Not that Simple: Email Delivery in the 21st Century

Florian Holzbauer
SBA Research

Johanna Ullrich
University of Vienna[*]

Martina Lindorfer
TU Wien

Tobias Fiebig
Max-Planck-Institut für Informatik

Abstract

Over the past two decades, the number of RFCs related to email and its security has exploded from below 100 to nearly 500. This embedded the Simple Mail Transfer Protocol (SMTP) into a tree of interdependent and delivery-relevant standards. In this paper, we investigate how far real-world deployments keep up with this increasing complexity of delivery- and security options. To gain an in-depth picture of email delivery apart from the giants in the ecosystem (Gmail, Outlook, etc.), we engage people to send emails to eleven differently configured target domains. Our measurements allow us to evaluate core aspects of email delivery, including security features, DNS configuration, and IP version support on the sending side across different types of providers.

We find that novel technologies are often insufficiently supported, even by large providers. For example, while 65.4% of email providers can resolve hosts via IPv6, only 44.3% can also deliver emails via IPv6. Concerning security features, we observe that less than half (41.5%) of all providers rely on DNSSEC validating resolvers, and encryption is mostly opportunistic, with 89.7% of providers accepting invalid certificates. TLSA, as a DNS-based certificate verification method, is only used by 31.7% of the providers in our study. Finally, we turned our eye to the impact modern standards have on unsolicited bulk email (SPAM). We found that greylisting is effective, reducing the SPAM volume by roughly half while not impacting regular delivery. However, and interestingly, SPAM delivery currently seems to focus on plaintext IPv4 connections, making IPv6-only, TLS-enforcing inbound email servers a more effective anti-SPAM measure—even though it also means rejecting a major portion of legitimate emails.

1 Introduction

Electronic mail (email) relies on the Simple Mail Transfer Protocol (SMTP) for delivery. This protocol was first specified in 1982 in RFC 821 and is now close to celebrating its 40th birthday [39]. SMTP had two design goals, namely to allow *reliable* and *efficient* delivery of emails. As with many protocols of the time, security and authenticity were not priorities [16]. In fact, anyone could relay emails through an SMTP server, which was the default configuration for many email servers – like Sendmail – until the late 1990s [3].

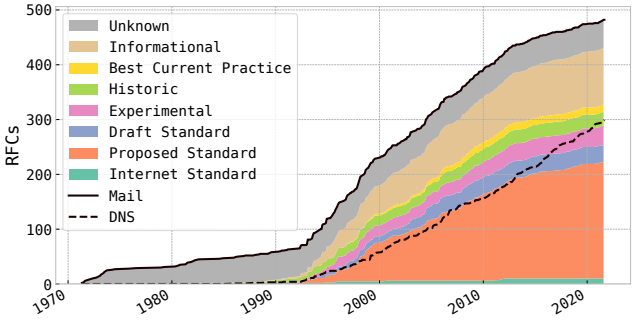


Figure 1: Overview of the explosion of email-related standards ("SMTP Camel"), compared to DNS-related standards.

However, the practical reality of the Internet led to increased security and authenticity requirements [16]. Since the mid-1990s, hundreds of protocols and extensions have been introduced to cover these gaps, as illustrated in Figure 1. In order to authenticate email, attempts mostly rely on the *Domain Name System* (DNS), which, in turn, suffers from authenticity issues. To address those issues, the *DNS Security Extensions* (DNSSEC) were introduced in 1999, which enabled signing DNS entries [1]. Besides authenticity, the original email protocol faced other security-related challenges, most notably confidentiality, as emails were exchanged in plaintext. In addition to end-to-end encryption approaches like Pretty Good Privacy (PGP) [7], this led to an extension of SMTP for Transport Layer Security (TLS) [18]. Finally, like all protocols on the Internet, SMTP was also affected by the introduction of IPv6.

All these factors have turned the *simple* from SMTP to *complex*. To outline this increase in complexity, we created the *SMTP Camel* in Figure 1 (after the famous DNS Camel of

[*] Christian Doppler Laboratory for Security and Quality Improvement in the Production System Lifecycle, Security & Privacy Group, Faculty of Computer Science

Bert Hubert, who illustrated the complexity of DNS with *"How many features can we add to this protocol before it breaks?"* [23]). Figure 1 visualizes RFCs related to email – and, for reference, DNS. We compiled this list by performing a title/keyword search on all RFCs on September 28, 2021.[1] In total, we found 481 email-related RFCs compared to 298 DNS-related ones. Among these, more than half of the RFCs belong to the standards track, representing mature standards. We see no development in draft standards as they were declared as deprecated in 2010 [21]. In June 2021, we reached a total of 225 proposed standards. Proposed standards only advance to Internet standards once they have *"widespread deployment of multiple implementations from different code bases"* [21]. Currently, only eleven email-related RFCs have met this requirement, and also the handling of this guideline by the Internet Engineering Task Force (IETF) varies. This indicates that the development of new standards has outpaced their implementation. Furthermore, since the latest email measurement study in 2020 [31], seven new email-related RFCs have been published.

In this paper, we investigate how the increasing number of additional standards has influenced email delivery in the wider ecosystem. Related work already demonstrated that adoption rates of email-related standards are low and implementations often rely on insecure defaults [8, 14, 17, 22, 26, 31, 37, 45]. However, previous work predominantly focused on large operators, such as Google (Gmail) or Microsoft (Outlook), and did not investigate fundamental aspects of email standards, like supported IP versions and the DNS infrastructure of sending systems. We take a step back and investigate the most fundamental aspects of email in transit across a wide sample going beyond major email providers.

To accomplish this, we introduced eleven target address configurations to verify how email providers implement email-related standards and protocols, i.e., we set up systems that – depending on the remote server's configuration and implementation – either do or do not receive measurement emails. Our measurement technique allows us to measure IP support, STARTTLS configuration, DNSSEC validation, and how different SMTP applications react to greylisting, an anti-SPAM technique by which incoming emails are initially rejected. Our focus is on protocols that influence email delivery once an email has been submitted. To increase the providers' coverage, we crowdsourced the sending of emails to participants recruited through mailing lists and social media.

As a result, we collect emails from three different sources, spanning (1) small participants in the email ecosystem, (2) large providers, and (3) unsolicited bulk email, aka SPAM. We are the first to discuss the impact of new and established standards on email delivery, as – in contrast to most related measurements – we rely on actively collecting emails, allowing us a more in-depth view of email server configurations.

[1] https://www.rfc-editor.org/search/rfc_search_detail.php

In summary, we make the following contributions:

- We introduce a new ranking method using passive data to find the top 15 email providers. Our results highly overlap with Liu et al. [32], while causing significantly less measurement overhead (see Section 3).
- We illustrate challenges in the interoperability between large centralized operators and smaller operators, including how the ability to deliver emails as the main objective limits the adoption of new network and security protocols. We describe how our datasets cover different actors in the email ecosystem in Section 4.
- We are the first to measure and connect the impact of protocol extensions in protocols email relies on – DNS(SEC) and IPv6 – to email delivery and the contrast between smaller and larger providers (see Section 5).
- We illustrate protocol support and compliance in the heavy-tail of the email ecosystem, i.e., in a large set of smaller email operators, and contrast this to earlier work and patterns found in large providers (see Section 6).
- Based on our results, we derive recommendations for email system operators on how they can utilize modern protocol compliance to – currently – reduce SPAM delivery (see Section 7).

Artifacts: Our measurement can be executed using any valid domain and a set of machines connected to the Internet. Along with our paper, we publish a setup-documentation and the scripts we used to receive and analyze emails sent to our systems at https://github.com/ichdasich/email-measurement-toolchain. For privacy reasons, we cannot publish our email dataset. This also applies to the SPAM dataset, as even SPAM may contain PII, for example in the recipient addresses.

2 Background: Protocols and Standards

In this paper, we focus on standards influencing email delivery between email servers, i.e., the Mail Transfer Agent (MTA). Email submission, e.g., the communication between Mail User Agent (MUA) and Mail Submission Agent (MSA), is not part of our study. We focus on IP- and DNS-related mechanisms that impact delivery. Interpretations of higher-level delivery security features, like the Sender Policy Framework (SPF) [27], DomainKeys Identified Mail (DKIM) [9], Authenticated Received Chain (ARC) [4], and Domain-based Message Authentication, Reporting, and Conformance (DMARC) [30] are out of scope for our study, as they only influence the receiver's decision on whether to accept incoming emails or not. We also did not include MTA Strict Transport Security (MTA-STS) in our study as this RFC was too recent

when we set up our infrastructure [33], but Section 3 describes how our work can be extended to include it in the future.

IPv4 [38] and IPv6 [10]. Since addresses in the 2^{32} bit address space of the Internet Protocol Version 4 (IPv4) are running out [41], Internet Protocol Version 6 (IPv6) with a 2^{128} bit address space was introduced in the late 1990s. Two concurrent IP versions introduce a great challenge in terms of interoperability on the network layer, especially as the adoption of IPv6 is still slow [25]. IP version support impacts email delivery *indirectly* via DNS support, i.e., the authoritative and recursive servers support the same IP version, and *directly*, i.e., in terms of whether the involved email servers both support the same IP version. Servers can support IPv4, IPv6, or both—also referred to as "dual-stack."

DNSSEC [5]. The DNS-Security Extensions (DNSSEC) provide authenticity to DNS responses by signing DNS entries via a keychain along the path of the DNS tree. A DNSSEC validating recursor responds with `SERVFAIL` in case of a validation error. As a consequence, the target domain cannot be resolved, and email delivery fails. Hence, in case of misconfigurations – common in system operations [11] – or attacks, the DNSSEC validation behavior of DNS resolvers at email-sending servers becomes important for email delivery. Similarly, DNSSEC is a prerequisite for DANE (see below).

STARTTLS [19]. The SMTP Service Extension for Secure SMTP over TLS (STARTTLS) enables TLS for email delivery. The connection is established on the same port as SMTP. The original SMTP handshake remains in cleartext. Sending- and receiving servers can (1) not support TLS, (2) support TLS and cleartext, (3) enforce TLS. TLS can be configured either in an (a) opportunistic or (b) strict manner. While opportunistic TLS configurations allow for encrypted connections not validating the remote certificate, strict configurations cause email delivery to fail in case of (1) invalid certificates, (2) not supporting mandatory ciphers, or (3) a connection to a non-TLS-supporting server. In turn, this can then impact email delivery, depending on whether a connection can be established or not.

DANE [20]. The DNS-Based Authentication of Named Entities (DANE) prevents MTA-to-MTA transport encryption from downgrade attacks, even in the absence of certificates signed by a certificate authority (CA); this is done through recording valid CA or end-entity certificates for a domain name via the TLSA DNS record. Trusting/guaranteeing the authenticity of TLSA records (i.e., preventing MITM and DNS cache poisoning scenarios) requires the use of DNSSEC, as described above. Several email server implementations, including Sendmail and Microsoft Exchange, do not yet support requesting TLSA records, in contrast to for example, Postfix and Exim [31].[2] DANE can be implemented similar to

[2]Microsoft announced support after our measurement period in Feb, 2022 (see https://techcommunity.microsoft.com/t5/exchange-team-blog/releasing-outbound-smtp-dane-with-dnssec/ba-p/3100920)

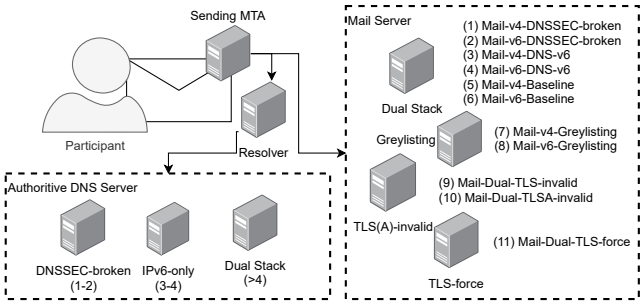


Figure 2: Overview of our measurement setup: 3 DNS servers serve 4 email servers with 11 differently configured target addresses.

TLS in an opportunistic or mandatory manner. Email delivery fails for both opportunistic and mandatory configurations if a signed TLSA record is available but certificate validation fails or for mandatory configurations if no TLSA record can be found.

Anti-SPAM (Greylisting [29]). Greylisting is one of the most simplistic approaches to reduce SPAM emails. It works by initially responding with SMTP code `4xx temporary failure`. While reputable servers usually re-attempt email delivery after several minutes, many SPAM senders do not keep enough state for this. For email delivery, greylisting introduces delays, and email delivery fails if an implementation does not attempt redelivery.

3 Methodology

Measurement Platform. Our measurement setup consists of four email servers running Postfix 3.6 [40] on OpenBSD 6.7 [36] in OpenBSD virtual machines (VMM). As we conduct non-performance bound network measurements, the exact type and model of the used hardware are not relevant to our measurement platform. Furthermore, we rely on three PowerDNS authoritative nameservers in version 4.3.1 to measure the impact of different DNS server setups. We configured a non-default TTL of 300 seconds for all entries in our DNS zones to minimize the impact of caching, i.e., a DNS resolver used by multiple study participants. This also affects our weekly spam domain rotations, pointing them at different measurement target addresses. However, we consider a maximum overlap of five minutes in comparison to a one-week measurement period negligible. IPv6 connectivity to our systems was provided via a Hurrican Electric IPv6 tunnel, while IPv4 connectivity was provided via dedicated IP space from the RIPE region. On these systems, we set up eleven email addresses, as shown in Figure 2. For each of these addresses, we applied different configuration states, which either enable or prevent remote servers from sending emails to them, depending on their own configuration state. This allows us to measure the remote servers' email delivery capabilities and protocol use by measuring whether they are able to deliver

```
measurement@v4-mail.example.com
measurement@v6-mail.example.com
measurement@v4-mail.v6only.example.com
measurement@v6-mail.v6only.example.com
measurement@v4-mail.dnssec-broken.example.com
measurement@v6-mail.dnssec-broken.example.com
measurement@v4-mail-greylisting.example.com
measurement@v6-mail-greylisting.example.com
measurement@mail-tls-force.example.com
measurement@mail-tls-invalid.example.com
measurement@mail-tlsa-invalid.example.com
```

Figure 3: List of email addresses for the 11 target configurations.

emails to these email addresses. We then asked participants to send *one* email with all measurement addresses in the `To:` field. If we do not receive a message at a specific target address but see in our baseline that the target is included in the `To:` header, we know that the respective feature is not supported. The target addresses can be easily extended to cover new protocols, e.g., MTA-STS [33] was introduced as a barrier against downgrade or interception attacks for domains that are unable to deploy DNSSEC. MTA-STS can be measured by adding two new target addresses in the future. One could implement the TLS-RPT standard to measure TLS reporting frequency, and the other could measure if providers still deliver emails in case of an enforced MTA-STS policy with non-matching MX records.

3.1 Target Address Configurations

We configured the following eleven different email addresses at the unique destination domains listed in Figure 3. Below, we describe the purpose of each of these addresses, i.e., which configuration parameters we tested with them:

IP Support. In order to test basic delivery behavior, we created for both IPv4 (`measurement@v4-mail.`, *Mail-v4-Baseline*) and IPv6 (`measurement@v6-mail.`, *Mail-v6-Baseline*) one address which is configured with no restrictions on delivery. Similarly, we created distinct IPv4- and IPv6 addresses for the DNS and greylisting measurements described below. Note that during our study, we noticed that our choice not to support STARTTLS on this system did indeed introduce an unexpected parameter in the case of senders that enforce STARTTLS use. In turn, this allowed us to detect six providers that enforce STARTTLS for outgoing emails.

DNS Recursion IPv6 Support. To test whether the recursive resolvers of an email sending host support IPv6, we created a subdomain that can only be resolved via IPv6, i.e., the zone had only `AAAA` glue records, and the hosts in the zone's `NS` records also only have `AAAA` records. Under that domain, we then again created two addresses for IPv4 and IPv6 delivery (`measurement@v4-mail.v6only.`, *Mail-v4-DNS-v6* and `measurement@v6-mail.v6only.`, *Mail-v6-DNS-v6*).

DNSSEC Validation. To test if the remote site validates DNSSEC, we set up a subdomain with a non-matching DS RRset in the parent, i.e., we provide a public key in the parent zone that does not match the key with which records are signed in our zone. Hence, a DNS recursive resolve validating DNSSEC is unable to validate DNSSEC for our domain and should therefore refuse to resolve it. Thus, an email server using a validating resolver cannot deliver emails to that domain. Under that domain, we again created two addresses for IPv4- and IPv6 delivery (`measurement@v4-mail.dnssec-broken.`, *Mail-v4-DNSSEC-broken* and `measurement@v6-mail.dnssec-broken.`, *Mail-v6-DNSSEC-broken*).

TLS Configuration. In order to test the TLS and TLSA behavior of sending hosts, we configured three email addresses that required the use of TLS to deliver emails:

- `measurement@mail-tls-force.` *Mail-Dual-TLS-force* on a correctly configured TLS enabled server.
- `measurement@mail-tls-invalid.` *Mail-Dual-TLS-invalid* on a server that provides a certificate with a non-matching CN/DNS0 entry.
- `measurement@mail-tlsa-invalid.` *Mail-Dual-TLSA-invalid* on a server that has a TLSA record configured, which does not match the supplied certificate.

This setup allows us to verify if systems (1) support STARTTLS, (2) perform opportunistic encryption, and (3) verify TLSA records. Due to a misconfiguration, these systems initially did not support TLS1.3. Hence, remote systems that only support TLS1.3 would be unable to deliver their emails. We were able to isolate the affected cases (76 emails from 29 providers) and reconstructed the actual state from the stored SMTP sessions, as the abort conditions differ between 'not supporting TLS,' 'rejecting the certificate/TLSA record,' and 'not having a matching cipher.'

Anti-SPAM (Greylisting). To identify RFC-compliant SMTP implementations, and as an additional control, we set up Postgrey that performs greylisting as an anti-SPAM measure (`measurement@v4-mail-greylisting.`, *Mail-v4-Greylisting* and `measurement@v6-mail-greylisting.`, *Mail-v6-Greylisting*). By configuring these addresses, we can test the impact of greylisting on average SPAM received and check whether legitimate email servers support multiple delivery attempts.

3.2 Email Collection and Recruitment

In order to provide different views on email delivery, we target three types of actors in the email ecosystem: (1) Regular providers by actively engaging users to send emails to our measurement system. (2) A set of top-ranked email providers

Table 1: Recruitment channels for study participants.

Type	Name	Description
Blogs	RIPE Labs	Article in RIPE's Research Blog/Newsfeed
	APNIC	Article in APNIC's Blog/Newsfeed
Social Media	Twitter	Tweets by researchers involved in the project
	LinkedIn	Posts by researchers involved in the project
	Reddit	Reddit post to /selfhosted
Mailing Lists	NANOG	North American Network Operator List
	INNOG	Indian Network Operator List
	AFNOG	African Network Operator List
	SAFNOG	South African Network Operator List
	DENOG	German Network Operator List
	NLNOG	Dutch Network Operator List
	IRTF-MAPRG	Network Research Interest Group at IETF/IRTF
	MAIL-OPS	Global Mail Operator List
Presentations	Internet.nl	Presentation at an organization promoting the adoption of security standards
Personal	-	Colleagues and personal networks, especially in the APNIC and LACNIC regions

by registering user accounts and sending emails. (3) Spammers by registering expired domains and collecting unsolicited emails targeting these domains.

Regular Providers. To collect emails, we actively engaged Internet users to participate in our study. We recruited participants via a social media campaign on Twitter, LinkedIn, and Reddit, via mailing lists focusing on email and network operators, blog articles promoted by Internet governance bodies, and our personal networks (see Table 1). Our recruitment message asked users to visit our website, which provided instructions on how the reader can participate in our study, what the purpose of our study is, and what data access and deletion rights they have. One critical aspect was to ensure that we would be able to distinguish whether an email to one of our measurement hosts was sent and not delivered or not sent at all. Thus, we instructed participants to add all measurement addresses to the `To:` field of a single email. In case a participant's provider performed pre-filtering, e.g., did not accept delivery to domains they cannot resolve, we removed affected emails from the dataset.

Large Providers. In order to rank email providers, we rely on the passively collected Farsight SIE DNS dataset [43]. This enables us to count email servers to which a lot of domains point their `MX` records, i.e., email servers used for a lot of domains. We assume that the number of domains using a provider's email servers correlates to the provider's size. For our ranking, we use DNSDB MX data extracted for November 2020, which includes data of 73,705,268 different `MX` lookups. We do not rank providers based on the amount of `MX` lookups, as low TTLs or different DNS resolver setups might bias the number of lookups. For each `MX`, we extract the public suffix, i.e., 'example.com' for 'mail.example.com' and 'example.co.uk' for 'mail.example.co.uk' using the Public Suffix List [35]. This results in 23,378,583 different public suffixes. We rank public suffixes of `MX` records by counting the number of different domains pointing their `MX` records towards them. We then register accounts at the top 15 providers according to this ranking to send emails to our target domains, as done in prior work [17, 22, 31, 32, 45]. This enables us to compare email delivery from regular providers with an exclusive set of large providers, but also to compare the results of our measurement pipeline to the results of prior work.

Table 2: Categories of domains from ExpiredDomains.

Category	Description
1990s	Domains with the first screenshot available on Archive.org between 1990 and 2000 (= "birth year")
alexa	Domains selected based on Alexa traffic rank
backlinks	Domains based on number of Majestic external backlinks
dmoz	Domains found in the latest snapshot of dmoz.org (~2017)
majestic	Domains with low Majestic million global rank
wiki	Domains with high numbers of Wikipedia links

Spammers. To collect SPAM emails, we registered expired domains that are still likely to receive SPAM. To do so, we relied on `expireddomains.net` for a list of domains [42]. To increase the likeliness that respective domains still receive SPAM, we chose them from different categories, based on their age ("birth year," i.e., the first entry in `Archive.org`), their popularity according to rankings from Alexa and Majestic, and the number of links from Wikipedia and the (now defunct) DMOZ content directory. Table 2 lists these categories; Table 3 lists the domains in each category, as well as the volume of SPAM we received during our measurements.

Once registered, we pointed `MX` records of respective domains at our target domains. To identify if domains still receive SPAM, we executed a three-week baseline measurement. During this period, all 50 re-registered domains pointed their `MX` records to the `MX` of *Mail-v4-Baseline*, i.e., our most basic configuration. We classified the domains' value for our measurement based on the amount of SPAM received as *high* (multiple times a week), *low* (once a week), and *none* (none received). To verify that received messages are SPAM, we consulted four active DNS blocklists: `bl.spamcop.net`, `ips.backscatterer.org`, `pbl.spamhaus.org` and `sbl.spamhaus.org`. We continuously verified the liveness of these blocklists by requesting IP 127.0.0.2 as a test record.

In total, we found 26% of domains receive SPAM on a regular basis, thus falling into category *high*. In the next step, we pointed high-value SPAM domains towards a set of our target addresses in a weekly rotation until each domain had been pointed at each target at least once. This allowed us to monitor the change in SPAM volume based on the corresponding test conditions. For these measurements, we relied on a reduced set of target addresses. As we only received individual emails and did not simultaneously measure all conditions for each sender, we did not differentiate IPv6 behavior for different target addresses. We only verified general IPv6 support (*Mail-v6-Baseline*), IPv4 sending for IPv6 only DNS (*Mail-*

Table 3: Re-registered domains for SPAM collection and the amount of SPAM emails we received for each of them.

	Category	Domain	Spam Frequency
1	1990s	anx-chicago-rawhide.com	low
2	1990s	intecconstruction.com	high
3	1990s	michael-rauch.com	-
4	1990s	mmf-maintenance.com	high
5	1990s	sapphire-controls.co.uk	high
6	1990s	stratos-bde.com	low
7	alexa	inkpreneur.com	-
8	alexa	jsmmf.org	-
9	alexa	kenyamalikmotors.com	-
10	alexa	lafdo.com	high
11	alexa	nepaltravelcentre.com	high
12	alexa	olakassen.com	-
13	alexa	onmylevelchey.com	-
14	backlinks	18Chaa.com	low
15	backlinks	521qiangwweisizu.com	-
16	backlinks	cretms.com	low
17	backlinks	fotiis.com	-
18	backlinks	g6china.com	-
19	backlinks	io365f.com	-
20	backlinks	io365i.com	-
21	backlinks	theproxylist.co.uk	-
22	backlinks	tuncayparlak.com	low
23	backlinks	vous-y-etes.com	-
24	dmoz	beechamsdrivingschool.co.uk	high
25	dmoz	bilder-touren-allgaeu.de	-
26	dmoz	costatehogrally.com	low
27	dmoz	djk-handball-coesfeld.de	-
28	dmoz	leben-ohne-alkohol.eu	low
29	dmoz	navesprefabricadassprint.com	high
30	dmoz	parissi.eu	-
31	dmoz	pringfieldfarms.co.uk	-
32	dmoz	printshopleeds.co.uk	low
33	dmoz	smugglegame.com	high
34	dmoz	sotralentz.es	high
35	dmoz	survivalschool.ch	high
36	dmoz	thermoboss.net	low
37	majestic	djmzengaman.com	-
38	majestic	eiecan.eu	-
39	majestic	hkmxdna.com	-
40	majestic	keerthiwrites.com	-
41	majestic	kientrucnghethuatduongdai.com	-
42	majestic	printspixelz.com	-
43	majestic	studiopaeez.com	low
44	majestic	thi-marprojects.be	high
45	wiki	catholic-church-corfu.org	low
46	wiki	grandeguerrafvg.org	-
47	wiki	iranairlinenews.com	-
48	wiki	mosul-network.org	-
49	wiki	unaf-foot.com	-
50	wiki	worldipcomgroup.com	low

v4-DNS-v6), DNSSEC behavior (*Mail-v4-DNSSEC-broken*), as well as our three TLS configurations.

3.3 Ethical Considerations

As our measurements focus on the technical aspects of the involved email setups, this study was not within the scope of our local human subject research ethics council. Nevertheless, we informed participants about the purpose of our data collection, which information we collected, and that they could withdraw from the study at any time. We received one request to be removed from the dataset and complied with this request immediately. In addition, we followed network measurement best practices as outlined in the Menlo report [6, 12].

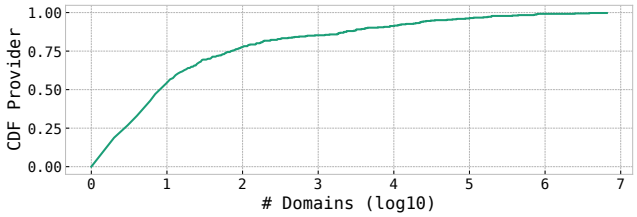


Figure 4: Validation of regular study participants tend to be/use small email providers. We match regular providers to the passive DNS ranking.

This means that we took the necessary technical precautions to protect the only Personally Identifiable Information (PII) we collect, i.e., the sending email addresses. We removed these addresses from our dataset as soon as possible before we started the aggregation of our collected data. Also, since the provider name might reveal PII, we do not publish or share provider names of smaller providers. For our measurements of large providers, we registered accounts ourselves and published their names for better comparison to related work, in accordance with common practice for email-related measurements [17, 22, 31, 32, 45].

4 Datasets

By following our approach, we collected three datasets covering (a) regular providers by volunteers sending emails to our measurement infrastructure, (b) large providers by registering accounts and sending emails ourselves, and (c) spammers by collecting unsolicited emails sent to re-registered domains.

(a) Regular Providers. Between July 4, 2020 and October 29, 2021 we received a total of 5,847 emails. After filtering emails that do not cover all eleven target addresses in the `To:` field, a total of 4,660 emails sent by 622 study participants remained for further analysis. There is a clear dominance of emails from European countries, see Table 5, a consequence of recruiting via our personal channels (e.g., on Twitter).

Multiple participants used the same infrastructure to send emails; beyond, emails of the same user might be sent by multiple servers in the same domain (e.g. `server1.domain.any` and `server2.domain.any`). Thus, we grouped the data set using the email servers' first-level domain (EHLO name) at the granularity of providers. This yields a total of 436 providers.

(b) Large Providers. Analysis of the Farsight SIE DNS dataset revealed the top 15 providers as presented in Table 4. We noticed a large gap in served domains even within the top 15 providers, ranging from 14.1% (Google) to 0.68% (1&1) of first-level domains (FLDs) in our passive DNS dataset. The top 15 providers jointly serve 33.8% of all FLDs with `MX` hosts. To gain an overview of provider sizes in our regular dataset, we matched regular providers with domains in the

Table 4: Top 15 providers based on passive DNS data. Providers greyed out have no online email service, e.g., *Above.com* is a domain broker.

										IP support				DNSSEC		Spam		TLS		
NR	Provider	2015 Durumeric [14]	2015 Foster [17]	2018 Hu [22]	2020 Lee [31]	2021 Tatang [45]	2021 Liu [32]	# Dom.	% Dom.	Mail-v4-Baseline	Mail-v6-Baseline	Mail-v4-DNS-v6	Mail-v6-DNS-v6	Mail-v4-DNSSEC-broken	Mail-v6-DNSSEC-broken	Mail-v4-Greylisting	Mail-v6-Greylisting	Mail-Dual-TLS-force	Mail-Dual-TLS-invalid	Mail-Dual-TLSA-invalid
1	Google	⋆	△	•	□	◇	⊙	9,148,093	14.08	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	Microsoft	⋆			□	◇	⊙	3,869,507	5.95	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	GoDaddy	⋆					⊙	2,453,911	3.78	✓				✓		✓		✓	✓	✓
4	OVHCloud	⋆					⊙	1,292,615	1.99	✓						✓		✓	✓	✓
5	Enom						⊙	871,527	1.34	✓				✓		✓		✓	✓	✓
6	One.com							797,194	1.23	✓						✓		✓	✓	
7	Namecheap						⊙	784,486	1.21	✓		✓		✓		✓		✓	✓	✓
8	Strato						⊙	762,923	1.17	✓	✓	✓	✓			✓	✓	✓	✓	
9	Yandex	⋆	△				⊙	759,482	1.17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	SiteGround						⊙	712,418	1.10	✓		✓		✓		✓		✓	✓	
11	H-email.net							575,451	0.89											
12	Above.com							469,500	0.72											
13	Beget						⊙	447,284	0.69	✓				✓		✓		✓	✓	✓
14	Tencent	⋆	△				⊙	442,064	0.68	✓		✓		✓		✓		✓	✓	✓
15	1&1							440,558	0.68	✓	✓	✓	✓			✓	✓	✓	✓	✓
Optimal Configuration										✓	✓	✓	✓			✓	✓	✓	✓	

Table 5: Number of countries/emails/AS per region. Our social media promotion led to an increased number of emails from European countries. We skipped large providers as geographical data has no impact on our provider ranking.

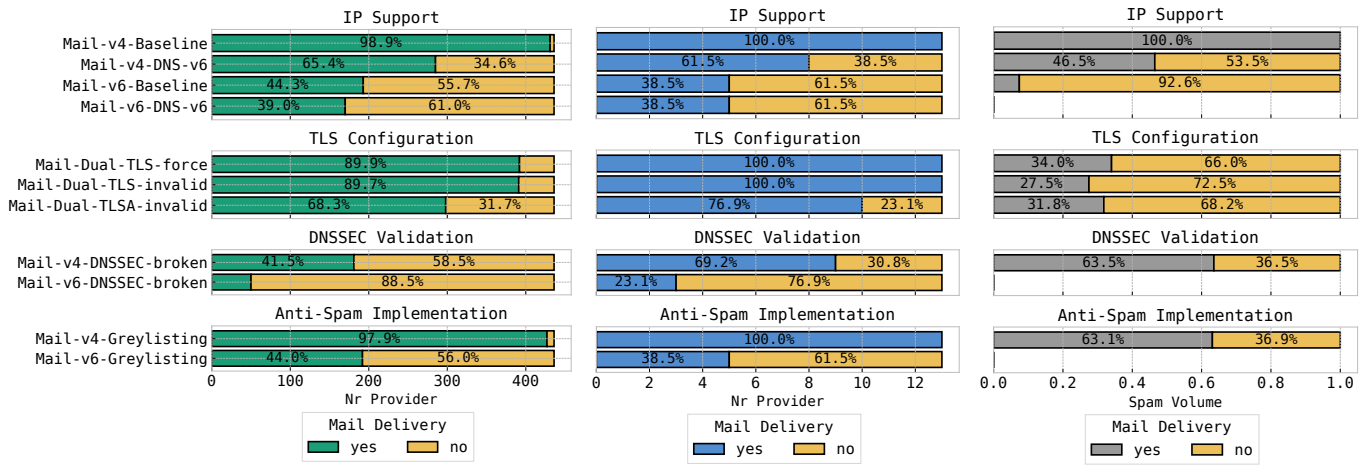
	Region	Africa	Asia	Europe	N. America	Oceania	S. America
Regular	Countries	5	12	30	2	1	3
	Emails	48	168	3,368	1,045	1	30
	ASes	5	19	202	60	1	3
SPAM	Countries	22	32	36	15	2	11
	Emails	95	2,056	1,963	2,437	17	204
	ASes	50	254	234	170	9	119

passive DNS dataset. Figure 4 shows the amount of FLDs pointing at each of the study participants' domains for email. 80% of regular providers have less than 150 domains relying on them for email service. Comparing our top 15 providers with previous work, we find the largest overlap, namely eleven providers, with Liu et al. [32], who used a five-step approach including MX records, Banner/EHLO messages, and TLS certificates to detect large email providers. Previous work relying on manual ranking results in less overlaps, namely six [14], three [17], two [31,45], and one [22] (see Table 4), and suggests that human perception of providers is different from their actual dominance in the email ecosystem.

(c) Spammers. We executed SPAM measurements in three phases. First, we conducted a baseline measurement from March 30, 2021 to April 6, 2021. Next, we pointed SPAM domains to our other target addresses in a weekly rotation. Finally, we did another baseline measurement to ensure that the baselines remained stable over our observation time. We received a total of 6,772 unsolicited emails. Thereof, 4,442 (65.7%) were classified as SPAM by one of our four DNS blocklists, suggesting that emails towards the re-registered domains are indeed SPAM. We included all received emails in our further analysis. In comparison to our regular provider dataset, SPAM emails are not dominated by a single region (see Table 5). In comparison to regular and large providers, we can only measure the SPAM volume and its reduction in dependence of the different configurations.

5 Results

For each of the three datasets, namely (a) *regular providers*, (b) *large providers*, and (c) *spammers*, Figure 5 shows the ratio of delivered to undelivered emails per target address. We provide the individual results for the top 15 providers, including a line indicating the optimal configuration, in Table 4. The optimal configuration includes IPv4- and IPv6 support for both email servers and DNS resolvers. Regarding TLS, providers should implement opportunistic STARTTLS, i.e., still use transport encryption when facing self-signed or expired certificates.



(a) *Regular Providers* (b) *Large Providers* (c) *Spammers*

Figure 5: Impact of different target address configurations on email delivery. For our investigation of spammers we skipped the IPv6 target addresses other than the baseline (this affects greylisting, DNSv6, DNSSEC).

However, they should validate TLSA records and reject email delivery in case of an invalid record. As a foundation for DANE and other DNS-based security standards, a provider should rely on a DNSSEC supporting and -validating resolver. Looking at the top 15 providers, we find major discrepancies for even the largest providers. We discuss our measurement results on IP support, TLS configuration, DNSSEC validation, and anti-SPAM implementation in the following sections.

5.1 IP Support

Email Servers. The *Mail-v4-Baseline* is configured without any restrictions on email delivery. For regular providers, however, this baseline is reduced by 5/436 (1.2%) as five providers enforced TLS causing undeliverability (see also Section 3.1). For large providers, the baseline is met by all providers. For spammers, the baseline is necessary to estimate the number of SPAM emails that are typically sent to the investigated domains. For all three populations, the delivery to *Mail-v6-Baseline* is reduced compared to the IPv4 baseline, implying limited deployment of IPv6 at email servers. Differences among regular and large providers remain small – the first received IPv6-only mails in 193/436 (44.3%) of the cases, the latter in 5/13 (38.5%) –, however, SPAM towards the IPv6 target is drastically reduced and accounts for 7.4% of the IPv4 baseline.

DNS Resolvers. Both targets, *Mail-v4-DNS-v6* and *Mail-v6-DNS-v6*, rely on an IPv6-only authoritative nameserver and allow to infer whether resolvers are capable of IPv6. The number of successfully delivered emails to *Mail-v4-DNS-v6* is consistently higher than for IPv6-only email servers (*Mail-v6-Baseline*) – 285/436 (65.4%) vs. 193/436 (44.3%) (regular providers), 8/13 (61.5%) vs. 5/13 (38.5%) (large providers), and 46.5% vs. 7.4% (spammers) – and lead to the conclusion that IPv6 support is more prevalent among DNS resolvers than among email servers. The difference is particularly remarkable for SPAM, and suggests that spammers rely on external DNS resolvers. In comparison to *Mail-v6-Baseline*, delivery towards *Mail-v6-DNS-v6* is, if at all, only slightly reduced – 193/436 (44.3%) vs. 170/436 (39%) (regular providers), and 5/13 (38.5%) vs. 5/13 (38.5%) (large providers) –, i.e., IPv6 support at the email server typically implies IPv6 support at the respective DNS resolver. For the regular providers, Table 6 shows interdependencies concerning IP support: Most dominant are dual stack implementations 171/436 (39.2%) resp. IPv4-only configurations for email and DNS 125/436 (28.7%), as well as IPv4-only email servers with dual stack DNS resolvers 116/436 (26.6%).

Table 6: DNS and email server IP support levels (IPv4 only, IPv6 only or dual stack) of regular providers; reads f.e. 22 (5.0%) have dual stack email servers, but IPv4-only DNS resolver.

		Email					
		IPv4		IPv6		Dual	
DNS	IPv4	125	28.7%	1	0.2%	22	5.0%
	IPv6	0	0.0%	0	0.0%	1	0.2%
	Dual	116	26.6%	0	0.0%	171	39.2%

Key Findings. In summary, we find that less than half of all regular email providers support IPv6 for their email deployments. Interestingly, IPv6 support for DNS is more frequent, even for providers that do not support IPv6 for their email servers. We conjecture that this is connected to – especially in smaller setups – using public resolvers like the commonly known Cloudflare (1.1.1.1) or Google (8.8.8.8) instances. Interestingly, we also find that 23/436 (5.3%) of the observed

providers *do* use IPv6 for their email setup while *not* using it for their DNS resolvers. Even though finding this case is not unsurprising – PowerDNS, for example, does not perform IPv6 resolution by default—it still means that these operators are not able to deliver emails to IPv6-only zones, even though their email servers support IPv6.

5.2 TLS Configuration

TLS Enforcement. If our target *Mail-Dual-TLS-force* enforces the use of TLS, 392/436 (89.9%) of the regular and all large providers behave accordingly. These numbers indicate a high prevalence of TLS capability among email servers. Concerning SPAM, TLS enforcement has a considerable effect and reduces the number of emails to 34.0%.

TLS Validation. In the presence of invalid certificates, as provided by *Mail-Dual-TLS-invalid*, a similar picture emerges for regular and large providers. As common practice suggests [13] providers regularly fall back on opportunistic STARTTLS. Just one of the regular providers is more strictly configured and rejects email delivery in the case of a certificate with a non-matching CD/DNS0 entry. TLSA mismatch as caused by *Mail-Dual-TLSA-invalid* should technically prevent opportunistic encryption from being used. However, we find that only 138/436 (31.7%) of regular providers and 3/13 (23.1%) of large providers honor the TLSA record and refuse delivery. When we turn our eye to SPAM delivery, we find that enforcing TLS has a significant impact on the number of received emails. On our two TLS-enforcing targets, only 27.5% (*Mail-Dual-TLS-Force*) and 31.8% (*Mail-Dual-TLSA-Invalid*) of the baseline values of emails are received.

Key Findings. The broad majority of providers support TLS. However, emails from 10.1% of regular providers in our dataset would be lost in case of enforcing it. Providers fulfilling TLS enforcement typically also fall back on opportunistic encryption in case of invalid certificates. TLSA – a method to move beyond opportunistic encryption, even in the absence of CA-signed certificates – is sadly ignored by the majority of providers. At the same time, TLS enforcement does not only increase security, but it also reduces SPAM by more than 65%. While spammers could implement TLS quickly, it still would force them to adopt more costly TLS handshakes.

5.3 DNSSEC Validation

Targets *Mail-v4-DNSSEC-broken* and *Mail-v6-DNSSEC-broken* allow to infer the prevalence of resolvers validating DNS records. For regular providers, 181/436 (41.5%) delivered emails to our first target. The remaining 255/436 (58.5%) of all providers conducted a thorough validation for DNSSEC. Among the large providers, DNSSEC validation appears less prevalent: Only 4/13 (30.8%) (IPv4) and 2/5 (40.0%) (IPv6) of providers validate DNSSEC. We suspect that operators refrain from deploying DNSSEC to avoid customers missing emails or being unable to send emails due to misconfigurations. Furthermore, we observed a significant SPAM reduction for domains with broken DNSSEC. We conjecture that this is due to common open resolvers that validate DNSSEC being regularly used by spammers. This suspicion was confirmed when we revisited our DNS servers' logs to identify the most commonly used DNS resolvers. Query logs are, however, not fully available as log rotations removed some logs due to high response numbers. Still this enabled us to identify the most commonly used DNS resolvers. We were able to match resolvers for 2839/4660 (61%) regular emails and for 3399/6772 (50.2%) of emails sent by spammers. We found 1,443 unique resolver IPs for regular providers and 1,774 for spammers. Relying on MaxMind's public GeoLite AS database, we lookuped AS information for each IP. This resulted in 259 unique ASes used for DNS resolution for regular providers and 269 for spammers. Comparing the DNS servers used by regular and large providers with those used by spammers revealed an overlap of 138 IPs and 62 ASes.

Key Findings. DNSSEC validation is performed in 255/436 (58.5%) (IPv4) and 143/193 (74.0%) (IPv6) and regular providers. The numbers for large providers are lower, i.e., 4/13 (30.8%) (IPv4) and 2/5 (40.0%) (IPv6). In comparison, previous work [8] found DNSSEC to be less common; however, those measurements focused on zones using DNSSEC. The numbers for DNSSEC validation among spammers are – surprisingly – comparable to those of large providers. However, this connects to spammers regularly using public resolvers that already validate DNSSEC.

5.4 Anti-SPAM (Greylisting)

The greylisting targets *Mail-v4-Greylisting* and *Mail-v6-Greylisting* provoked an error in delivery the first time and accepted the email in a second – delayed – attempt. Legitimate providers reattempt to deliver emails in case of a failure, and our measurements indeed show that this is the case. Only 4/436 (0.9%) (IPv4) and 1/193 (0.5%) (IPv6) of the regular providers refrain from retransmission, and no large provider does so. However, greylisting reduces the number of received SPAM emails by 36.9%. Interestingly, this makes greylisting a less effective anti-SPAM measure than enforcing TLS.

Key Findings. Greylisting reduces the SPAM volume by 36.9% and does not introduce delivery problems for legitimate email. However, greylisting has less impact than TLS enforcement, which reduces SPAM by over 65%.

6 Related Work

In the past years, email has been receiving significant attention from the research community. In this section, we systematize eleven email-related measurement studies from 2014 onward.

Table 7: Measured adoption rates by related work. Percentages are collected for domains with MX records. SPF, DKIM and DMARC are included for comparison only as they merely influence the receiver's decision to accept incoming emails.

Citation	Year	Active Meas.	Domains	Sample Size	SPF	DKIM	DMARC	DNSSEC	DANE	TLS (inc.)
Adkins et al. [2]	2014		Facebook	/	-	-	-	-	-	76%
Foster et al. [17]	2015		Alexa	1M	42.3%	-	1%	3.4%	-	-
Foster et al. [17]	2015		Adobe	1M	43.6%	-	0.9%	2.8%	-	54%
Durumeric et al. [14]	2015	•	Gmail	/		-	-	-	-	80%
Durumeric et al. [14]	2015		Alexa	1M	47%	-	1.1%	-	-	81.8%
Hu et al. [22]	2018		Alexa	1M	44.9%	-	5.1%	-	-	-
SIDN [44]	2019		.nl	5.9M	44.2%	18.6%	8%	53%	-	62%
Kambourakis et al. [26]	2019/20	•	Custom	3236	80.7%	59.4%	51.3%	23.2%	17.6%	97.6%
Lee et al. [31]	2020		Alexa	100K	-	-	-	-	0.5%	-
Tatang et al. [45]	2021		*x*	2.04M	50%	13%	11%	-	-	-
Yajima et al. [34]	2021		Tranco	10K	88.7%	-	54.1%	7.7%	0.8%	-
Our work	2020/21	•	Custom	417	91.3%	63%	53.5%	57.4%∗	21.6%∗	89.9%

∗: We can only verify the percentage of DNSSEC resolvers and TLSA validating email servers.
•: Studies with active measurements
x: Mix of Alexa top 1M, Tranco, Majestics

We find that these studies use different sample sets and measurement methodologies. Sample sets range from top 1M domain lists to email collections with sample sizes from a million domains to a few thousand. However, using different methodologies, they all ultimately report comparable adoption rates of security-related email protocols, including SPF, DKIM, and DMARC. Hence, we compare their adoption rates and findings to our results in Table 7 to validate our methodology and provide a comprehensive picture of current providers' email delivery capabilities. Related work on email delivery so far primarily focused on large providers and did not consider the transport perspective – especially IPv6 and DNS – highlighting the gap our work fills.

Adoption Rates. Looking at the reported adoption rates from related work, we do find an upward trend in adoption, especially for security-related standards. We can also observe the difference in adoption rates per region. For example, `.nl` sees a 53% adoption rate of DNSSEC, which is significantly higher than the, e.g., 7.67% adoption rate for DNSSEC for Tranco Top 10K domains reported by Yajima et al. [34]. We attribute this high adoption rate to the Registrar Scorecard, a campaign incentivizing the deployment of standards by the Dutch domain name registrar SIDN, responsible for the .nl top-level domain [44]. In contrast to the number of DNSSEC-enabled zones, we find the number of validating resolvers to be considerably higher. We find a 57.35% of participants in our study rely on DNSSEC-validating resolvers, mostly due to common public resolvers, for example, the popular 8.8.8.8 resolver offered by Google.

Large Providers. Related work uses several methods for identifying and ranking large email providers (see Table 8): Durumeric et al. [14], Hu et al. [22], and Tatang et al. [45] used manual rankings by relying on their own expertise. However, this might induce bias towards the researcher's experience and location. Foster et al. [17], and Lee et al. [31] relied on email address domains from the leak of Adobe user records

Table 8: Overview of large provider sets used in related work.

Year	Rel. W.	Overlap	Size	Method
2015	Durumeric et al. [14]	6	19	Manually
2015	Foster et al. [17]	3	22	Adobe leak
2018	Hu et al. [22]	1	35	Manually
2020	Lee et al. [31]	2	29	Adobe leak
2021	Tatang et al. [45]	2	25	Manually
2021	Liu et al. [32]	11	15	Custom
2021	Our work		15	passive DNS

in 2013 [28] to rank email providers. However, this approach is limited to a one-time data dump and in completeness as it cannot detect different domains pointing their `MX` records at the same provider. Liu et al. [32] proposed a more comprehensive approach to detect and rank email providers in 2021. One of their major components is certificate information gathered through Internet-wide SMTP handshakes. In contrast, we introduce a new ranking method based on already existing passive DNS data from DNSDB (see Section 3). Based on this ranking we list the top 15 providers in Table 4. Our method thereby overlaps highly with the results of Liu et al., while introducing significantly less measurement overhead and revealing additional providers.

Sender-side Evaluation. We only found two related measurement studies relevant to the sender-side aspects of email delivery [8, 31]. Chung et al. [8] performed a study focusing on DNSSEC adoption independent of email delivery setups in 2017. They set up ten differently misconfigured target domains (missing, incorrect, expired RRSIGS; missing DNSKEYs; incorrect DS; etc.), collecting data from 4,427 DNSSEC capable resolvers (DO bit set) from the Luminati proxy service. They found that 3,635 (81.1%) failed to validate DNSSEC responses. Only 543 (12.2%) resolvers did handle all ten different scenarios correctly. As we did not focus on DNSSEC validation specifically, but only wanted to test if validation was attempted, we relied on a single DNSSEC

setup for our measurement. Similar to us, Lee et al. [31] used 14 target domains to measure DNSSEC, STARTTLS, and DANE validation in 2020. However, they only measured the top 29 providers ranked by email addresses in the Adobe leak. The measurement setup is similar to ours, but contrary to Lee et al., we actively engaged participants to send emails to our target domains. Hence, we were able to cover a wider range of providers. Our set of large providers also differs from Lee et. al. as we used a more comprehensive ranking method, similar to that of Liu et al. [32]. Other studies evaluate email-related protocols from the receiver's perspective [2,14,17,26], i.e., evaluating emails once they are successfully delivered. For example, studying DNS TXT records between 2015 and 2018, van der Toorn et al. [46] observed a rise in the adoption of email security standards, such as SPF and DKIM, and attributed this to stricter policies from large email providers. However, this line of work generally finds similar problems on the receiver side as we observed on the sender side, e.g., the high complexity of standards, generally low adoption, and therefore, low validation rates. Durumeric [14] found that SPF network ranges are usually configured overly broad, e.g., nearly 30% of domains allow IPv4 address ranges of more than a /16 to originate emails. Furthermore, SPF inclusions are not used carefully, and a multitude of domains trust the same handful of cloud providers. Hu et al. [22] found that 34 of 35 (97%) of popular email providers deliver forged emails to inboxes even if validation of either one or multiples of SPF/DKIM/DMARC failed. Tatang et al. [45] compiled a list of DKIM selectors and found that domains do not only commonly share the same selector, but also the same key.

Standard Complexity. In 2021, Yajima et al. [34] first discussed how standards' complexity influences their adoption rate. They measured DNS-based security mechanisms and found that setup difficulty influences the adoption rate. Their rating of setup difficulty awards points for the following configuration aspects: DNS record (1pt); DNS server configuration (2pt); email server configuration (2pt); web server configuration (2pt); required third party (3pt). DNSSEC and DANE score the highest with 6 points. While DANE is a relatively new standard introduced in 2012, DNSSEC was introduced in 1999 and still faces a relatively low adoption and validation rate. Potential causes include a (perceived) high risk of service disruptions due to misconfigurations – even in 2021, we still regularly see outages of top-level domains due to misconfigured DNSSEC [24] – and complexity in maintaining DNSSEC. Further investigating the complexity of DNSSEC key material handling, Chung et al. [8] found that a majority of domains roll keys too infrequently, use weak keys, or do not perform rollovers correctly.

7 Discussion

Successful system operation includes design, implementation, and maintenance. In a world of ubiquitous networking, systems like the email ecosystem cannot be redesigned from scratch, but have to be carefully adapted. This means that successful further development has to consider the impact of improvements on the existing ecosystem. Hence, our measurement provides a perspective on the current state of email.

Our measurements pinpoint an apparent gap between the email ecosystem as standardized by the IETF and its actual deployment. Recently introduced standards such as TLSA (validation) have not made it into practice. Thus, our results suggest that the development of new email standards has to be accompanied by strategies fostering their actual deployment.

7.1 Heavy-tail Email

A pattern that emerges in our measurements as well as in the work of, e.g., Liu et al. [32] is the heavy-tail nature of email: As Table 4 shows, a small portion of operators provide email services to the majority of users and domains on the Internet. Our investigation of related work also shows that studies often focus only on this top part of email providers. However, when we want to understand the email ecosystem, the major challenge is identifying and measuring the diverse tail of email providers and small self-hosted email instances. This becomes particularly challenging if – like in our measurements – user participation is necessary, and might lead to a situation where smaller providers are less investigated with potentially negative impact on their security, resilience, etc.

In a more techno-philosophical dimension, this development also raises concerns in the context of centralization. For example, in 2021 Fiebig et al. measured the migration of universities to large cloud providers, including their email infrastructures [15]. Centralization might accelerate the adoption of standards (e.g., if the relevant players are directly involved in standardization), but this can also potentially enforce the deployment of burdensome standards by small operators, effectively creating a walled garden. Beyond, failure of a single large provider, either due to an accidental error or a deliberate attack, affects a large share of users/domains, emphasizing the importance of decentralization and diversity for the resilience of the overall email ecosystem.

What we certainly highlight – if we want to keep a distributed Internet – is that future development efforts should not only focus on improving standards themselves, but also make it easier to follow these standards and enable operators to run their email infrastructure in full standard compliance. We encourage RFCs drafted by the IETF to be accompanied by *technical and organizational measures facilitating implementation*, reducing the gap between standardization and deployment.

7.2 Delivery vs. Adoption

Looking at the large provider dataset in our study, we find that currently especially large providers prioritize email delivery over security, e.g., DNSSEC validation is enabled for

Google's public DNS service, but not for the resolvers Gmail relies on. This is understandable from an operational standpoint but suggests that security is still considered subordinate to functional goals. We conjecture that Google prioritizes the deliverability of emails over strict enforcement of DNSSEC. The status-quo appears to represent an upside-down world: Precisely for large providers, the deployment of a new security feature appears manageable; yet, they refrain from doing so in a strict manner. At the same time, small operators implement the respective features at a disproportionate operational overhead.

This divergence of the email ecosystem ultimately creates challenges, as new security features often do address actual problems. Hence, the operations community must discuss how this divide can be addressed in the future. The Registrar Scorecard has already proven that financial incentives are successful [44]. Thus, we suggest including the design of such systems already during standardization. The Internet Governance Forum also recommends financial incentives by translation of standards into business cases [47]. However, this poses various challenges, among others the collaboration of multiple stakeholders, funding, and the operation of respective evaluation systems, which have to be solved by future work.

7.3 Standard Deployment and SPAM

In our study, we find that TLS enforcement and IPv6-only delivery have a significant impact on the amount of SPAM systems receive. While IPv6-only delivery naturally has a significant negative impact on legitimate emails being delivered, this impact is smaller when enforcing TLS. According to our measurements, emails from about 10% of regular providers would be affected. However, it is hard to determine an adoption threshold for which enforcement of standards is justified. On the one hand, TLS is an old and well-understood standard, fully supported by large providers which represent the driving force in standard deployment; also the implementation effort is low compared to other standards like DNSSEC or DANE. On the other hand, it is unclear why 10.1% of these providers have not implemented (START)TLS. If this is the case because delivery is still possible without, enforcement of TLS should take place; if the reasons are rooted in structural aspects (e.g., lacking support for certain types of systems or adequately educated staff), we suggest to target these root causes first, again requiring additional technical and organizational measures accompanying RFCs.

8 Conclusion

We investigated email delivery, especially in terms of protocol use (IPv4 vs. IPv6, recursive DNS servers' configuration, TLS sending support) and thereby complement existing related work, which mostly investigated the receiving side of the email ecosystem. Together with a review of related work on email delivery, this allows us to paint a comprehensive picture of the complexity of email delivery in 2021.

We find that 'new' protocols and extensions relevant to email delivery, like IPv6 and DNSSEC, lack adoption. The overall ecosystem is slow in this regard, especially since large email providers prioritize email delivery and – while trying to offer as many options as possible to receive emails – take a conservative stance when trying to deliver emails to others. This highlights the importance of including the heavy-tail of smaller providers in email-related measurements. Our results show that standard deployment is lower than it could be. At the same time, we know that financial incentives work well to increase deployment rates. Hence, we suggest that such incentive systems should accompany Internet standards. However, continuous funding appears to be difficult; thus, future work should also address the impact of non-financial incentives.

Acknowledgements

This material is based upon work partially supported by (1) the Christian-Doppler-Laboratory for Security and Quality Improvement in the Production System Lifecycle; the financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development and the Christian Doppler Research Association are gratefully acknowledged; (2) SBA Research (SBA-K1), a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the province of Vienna. The COMET Programme is managed by FFG; (3) Project 877110 2big2fail funded by the Program "BRIDGE 1" (FFG); (4) Project FO999887504 DynAISEC funded by the Program "ICT of the Future"—an initiative of the Austrian Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology; (5) the European Commission through the H2020 project CyberSecurity4Europe (Grant No. #830929); and (6) by the Vienna Science and Technology Fund (WWTF) through project ICT19-056.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of their host institutions or those of the European Commission.

References

[1] Donald E. Eastlake 3rd. Domain Name System Security Extensions. RFC 2535, RFC Editor, March 1999. http://www.rfc-editor.org/rfc/rfc2535.txt.

[2] M. Adkins. The Current State of SMTP STARTTLS Deployment, 2014. Retrieved Sept. 16, 2021 from https://www.facebook.com/notes/1453015901605223.

[3] Eric Allman. sendmail 8.9.0 released. Retrieved Sept. 20, 2021 from https://www.sendmail.org/~ca/email/releases/sm890announce.html.

[4] Kurt Andersen, Brandon Long, Seth Blank, and Murray Kucherawy. The Authenticated Received Chain (ARC) Protocol. RFC 8617, July 2019. https://www.rfc-editor.org/info/rfc8617.

[5] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose. DNS Security Introduction and Requirements. RFC 4033, RFC Editor, March 2005. http://www.rfc-editor.org/rfc/rfc4033.txt.

[6] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The Menlo Report. *IEEE Security & Privacy*, 10(2):71–75, 2012.

[7] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer. OpenPGP Message Format. RFC 4880, RFC Editor, November 2007. http://www.rfc-editor.org/rfc/rfc4880.txt.

[8] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, and Christo Wilson. A longitudinal, end-to-end view of the DNSSEC ecosystem. In *Proceedings of the USENIX Security Symposium (USENIX Security 17)*, 2017.

[9] D. Crocker, T. Hansen, and M. Kucherawy. DomainKeys Identified Mail (DKIM) Signatures. STD 76, RFC Editor, September 2011. http://www.rfc-editor.org/rfc/rfc6376.txt.

[10] Dr. Steve E. Deering and Bob Hinden. Internet Protocol, Version 6 (IPv6) Specification. RFC 8200, July 2017. https://rfc-editor.org/rfc/rfc8200.txt.

[11] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. Investigating system operators' perspective on security misconfigurations. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2018.

[12] David Dittrich and Erin Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, 2012. https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.

[13] Viktor Dukhovni. Opportunistic Security: Some Protection Most of the Time. RFC 7435, December 2014. https://rfc-editor.org/rfc/rfc7435.txt.

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

[15] Tobias Fiebig, Seda Gürses, Carlos H Gañán, Erna Kotkamp, Fernando Kuipers, Martina Lindorfer, Menghua Prisse, and Taritha Sari. Heads in the clouds: Measuring the implications of universities migrating to public clouds. *arXiv preprint arXiv:2104.09462*, 2021.

[16] Tobias Fiebig, Franziska Lichtblau, Florian Streibelt, Thorben Krüger, Pieter Lexis, Randy Bush, and Anja Feldmann. Learning from the past: designing secure network protocols. In *Cybersecurity Best Practices*. Springer, 2018.

[17] Ian D Foster, Jon Larson, Max Masich, Alex C Snoeren, Stefan Savage, and Kirill Levchenko. Security by any other name: On the effectiveness of provider based email security. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2015.

[18] P. Hoffman. SMTP Service Extension for Secure SMTP over TLS. RFC 2487, RFC Editor, January 1999. http://www.rfc-editor.org/rfc/rfc2487.txt.

[19] P. Hoffman. SMTP Service Extension for Secure SMTP over Transport Layer Security. RFC 3207, RFC Editor, February 2002. http://www.rfc-editor.org/rfc/rfc3207.txt.

[20] P. Hoffman and J. Schlyter. The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA. RFC 6698, RFC Editor, August 2012. http://www.rfc-editor.org/rfc/rfc6698.txt.

[21] R. Housley, D. Crocker, and E. Burger. Reducing the Standards Track to Two Maturity Levels. BCP 9, RFC Editor, October 2011. http://www.rfc-editor.org/rfc/rfc6410.txt.

[22] Hang Hu and Gang Wang. End-to-end measurements of email spoofing attacks. In *Proceedings of the USENIX Security Symposium (USENIX Security 18)*, 2018.

[23] Bert Hubert. DNS-Camel, 2018. Retrieved Jan. 13, 2022 from https://blog.apnic.net/2018/03/29/the-dns-camel/.

[24] IANIX. Major DNSSEC Outages and Validation Failures, November 2021. Retrieved Nov. 16, 2021 from https://ianix.com/pub/dnssec-outages.html.

[25] Siyuan Jia, Matthew Luckie, Bradley Huffaker, Ahmed Elmokashfi, Emile Aben, Kimberly Claffy, and Amogh Dhamdhere. Tracking the deployment of IPv6: Topology, routing and performance. *Computer Networks*, 165:106947, 2019.

[26] G. Kambourakis, G. Draper, and I. Sanchez. What Email Servers Can Tell to Johnny: An Empirical Study of Provider-to-Provider Email Security. *IEEE Access*, 8:130066–130081, 2020.

[27] S. Kitterman. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208, RFC Editor, April 2014. http://www.rfc-editor.org/rfc/rfc7208.txt.

[28] Brian Krebs. Adobe To Announce Source Code, Customer Data Breach, October 2013. Retrieved Jun. 6, 2022 from https://krebsonsecurity.com/2013/10/adobe-to-announce-source-code-customer-data-breach/.

[29] M. Kucherawy and D. Crocker. Email Greylisting: An Applicability Statement for SMTP. RFC 6647, RFC Editor, June 2012. http://www.rfc-editor.org/rfc/rfc6647.txt.

[30] M. Kucherawy and E. Zwicky. Domain-based Message Authentication, Reporting, and Conformance (DMARC). RFC 7489, RFC Editor, March 2015. http://www.rfc-editor.org/rfc/rfc7489.txt.

[31] Hyeonmin Lee, Aniketh Girish, Roland van Rijswijk-Deij, Taekyoung Kwon, and Taejoong Chung. A Longitudinal and Comprehensive Study of the DANE Ecosystem in Email. In *Proceedings of the USENIX Security Symposium (USENIX Security 20)*, 2020.

[32] Enze Liu, Gautam Akiwate, Mattijs Jonker, Ariana Mirian, Stefan Savage, and Geoffrey M Voelker. Who's Got Your Mail? Characterizing Mail Service Provider Usage. In *Proceedings of the ACM Internet Measurement Conference*, 2021.

[33] D. Margolis, M. Risher, B. Ramakrishnan, A. Brotman, and J. Jones. SMTP MTA Strict Transport Security (MTA-STS). RFC 8461, RFC Editor, September 2018. http://www.rfc-editor.org/rfc/rfc8461.txt.

[34] Yoshiro Yoneya Masanori Yajima, Daiki Chiba and Tatsuya Mori. How prevalent is the operation of DNS security mechanisms? Retrieved Sept. 15, 2021 from https://indico.dns-oarc.net/event/39/contributions/867/.

[35] Mozilla. Public Suffix List, 2021. Retrieved Nov. 24, 2021 from https://publicsuffix.org/list/public_suffix_list.dat.

[36] OpenBSD. OpenBSD 6.7. Retrieved Oct.12, 2021 from https://www.openbsd.org/67.html.

[37] Damian Poddebniak, Fabian Ising, Hanno Böck, and Sebastian Schinzel. Why TLS is better without STARTTLS: A Security Analysis of STARTTLS in the Email Context. In *Proceedings of the USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2021.

[38] Jonathan B. Postel. Internet Protocol. RFC 791, September 1981. https://www.rfc-editor.org/info/rfc791.

[39] Jonathan B. Postel. Simple Mail Transfer Protocol. STD 10, RFC Editor, August 1982. http://www.rfc-editor.org/rfc/rfc821.txt.

[40] Postfix. Postfix stable release 3.6.0. Retrieved Oct. 12, 2021 from http://www.postfix.org/announcements/postfix-3.6.0.html.

[41] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. A primer on IPv4 scarcity. *ACM SIGCOMM Computer Communication Review*, 45(2):21–31, 2015.

[42] Marco Schmidt. Expired Domains, 2021. Retrieved March 15. 2021 from https://www.expireddomains.net/.

[43] Farsight Security. Passive DNS historical internet database: Farsight DNSDB, 2021. Retrieved Nov. 24, 2021 from https://www.farsightsecurity.com/solutions/dnsdb/.

[44] SIDN. Registrar Scorecard yields great results. Retrieved Sept. 16, 2021 from https://www.sidn.nl/en/news-and-blogs/registrar-scorecard-yields-great-results.

[45] Dennis Tatang, Florian Zettl, and Thorsten Holz. The Evolution of DNS-based Email Authentication: Measuring Adoption and Finding Flaws. In *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses*, 2021.

[46] Olivier van der Toorn, Roland van Rijswijk-Deij, Tobias Fiebig, Martina Lindorfer, and Anna Sperotto. TXTing 101: Finding Security Issues in the Long Tail of DNS TXT Records. In *Proceedings of the International Workshop on Traffic Measurements for Cybersecurity (WTMC)*, 2020.

[47] De Natris Consult Wout de Natris. Setting the Standard for a more Secure and Trustworthy Internet, 2020. Retrieved from https://www.intgovforum.org/multilingual/index.php?q=filedepot_download/9615/2023.