cases where the same key is used multiple times are not critical because most domains with duplicate keys belong to the same organization.

5.5.6 Threats to Validity. A first limitation of the email analysis method is that our email archive did not contain emails from all the top 1 million domains. Thus, our analyzed email dump is biased because we do not have representative email users, especially with the communication partners in the Linux kernel email dump. A second limitation is that our generated selector list contains only 3,498 selectors. Thus, it did not include any randomly generated selector. Better results might need more extensive lists. However, we tried to get as comparable results as possible by combining both approaches to the best of our ability at a reasonable expense.

5.6 Combination of Authentication Methods

Based on the scan from August 2019 of the Alexa list, 563,474 domains have SPF records, 94,244 domains have DMARC records, and 113,855 domains have DKIM records, including invalid records. Finally, we consider the use of the studied technologies in combination based on the scan results from August 2019. Only 30,425 (about 3%) domains use DMARC, SPF, and DKIM together. The percentage is comparatively low and should be much higher to make fake emails more difficult. SPF and DKIM use almost 10% (98,276 domains) of all investigated domains. DMARC and SPF use about 9% (90,895 domains) and DMARC and DKIM about 3% (30,958). This shows that if DMARC is used in almost all cases, SPF is also implemented (more than 96%). If DKIM is used, SPF is often also available (about 86%). DMARC in combination with DKIM is a rather rare case quite comparable with all three methods together. This confirms the results of our first case study (see Section 3) that generally, when a method is used, it is SPF. If at least two methods are used, SPF is in most cases one of these methods.

6 DISCUSSION AND LESSONS LEARNED

As discussed in Section 3.2, using either SPF or DKIM alone makes spoofing attacks more difficult, but does not completely prevent them. We observed that SPF is the most frequently implemented method in practice, and even the adoption rate of around 50% valid records might not be satisfactory. Nevertheless, obviously the adoption rate has improved since the last measurement studies which were performed about five years ago [11, 13]. This increase is statistically significant showing that more and more domain operators are using SPF. Another takeaway is that particularly popular (highly ranked domains) use SPF much more than all the others. There even seems to be a certain saturation in the Top 100 domains in particular, where 90% of the domains use SPF, as we could not detect any changes in use over the entire measurement period.

The analysis of adoption rates of governmental domains of different countries revealed strong differences, which are likely caused by binding legal directives which enforce an adoption of security standards. It seems like political regulations help to increase the security of emails significantly. In certain areas, such a legal enforcement could enable faster adoption and reduce the risk of email spoofing and related attacks. We recommend domain holders to implement at least one DNS TXT record with the DMARC *Reject* policy to prevent misuse of the domain, even without a mail server they run themselves.

DKIM and DMARC implementation is only about a tenth. Nevertheless, we were also able to measure a significant increase at DMARC. This proves that there is a lot of progress in this area as well. However, the bottom line shows that many services only implement SPF so that impersonation is still possible for a large fraction of domains. In the end, we are still far away from stopping the spam problem with this methods. However, it is encouraging that the major mail providers are implementing the techniques much better than the average services available on the Internet.

DMARC only offers adequate protection against spoofing in combination with SPF and DKIM. The number of services which implement all three email-based mail authentication methods is shallow at about three percent. If we again compare the results of our surveyed mail providers with all the top 1 million domains, we notice that the email providers are significantly better positioned than the overall average.

Even though we could show that more and more people are taking care of DNS based email authentication, the overall spread on the top 1 million lists is still improvable as it was five years ago. Considering only the top 100 or top 1.000 domains, a certain saturation is reached there and thus the use of the methods in this cases is satisfactory. There are also clear differences per TLD, for example *.cn* with a huge increase. We hope that this work will be noticed as a kind of wake up call for all domain operators to check their email authentication methods and configurations, and that a further increase in usage will be achieved in the near future.

We argue that our measurements should be continued regularly to record further changes over time. In addition, we should try to raise awareness of the issue of email security among domain operators. More qualitative studies like the work by Hu et al. [18] on the use of security mechanisms may also provide further insights and findings in the field of email security in the next years.

6.1 Ethical Considerations

In scope of our study, we analyzed publicly available information for SPF and DMARC. We have not exploited misconfigurations we found and reported them to the affected parties. When investigating DKIM, we did not collect public information. An attacker could exploit especially weak keys. For this reason, we did not state the domains which use a 384-bit key as such short keys can easily be factorized even with limited resources. Even though domains with 768-bit DKIM keys are weak, we think that the disclosure of these domains is justifiable, as an attacker would need significantly more resources to factorize these keys successfully.

7 RELATED WORK

Email security measurements. There are a few measurements works in the area of email security, and also the aspect of authentication was considered in prior work, in particular for SPF and DMARC. To the best of our knowledge, the first paper dealing explicitly with SPF and other anti-phishing protocols was published by Gorling in 2007 [16] one year after publication of SPFv1 in RFC 4408. An analysis among the *.se* domains revealed that only 1.6% out of 385,862 domains implemented SPF. An exclusion of domains without MX record increased the number minimally to around 1.9%.

Back in 2015, Durumeric et al. [11] and Foster et al. [13] published studies on the security of the email system. Durumeric et al. covered the authentication methods SPF/DMARC. They analyzed the security configurations of top email providers based on SMTP connections from and to the Google mail server. Foster et al. evaluated the security of the email system and mechanisms that can protect the confidentiality, authenticity, and integrity of email messages. In particular, they analyzed the support of SPF and DMARC of the most common email providers. The analysis also covers the TLS protocol and the evaluation of the certificates of the email server. We do not only update the current state but extend these works by especially observing the evolution of the protocols over time to recognize that continuous progress is made but the overall adoption is still improvable. Thus, to give this a further push, we argue that it is time to publish more work in this area. The DKIM analysis examines the topic from a different perspective and is therefore not comparable directly. Recent work from Hu and Wang from 2018 [19] contains an analysis of spoofing attacks and which mechanisms can prevent these kinds of attacks. In particular, 35 email providers were analyzed and checked which protocols they have implemented to stop phishing or spoofing attacks. We also extend this work with deeper analyses and focus more on the development than the current status. The most recent work is also from Hu et al. from 2018 [18]. In this work, the authors analyze why the adoption rate of anti-spoofing protocols is still low; for this purpose, a user study with nine administrators was conducted. In contrast, we perform a large-scale empirical study on SPF, DMARC, and DKIM adoption in practice. We think that in the future more qualitative studies should be done on this topic.

Although there are various good measurement studies available on this topic, they are not as extensive as our study. Moreover, we argue that it is necessary to revisit this topic since previous measuring efforts are already several years old and the Internet with DNS and email communication has evolved rapidly. We intend to continue the existing research and especially to examine the evolution over time of the investigated protocols. Especially because fake emails are still a major Internet problem and through the publication of binding political directives and recommendations, a lot changed recently. Thus, it is reasonable to update and also enhance our understanding in this area. Additionally, awareness of these protocols needs to be improved. For example, we are not aware of any study of DKIM keys used in practice. Other existing studies always analyze DKIM from a different point of view (e.g., by analyzing email dumps) as we do and there is no one capable of giving a lower bound of DKIM usage in general. However, further email attacks like email header injection [6] or spam emails without spoofing [10] are not considered. Research dealing with email security on the application side (S/MIME, OpenPGP) is also available [28, 30], but our study examines the server-side.

DNS measurements. Many works measure various aspects of DNS. For example, Rijswijk-Deij et al. discussed the challenges for active measurements in DNS [43]. Some works look at the infrastructure of DNS and analyze security and privacy aspects [1, 24], as well as works that study DNS manipulations in particular [26, 29] or works focusing on DNS security extensions [7–9, 37]. Others consider misconfigured servers, DNS tunneling or certain resource record types [31, 38, 39]. In our work, we measure a still so far neglected but important DNS aspect of DNS-based email authentication.

8 CONCLUSION

In this paper, we empirically examined different DNS-based email authentication methods. We found that progress was made in terms of deployment (significant increase in use of the mechanisms). Additionally, we revealed different issues such as misconfigurations, weak keys, or shared duplicate keys. We expect this work with a focus on email authentication to further improve the adoption of SMTP authentication extensions by demonstrating the still poor overall adoption.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy - EXC 2092 CASA – 390781972.

REFERENCES

[1] Bernhard Ager, Wolfgang Mühlbauer, Georgios Smaragdakis, and Steve Uhlig. 2010. Comparing DNS Resolvers in the Wild. In *ACM SIGCOMM Internet Measurement Conference (IMC).*

[2] Alexa. [n.d.]. Alexa top 1 million domain list. https://s3.amazonaws.co/alexa-static Accessed: 2019-05-01.

[3] atlassian-dkim [n.d.]. Configuring Jira Cloud to send emails on behalf of your domain. https://confluence.atlassian.com/adminjiracloud/configuring-jira-cloud-to-send-emails-on-behalf-of-your-domain-900996548.html Accessed: 2019-08-01.

[4] BOD-18-01 [n.d.]. Binding Operational Directive 18-01: Enhance Email and Web Security. https://cyber.dhs.gov/bod/18-01/. Accessed: 2020-05-21.

[5] Kevin Borgolte, Christopher Kruegel, and Giovanni Vigna. 2015. Meerkat: Detecting website defacements through image-based object recognition. In *USENIX Security Symposium.*

[6] Sai Prashanth Chandramouli, Pierre-Marie Bajan, Christopher Kruegel, Giovanni Vigna, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. 2018. Measuring E-mail Header Injections on the World Wide Web. In *ACM Symposium on Applied Computing.*

[7] Taejoong Chung, Roland van Rijswijk-Deij, Bala Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. 2017. A Longitudinal, End-to-End View of the DNSSEC Ecosystem. In *USENIX Security Symposium.*

[8] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. 2017. An End-to-End View of DNSSEC Ecosystem Management. *USENIX ;login:* 42, 4 (Winter 2017).

[9] Taejoong Chung, Roland van Rijswijk-Deij, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. 2017. Understanding the Role of Registrars in DNSSEC Deployment. In *ACM SIGCOMM Internet Measurement Conference (IMC).*

[10] K. Dan, N. Kitagawa, S. Sakuraba, and N. Yamai. 2019. Spam Domain Detection Method Using Active DNS Data and E-Mail Reception Log. In *Computer Software and Applications Conference (COMPSAC).*

[11] Zakir Durumeric, David Adrian, Ariana Mirian, James Kasten, Elie Bursztein, Nicolas Lidzborski, Kurt Thomas, Vijay Eranti, Michael Bailey, and J Alex Halderman. 2015. Neither snow nor rain nor MITM...: An empirical analysis of email delivery security. In *ACM SIGCOMM Internet Measurement Conference (IMC).*

[12] EU DMARC [n.d.]. CERT-EU Security Whitepaper 17-001 DMARC - Defeating E-Mail Abuse. http://cert.europa.eu/static/WhitePapers/Updated-CERT-EU_Security_Whitepaper_DMARC_17-001_v1_2.pdf Accessed: 2020-05-21.

[13] Ian D Foster, Jon Larson, Max Masich, Alex C Snoeren, Stefan Savage, and Kirill Levchenko. 2015. Security by any other name: On the effectiveness of provider based email security. In *ACM Conference on Computer and Communications Security (CCS).*

[14] freshdesk [n.d.]. Digitally Sign Emails with DKIM. https://support.freshdesk.com/support/solutions/articles/223779-digitally-sign-emails-with-dkim Accessed: 2020-05-21.

[15] gmail [n.d.]. Gmail now has 1.5 billion active users. https://www.digitalinformationworld.com/2018/10/gmail-now-has-1-5-billion-active-users.html# Accessed: 2020-05-21.

[16] Stefan Görling. 2007. An overview of the Sender Policy Framework (SPF) as an anti-phishing mechanism. *Internet Research* 17, 2 (2007), 169–179.

[17] Ralph Holz, Johanna Amann, Olivier Mehani, Matthias Wachs, and Mohamed Ali Kaafar. 2016. TLS in the Wild: An Internet-wide Analysis of TLS-based Protocols for Electronic Communication. In *Symposium on Network and Distributed System Security (NDSS).*

[18] Hang Hu, Peng Peng, and Gang Wang. 2018. Towards understanding the adoption of anti-spoofing protocols in email systems. In *2018 IEEE Cybersecurity Development (SecDev).*

[19] Hang Hu and Gang Wang. 2018. End-to-end measurements of email spoofing attacks. In *USENIX Security Symposium.*

[20] Scott Kitterman. 2014. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208. https://doi.org/10.17487/RFC7208

[21] Scott Kitterman. 2018. Cryptographic Algorithm and Key Usage Update to DomainKeys Identified Mail (DKIM). RFC 8301. https://doi.org/10.17487/RFC8301

[22] Murray Kucherawy, Dave Crocker, and Tony Hansen. 2011. DomainKeys Identified Mail (DKIM) Signatures. RFC 6376. https://doi.org/10.17487/RFC6376

[23] Murray Kucherawy and Elizabeth Zwicky. 2015. Domain-based Message Authentication, Reporting, and Conformance (DMARC). RFC 7489. https://doi.org/10.17487/RFC7489

[24] Marc Kührer, Thomas Hupperich, Jonas Bushart, Christian Rossow, and Thorsten Holz. 2015. Going wild: Large-scale classification of open DNS resolvers. In *ACM SIGCOMM Internet Measurement Conference (IMC).* ACM.

[25] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Symposium on Network and Distributed System Security (NDSS).*

[26] Baojun Liu, Chaoyi Lu, Haixin Duan, Ying Liu, Zhou Li, Shuang Hao, and Min Yang. 2018. Who is answering my queries: understanding and characterizing interception of the DNS resolution path. In *USENIX Security Symposium.*

[27] Majestics. [n.d.]. The Majestics Million. https://majestic.com/reports/majestic-million Accessed: 2019-05-01.

[28] Jens Müller, Marcus Brinkmann, Damian Poddebniak, Hanno Böck, Sebastian Schinzel, Juraj Somorovsky, and Jörg Schwenk. 2019. "Johnny,

you are fired!" – Spoofing OpenPGP and S/MIME Signatures in Emails. In *USENIX Security Symposium*.

[29] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. 2017. Global Measurement of DNS Manipulation. In *USENIX Security Symposium*.

[30] Damian Poddebniak, Christian Dresen, Jens Müller, Fabian Ising, Sebastian Schinzel, Simon Friedberger, Juraj Somorovsky, and Jörg Schwenk. 2018. Efail: Breaking S/MIME and OpenPGP email encryption using exfiltration channels. In *USENIX Security Symposium*.

[31] Adam Portier, Henry Carter, and Charles Lever. 2019. Security in Plain TXT. In *Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*.

[32] Proofpoint [n.d.]. Federal DMARC Adoption Rates Increase Significantly to Address BOD 18-01 Deadline. https://www.proofpoint.com/us/corporate-blog/post/federal-dmarc-adoption-rates-increase-significantly-address-bod-18-01-deadline. Accessed: 2020-05-21.

[33] Vera Rimmer, Davy Preuveneers, Marc Juarez, Tom Van Goethem, and Wouter Joosen. 2018. Automated Website Fingerprinting through Deep Learning. In *Symposium on Network and Distributed System Security (NDSS)*.

[34] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D Strowes, and Narseo Vallina-Rodriguez. 2018. A long way to the top: significance, structure, and stability of internet top lists. In *ACM SIGCOMM Internet Measurement Conference (IMC)*.

[35] sendgrid-dkim [n.d.]. How to set up domain authentication. https://sendgrid.com/docs/ui/account-and-settings/how-to-set-up-domain-authentication/ Accessed: 2020-05-01.

[36] Standard DKIM [n.d.]. Domain with sendinblue. https://discourse.mailinabox.email/t/domain-with-sendinblue/1961. Accessed: 2020-05-01.

[37] Dennis Tatang, Robin Flume, and Thorsten Holz. 2021. A First Large-scale Analysis on Usage of MTA-STS. In *Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*.

[38] Dennis Tatang, Florian Quinkert, and Thorsten Holz. 2019. Below the Radar: Spotting DNS Tunnels in Newly Observed Hostnames in the Wild. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*.

[39] Dennis Tatang, Carl Schneider, and Thorsten Holz. 2019. Large-Scale Analysis of Infrastructure-Leaking DNS Servers. In *Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*.

[40] UK DMARC [n.d.]. DMARC Required For UK Government Services By October 1st . https://dmarc.org/2016/06/dmarc-required-for-uk-government-services-by-october-1st/ Accessed: 2020-05-01.

[41] US presidents [n.d.]. Nearly all 2020 presidential candidates are not using a basic email security feature. https://techcrunch.com/2019/04/30/dmarc-presidential-candidates/ Accessed: 2020-05-01.

[42] Valimail Email Fraud [n.d.]. Research: Crisis of Fake Email Continues to Plague Industries Worldwide . https://www.valimail.com/blog/q2-2018-report-fake-email-crisis/ Accessed: 2020-05-01.

[43] R. van Rijswijk-Deij, M. Jonker, A. Sperotto, and A. Pras. 2016. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements. *IEEE Journal on Selected Areas in Communications* (2016).

A APPENDIX

Table 7: Comparision according the adoption of SPF, DKIM and DMARC of popular email providers (Legend: ✓ Protocol supported, ✗ Protocol not supported)

Email Provider	SPF		DKIM		DMARC	
	SUP	Policy	SUP	Key Len	SUP	Policy
aol.com	✓	softfail	✓	2048	✓	reject
daum.net	✓	softfail	✗	-	✗	-
fastmail.com	✓	neutral	✓	2048	✓	none
firemail.de	✓	neutral	✗	-	✗	-
freemail.hu	✓	softfail	✓	2048	✗	-
freenet.de	✓	softfail	✗	-	✗	-
gmail.com	✓	softfail	✓	2048	✓	none
gmx.de	✓	hardfail	✗	-	✗	-
hotmail.com	✓	softfail	✓	2048	✓	none
inbox.lv	✓	softfail	✓	1024	✓	quarantine
interia.pl	✓	hardfail	✓	1024	✗	-
mail.de	✓	neutral	✓	1024	✓	none
mail.ru	✓	softfail	✓	1024	✓	reject
naver.com	✓	softfail	✓	2048	✗	-
op.pl	✓	hardfail	✓	1024	✗	-
outlook.com	✓	softfail	✓	2048	✓	none
protonmail.com	✓	softfail	✓	1024	✓	quarantine
runbox.com	✓	hardfail	✓	2048	✓	none
sapo.pt	✓	softfail	✗	-	✗	-
seznam.cz	✓	neutral	✓	1024	✓	none
t-online.de	✗	-	✗	-	✗	-
tutanota.com	✓	hardfail	✓	2048	✓	none
web.de	✓	hardfail	✗	-	✗	-
yahoo.com	✓	neutral	✓	2048	✓	reject
zoho.eu	✓	hardfail	✓	1024	✓	reject

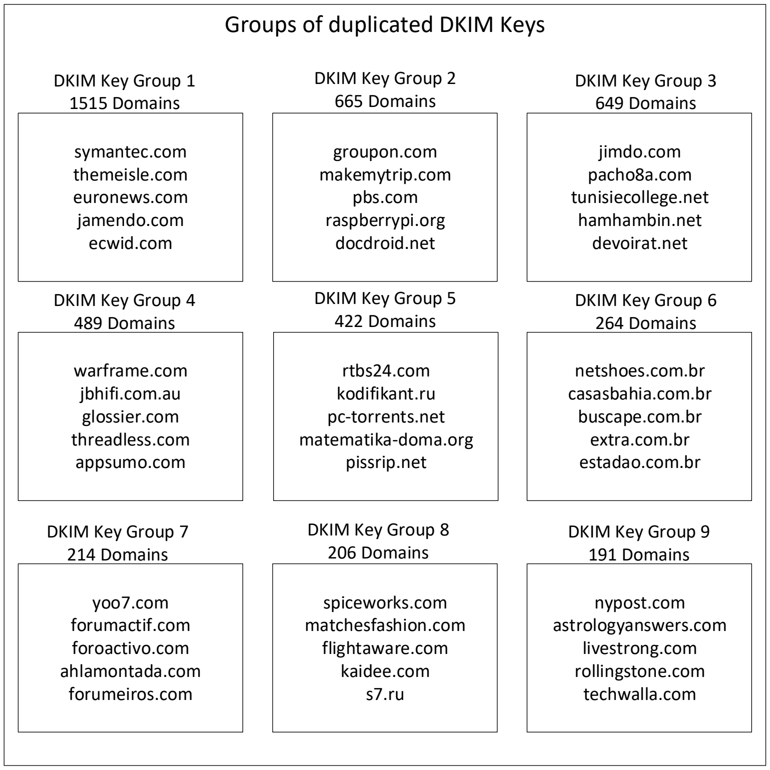

Figure 1: Top 9 biggest groups of domains sharing the same DKIM key