BREAKSPF: How Shared Infrastructures Magnify SPF Vulnerabilities Across the Internet

Chuhan Wang †, Yasuhiro Kuranaga †, Yihang Wang †, Mingming Zhang ‡,
Linkai Zheng †, Xiang Li †, Jianjun Chen †‡✉, Haixin Duan †‡§✉, Yanzhong Lin ¶, and Qingfeng Pan ¶

†*Tsinghua University*, ‡*Zhongguancun Laboratory*, §*Quan Cheng Laboratory*, ¶*Coremail Technology Co. Ltd*

***Abstract*—Email spoofing attacks pose a severe threat to email systems by forging the sender's address to deceive email recipients. Sender Policy Framework (SPF), an email authentication protocol that verifies senders by their IP addresses, is critical for preventing email spoofing attacks. However, attackers can bypass SPF validation and launch convincing spoofing attacks that evade email authentication. This paper proposes BreakSPF, a novel attack framework that bypasses SPF validation to enable email spoofing. Attackers can actively target domains with permissive SPF configurations by utilizing cloud services, proxies, and content delivery networks (CDNs) with shared IP pools. We leverage BreakSPF to conduct a large-scale experiment evaluating the security of SPF deployment across Tranco top 1 million domain names. We uncover that 23,916 domains are vulnerable to BreakSPF attacks, including 23 domains that rank within the top 1,000 most popular domains. The results underscore the widespread SPF configuration vulnerabilities and their potential to undermine the security of email systems. Our study provides valuable insights for detecting and mitigating SPF vulnerabilities and strengthening email system security overall.**

I. INTRODUCTION

Email service is one of the popular services on the internet [1]. Because of its critical position, email service has become an important target for attackers, which are often abused to conduct phishing attacks [2] and malware distribution [3]. Email spoofing attacks are a critically crucial cyber threat that can have devastating consequences for individuals and organizations, in which attackers exploit the trust individuals place in familiar senders to achieve their malicious objectives. By successfully forging the sender's identity, attackers can launch sophisticated phishing attacks and business email compromises that can cause financial loss, compromise sensitive data, and damage reputations [4]–[6].

To address the issue of insufficient authentication in the standard SMTP protocol [7], [8], researchers have proposed several solutions, including Sender Policy Framework (SPF) [9], DomainKeys Identified Mail (DKIM) [10], Domain-based Message Authentication, Reporting, and Conformance (DMARC) [11], and the Authenticated Received Chain (ARC). These protocols work together to form *the email authentication chain* [12] and ensure a complete authentication process.

SPF validation is the basic and critical step in the entire email authentication chain. If the SPF protocol can be bypassed, email authentication chains are no longer resistant to email spoofing attacks. Attackers can send realistic spoofing emails that pass the verification of email authentication protocols. A recent study [13] shows that SPF is the most commonly used email authentication protocol. 69.8% in MX domains from the Alexa Top 1M domain list have deployed SPF. The adoption rate of SPF is significantly greater than that of other protocols, including DKIM (37.0%) and DMARC (15.1%), which shows that SPF plays an indispensable role in protecting users from email spoofing attacks.

SPF is an IP-based authentication protocol that binds senders' IP addresses with the identity to be authenticated. However, this trust model is fragile because anyone who controls the IP addresses listed in an email domain's SPF record can send spoofing emails on behalf of that domain.

Moreover, SPF vulnerabilities can be magnified with the emergence of shared infrastructure. First, more and more organizations and institutions host their email services to professional email service providers [14]. Most email providers require their clients to include SPF records of email providers in their own SPF records, which directly leads to the centralization of SPF deployment. Namely, a large number of email service SPF records rely on several large email providers. This trend runs counter to the fundamental principle of SPF, which is designed to establish identity authentication based on IP addresses. A single IP address may be able to send emails on behalf of thousands of domains, exacerbating the risks SPF protocols face. Second, the era of cloud services has lowered the barrier for attackers to obtain IP addresses. There are many ways for attackers to obtain and use some IP addresses that do not belong to them, such as cloud servers, proxy services, and serverless functions. These challenges have implications for the security of SPF protocols and call for a reevaluation of the current approaches used to authenticate email identities.

Research Gap. Previous studies [7], [15], [16] discussed the SPF risks at the theoretical level. These studies pointed out that some SPF records are configured too broadly and include too large subnets. However, they did not further analyze the attack scenarios when the IP range configuration in SPF is too broad. The success of exploiting SPF vulnerabilities for sending spoofing emails depends on two factors: (1) whether configuration issues exist in SPF records and (2) whether

✉ Corresponding authors:{jianjun, duanhx}@tsinghua.edu.cn.

Network and Distributed System Security (NDSS) Symposium 2024
26 February - 1 March 2024, San Diego, CA, USA
ISBN 1-891562-93-2
https://dx.doi.org/10.14722/ndss.2024.23113
www.ndss-symposium.org

the IP addresses included in the SPF records can be obtained by the attacker. Previous research focused on analyzing SPF records for syntax, which is not comprehensive enough to identify real vulnerable SPF records. For example, it is difficult to judge whether such an SPF record (`"v=spf1 ip4:107.21.107.7/16 mx -all"`) is vulnerable only by checking the record itself. However, we found attackers can obtain an IP address included in this SPF record from Amazon cloud service, which confirms that this SPF record is vulnerable. This kind of SPF vulnerability is difficult to discover through simple SPF measurements. We believe the potential systemic security risks in the currently deployed SPF records have been overlooked.

Our Study. In this paper, we performed the first systematic analysis of SPF vulnerabilities from the perspective of IP address availability. We designed an attack framework called BreakSPF that utilized IP addresses from shared infrastructure to exploit overly permissive SPF configuration vulnerabilities. With the BreakSPF framework, attackers can perform email spoofing attacks using any IP address sourced from public shared infrastructures. Such an attack can circumvent the protections of existing email authentication chains. To build the BreakSPF framework, we have solved the following challenges:

(i) How can a large pool of usable IP addresses be gathered to carry out BreakSPF attacks? To amass a substantial number of IP addresses for the BreakSPF attack framework, we surveyed shared infrastructures where attackers can obtain IP addresses and categorized them into five types, including cloud servers, proxy services, serverless functions, CI/CD platforms, and CDN services.

(ii) How do we utilize these shared IP addresses to launch email spoofing attacks? We proposed a novel *cross-protocol email spoofing attack technique*, incorporating CDN services and HTTP proxy services into the BreakSPF attack framework. It leverages the similarities between HTTP and SMTP protocols and the robustness of email servers since email servers will interpret HTTP request headers as illegal SMTP commands. Attackers can send crafted HTTP packets to make HTTP proxy services, and CDN services act as attack nodes, forwarding spoofing emails to the victim's email server. This technique can expand the types of shared infrastructure that BreakSPF can utilize.

(iii) How to accurately and efficiently find vulnerable SPF records affected by a particular IP address? First, intricate dependencies between domains, as well as between domains and IP addresses, pervade the SPF ecosystem. To pinpoint all vulnerable domains affected by an attacker-controlled IP address, we need to recursively gather the SPF records of all domains and construct complete SPF dependency trees. This enables mapping each IP address to the relevant ancestor domain nodes in the tree. Second, since the experiment involves millions of domain names and our access to some IP addresses is time-restricted, we must condense the search space and optimize search efficiency. To quickly retrieve vulnerable domains, we developed an algorithm to parse, store, and query SPF records. We used the algorithm to parse the SPF records of all tested domains, and we constructed an SPF reverse database mapping IP addresses to the relevant domains with the SPF dependency tree. With our designed query algorithm for the SPF reversed database, we could quickly retrieve all vulnerable domains impacted by a given IP address.

Key Findings. We collected 87,430 IP addresses from five types of shared infrastructure settings across the Internet and used them to conduct a large-scale BreakSPF experiment based on Tranco top 1 million domains. We sent several crafted emails to prominent email services to validate attack effectiveness, as shown in Figure 12. The results demonstrated that BreakSPF can bypass SPF and DMARC verification, enabling spoofed emails to enter inboxes of popular email services. Our experiments uncovered prevalent security risks raised by SPF vulnerabilities. We detected 23,916 vulnerable domain names, with 23 in the top 1,000 (e.g., microsoft.com, qq.com) and 188 in the top 10,000. We also proved centralized SPF dependencies can increase SPF vulnerability impact from the perspectives of providers and individual IPs. For example, we found four vulnerable email providers can impact over 1k domains each, and a special IP is exploitable for spoofing emails on behalf of over 10k domains. From the experimental results, we find that a small number of IPs are relied upon by a large number of domains, which implies that an attacker only needs a very low cost to conduct large-scale phishing spoofing attacks. These findings indicate that the BreakSPF attack model is indeed possible in real life and may have been exploited by attackers.

We responsibly disclosed the above vulnerabilities to the relevant domain administrators via emails and vulnerability report platforms like HackerOne. Tencent and Shopee have acknowledged and fixed our reported issues. We proposed three mitigation strategies, including port management, online detection services, and DMARC reports. We have developed an online detection tool to assist email administrators in promptly identifying SPF configuration issues. We believe our efforts will help reduce email spoofing, raise SPF configuration awareness, and improve email security overall.

Contributions. The contributions of the paper are as follows:

- We conducted the first systematic analysis of SPF vulnerabilities from the perspective of IP availability.
- We proposed a cross-protocol attack model that enables attackers to leverage HTTP proxy and CDN services to send spoofing email packets through the HTTP protocol.
- We have collected a comprehensive set of IP addresses (87,430) from five types of shared infrastructure settings across the Internet, which can be utilized for conducting BreakSPF attacks.
- We found that shared infrastructures magnify SPF vulnerabilities. Our experimental results highlight the widespread prevalence of SPF vulnerabilities on the internet and the high success rate of BreakSPF attacks.
- We proposed four mitigation strategies and disclosed our findings to relevant organizations and email service providers.

II. Background

A. SPF

SPF (Sender Policy Framework) [9] is an IP-based email authentication standard that helps protect senders and recipi-

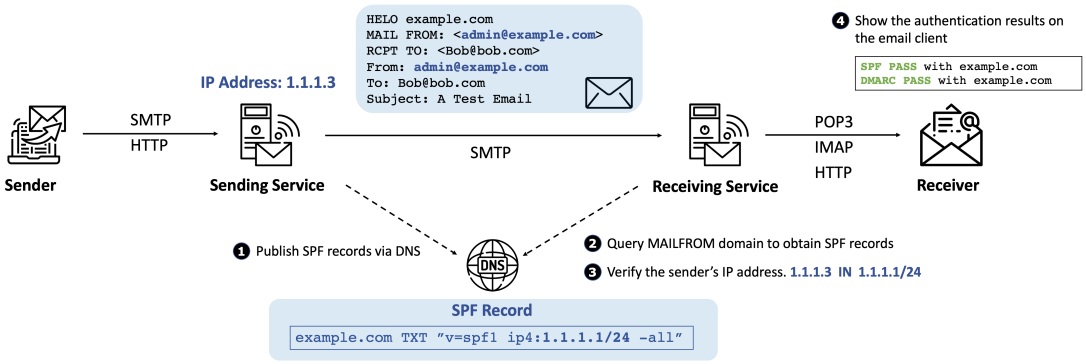


Fig. 1. SPF Verification Workflow.

ents from spam, spoofing, and phishing. Figure 1 shows the three steps of the SPF verification workflow. (1) By adding an SPF record to Domain Name System (DNS) TXT resource records, email administrators can provide a public list of senders that are approved to send emails from the current domain. (2) When a receiving email service supporting SPF protocol receives an email, it will extract the domain name in SMTP `MAILFROM` command and query the SPF record in the TXT record of the corresponding domain name. (3) Then, the IP address of the sending server is compared with the IP address lists in the SPF record. If the match is successful, the email passes the SPF verification.

SPF Record. The SPF record is a single string deployed in a DNS TXT resource record. RFC 7208 [9] specified that multiple SPF records are not permitted for the same domain name. SPF records start with "v=spf1" and consist of various informational elements that are represented by multiple `[qualifier]mechanism:value` pairs. Various mechanisms and qualifiers are defined in the protocol. The important mechanisms include:

- `all` represents the whole IP address space. "all" is generally after a qualifier (e.g., "-"). Mechanisms listed after "all" will be ignored.
- `include` represents an SPF record of another domain that is included in the current SPF record. The receiving services need to evaluate included SPF records recursively.
- `redirect` should be the last term in an SPF record. If other mechanisms fail to match, receiving services need to query the redirected SPF records to validate the current IP address.
- `ip4` represents a definite IPv4 address or IPv4 address segment. An IPv4 address segment contains an IPv4 address followed by a CIDR prefix (e.g., 192.168.1.0/24). The default `ip4` CIDR prefix is 32.
- `ip6` represents a definite IPv6 address or IPv6 address segment (e.g., 2001:db8::cd30/64). The default `ip6` CIDR prefix is 128.
- `mx` represents MX hosts of the current domain.

The qualifiers in SPF records contains `+`, `-`, `?` and `~`. Qualifiers are used to indicate what SPF verification result will be caused if the current mechanism is matched. The qualifier is optional and defaults to `+`.

- `+` represents that if the IP address matches the current mechanism, the SPF verification result is "pass".
- `-` represents that if the IP address fails to match the current mechanism, the SPF verification result is "fail". It is the opposite of `+`.
- `?` leads to a neutral result, which means the sending service is not asserting whether the host is authorized.
- `~` leads to a softfail result, which means the host is probably not authorized.

```
example.com.   TXT   "v=spf1 +mx
ip4:1.1.1.1/24 ip6:2001:db8::cd30/128
-ip4:2.2.2.2/24 include:spf.example.com
-all"
```

Fig. 2. An example of SPF Records.

Figure 2 shows an example of SPF records. It specifies that emails from the domain should be sent from an IPv4 address range (1.1.1.1/24), a specific IPv6 address (2001.db8::cd30), or IP addresses corresponding to the MX records. It also explicitly denies a specific IPv4 address range (2.2.2.2/24) and includes the SPF record from another domain (spf.example.com). The "-all" mechanism signifies a strict policy, stating that the email should be rejected if the SPF check fails.

B. Email Authentication Chains

To make up for the lack of authentication mechanisms of the Simple Mail Transfer Protocol (SMTP) [17], various email authentication protocols have been proposed these years. SPF, DKIM, DMARC, and ARC are currently the common solutions among these protocols. These protocols protect the authenticity of emails from different aspects. SPF verifies the identity of email senders by the IP address of the senders' mail transfer agent (MTA). DKIM ensures the integrity of the content by validating the senders' DKIM signature. DMARC makes an alignment test between the authentication identifiers validated by SPF and DKIM, and MIME From to ensure the authenticity of senders. These email security extension protocols constitute the email authentication chain and cooperate to complete email identity authentication, improving the email ecosystem's credibility.

DKIM. DomainKeys Identified Mail (DKIM) [10], is an email authentication protocol based on digital signatures, which can

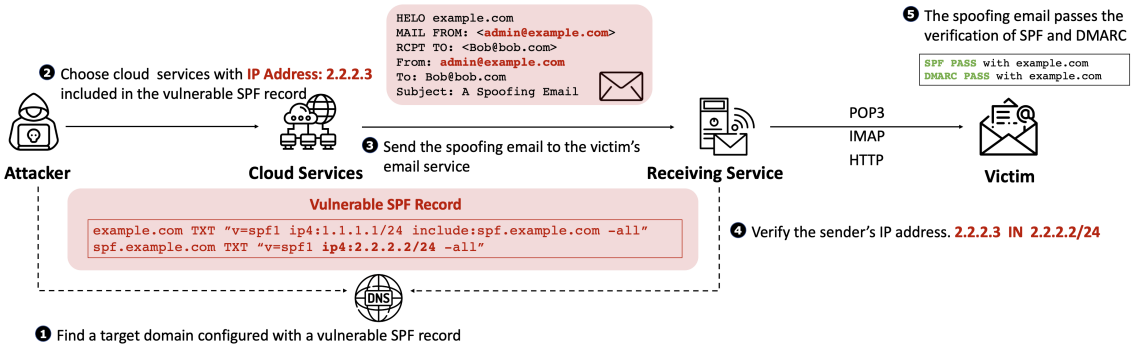


Fig. 3. BreakSPF Attack Model.

help the recipient's email server detect whether the email has been maliciously tampered with during transmission. The sending email service will calculate a DKIM signature according to the email content with its private key and add the DKIM signature in the email header. The recipient server will look up the public key of the sender's domain through DNS and verify the correctness of the DKIM signature through the public key.

DMARC. Domain-based Message Authentication, Reporting and Conformance (DMARC) [11] is a supplemental email authentication protocol to the SPF and DKIM protocols that aligns the domain name in the MIME From header with the authenticated identifier verified by either the SPF or DKIM protocol. Note that DMARC only requires either SPF or DKIM to pass. DMARC also provides a reporting mechanism that informs the domain owner about who is sending emails on behalf of that domain and the authentication results of these emails. The domain owner can use the DMARC reports to analyze the effectiveness of their email authentication measures and take steps to address any issues.

ARC. The authenticated received chain protocol(ARC) [18] is an authentication protocol used to verify the identity of forwarded emails, which solves the problem of SPF, DKIM, and DMARC protocol verification results being disrupted during email forwarding. The ARC protocol requires each hop's MTA to add an ARC signature in email headers and indicate the current SPF, DKIM, and DMARC verification results when forwarding an email. In this way, each hop's ARC signature forms a signature chain to ensure that the email's identity verification information is legal and trustworthy during transmission, improving the reliability of email delivery and reducing the possibility of false positives as spam.

III. Attack Model and Cross-protocol Attack

In this section, we will introduce the attack model of BreakSPF and propose a novel cross-protocol attack that can utilize HTTP services to send emails, which expands the IP pools attackers can exploit.

A. BreakSPF Attack Model

The objective of the BreakSPF attack is to send spoofing emails to arbitrary victims, posing as popular domains, while ensuring that these spoofing emails pass SPF and DMARC authentication. While it is widely acknowledged that configuring SPF with excessively broad IP address ranges can pose a

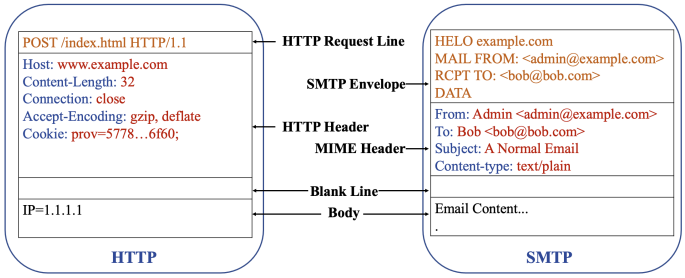
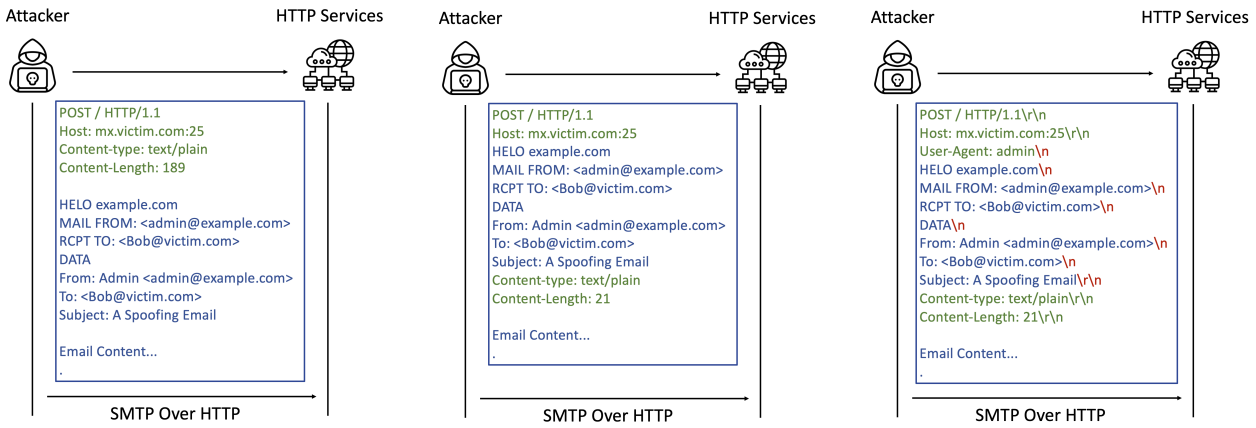


Fig. 4. The Similarities between HTTP and SMTP.

security risk, few efforts have been made to evaluate whether realistic attackers can exploit this vulnerability. Thus, we propose the BreakSPF attack model, which translates vulnerable SPF configuration problems into realistic email spoofing attacks. The vulnerabilities of SPF deployment are exploited by this attack model, which circumvents the protection offered by the current email authentication chains. Figure 3 illustrates the attack model of BreakSPF. It comprises a popular domain (e.g., example.com) configured with a vulnerable SPF record containing a wide range of IP addresses, an attacker capable of controlling multiple shared infrastructures, and arbitrary victims with their email services (such as victim.com).

In the BreakSPF attack model, we assume that (1) attackers have access to a wide variety of public services that allow them to acquire enough IP pools to bypass SPF validation, detailed in Section VI, and (2) attackers are able to identify the popular domains that contain vulnerable SPF records with the IP addresses they currently control, detailed in Section IV. The attacker is not required to have the ability to act as an active Man-in-the-Middle (MitM) attacker and change the DNS entries or perform other DNS spoofing attacks.

As shown in Figure 3, the BreakSPF attack model contains the following steps: (1) attackers find a target domain configured with a vulnerable SPF record, (2) attackers choose public services with IP addresses included in the vulnerable SPF record, (3) attackers utilize the chosen public service to send spoofing emails to the victim, (4) the email service of the victim verify the sender's IP address according to the domain in the SMTP `MAILFROM` command, and the SPF verification of this kind of spoofing emails is pass, (5) the victim will receive a convincingly realistic yet forged email that successfully passes SPF and DMARC authentication.



(a) SMTP Embedded as HTTP Body (A_1) (b) SMTP Embedded as HTTP Request (A_2) (c) SMTP Embedded as HTTP Header (A_3)

Fig. 5. Cross-Protocol Email Spoofing Attack Techniques.

B. Cross-protocol Email Spoofing Attack

We propose a novel cross-protocol email spoofing attack to expand the pool of IP addresses that can be used for BreakSPF attacks. This cross-protocol email spoofing attack utilizes HTTP services that offer HTTP forwarding functionality, such as HTTP proxy services and CDN services, to send email packets.

The cross-protocol email spoofing attack leverages the similarities between HTTP and SMTP protocols, as well as the fault tolerance of email servers. First, HTTP and SMTP are pure text-based protocols with similar structures, as illustrated in Figure 4. The data structure of HTTP and SMTP both consist of header and body sections, and their header fields are composed in a similar format, i.e., `<header name>:<data>`. The Multipurpose Internet Mail Extensions (MIME) protocol [19], initially designed for transmitting various email data formats, is also widely used for data transmission in HTTP protocol. Second, the communication processing logic of the email server has high robustness, which allows it to receive and ignore unidentified SMTP commands. Due to the aforementioned factors, an attacker can perform email spoofing attacks by sending HTTP request that embeds with email messages to a targeted email server.

In our analysis, we identify three types of cross-protocol email spoofing techniques, including SMTP Embedded as HTTP Body (A_1), SMTP Embedded as HTTP Request (A_2), and SMTP Embedded as HTTP Header (A_3). Attackers can send spoofing emails embedded in HTTP requests (as shown in Figure 5) using HTTP proxies or CDN services. In the A1 attack, we embed the entire data of SMTP communication as the HTTP body. Such HTTP packets conform to the rules of HTTP syntax and are not rejected by HTTP services. This technique requires the SMTP service to tolerate many SMTP command errors. In the A2 attack, we integrate SMTP commands and MIME headers into the HTTP headers to reduce the occurrence of SMTP command errors. However, due to significant differences between the HELO command, DATA command, and HTTP header fields, certain HTTP proxies and CDN services may consider this kind of packet as an incorrect data format and terminate the transmission. The A3 attack optimizes the A2 attack by embedding SMTP commands and MIME headers into a single HTTP header. Based on the HTTP protocol [20], HTTP services utilize CRLF (Carriage Return Line Feed, "`\r\n`") as the end-of-line marker. However, most SMTP services support both "`\n`" and "`\r\n`" as line break characters. We leverage the inconsistencies of line break interpretation between HTTP and SMTP services to construct this attack, bypassing defense strategies implemented by some proxy services against A2 attacks.

In the BreakSPF attack model, attackers utilize cross-protocol attack techniques to control CDN services and HTTP proxy services to launch email spoofing attacks. Since most CDN services support arbitrary origin servers and port configurations, we can configure the CDN's origin server as the MX record of the target email service and the original port as 25. We only need to send a crafted POST request to the domain name configured with CDN, and CDN will automatically forward this request to the target email service. Although such emails will contain some HTTP headers, they can still be accepted by email servers due to their inherent tolerance. For HTTP proxy services, attackers need to modify the HTTP request line and Host header based on the type of HTTP proxy. Experiment results about cross-protocol attacks will be discussed in Section VII.

IV. Exploitation Workflow of BreakSPF Attack

This section will introduce the exploitation workflow of the BreakSPF attack and techniques and potential challenges for each procedure. According to Figure 6, we divide the workflow of BreakSPF into the following six steps.

A. Domain Collection

First, we obtained a list of potential attack target domains from the Tranco domain list [21]. In addition, we collected the subdomains of Tranco Top 1M domains from a passive DNS dataset similar to Farsight DNSDB, provided by QiAnXin Company, since SPF records may be configured on subdomains for certain domains. The passive DNS dataset is collected from public DNS resolvers known as 114DNS, the largest DNS provider in China [22]. Our experiments involve a total of 7,183,870 domains, which include Tranco Top 1M domain names and their subdomains.

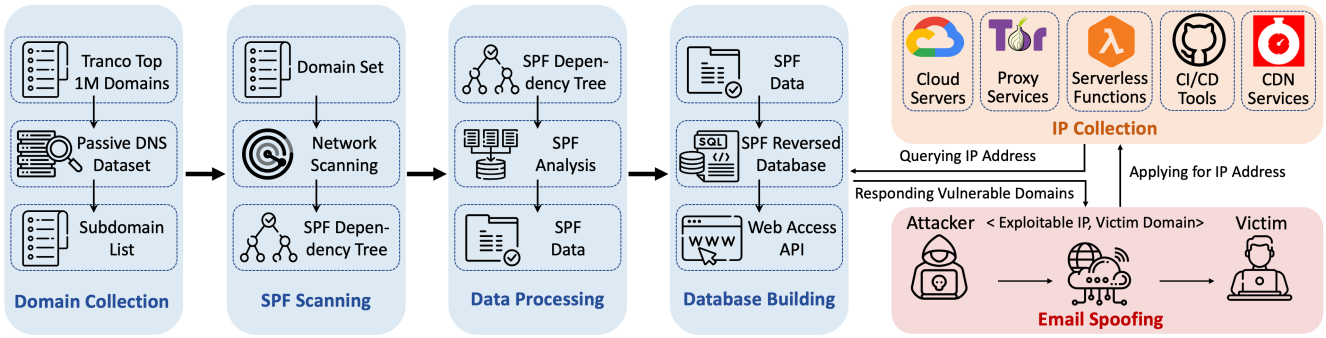


Fig. 6. Exploitation Workflow of BreakSPF Attack.

B. SPF Scanning

We then proceeded to scan the SPF records for the domain set by querying their TXT resource records with `XMap` [23] and extracting the DNS responses that began with "v=spf1" as the prefix. Because of the `redirect` and `include` mechanisms (as introduced in Section II) used in SPF records, each SPF record can establish dependency relationships with other domains. In this way, the complete SPF configuration of a domain constructs an SPF dependency tree. Each SPF record is a node of this tree structure. If we only scan the root node of the SPF dependency tree, it is hard to find vulnerable SPF records. To evaluate the status of SPF deployment more comprehensively, we need to traverse the tree structure recursively.

We recorded the domains with multiple SPF records during the recursion since they are invalid configurations according to RFC 7208 [9]. After that, we further parsed the SPF records according to the SPF syntax rules [9], extracted the domain names corresponding to `include` and `redirect` mechanisms, and scanned them recursively using the depth-first search (DFS) algorithm. We recorded all scanning results to avoid performing duplicate searches. We set the recursion depth to 10 during the scanning process. If the recursion depth exceeded 10, we stopped scanning and considered the SPF record of that domain invalid.

C. Data Processing

After scanning SPF records, we need to process the results of the SPF scanning. First, we can perform four types of analysis based on the SPF scanning results: adoption rate of SPF, grammatical analysis of SPF records, `include` mechanism analysis, and IP coverage of SPF records. Details of these analyses will be discussed in Section V. These results will provide essential data support for the BreakSPF attack.

Next, we established a reverse query mechanism for the SPF dependency tree. We re-parsed the SPF records and traversed the SPF dependency tree to record each node's ancestors. This mechanism allows us to know which domains include the SPF records of a target domain, which is critical for the subsequent attack process. Using this mechanism, we can determine which domain names a vulnerable SPF record can affect and which email providers are used by popular domain names.

D. Database Building

The most critical step in the attack process is creating mappings from the IP addresses in the SPF records to their corresponding domain names. After establishing such a correlation database, an attacker can quickly determine whether a controlled or compromised IP address is included in the SPF record of a well-known domain name.

We parsed each SPF record and extracted IPv4 addresses from the "`ip4:`" tag. Considering the IPv4 address space is too large, we optimized the storage mode using a tree structure. We converted an IP address into a 32-bit integer and an IP address block into an integer range. We used the first number in this range as the key of the database and stored the domain name and CIDR prefix length on this key. For example, the SPF record of example.com contains an IP address range of 192.168.0.0/16, and we will store {"domain":"example.com";"cidr":"16"} in the database entry corresponding to 3,232,235,520.

In this experiment, we ignored IPv6 addresses for the time being, since the IP addresses declared in SPF records are still dominated by IPv4 addresses. According to our measurement, only 2.2% domains configure IPv6 addresses in their SPF records.

We designed a query mode for a single IP address in the SPF reversed database and provided a web application programming interface (Web API). Attackers can access this web interface through the IP address they control to obtain information about which domain names the current IP address can represent to send spoofing emails. The web server provides both `GET` and `POST` request interfaces. The backend function will analyze the IP address of requests sent in `GET` methods and the IP data sent in the `POST` body. When the backend function obtained the IP address submitted by the attacker, it traversed the CIDR prefix length from large to small (32 to 1), performed an AND operation on the IP address and the subnet mask, and converted the obtained subnet prefix into an integer as a key to query in the database. Then, the database returned a JSON format response. The response may contain multiple domain names, and we iterate through them individually. If the CIDR prefix length corresponding to the current domain name in the JSON response is less than or equal to the previously enumerated CIDR prefix length, it is considered a successful hit. The backend function recorded the hit and analyzed the next domain until the program completed

the full cycle. Finally, the backend function de-duplicated the results according to the domain names and returned the results to attackers.

E. IP Collection

After setting up the reversed SPF database, attackers can easily launch the BreakSPF attack by obtaining feasible IP addresses and sending spoofed emails based on the query results from the database. The use of shared infrastructures, such as cloud services, enables attackers to acquire large volumes of IP addresses.

To conduct a comprehensive assessment of the SPF vulnerability status, we tried to obtain as many IP addresses as possible. The greater the diversity of IP addresses, the more IP addresses we can cover, and the better our experimental results will be. Therefore, we sorted out a list of the current ways attackers can obtain public IP addresses on the Internet, which includes cloud servers, proxy services, serverless functions, CI/CD tools, and CDN services. There are differences between these categories in terms of acquiring and using IPs to send spoofing emails. We explain the details of each category in Section VII. After obtaining IP addresses using the above methods, we leveraged the web API of the SPF reversed database to query and identify domains vulnerable to spoofing using these IP addresses. Meanwhile, the backend function of the web API recorded the query results for our subsequent data analysis.

Our system is designed to be extensible, allowing for the inclusion of additional IP acquisition methods within the existing framework if they are discovered in the future.

F. Email Spoofing Attack

The final step of the BreakSPF attack model is conducting email spoofing attacks. Attackers need to select a domain name influenced by the obtained IP address to send spoofing emails. Attackers can use a programming language to establish an SMTP connection with the victim's email service, acting as a Message Transfer Agent (MTA). Since the sender's IP address is included in the SPF record of the sender's domain name, these carefully crafted spoofing emails can pass SPF and DMARC verification, making them difficult to detect even for technical experts.

V. The Deployment Status of SPF

As the BreakSPF attack framework requires a scan of the current deployment status of SPF, we will introduce the deployment status of SPF in this section. Understanding the deployment status of SPF can help us analyze the feasibility and scope of the BreakSPF attack.

A. Adoption Rate of SPF

Table I shows the deployment status of SPF records among Tranco top million domains. It can be seen that 60.9% of the top million domains have deployed SPF records, and 55.9% have deployed valid SPF records. We also need to consider that not all domains in the Tranco top million provide email services, so we also measured the configuration of SPF among domains with email services. According to RFC 5321 [24], we consider domains that have configured MX records or that return an SMTP banner on port 25 in their A records as domains that provide email services. The adoption and valid rates among email domains are 79.4% and 72.7%, respectively. The adoption rate of SPF has significantly improved compared to previous measurement studies [7], [15], [16], particularly among domains that provide email services.

TABLE I. SPF Adoption Rate among Tranco Top 1 Million Domains.

Status	Top1M Domains # (%)	Email Domains[1] # (%)
Total domains	1000000 (100.0 %)	738310 (100.0 %)
w/ SPF	609,236 (60.92 %)	586,316 (79.41 %)
w/ valid SPF	559,296 (55.93 %)	536,976 (72.73 %)
Soft Fail	311,277 (31.13 %)	305,326 (41.35 %)
Hard Fail	205,181 (20.52 %)	189,984 (25.73 %)
Neutral	25,997 (2.60 %)	25,266 (3.42 %)
Pass	742 (0.07 %)	670 (0.09 %)
w/ Include	417,144 (41.71 %)	410,899 (55.65 %)
w/ Redirect	13,737 (1.37 %)	13,520 (1.83 %)

[1] Email domains refer to domains configured with email services, including domains configured with MX records and domains that provide email services on port 25 at their A records.

B. Grammatical analysis of SPF records

Although SPF is an important security protocol for verifying the sender's identity, its effectiveness relies on proper deployment. However, we have identified that 8.4% of SPF records suffer from grammar errors, which undermine the protective capabilities of SPF. We have discovered five common types of SPF deployment misconfiguration, as detailed in Table II.

Notably, the prevalent issue in SPF deployment is the occurrence of too many DNS lookups, constituting 63% of all identified grammar errors. RFC 7208 [9] specifies that it will return a `permerror` result when a single SPF resolution process involves more than ten DNS queries, Furthermore, the misconfiguration of multiple SPF records is a frequently encountered issue that also leads to `permerror` outcomes. Format errors encompass the presence of redundant or missing spaces, illegal commas, as well as misconfigured IP addresses and CIDR prefix length (e.g., 68.0.3.1/96). Spelling errors pertain to the misspelling of mechanism names, such as substituting "ip4" with "ipv4" or "ip6" with "ipv6". The coexistence of `all` mechanism and `redirect` mechanism is an invalid configuration since the redirect mechanism must be ignored when an all mechanism exists according to RFC 7208, contradicting the intended configuration. Last, our experiment results show that 742 domains even set the sender policy as "pass", meaning that attackers can send spoofing emails that can pass SPF verification using any IP address. It is even worse than the above misconfiguration.

C. Include Mechanism Analysis

Analyzing the include mechanism helps us to understand the intricate dependency relationships concealed within SPF configurations. As more and more organizations and institutions use professional email service providers or email marketing services, the `include` mechanism has been widely

TABLE II. ANALYSIS OF SPF MISCONFIGURATION.

Misconfiguration Type	# Domain	%
Too Many DNS Lookups	32,254	63.15%
Double SPF Records	15,700	30.74%
Format Errors	2,838	5.56%
Spelling Errors	986	1.93%
Coexisting `all` and `redirect`	612	1.20%
Total	51,076	100.00%

used. Email providers generally require customers to include their domain in the SPF record. According to our measurement results, among the Tranco top 1 million domains, 448,046 domains use the `include` mechanism in their SPF records, accounting for 73.5% of all domains deployed SPF protocol.

We analyzed the SPF records of all domains and identified the top ten most widely used email providers based on the number of times their SPF records were included. We first recursively scanned and traversed each domain and all the domains included in their SPF records. We counted the most frequently included domains using this approach. Then, we aggregated these domains by their second-level domains (SLD), as these domains may come from the same provider, such as *_netblocks.google.com* and *_netblocks2.google.com*. Finally, the results are shown in Table III. The third column represents how many Tranco Top 1M domains and subdomains include SPF records of this email provider. The fourth column represents the percentage of the influenced domains to all domains deployed SPF protocol.

This information could be useful for identifying email providers being targeted in the BreakSPF attack and understanding which email providers are most commonly used overall. These data also reflect the phenomenon of excessive centralization in the actual deployment of SPF, which may bring potential risks. Once the SPF record management of these email providers mentioned above is vulnerable, it may affect thousands of domains simultaneously.

TABLE III. TOP 10 EMAIL PROVIDERS BASED ON INCLUDE MECHANISM ANALYSIS.

Rank	Email Providers	# Included	%
1	outlook.com	181,544	20.07%
2	google.com	142,317	15.73%
3	amazonses.com	44,466	4.92%
4	sendgrid.net	44,200	4.89%
5	mandrillapp.com	38,437	4.25%
6	mcsv.net	38,260	4.23%
7	mailgun.org	34,790	3.85%
8	zendesk.com	30,869	3.41%
9	mailchannels.net	20,837	2.30%
10	salesforce.com	20,692	2.29%

D. IP Coverage of SPF Records

We analyzed the number of IP addresses covered by the SPF records of each domain. Figure 7 shows the relationship between the number of domains and the number of IP addresses included in SPF records, ranging from 2^0 to 2^{32}. The x-axis represents the number of IP addresses included in the SPF records, calculated with the log_2^{number} function since SPF records use CIDR prefix length to configure IP address ranges. The data analysis here excludes SPF records that have not configured any IP addresses, such as "`v=spf1 -all`". Our findings revealed that 51.7% of domains have SPF records that include more than 65,536 (2^{16}) IP addresses. This is a considerably large range, considering that most email domains do not require such a vast number of IP addresses.

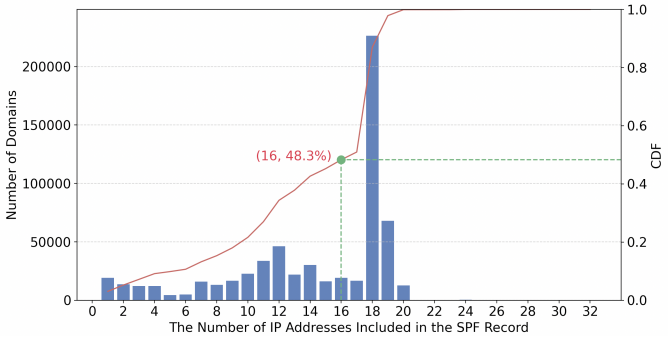


Fig. 7. IP Coverage Analysis of SPF Records. The x-axis represents the number of IP addresses included in the SPF records, calculated with the log_2^{number} function. For example, 51.7% of domains have SPF records that contain more than 65,536 (2^{16}) IP addresses.

VI. SHARED IPS COLLECTION

The BreakSPF attack framework needs sufficient IP addresses to verify the feasibility and effectiveness of the attack. We collected shared IP pools that can be used to launch Break SPF attacks on the Internet and categorized them into five types, including cloud servers, proxy services, serverless functions, CI/CD platforms, and CDN services.

A. Overview

By collecting IP addresses from the above five types of services, we obtained a total of 87,430 IP addresses and used these IP addresses to access the Web API provided by our attack framework. The details are shown in Table IV.

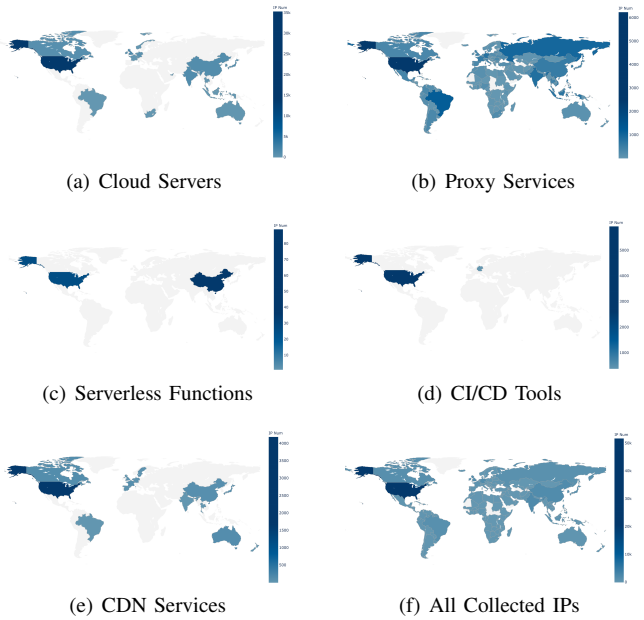
IP diversity. We analyzed IP distribution in different network blocks and ASes for each type of service. All collected IP addresses come from 201 /8 subnets, 11,162 /16 subnets, and 49,471 /24 subnets. We use a Python extension module called `pyasn` to analyze Autonomous System Number (ASN) corresponding to these IPs. These IPs come from 4,383 ASN. In addition, we further analyzed the geographical distribution of these IP addresses by country. All collected IPs cover 181 countries and regions around the world. The geographical distribution of all IPs is shown in Figure 8. This indicates that the collected IP addresses exhibit excellent geographical distribution characteristics, facilitating the analysis of SPF configuration issues in different countries and regions. The IP addresses from proxy services demonstrate the most favorable geographical distribution.

Cost. Our experiments demonstrated that attackers can obtain a large number of available IP addresses at a very low cost, with an average acquisition cost per IP address being less than $0.01. Attackers can leverage these IP addresses to carry out

TABLE IV. OVERVIEW OF THE BREAK SPF EXPERIMENT.

Services		IP Obtained	Unique IPs	Successful Hit	IP diversity				Port	
					/8	/16	/24	ASN	25	465
Cloud Servers	Alibaba	1,028	909	887	19	55	721	2	◑	●
	Amazon	9,680	9,679	8,788	21	449	7,304	2	◑	●
	Azure	33,580	30,498	6,255	22	376	10,998	1	◑	●
	Digitalocean	987	976	967	34	55	822	1	●	●
	Google	1,036	216	216	7	88	215	1	◑	●
	Linode	1,017	989	977	28	45	426	1	●	●
	Tencent	1,009	996	944	25	65	730	2	◑	●
	Vultr	307	282	277	31	46	232	1	◑	●
Proxy Services	VPN	389	339	309	102	282	306	101	◑	●
	Open Proxy	68,653	3,061	13,704	189	1,811	2,713	1,985	●	●
	RESIP	30,000	23,876	22,468	193	8,063	16,533	2,851	●	●
	Tor	1,213	1,208	1,068	108	378	592	238	◑	◑
Serverless Function	Alibaba	3,269	39	33	4	13	33	2	●	●
	Amazon	100	3	1	2	3	3	1	●	●
	Azure	1,879	13	0	1	3	4	1	●	●
	Baidu	60	3	3	2	2	3	1	●	●
	Google	46	4	4	2	2	4	1	●	●
	Huawei	234	6	6	5	5	6	3	●	●
	Tencent	7,398	62	32	8	9	38	2	●	●
CI/CD Platforms	Circleci	4,446	377	329	13	147	372	1	●	●
	Github	5,000	3,648	1,388	14	148	2,578	1	●	●
	Vercel	3,209	3,198	2,196	4	50	2,405	1	●	●
CDN Service	Gcore	13,514	200	87	18	35	74	1	●	●
	Verizon	11,157	1,097	989	4	4	13	1	●	●
	Alibaba	14,615	549	546	11	12	23	5	●	●
	Fastly	16,917	5,127	4,838	9	9	113	1	●	●
	Tencent	14,385	70	61	23	33	48	10	●	●

●: This means that the current IP source opens port 25 for outbound communication.
◑: For cloud hosting, this means that the provider uses the default blocking policy, but users can open port 25 by submitting an application form. For proxy services, this means that part of the proxy nodes have opened port 25.
●: This means that the current IP source blocks port 25 for outbound communication.



(a) Cloud Servers (b) Proxy Services (c) Serverless Functions (d) CI/CD Tools (e) CDN Services (f) All Collected IPs

Fig. 8. Global Distribution of Collected IPs from Different Sources in Our Experiment.

subsequent email spoofing attacks and phishing campaigns. The costs are almost negligible compared to the potential gains from such attacks. Many cloud service providers offer a certain amount of free usage, such as Amazon and Azure. Additionally, all cloud service providers support a pay-as-you-go model, where charges are typically based on the duration of cloud server usage. Attackers can rapidly rotate through a substantial number of IP addresses in a short period under this operational model. Serverless functions, CI/CD platforms, and CDN services also provide users with certain free allowances, enabling attackers to execute their attacks using these free resources. For instance, Github Action allows each account to use 2000 minutes for free per month. As for proxy services, open HTTP proxies and the Tor network are freely available, while residential proxy services usually charge based on traffic usage, allowing attackers to obtain over 30,000 available IP addresses for only $10.

B. Cloud Servers

Cloud computing host is the most common cloud service model. Ordinary users can apply for a cloud host on cloud service platforms such as Amazon Web Service [25], Google Cloud [26], and Microsoft Azure [27]. When applying, they can also obtain an IP address. Cloud service providers usually have a large number of IP address resources, which allows attackers to obtain IP addresses easily. Most cloud services

support a pay-as-you-go payment model, allowing users to request cloud servers based on their flexible needs. Cloud service providers offer APIs for operating cloud servers, which allows users to request and release servers quickly. This feature makes it possible for us to integrate cloud service providers into our BreakSPF attack framework.

Firstly, we will automatically request cloud servers by calling the APIs provided by cloud service providers. Then, we will obtain the IP address assigned to this server through the cloud service providers' API and send it to the query interface provided by our BreakSPF attack framework. The web API will return the query result of this IP in the SPF reverse query database. If the IP address has a potential impact on well-known domains, such as the Tranco Top 10,000 domains, we will temporarily refrain from releasing the current IP address to conduct email spoofing experiments. We will send spoofing emails from this IP address to our email service to validate the results of our attacks. Screenshots of the attack results will also facilitate our subsequent vulnerability reporting efforts.

In the eight popular cloud server providers we chose in the experiment, Linode opens port 25 for outbound communication, while other providers have blocked port 25 by default. Meanwhile, all cloud service providers, except for DigitalOcean, have the option to request the opening of port 25 through a ticket submission process.

C. Proxy Services

Proxy services are network applications that act as an intermediary between clients and target servers. Proxy services are often used to circumvent network censorship or IP-based geolocation restrictions. Proxy services usually provide a large number of egress IP addresses to facilitate user switching, and this feature can be used for our experiments. We collected four commonly used proxy services: open proxy, VPN, Tor, and residential proxy (RESIP). From Table IV and Figure 8, we can see that the proxy service has the best IP diversity, which involves a very large number of AS and has a global geographic distribution.

Open Proxy. Open proxies are a type of proxy servers that allow anyone on the Internet to connect to and use the proxy server without any authentication or authorization. We crawled all the open proxy servers from 8 open-source repositories [28]–[35] and further classified them into 3 types of HTTP-related proxies (HTTP proxy, Transparent HTTP Proxy, and HTTP Tunnel) and SOCKS proxy (SOCKS4 and SOCKS5). Figure 9 shows the differences between 3 HTTP-related proxies, where the red portion represents the destination IP and port number to be connected to through the proxy, the blue portion represents the HTTP request method, and the green portion represents the Host header (usually optional).

These repositories regularly update the available proxy lists, and the proxy file contents in the repositories are automatically updated by bots. However, it's important to note that not all of these proxies are suitable for attacks. Some proxies may become invalid in a short period of time, while others may restrict the destination IPs and ports that can be connected. Additionally, certain proxies, such as HTTP Proxy and Transparent HTTP Proxy, only allow network traffic for HTTP protocol.

```
GET http://smtp.victim.com:25/ HTTP/1.1\r\n
Host: ${proxy_host}:${proxy_port}\r\n
\r\n
```

(a) HTTP Proxy

```
GET / HTTP/1.1\r\n
Host: smtp.victim.com:25\r\n
\r\n
```

(b) Transparent HTTP Proxy

```
CONNECT smtp.victim.com:25 HTTP/1.1\r\n
Host: ${proxy_host}:${proxy_port}\r\n
\r\n
```

(c) HTTP Tunnel

Fig. 9. Three types of HTTP Proxy.

To reveal the target address limitation on the proxy servers, we set up HTTP servers listening on ports 80 and 25 and an SMTP server on port 25. We connected to these three types of servers through the proxy servers to explore their connectivity to the target servers. If a proxy can connect to port 25 of our controlled servers, and the frontend and backend IPs of the proxy remain stable, we consider it suitable for BreakSPF attacks.

TABLE V. OVERVIEW OF OPEN PROXY USABILITY

Proxy Type	Total	HTTP/80	HTTP/25	SMTP/25
HTTP(s) Proxy	39,000	1,552	1,035	N/A
Transparent Proxy	39,000	1,513	1,136	N/A
HTTP(s) Tunnel	39,000	1,307	832	536
Socks4	10,449	360	243	275
Socks5	10,775	127	66	59

As shown in the table V, after conducting dedicated experiments, we identified a total of 1,552 HTTP proxies, 1,513 HTTP transparent proxies, 1,307 HTTP tunnels, 360 SOCKS4 proxies, and 127 SOCKS5 proxies that successfully connected to port 80 on our server. This indicates that these proxy servers are indeed providing proxy services and can be used freely without authentication.

As mentioned earlier, proxy servers may impose restrictions on target ports. Thus, we also tested the connectivity to port 25, revealing that 1,035 HTTP proxies, 1,136 HTTP transparent proxies, 832 HTTP tunnels, 243 SOCKS4 proxies, and 60 SOCKS5 proxies successfully connected to port 25 on our server via HTTP protocol.

It is evident that the number of proxy servers allowing connections to port 25 is significantly lower than those allowing connections to port 80. This is because port 25 is the default port for SMTP services, so many proxy servers restrict access to prevent abuse, but there still is a considerable number of open proxies still allow connections to port 25.

VPN. Virtual private network (VPN) service is a commonly used proxy service that allows users to browse the internet anonymously and securely and hide their IP address and online activity. VPNGate [36] is an academic project maintained by

the University of Tsukuba, Japan. VPNGate provides thousands of VPN exit nodes for users to use freely. Users need to install the SoftEther VPN, an open-source server that is free to use and does not require user registration. In our experiment, we collect around 400 IP addresses from VPNGate to conduct the BreakSPF experiment.

Tor. The onion router (Tor) [37], is a free and open-source software designed to provide users with online anonymity and privacy. The official website of Tor published a list of the exit node's IP addresses [38]. In total, we acquired close to 1,200 IP exit nodes from the Tor network. We also tested the port openness of Tor network exit nodes. We carefully selected 200 IP addresses from the 1,200 Tor IP addresses for more comprehensive testing. Our findings showed that 49 out of the 200 IP addresses had an open port 465, and similarly, 43 out of the 200 IP addresses had an open port 587. The degrees of openness between the two were nearly identical. However, only 9 out of the 200 IPs had an open port 25, indicating that most Tor servers close port 25 connections to prevent their servers from being used for forwarding spam or other malicious activities.

Residential Proxy. Residential proxy networks typically use home networks to provide proxy services for users and are often used for web crawling, data mining, and automated account registration. They can help users bypass geographical restrictions and access content that is not available in their region. A previous study [39] has shown that residential proxy networks tend to maintain an expansive IP pool with a high degree of diversity. The characteristics of residential proxy networks are exactly what BreakSPF experiments require. We chose a residential proxy service that supports SOCKS proxy [40]. This residential proxy service supports automatic IP rotation, so we made 30,000 requests to our web server through the residential proxy services and obtained 23,876 distinct IP addresses. However, the residential proxy service has relatively strict restrictions on port management, generally restricting communication on ports 25 and 465.

D. Serverless Functions

A serverless function, also known as Function-as-a-Service (FaaS), is a new type of cloud computing execution model. With this technology, developers can implement programs and deploy web applications without the hassle of managing server settings. These platforms assign public IP addresses to each function instance, which customers can use to communicate with backend services like object storage buckets or servers. We conducted experiments using serverless functions on several popular cloud providers, as listed in Table IV. During the experiments, we deployed testing programs in a Python 3 environment to send emails to our controlled email servers. We rotate our testing accounts and service regions to obtain various IP addresses. However, we found serverless service providers (SSPs) only provide a small number of egress IP addresses for their users. A similar view was also confirmed by the work of Xiong et al [41]. Although the IP addresses we can collect through serverless are much smaller than those of other shared infrastructures, serverless has a more relaxed port management policy than cloud services, which allows attackers to use serverless functions to send spoofing emails successfully.

E. CI/CD Platforms

Continuous Integration and Continuous Deployment (CI/CD) tools are used to automate the software development process. They are used to integrate code changes, build the code, run automated tests, package the code into a deployable format, and deploy the code to various environments, such as staging and production. Due to functional requirements, CI/CD platform will also provide network connectivity. We found these CI/CD platforms do not impose strict restrictions on ports. Attackers can utilize the IPs from CI/CD platforms to send spoofing emails. In our experiment, we collected the IPs from three CI/CD platforms, including GitHub Actions [42], Circleci [43], and Vercel [44]. CI/CD platforms usually allow users to submit their own deployment scripts and run them. Therefore, we can deploy the test code to the CI/CD platforms and then continuously collect the IP address of the CI/CD by setting up scheduled tasks to trigger periodically once every five minutes. In addition, we found that only the CircleCI platform restricts outgoing connections on port 25. By analyzing the exit IP addresses of CircleCI, we can know that CircleCI uses Amazon cloud service as infrastructure.

F. CDN

Content Delivery Network (CDN) is a distributed network composed of server clusters located in different geographic locations. It helps client websites achieve load balancing, reduce network latency, and defend against DDoS attacks. CDNs reduce web access latency by redirecting users to cache servers closer to the user and reducing the load on the original web server. CDNs typically only forward HTTP traffic between the HTTP client and the origin server, not SMTP traffic. However, due to the similarities between the HTTP and SMTP protocols, we found that some sophisticatedly crafted HTTP traffic can also be relayed to the origin server by CDNs, and can be recognized by the SMTP server.

To leverage the extensive global nodes offered by CDNs, we first identified CDN providers that allow the origin server to be directed to port 25, as some providers impose port restrictions on the origin server. Upon investigation, we discovered that Gcore, Verizon, Alibaba, Fastly and Tencent CDN providers fit this requirement. Next, we deployed websites on these providers and configured the origin server address to correspond with our controlled server's SMTP port 25. Lastly, we initiated the crafted HTTP(SMTP) requests from the websites hosted on the CDN and monitored our controlled SMTP server for any incoming connection HTTP(SMTP) requests from the given CDN providers.

CDNs are different from other shared infrastructures in that an attacker does not control the egress IP address of the CDN. However, the CDN egress node usually has a great correlation with the geographic location of the original site. The attacker can deploy the original site in the adjacent geographic location of the target email service, collect the IP address of the CDN exit node, a further query which vulnerable domains the IP can affect through the BreakSPF attack framework, and then send spoofing emails to the victim. If the attacker can register the account of the target email service, he can first send emails through the CDN while recording the exit IP of the CDN.

VII. BreakSPF Experiment Results

A. Overview

SPF vulnerabilities are prevalent on the Internet. From our experiments, we uncover that managing SPF records is a challenging task, which potentially leads to prevalent BreakSPF attacks in the wild. According to the results, we find BreakSPF can affect a total of 23,916 domains, with 23 of them belonging to the top 1,000 domains in Tranco ranking and 1,653 domains in the top 100,000. We present the top 10 well-known domains influenced by the BreakSPF attack in Table VI, which includes prominent domains like *microsoft.com, tencent.com, trendmicro.com.*

The centralization of SPF deployment magnifies SPF vulnerabilities. Our further analysis revealed that a single IP address could influence thousands of domain names due to the centralization of SPF deployment. For instance, we found a single IP address can be used to send spoofing emails on behalf of more than 10,000 domain names. The misuse of the `include` mechanism has resulted in this phenomenon, where thousands of domain names reference the same domain name in their SPF records. This configuration is prevalent among domain names that utilize the same email provider. However, inadequate management of such a critical SPF record allows attackers to obtain IP addresses that are included in this SPF record. Table VII presents the top 10 domain groups that are influenced by a single IP address, as identified in our study.

CIDR Prefix Length Analysis of Vulnerable SPF records. This study analyzes the vulnerable CIDR prefix length of all domains influenced by the BreakSPF attack. Figure 10 shows the relationship between the cumulative distribution of vulnerable SPF domains and CIDR prefix length. The results indicate that the CIDR prefix length of most vulnerable domains is between /16 and /24, with a significant portion of domains having a prefix length shorter than /16. Notably, a prominent peak was observed in the distribution of CIDR prefix length at /16, which may stem from its common usage in SPF records despite being a potentially insecure configuration.

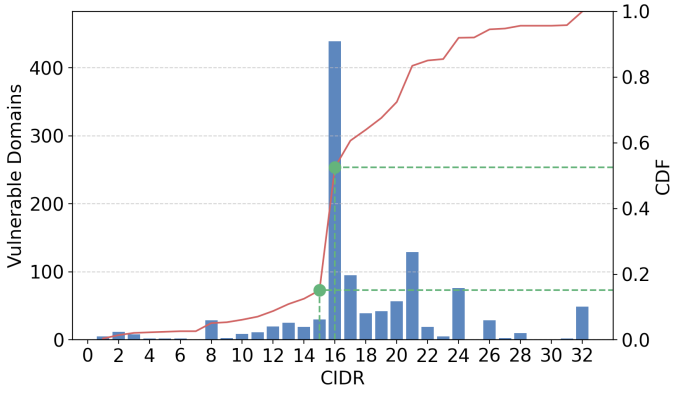


Fig. 10. CIDR Prefix Length Analysis of SPF-vulnerable Domains.

We have categorized the BreakSPF attacks based on the attacker's control over shared infrastructure into three types: (1) BreakSPF with fixed IP addresses, (2) BreakSPF with changing IP addresses, and (3) BreakSPF with cross-protocol attacks. Below, we present the experimental results for each category:

TABLE VI. Top 10 Well-known Domains Influenced by BypassSPF Attack.

Domain	Rank	IP	Source
microsoft.com	5	20.*.*.30	CI/CD Platforms
qq.com	11	114.*.*.86	Cloud Servers
csdn.net	76	114.*.*.86	Cloud Servers
huanqiu.com	110	114.*.*.86	Cloud Servers
godaddy.com	142	72.*.*.69	Tor
rednet.cn	306	114.*.*.86	Cloud Servers
mama.cn	311	114.*.*.86	Cloud Servers
zhihu.com	420	114.*.*.86	Cloud Servers
ieee.org	523	201.*.*.173	RESIP
ucla.edu	610	131.*.*.85	VPN

B. BreakSPF with Fixed IP Addresses

In this category of attacks, the attacker can maintain long-term control over a specific IP address and act as an MTA to send spoofed emails to the victim's email service directly. The shared infrastructure primarily involves cloud servers and proxy services. Common spam defense strategies, such as greylisting mechanisms, are unable to mitigate such attacks effectively.

Cloud Servers. Through IP addresses from cloud servers, we have achieved 19,327 successful hits, impacting a total of 5,462 domain names. Among these, ten domain names rank within the top 1,000, including well-known companies such as Tencent and Trendmicro. The advantage of using cloud services is that attackers can maintain continuous control over the same IP address. If an IP address is found to affect prominent domain names, attackers can retain this IP address and launch stable email spoofing attacks against any targeted victims.

Case Study of Cloud Servers. S[1] is a Chinese email provider that offers commercial promotion services such as emails and SMS to many businesses. Our experiment found that the IP addresses of Alibaba Cloud servers and Huawei Cloud servers we applied for were included in the SPF records of S's subdomains. Figure 11 shows the SPF dependency tree of S. Companies using S's email promotion services will add "`include spf.send****.org`" in their own SPF records. This allows attackers to send spoofing emails on behalf of all of S's customers. 391 domains in the Tranco Top 1M domains include S's SPF records, including *shopee.ph*, a well-known e-commerce online shopping platform in Southeast Asia. We queried the ASN of the IP addresses in S's SPF records and found that S used services from three cloud service providers, including Alibaba Cloud, Ucloud, and Huawei Cloud. This case demonstrates that email providers will also use public cloud infrastructure, which magnifies SPF vulnerabilities. Attackers can easily send spoofing emails by applying for cloud servers.

Proxy Services. With the proxy service, we obtained 24,053 successful collisions and found a total of 2,707 vulnerable

[1]As of the time we submitted our paper, this company had not responded to our vulnerability report, so we anonymized it.

TABLE VII. TOP 10 DOMAIN GROUPS INFLUENCED BY SINGLE IP.

Rank	IP	# Domain[1]	Source	Provider	Representative Domain
1	162.*.*.128	11,408	Proxy Service	HTTP Proxy	websitewelcome.com
2	114.*.*.153	4,604	Cloud Server	Tencent	qq.com
3	213.*.*.46	4,580	Proxy Service	HTTP Proxy	batmanapollo.ru
4	116.*.*.140	1,189	Proxy Service	RESIP	mailcontrol.com
5	161.*.*.149	411	Cloud Server	Alibaba	shopee.ph
8	80.*.*.207	240	Proxy Service	Tor	mailbox.org
9	154.*.*.131	131	Proxy Service	RESIP	netblocks.aserv.co.za
10	185.*.*.2	110	Proxy Service	Tor	octopuce.fr
11	133.*.*.61	97	Proxy Service	HTTP Proxy	myasp.jp
13	81.*.*.68	74	Proxy Service	HTTP Proxy	jino.ru

1. Affected Domains: represents the total number of domains that can be affected by the current IP.

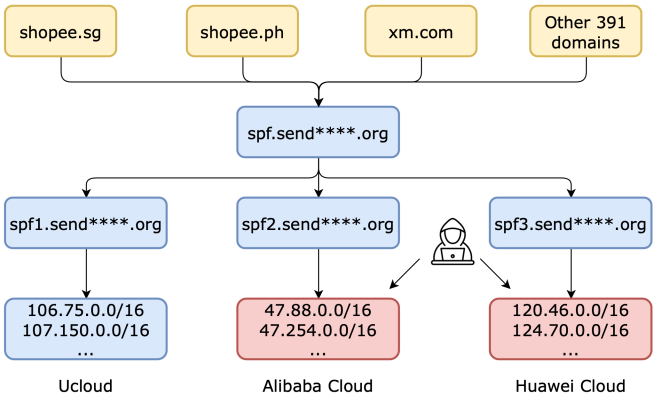


Fig. 11. SPF Dependency Tree of S.

domains, involving many well-known domains including *godaddy.com* and *ieee.org*.

C. BreakSPF with Changing IP Addresses

In this category of attacks, the attackers do not have the ability to determine the specific outgoing IP address for each external connection. Instead, they can only temporarily control the IP address through specific functionalities or methods. As a result, the attackers must dynamically assess which domains will be affected based on the outgoing IP address for each connection. However, due to the continuous variation of IP addresses, it becomes challenging to defend against these attacks using traditional IP blacklisting methods. The dynamic nature of the IP addresses used by the attackers makes it difficult to maintain an up-to-date blacklist and effectively block their malicious activities. This category of attacks involves public infrastructure, including serverless functions and CI/CD platforms.

Serverless Functions. Although the number of outbound IP addresses obtained from serverless is relatively small, 5064 domains are affected by serverless outbound IP addresses. This indicates that serverless is an efficient method for conducting email spoofing attacks. Due to the limited number of outbound IP addresses from serverless, attackers can launch successful email spoofing attacks with only a few attempts.

CI/CD Platforms. We collected 7,223 IP addresses from CI/CD platforms and achieved a total of 3,913 successful hits that could affect 145 domains, including *trendmicro.com*, a well-known cyber security company.

Case Study of GitHub Actions. During our initial scan of SPF records in April 2022, we discovered that Microsoft's subdomain (*_spf1-meo.microsoft.com*) SPF record contained IP address ranges (e.g., 20.192.0.0/10) associated with the GitHub Actions mechanism, and this subdomain was included in microsoft.com. However, when we obtained IP addresses from Github and tried to conduct a BreakSPF attack, we found that this issue had been fixed. Nevertheless, this case illustrates that even well-known technology companies like Microsoft may encounter issues in properly managing SPF records.

D. BreakSPF with Cross-protocol Attacks

This category of attacks involves shared infrastructure, including open HTTP proxies and CDN services. In this category of attacks, the attackers do not directly control any IP addresses. Instead, they utilize cross-protocol attack techniques to embed SMTP data into HTTP data packets, which are then forwarded to the victim's email service by the HTTP proxies and CDN's exit nodes. The covert nature of these attacks makes them difficult to trace and detect.

Open HTTP Proxy. We found 17,065 domain names were influenced by the IP addresses from open HTTP proxies. Among them, we found the largest vulnerable domain group, which is caused by the SPF records of *websitewelcome.com*. 11,344 domains contain SPF records of *websitewelcome.com*. We also tested the three cross-protocol attack techniques (introduced in Section III) on three types of HTTP-related proxy models: ordinary HTTP proxy, HTTP transparent proxy, and HTTP tunnel. After the HTTP tunnel establishes a connection, there is actually not much difference with the SOCKS proxy. As long as it can connect to port 25 of the destination email service, it can be used to deliver SMTP packets directly. While ordinary HTTP Proxy and transparent proxy mainly provide HTTP services, we need to use cross-protocol attack techniques. We evaluate the availability of three types of HTTP proxies for cross-protocol attacks, and the results are shown in the table VIII. The numbers in the table represent the count of unique proxy services that can successfully send emails using the three cross-protocol techniques with different types of HTTP proxy models. Since most of the HTTP proxies may be exploited through multiple attack techniques, the total number is not simply the sum of the numbers in the previous columns.

TABLE VIII. RESULT OF CROSS-PROTOCOL ATTACK TECHNIQUES ON OPEN HTTP PROXIES

Proxy Protocol	A_1	A_2	A_3	Total
HTTP(s) Proxy	1,035	970	829	2,407
Transparent Proxy	1,136	808	844	1,980
HTTP Tunnel	N/A	N/A	N/A	467

CDN Services. A total of 6854 successful hits were reached for the collected CDN exit node IPs, and 564 domains with SPF vulnerabilities were found, including *fastly.com*. Fastly, a CDN provider includes the IP addresses of its CDN exit nodes in its own SPF records. Our tests found that all CDN services can perform A1 cross-protocol attacks because A1 attacks belong to normal HTTP packets. However, CDNs usually have defensive measures against A2 attacks because CDNs usually process the headers and add colons after SMTP HELO commands and DATA commands. We found that only Tencent CDN service suffers from our proposed A3 attack. the A3 attack has a lot of flexibility, and the attacker can even initiate it through GET requests. Considering the node diversity of CDNs, it may still pose great harm.

VIII. DISCUSSION

A. Root Causes

IP addresses are not suitable for identity authentication. The ease of obtaining IP addresses makes IP addresses not a good choice for identity verification, especially with the advent of the cloud service era. Attackers can obtain a large number of IP addresses at a very low cost, as discussed above. In addition, some email providers will use public cloud service infrastructure (e.g., SendCloud), so they will include many IP address segments of cloud service providers in their SPF records. If the email provider does not strictly manage its SPF records, attackers can send spoofing emails to all users of the email provider just by applying for the same cloud service IP addresses.

Shared infrastructures magnify SPF vulnerabilities. Most email providers now provide services to users through the `include` mechanism. Users need to include the SPF record of the email provider in their own SPF record. This leads to a large number of email service SPF records relying on several large email providers, and the same IP address may be able to send emails on behalf of thousands of domain names, which increases the security risk of SPF protocol. The misuse of the include mechanism is contrary to the original intention of SPF design, which makes it impossible for different domain names to be distinguished by their sending IP addresses. Besides, in this configuration mode, the modification of SPF records has a certain lag. Email providers may use a new domain name to configure the SPF record, but its customer does not modify it in time, or its customer just includes the new SPF record but does not delete the old one, which will also cause problems.

Difficulties in SPF management. Because SPF is verified based on the IP address of the sending server, the IP address of the sending server may change with the change of assets. The email administrator needs to modify the SPF record in time according to the current situation of the sending server. For the sake of robustness, email administrators may not delete obsolete SPF records in time. Previous work [13] pointed out that similar problems also exist in the DKIM key update process. Some email administrators do not understand the SPF record configuration enough, which may also cause configuration problems, including misconfiguration or configuration of an IP address segment with a large CIDR prefix length. Currently, SPF records of most email providers contain a large number of IP addresses, but there may not be so many email servers actually used.

B. Attack Feasibility

In our experiment, we also sent spoofing emails to some renowned email services to verify the feasibility of BreakSPF attacks, as shown in Figure 12. BreakSPF attacks permit attackers to send emails that pass SPF verification. Since DMARC validation necessitates only the fulfillment of either SPF or DKIM, spoofed emails bypassing SPF can likewise sail through DMARC verification. As these spoofed emails lack DKIM signatures, the recipient's email service interprets them as originating from a domain sans DKIM configuration, which generally doesn't impede the delivery of these emails to users' inboxes.

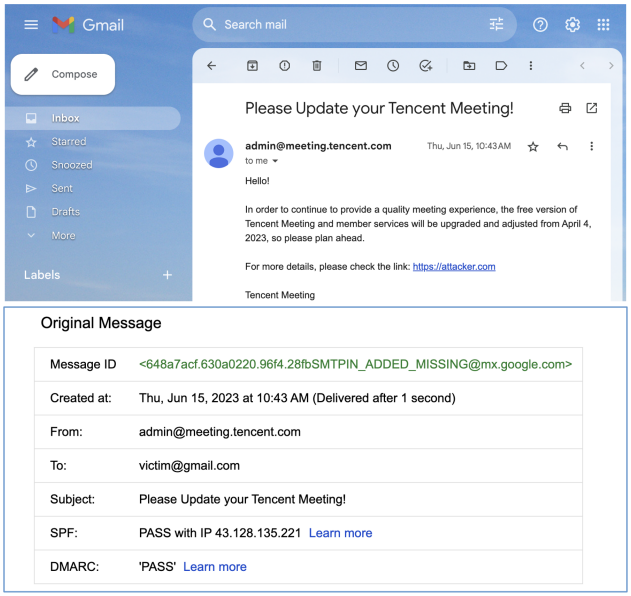


Fig. 12. A spoofing email sent to Gmail impersonating *admin@meeting.tencent.com*. The spoofing email passed the verification of SPF and DMARC.

We also acknowledge the existence of other spam detection methods may affect the effectiveness of BreakSPF attacks. However, the impact is limited. Spam filters have long been the primary defense against email attacks. Different email contents can significantly influence the outcomes of spam filters. This defense is probabilistic and may not entirely protect against the BreakSPF attack we proposed. Implementing overly strict filtering policies can lead to many false positives, which may negatively impact the receipt of legitimate emails by users. Greylisting technology [45] has a certain defensive effect on BreakSPF attacks, especially for those methods where the attacker can only temporarily control the IP address, such as serverless functions and CI/CD tools. However, if the IP

address that initiates the attack comes from cloud services controlled by attackers, attackers can resend the spoofing email after a certain time interval to bypass the protection of greylisting. Besides, Outlook email has also incorporated sender reputation and security threat intelligence as additional methods to inspect and identify spoofed emails [46]. These mechanisms require additional information, lack a unified protocol standard, and may not be implemented by all email services. It is important to note that SPF remains a critical component in the email authentication chain and is widely deployed. Many spam detection mechanisms (e.g., SpamAssassin[2]) will consider the results of SPF and DMARC validations. We believe that addressing our identified SPF vulnerabilities contributes to enhancing the overall security of the email ecosystem.

C. Ethical Consideration

In the experiment, we set up our own email service as the victim and did not send spoofing emails to any real users. For vulnerable domains, we only sent test emails to validate our attack model. We have also tried our best to contract with them. For all the cloud services involved in the experiment, we are complying with its user rules and paying the corresponding fees. We only use the normal functions of these cloud services, which will not affect the normal use of these cloud services. For DNS services involved in the experiment, we only query the SPF record of each domain name once, which will not affect the normal DNS services of the involved domains.

D. Responsible Disclosure.

We have tried to disclose vulnerabilities in two ways, to our best effort. First, we directly submitted vulnerability reports to the domain vendors that hold Security Response Center (SRC) or have cooperation with HackerOne[3], such as Tencent, Shopee, and Trendmicro. So far, Tencent and Shopee have acknowledged and fixed the reported issues. Tencent SRC claims it is challenging to sort out an accurate list of IP addresses since there are many upstream businesses and scenarios, which indicates the challenges in mitigating the SPF vulnerabilities for large companies.

Second, in the case of general domains, we tried to contact the domain administrators or security departments by sending reports to five designated email addresses, namely security@, abuse@, postmaster@, support@, and info@, according to RFC 2142 [47] and Stock et al. [48]. We have sent vulnerability report emails to all affected domains, providing detailed information on the affected domain names and the potentially compromised IP addresses. We are now receiving feedback from domain administrators and discussing with them. So far, we've received around 500 emails, with 60 domains acknowledging and thanking us for our reports. Some domain administrators responded to our emails and asked us for further details about the vulnerability. Additionally, we have received around 420 automated responses.

Before we submitted the paper, we re-tested the vulnerable domains and found the 7945 domains had already fixed their SPF vulnerability. Administrators of all vulnerable domains have at least eight months to fix the vulnerabilities before the paper is officially published.

[2]https://spamassassin.apache.org/
[3]https://www.hackerone.com/

IX. Mitigation

It is important for email service providers to monitor their SPF records regularly to prevent BreakSPF attacks, maintain the security of their services and protect their users. Besides, we think there are some effective mitigation strategies.

A. Port Management

To send spoofing emails via cloud service IP addresses, attackers need to establish a connection with several specific ports of the victim's email servers, such as 25, and 465 ports. Therefore, strengthening port management for cloud services can effectively prevent attackers from cloud IP abuse, and thus prevent BreakSPF attacks. The results of our experiments show that most cloud hosts restrict port 25, but no vendor restricts egress communication to the 465 port. Proxy services do relatively well on this point. Most proxy services have restrictions on connections to these ports.

B. Online Detection Services

We developed an online SPF vulnerability detection service based on the code of this experiment. The service can be accessed at *https://breakspf.cloud*. Email administrators only need to provide a domain name with an SPF record deployed to our online detection service, and the online detection service will automatically query the SPF record corresponding to the domain name, and perform grammar analysis on the SPF record to determine whether there exist grammar problems in the SPF record. At the same time, it will judge whether the current SPF record contains the IP addresses we obtained through the cloud service in our experiment, and calculate the intersection of the current SPF record and SPF records of other well-known domain names in our database.

C. DMARC Reports

DMARC [11] provides a good feedback mechanism called DMARC Reports. DMARC allows email administrators to configure an email address in the DMARC record for receiving DMARC feedback reports. Recipients who receive emails from this domain name will periodically collect the validation results of emails sent from that domain and send aggregated reports to the email address in the DMARC record. DMARC reports will include the number of emails sent from each IP address and the corresponding verification results. DMARC reports can help email administrators analyze whether there are problems with the current deployment of email authentication protocols and fix them in time. If the attacker's target supports sending DMARC reports, administrators can periodically check DMARC reports to detect if there exist emails sent from uncommonly used IP addresses. Therefore, improving the adoption rate of the DMARC protocol and the support for DMARC reports can mitigate BreakSPF attacks.

X. Related Work

A. Email spoofing attacks

Email services have long been exposed to the threat of email spoofing attacks. In recent years, there have been several works about email spoofing attacks. Hu et al. [7] conducted end-to-end email spoofing tests to verify the effectiveness of these email authentication protocols. Another work from Hu et al. [49] discussed problems in the design of SPF, DKIM, and DMARC protocols through a user study with 9 email administrators. Chen et al. [8] pointed out that inconsistencies between different components of email systems could be exploited by attackers to send spoofing emails. They discovered a series of evasion exploits and conduct experiments on 10 popular email providers and 19 email clients. Shen et al. [12] proposed a total of 14 different email spoofing attack techniques, and conduct large-scale measurements and analysis on 30 popular email services and 23 email clients. They demonstrated the security of chain structure-based authentication mechanisms in the email ecosystem depends on the weakest link.

Their works are based on the characteristics of email authentication protocols or the vulnerabilities brought by the cooperation of protocols to bypass SPF protection. Our work exploits the vulnerabilities in the SPF protocol deployment to construct spoofing emails. The spoofing emails sent using our attack method can pass the verification of SPF protocol and DMARC protocol, which are completely indistinguishable from normal emails.

B. Measurement of email authentication protocols

In recent years, there have been many studies measuring the current state of deployment of these email authentication protocols [13], [15], [16], [50], [51]. In 2015, Foster et al. [15] and Durumeric et al. [16] conducted two concurrent studies on the deployment status of email authentication protocols, including SPF, DKIM, DMARC, and STARTTLS. Foster et al. analyzed the adoption rate of these email security protocols based on the Alexa Top Million domains. In addition to Top Million domains, Durumeric et al. also analyzed over a year of SMTP connection data from Gmail. In 2021, Deccio et al. [50] designed a customized DNS server, measured the adoption rate of email authentication protocols through actual email interactions, and analyzed the email sender verification process with the data from the DNS server. In 2022, Bennett et al. [51] discovered two buffer overflow vulnerabilities in libSPF2 and performed a large-scale measurement of both vulnerabilities. Wang et al. [13] performed a large-scale measurement of DKIM deployment and analyzed security issues of DKIM records and DKIM signatures.

These measurement studies focus on the adoption rates of SPF, DKIM, and DMARC, as well as the invalid deployments in actual deployments. Our work analyzes the current state of SPF deployment from a different perspective than just syntactic analysis of SPF records, and we reveal that even a properly configured SPF record may still have security risks.

C. Cloud IP reuse

Pauley et al. [52] conducted a large-scale measurement of the IP reuse problem in public clouds, and they proposed the cloud squatting attack, which can exploit latent configurations of previous tenants to obtain sensitive information by applying for cloud servers. They tested 1.5 million unique IP addresses from the Amazon cloud service and received over 5 million cloud messages. The similarity is that we are both concerned about the risks associated with the abuse of public cloud service resources. The difference is that they are concerned about the problems that can be caused by the previous configuration of the tenant using that IP address, so they are more concerned about the packets that can be received by the current IP address. We are concerned with email spoofing attacks that can be initiated proactively when an attacker has access to a large number of IP addresses. Thus, our work covers many more types of cloud services.

XI. Conclusion

In this paper, we analyzed the systemic risks associated with SPF configurations in the network. We proposed the BreakSPF attack framework, which enables attackers to efficiently and accurately discover domains with SPF vulnerabilities and launch email spoofing attacks. We conducted a large-scale BreakSPF experiment based on the Tranco top 1 million domains and found that 23,916 domains were affected by the BreakSPF attack. Furthermore, we proposed novel cross-protocol attacks that amplify the impact of SPF vulnerabilities.

Our work highlights the vulnerabilities in the email authentication chain and demonstrates that shared infrastructure can magnify these weaknesses. The current email authentication chain establishes trust based on IP addresses, which may not be an optimal choice. Therefore, we need to explore better approaches to address the issue of email spoofing. With the emergence of cloud services, an increasing number of services are being deployed on shared infrastructure, leading to a shift in the trust model and potentially challenging previously established security mechanisms. We hope this research will raise awareness in the technical community regarding the security of the email authentication chain and the issue of shared infrastructure.

Acknowledgement

We sincerely thank all the anonymous reviewers and our shepherd for their valuable reviews and comments to improve this paper and especially for our shepherd's thoughtful and patient guidance. We thank the domain administrators who responded and fixed the vulnerabilities we reported. We are grateful for the support from Qi-Anxin and Coremail. We also thank Chenrui Li for assisting in editing this paper. This work was partly supported by the National Natural Science Foundation of China (62272265) and the Taishan Scholars Program. Any opinions, findings, conclusions, or recommendations expressed in this paper do not necessarily reflect the views of sponsors.

References

[1] "How many email users are there?" https://99firms.com/blog/how-many-email-users-are-there/, accessed: Apr 13, 2023.

[2] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupé, and G. Ahn, "Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale," in *29th USENIX Security Symposium, USENIX Security 2020, August 12-14, 2020*, S. Capkun and F. Roesner, Eds. USENIX Association, 2020, pp. 361–377. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/oest-sunrise

[3] S. Wen, W. Zhou, J. Zhang, Y. Xiang, W. Zhou, W. Jia, and C. C. Zou, "Modeling and analysis on the propagation dynamics of modern email malware," *IEEE Trans. Dependable Secur. Comput.*, vol. 11, no. 4, pp. 361–374, 2014. [Online]. Available: https://doi.org/10.1109/TDSC.2013.49

[4] "Business e-mail compromise the 12 billion dollar scam," https://www.ic3.gov/Media/Y2018/PSA180712, accessed: Jan 28, 2021.

[5] "Research: Crisis of fake email continues to plague industries worldwide," https://www.valimail.com/press/research-crisis-of-fake-email-continues-to-plague-industries-worldwide-2/, accessed: Jan 28, 2021.

[6] "Spam and phishing in q1 2019," https://securelist.com/spam-and-phishing-in-q1-2019/90795/, accessed: Jan 28, 2021.

[7] H. Hu and G. Wang, "End-to-end measurements of email spoofing attacks," in *27th USENIX Security Symposium, USENIX Security 2018, Baltimore, MD, USA, August 15-17, 2018*, W. Enck and A. P. Felt, Eds. USENIX Association, 2018, pp. 1095–1112. [Online]. Available: https://www.usenix.org/conference/usenixsecurity18/presentation/hu

[8] J. Chen, V. Paxson, and J. Jiang, "Composition kills: A case study of email sender authentication," in *29th USENIX Security Symposium, USENIX Security 2020, August 12-14, 2020*, S. Capkun and F. Roesner, Eds. USENIX Association, 2020, pp. 2183–2199. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/chen-jianjun

[9] S. Kitterman, "Sender policy framework (SPF) for authorizing use of domains in email, version 1," *RFC*, vol. 7208, pp. 1–64, 2014. [Online]. Available: https://doi.org/10.17487/RFC7208

[10] D. Crocker, T. Hansen, and M. S. Kucherawy, "Domainkeys identified mail (DKIM) signatures," *RFC*, vol. 6376, pp. 1–76, 2011. [Online]. Available: https://doi.org/10.17487/RFC6376

[11] M. S. Kucherawy and E. D. Zwicky, "Domain-based message authentication, reporting, and conformance (DMARC)," *RFC*, vol. 7489, pp. 1–73, 2015. [Online]. Available: https://doi.org/10.17487/RFC7489

[12] K. Shen, C. Wang, M. Guo, X. Zheng, C. Lu, B. Liu, Y. Zhao, S. Hao, H. Duan, Q. Pan, and M. Yang, "Weak links in authentication chains: A large-scale analysis of email sender spoofing attacks," in *30th USENIX Security Symposium, USENIX Security 2021, August 11-13, 2021*, M. Bailey and R. Greenstadt, Eds. USENIX Association, 2021, pp. 3201–3217. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/shen-kaiwen

[13] C. Wang, K. Shen, M. Guo, Y. Zhao, M. Zhang, J. Chen, B. Liu, X. Zheng, H. Duan, Y. Lin, and Q. Pan, "A large-scale and longitudinal measurement study of DKIM deployment," in *31st USENIX Security Symposium, USENIX Security 2022, Boston, MA, USA, August 10-12, 2022*, K. R. B. Butler and K. Thomas, Eds. USENIX Association, 2022, pp. 1185–1201. [Online]. Available: https://www.usenix.org/conference/usenixsecurity22/presentation/wang-chuhan

[14] E. Liu, G. Akiwate, M. Jonker, A. Mirian, S. Savage, and G. M. Voelker, "Who's got your mail?: characterizing mail service provider usage," in *IMC '21: ACM Internet Measurement Conference, Virtual Event, USA, November 2-4, 2021*, D. Levin, A. Mislove, J. Amann, and M. Luckie, Eds. ACM, 2021, pp. 122–136. [Online]. Available: https://doi.org/10.1145/3487552.3487820

[15] I. D. Foster, J. Larson, M. Masich, A. C. Snoeren, S. Savage, and K. 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, Denver, CO, USA, October 12-16, 2015*, I. Ray, N. Li, and C. Kruegel, Eds. ACM, 2015, pp. 450–464. [Online]. Available: https://doi.org/10.1145/2810103.2813607

[16] Z. Durumeric, D. Adrian, A. Mirian, J. Kasten, E. Bursztein, N. Lidzborski, K. Thomas, V. Eranti, M. Bailey, and J. A. Halderman, "Neither snow nor rain nor MITM...: an empirical analysis of email delivery security," in *Proceedings of the 2015 ACM Internet Measurement Conference, IMC 2015, Tokyo, Japan, October 28-30, 2015*, K. Cho, K. Fukuda, V. S. Pai, and N. Spring, Eds. ACM, 2015, pp. 27–39. [Online]. Available: https://doi.org/10.1145/2815675.2815695

[17] J. Postel, "Rfc0821: Simple mail transfer protocol," 1982.

[18] K. Andersen, B. Long, S. Blank, and M. S. Kucherawy, "The authenticated received chain (ARC) protocol," *RFC*, vol. 8617, pp. 1–35, 2019. [Online]. Available: https://doi.org/10.17487/RFC8617

[19] N. Freed and N. S. Borenstein, "Multipurpose internet mail extensions (MIME) part one: Format of internet message bodies," *RFC*, vol. 2045, pp. 1–31, 1996. [Online]. Available: https://doi.org/10.17487/RFC2045

[20] R. T. Fielding, J. Gettys, J. C. Mogul, H. F. Nielsen, L. Masinter, P. J. Leach, and T. Berners-Lee, "Hypertext transfer protocol - HTTP/1.1," *RFC*, vol. 2616, pp. 1–176, 1999. [Online]. Available: https://doi.org/10.17487/RFC2616

[21] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, ser. NDSS 2019, Feb. 2019.

[22] "114 dns." https://www.114dns.com/, accessed: April 13, 2023.

[23] X. Li, B. Liu, X. Zheng, H. Duan, Q. Li, and Y. Huang, "Fast IPv6 Network Periphery Discovery and Security Implications," in *Proceedings of the 2021 IEEE/IFIP International Conference on Dependable Systems and Networks*, ser. DSN '21, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9505062

[24] J. C. Klensin, "Simple mail transfer protocol," *RFC*, vol. 5321, pp. 1–95, 2008. [Online]. Available: https://doi.org/10.17487/RFC5321

[25] "Cloud services - amazon web services (aws)." https://aws.amazon.com/, accessed: January 30, 2023.

[26] "Cloud computing services — google cloud." https://cloud.google.com/, accessed: January 30, 2023.

[27] "Cloud computing services — microsoft azure." https://azure.microsoft.com/en-us/, accessed: January 30, 2023.

[28] "ProxyList.to Proxy List," ProxyList.to. [Online]. Available: https://github.com/proxylist-to/proxy-list

[29] M. Güvençli, "Free HTTP Proxy List." [Online]. Available: https://github.com/mertguvencli/http-proxy-list

[30] monosans, "Proxy-list." [Online]. Available: https://github.com/monosans/proxy-list

[31] jetkai, "Socks4/5 & http/s proxies // online + archive." [Online]. Available: https://github.com/jetkai/proxy-list

[32] prxchk, "Best Free Proxy Servers." [Online]. Available: https://github.com/prxchk/proxy-list

[33] roosterkid, "Update July 15, 2021." [Online]. Available: https://github.com/roosterkid/openproxylist

[34] Z. S.R, "Proxy-Master." [Online]. Available: https://github.com/MuRongPIG/Proxy-Master

[35] SpeedX, "PROXY-LIST." [Online]. Available: https://github.com/TheSpeedX/PROXY-List

[36] D. Nobori and Y. Shinjo, "VPN gate: A volunteer-organized public VPN relay system with blocking resistance for bypassing government censorship firewalls," in *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2014, Seattle, WA, USA, April 2-4, 2014*, R. Mahajan and I. Stoica, Eds. USENIX Association, 2014, pp. 229–241. [Online]. Available: https://www.usenix.org/conference/nsdi14/technical-sessions/presentation/nobori

[37] R. Dingledine, N. Mathewson, and P. F. Syverson, "Tor: The second-generation onion router," in *Proceedings of the 13th USENIX Security Symposium, August 9-13, 2004, San Diego, CA, USA*, M. Blaze, Ed. USENIX, 2004, pp. 303–320. [Online]. Available: http://www.usenix.org/publications/library/proceedings/sec04/tech/dingledine.html

[38] "Tor exit list," https://check.torproject.org/torbulkexitlist, accessed: Apr 13, 2023.

[39] X. Mi, X. Feng, X. Liao, B. Liu, X. Wang, F. Qian, Z. Li, S. A. Alrwais, L. Sun, and Y. Liu, "Resident evil: Understanding

residential IP proxy as a dark service," in *2019 IEEE Symposium on Security and Privacy, SP 2019, San Francisco, CA, USA, May 19-23, 2019*. IEEE, 2019, pp. 1185–1201. [Online]. Available: https://doi.org/10.1109/SP.2019.00011

[40] "Residential-proxy." https://www.zishonproxy.com/residential-proxy/, accessed: April 13, 2023.

[41] J. Xiong, M. Wei, Z. Lu, and Y. Liu, "Warmonger: Inflicting denial-of-service via serverless functions in the cloud," in *CCS '21: 2021 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, Republic of Korea, November 15 - 19, 2021*, Y. Kim, J. Kim, G. Vigna, and E. Shi, Eds. ACM, 2021, pp. 955–969. [Online]. Available: https://doi.org/10.1145/3460120.3485372

[42] "Understanding github actions." https://docs.github.com/en/actions/learn-github-actions/understanding-github-actions, accessed: April 13, 2023.

[43] "The continuous integration platform preferred by over 1 million engineers." https://circleci.com/, accessed: April 13, 2023.

[44] "A better way to build software." https://vercel.com/workflow, accessed: April 13, 2023.

[45] M. S. Kucherawy and D. Crocker, "Email greylisting: An applicability statement for SMTP," *RFC*, vol. 6647, pp. 1–17, 2012. [Online]. Available: https://doi.org/10.17487/RFC6647

[46] "Anti-spoofing protection in eop," https://learn.microsoft.com/en-us/microsoft-365/security/office-365-security/anti-phishing-protection-spoofing-about, accessed: July 17, 2023.

[47] D. Crocker, "Mailbox names for common services, roles and functions," *RFC*, vol. 2142, pp. 1–6, 1997. [Online]. Available: https://doi.org/10.17487/RFC2142

[48] B. Stock, G. Pellegrino, F. Li, M. Backes, and C. Rossow, "Didn't you hear me? - towards more successful web vulnerability notifications," in *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, 2018. [Online]. Available: https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_01B-1_Stock_paper.pdf

[49] H. Hu, P. Peng, and G. Wang, "Towards understanding the adoption of anti-spoofing protocols in email systems," in *2018 IEEE Cybersecurity Development, SecDev 2018, Cambridge, MA, USA, September 30 - October 2, 2018*. IEEE Computer Society, 2018, pp. 94–101. [Online]. Available: https://doi.org/10.1109/SecDev.2018.00020

[50] C. Deccio, T. Yadav, N. Bennett, A. Hilton, M. Howe, T. Norton, J. Rohde, E. Tan, and B. Taylor, "Measuring email sender validation in the wild," in *Proceedings of the 17th International Conference on emerging Networking EXperiments and Technologies*, 2021, pp. 230–242.

[51] N. Bennett, R. Sowards, and C. T. Deccio, "Spfail: discovering, measuring, and remediating vulnerabilities in email sender validation," in *Proceedings of the 22nd ACM Internet Measurement Conference, IMC 2022, Nice, France, October 25-27, 2022*, C. Barakat, C. Pelsser, T. A. Benson, and D. R. Choffnes, Eds. ACM, 2022, pp. 633–646. [Online]. Available: https://doi.org/10.1145/3517745.3561468

[52] E. Pauley, R. Sheatsley, B. Hoak, Q. Burke, Y. Beugin, and P. D. McDaniel, "Measuring and mitigating the risk of IP reuse on public clouds," in *43rd IEEE Symposium on Security and Privacy, SP 2022, San Francisco, CA, USA, May 22-26, 2022*. IEEE, 2022, pp. 558–575. [Online]. Available: https://doi.org/10.1109/SP46214.2022.9833784