Decoding DNS Centralization: Measuring and Identifying NS Domains Across Hosting Providers

Qihang Peng[1,*], Mingming Zhang[2,*], Deliang Chang[3,✉],
Jia Zhang[1,✉], Baojun Liu[1], Haixin Duan[1]
[1]Institute for Network Sciences and Cyberspace, BNRist, Tsinghua University, China,
[2]Zhongguancun Laboratory, China, [3]QI-ANXIN Technology Research Institute, China
pqh24@mails.tsinghua.edu.cn, zhangmm@mail.zgclab.edu.cn,
changdeliang01@qianxin.com, {zhangjia2017, lbj, duanhx}@tsinghua.edu.cn

***Abstract*—The Domain Name System (DNS) is designed to be distributed, which aims to provide services with low latency and great reliability. However, after decades of development and changes in Internet business models, various aspects of the DNS ecosystem have begun to show signs of centralization. To investigate the centralization from the viewpoint of hosting service providers, we develop an automated method based on similarity among NS domains and co-hosting relationship to identify the hosting providers for authoritative name servers, so that we can identify hosting providers in DNS zone file to count the number of domains which a hosting provider host. This tool demonstrates greater accuracy than previous methods and our testing demonstrates the ability to identify hosting service providers for most domains in real-world measurement tasks. Using this tool, we conducted measurements on the dataset combined with `.com`, `.net` and `.org` TLD zones. We find that the top 10 providers collectively host over 54.19% of domains while top 100 providers host over 82.99% domains, which shows a significant level of centralization in hosting service providers within the DNS. Through an analysis of NSone's NS domains and a statistical examination of top providers' NS domains, we find that directly identifying the base domain as the provider is inappropriate. Furthermore, we discover relationships among hosting providers and between hosting providers and infrastructure that are more complex than previously anticipated. Finally, based on our research findings, we offer corresponding suggestions to mitigate the continued development of centralization.**

I. Introduction

DNS, as a critical infrastructure of the Internet, is designed as a distributed system to effectively enhance the robustness and relative security of Internet services. However, with the development of the Internet, DNS is widely perceived to exhibit significant centralization, which poses potential yet tangible risks. In previous studies, researchers have measured DNS centralization across multiple layers, including traffic [1], infrastructure [2], [3], market share [4]–[6] and resolvers [5]–[7]. However, little attention has been paid to developments at the hosting provider level.

Hosting providers should be treated with serious attention cause providers are responsible for managing and operating the hosting service. When a provider's service experiences failures, the availability of the domains it hosts may be adversely affected. Besides, if a provider's services have exploitable security vulnerabilities, all domains it hosts may be affected by such attacks. Moreover, since a provider may deploy services across multiple infrastructures, these negative impacts can extend beyond a single infrastructure. In recent years, researchers have proposed several effective attacks targeting providers, We list some of them in table I. Therefore, it is essential to treat the issues about hosting providers seriously.

TABLE I: Example of attacks and affected top providers

Attack Name	Affected Top Provider
Zombie awakening [8]	Amazon, Godaddy, NSone
XDAuth [9]	Amazon, NSone
Disablance [10]	Cloudflare

Identifying hosting providers is the first step in studying them. In DNS measurement scenarios, it is necessary to determine the provider associated with each NS domain. However, this identification process is more challenging than it appears. On one hand, NS domains do not always explicitly indicate their provider; for example, `dns2.hichina.com` belongs to Alibaba, and `rs61a.registrar-servers.com` is associated with Namecheap. On the other hand, the relationship between providers and NS domains can change due to business strategies and commercial acquisitions. For instance, `ns.linode.com` originally belonged to Linode, but after Akamai's acquisition of Linode in 2022, control and hosting services over this domain were transferred to Akamai.

In the past, researchers have employed various methods to identify hosting providers. Common approaches include determining the provider by querying the autonomous system (AS) of the IP address associated with the NS domain [11], retrieving registration information via the WHOIS protocol, or combining WHOIS data with certificates of the NS domain [8]. However, these methods have certain limitations: The AS-based approach may blur the distinction between infrastructure and hosting providers, leading to potentially inaccurate identification results. Meanwhile, the WHOIS-based approach may fail to capture the latest provider information due to delays in registration updates.

Our study. We propose a new fingerprinting approach based on co-hosting relationships, where the proportion of shared

*They contributed equally to this work.

hosting among NS domains is used to determine whether they belong to the same provider. Based on this, we design an identification method that leverages clustering based on co-hosting similarity. This method requires only the TLD zone file as input and clusters NS domains based on structural similarity and co-hosting similarity to form NS groups. Each group's final provider is then determined using WHOIS data, certificates, and base domains.

Using this method, we analyze the dataset combined with TLD zone files from the `.com` zone, `.net` zone and `.org` zone from September 14, 2024. Further, We begin by observing a large number of unusual base domains within NSone, investigating the NS domains associated with NSone. Additionally, we conduct a preliminary analysis of the NS domains for the top 10 providers. Finally, We begin by examining the discrepancy between the top 10 providers and the top 10 ASes in the centralization measurements. From this perspective, we summarize the patterns of how providers deploy services on infrastructure and quantitatively measure the usage of infrastructure by leading providers.

Our findings. We first find a significant centralization phenomenon among hosting providers, with the top 10 providers with only 4.05% NS domains hosting over 54.19% of all domains. We also measure centralization at the infrastructure level for comparison: the top 10 autonomous systems collectively host more than 50% of the domains.

We then find a high degree of consistency in the hosted content between NS domains previously considered to belong to independent providers and other hosting providers. This indicates that the relationships between providers are more complex than commonly perceived. Based on this fact, we found that among the domains hosted by multiple NS base domains, only one-fifth are truly hosted by more than two providers, accounting for just 0.58% of all. This could increase the risk of single points of failure, and users may unknowingly contribute to the further centralization of the system.

Finally, We find the presence of a highly complex many-to-many relationship between hosting providers and infrastructure. Additionally, we perform basic security attribute measurements on the hosting services deployed on AS 16509. The results reveal that even services hosted on the same infrastructure exhibit differences in attributes, demonstrating the potential influence of hosting providers on service quality. In the end, we conducted segmented measurements of the number of infrastructures used by hosting providers across different tiers. The results show that as the scale of providers decreases, the proportion of those using multiple infrastructures for services first increases and then decreases. The top three providers, which host 37.65% of all domains, rely exclusively on their own infrastructures to provide services. This approach may pose potential risks, and such risks are likely to intensify as centralization increases.

II. Background

A. DNS

The domain namespace (DNS) is a hierarchical structure composed of multiple levels of zones, including the root, top-level domains, i.e. TLD (e.g., `.com`), second-level domains, i.e. SLD (e.g., `example.com`), and so on [12], [13]. Each level's content is stored by the corresponding name servers, which also maintain the hosting records for their subdomains. For example, root servers maintain the hosting records of .com and the TLD nameservers of `.com` maintain the hosting records of `example.com`.

If a domain owner wants their domains to be resolved to an IP, they need to find a name server (NS), which stores the domain and IP address information for hosting. The domain of the NS server is called the NS domain, and the hosted domain will change its NS record to the corresponding NS domain.

There are various protocols and configurable options in DNS, some of which are set by the NS administrator. For example, the AXFR protocol is used for replicating entire DNS zone data between primary and secondary servers [14]. However, enabling AXFR may allow unauthorized attackers to access the entire zone's information, posing a risk of privacy leakage. EDNS is an extended standard that enhances DNS functionality [15], [16], supporting larger packet sizes and optional features, and is a prerequisite for enabling another security measure, DNSSEC [17]–[19]. DNS Cookie is a lightweight security mechanism that uses client-server encrypted tokens to authenticate requests [20].

B. DNS Hosting Services

If users want their domain to be resolvable by other users through the DNS, they need to obtain DNS hosting services and delegate their domain to an authoritative server. In the following text, we will refer to these domains owned by users as user domains. In this process, the user can either set up their authoritative server, known as self-hosting, or opt for a third-party DNS hosting service provider.

Once a user domain is hosted on a particular authoritative server, the NS record of the domain points to the domain of that authoritative server, indicating that the server is hosting the domain. To enhance robustness, a domain's NS records may point to multiple NS, allowing users to obtain the domain's IP address from alternative NSes if one fails. An NS does not necessarily point to a single authoritative server; it may point to multiple servers, employing effective load-balancing strategies to reduce the load on each server, thereby improving fault tolerance and resistance to DDoS attacks.

On the internet, users typically have two options for hosting their domains: the first is to choose a third-party hosting provider; the second is to set up their authoritative servers for hosting. In comparison, opting for third-party hosting is generally safer, offers lower latency, and requires less operational effort—this is because third-party hosting service providers usually have more specialized teams for maintenance and operations. However, some users may still choose to self-host their domains for various reasons. In general, self-hosted

services tend to be small in scale, hosting relatively fewer domains. The quality and security of such services depend entirely on the individual administrator, making them more likely to maintenance neglect, delayed upgrades, and failure to apply the latest security patches. Overall, users tend to prefer third-party hosting services.

C. Co-hosting relationship

Co-hosting occurs when a public DNS hosting provider assigns customer domains a group of NS domains that may belong to multiple providers sharing DNS infrastructure (e.g., servers) or having business partnerships. These NS domains often follow consistent naming patterns. For example, Bridge Pay's website `bridgepaynetwork.com` uses NS domains like `dns1.p03.nsone.net` and `ns01.bridgepayns.com`. We term such NS domains as co-hosting NS domains and providers with nameservers showing this behavior as co-hosting providers.

By analyzing authoritative NS records from TLD zone files, we observed that providers often assign customer domains to NS domains with specific naming patterns. Consequently, NS domains within the same provider typically manage similar groups of customer domains, exhibiting a high degree of co-hosting similarity. Continuing with the example of `nsone.net` and `bridgepayns.com`, we find that the domains hosted by NS domains belonging to NSone and those hosted by the series of NS domains `ns01.bridgepayns.com` to `ns04.bridgepayns.com` are highly consistent. Ultimately, through various methods, we confirm that this series of NS domains with `bridgepayns` as the base domain also belongs to NSone. Essentially, it represents a specialized domain hosting service purchased by Bridge Pay from NSone. Based on this observation, we developed a method to identify NS domains from the same provider by evaluating whether two NS domains or groups share a significant proportion of co-hosted customer domains.

III. Methodology

We designed a new method that groups NS domains belonging to the same hosting provider based on textual, infrastructure, and user domain co-hosting similarities. Each group is assigned a hosting provider label, indicating that all NS domains within the group belong to that provider. We first provide an overview of this method and then describe the process of each step in detail.

A. Framework Overview

Figure 1 illustrates the major processes of our framework. It takes input from a DNS zone file and first merges similar NS domains into NS groups. Then it utilizes the hosting information in the zone file to delineate co-hosting relationships among NS domains. Through iterative rounds of grouping and merging based on shared hosting, the method ultimately produces several NS groups, each labeled with a provider tag representing the common provider for all NSes in that group.

B. ***Step1***: Clustering NS based on Textual Pattern Similarity

We collect and analyze NS domains from various providers, using their textual pattern similarity to perform the first round of merging. The process is shown in Figure 2.

We first explain our understanding of textual pattern similarity for NS domains. Specifically, we use `tldextract` [21] to divide domains into three parts: subdomains (subs), base domains (base), and effective TLDs (eTLDs). We define NS1 and NS2 as similar if and only if:

1) They have the same number of subdomain levels, (hereafter denoted as SL);

2) They have the same core fields that are defined as:

$$core = \begin{cases} base & \text{level(sub)} \leq 1. \\ \text{last sub} + \text{base} & \text{level(sub)} \geq 2. \end{cases}$$

3) They share the same TLD, or their TLD don't appear in our 'low-confidence TLD'list.

For example, in Figure 2, only the first and the second NS domains will be treated as similar. The third NS has a different number of subdomains, and the part of the fourth NS domain that requires comparison differs from that of the first and second NS domains, while the fifth NS domain's TLD appear in our list of low-confidence TLDs.

We will explain in detail the rationale behind these criteria:

The rationale for the first criterion is that having the same level of subdomain is a basic guarantee of structural similarity in NS domains.

The rationale for the second criterion lies in the fact that the base domain is a key indicator of whether NS domains belong to the same hosting provider. When the subdomain contains multiple segments, we found that hosting providers often use the final segment to distinguish between specific product lines, such as in `*.ns.cloudflare.com` vs `*.secondary.cloudflare.com`. To allow for smaller NS groups in the initial step, enabling further rounds of precise merging based on co-hosting relationships, we added a restriction for multi-segment subdomains.

The third criterion essentially aims to exclude certain special cases. In our study, we identified some NS domains with textual formats highly similar to those of well-known hosting providers, but with different TLDs, such as `zeus.ns.cloudflare.com2` and `ns3.dreamhost.cm`. While we did not investigate the reasons behind this phenomenon, we manually maintain a list to exclude these low-confidence TLD domains, thus avoiding potential misclassification.

After this part, several preliminary NS groups are formed within the NS list. Let the i-th NS group be denoted as G_i.

C. ***Step2***: Clustering NS based on Co-hosting Similarity

We designed this process based on the similarity of domains hosted within the same provider. We abstract each NS group as a graph. Let the number of domains hosted by G_i be denoted as N_i, and the number of domains co-hosted by G_i and G_j be denoted as $Co_{i,j}$. The similarity between G_i and G_j is

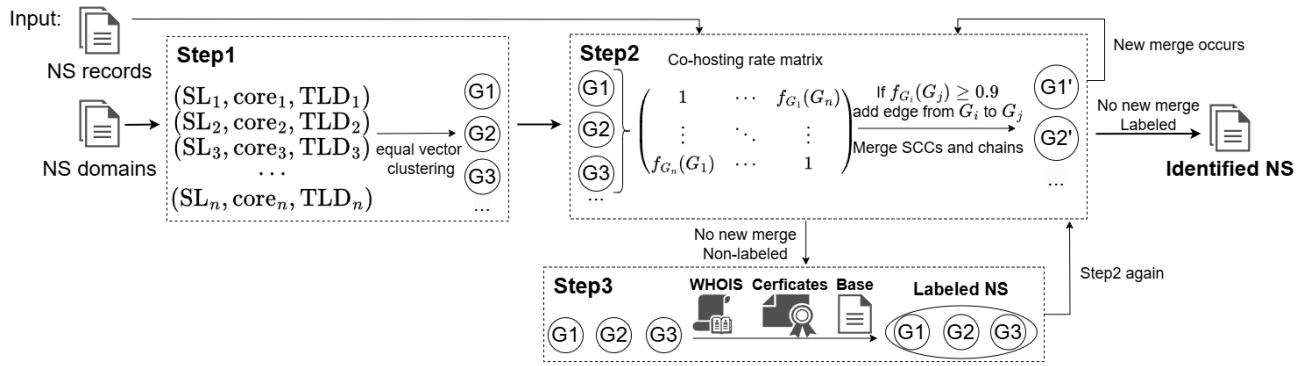

Fig. 1: The workflow of the method.

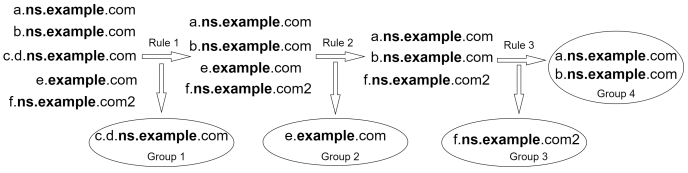

Fig. 2: Example process of step1

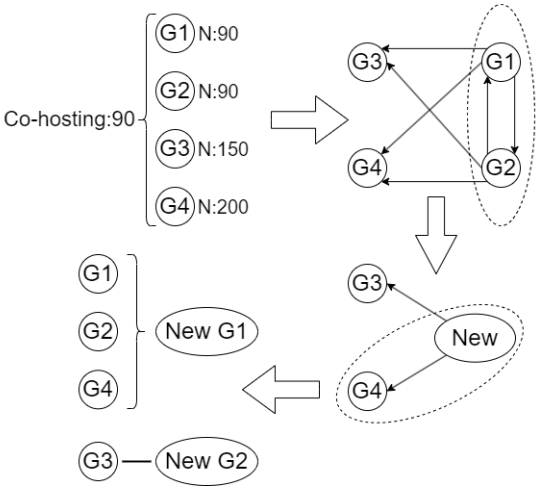

Fig. 3: Example process of step2

defined as $f_{G_i}(G_j) = \frac{Co_{i,j}}{N_i}$. If $f_{G_i}(G_j) \geq 0.9$ we add a directed edge from G_i to G_j.If there are strongly connected components (SCC) in the graph, we merge the nodes within those components into a new group. If there are no SCCs, we merge each node with the group which has the most N_i among the groups the node points to. This process is iterative and continues until there are no more points to merge in the graph. The process is shown in Figure 3.

The 90% threshold was chosen after extensive experimentation with the method's two clustering steps and proved to be optimal. When this threshold drops below 80%, more erroneous merges occur, primarily reflected in the merging of smaller providers' NS domains into those of larger providers. Raising this threshold reduces such errors, but when set above 95%, it prevents many providers' NS domains from clustering together (e.g., Cloudflare, Akamai). Thus, we selected 90% based on the method's overall performance.

D. *Step3: Group Labeling with Supplemental Information*

The main goal of this part is to assign a primary provider to each group. In this method, we use WHOIS data, certificate data and base domains of NS for labeling. Specifically, we assign preset score values to each of these three types of information. For each NS domain in the group, we extract its corresponding information in these three categories, adding the preset score values multiplied by the number of user domains hosted by the NS domain. Finally, we select the field with the highest score as the provider for this group. However, this module is open-ended, and we can integrate other methods (e.g., LLM analysis) as well. After assigning the corresponding provider label to each group, this module will also perform a final merge of all groups labeled with the same provider.

1) WHOIS: We obtained domain registration information through the WHOIS protocol. for all NS domains and their corresponding base domains. Although the WHOIS definition document [22] does not specify the components of domain registration information data through the WHOIS protocol, it generally includes information regarding a domain's Registrant Organization, Administrative Organization, and Technical Organization, providing clues about the domain's associated provider. We assigned corresponding scores to each field type allowing us to accumulate scores for each provider.

During the processing, we observe that for privacy protection reasons, WHOIS data providers may offer data with hidden information, such as Registrant Organization marked as `Registration Private`. Therefore, we implement additional filtering rules to exclude such WHOIS data.

2) Certificates: According to the digital certificate standard X.509v3 published in RFC 5280 [23], digital certificates contain an `OrganizationName` that reflects the managing organization of a domain. We use the zgrab2 tool [24] to establish TLS connections with NS domains and their corresponding base domains, obtain their certificates, and extract the organization names. Similarly, we assign a corresponding score value to this field type.

3) Base domain: Given that some NS domains may include provider information in their base domain, we incorporate this method to determine the provider of an NS domain. However, since providers often use multiple base domains (e.g., Alibaba uses both `alidns.com` and `hichina.com`, and Akamai, through acquisitions, uses `linode.com` in addition to `akam.net`), some base domains may not accurately reflect the true provider. Thus, we assign a lower confidence level to this information, using it only as a backup when WHOIS and certificate information do not indicate the provider.

4) Dealing with String: To address the potential inconsistencies in hosting provider name fields (such as `GoDaddy LLC, GoDaddy`), we applied a string processing method. The input to this method is a series of strings representing hosting providers. The essence of the method is to group similar strings together through a series of comparisons and treat them as representing the same hosting provider. We first maintained a list containing terms representing company names and types, such as 'ltd', 'Co', 'LLC', etc. When processing each string, we first match it with the list and remove the matched parts. Next, we attempt to remove other content(including space) in the string, retaining only the alphanumeric characters (with letters converted to lowercase). Next, we compute the edit distance between each pair of processed strings. If the edit distance is less than 15%, which is an empirical value that allows for more correct merges while minimizing incorrect ones, of the length of the longer string, the two strings are considered synonymous and are merged using a union-find data structure. Then, the corresponding groups are merged accordingly.

E. *Step2 again*

Although the previous string processing and merging step partially consolidated homogeneous groups (i.e., groups with different label representations that actually refer to the same provider), it did not achieve complete merging. Therefore, we perform another round of clustering based on co-hosting relationships, following the same procedure as in Step 2, including iterative processing until no further merges occur. The primary goal of this additional merging step is to further consolidate homogeneous groups, facilitating the analysis and interpretation of statistical results.

IV. Evaluation and comparison of the tools

In this section, we present a comparison between several methods and our newly proposed method. We selected the following three methods for comparison: WHOIS, ASN and WHOIS + certificates. These methods have been commonly used in previous research to determine the provider associated with an NS domain [8], [11].

We constructed a test set for comparison purposes. This test set consists of two parts: the first part includes 580 NS domains from several large providers, such as Cloudflare, Google Cloud, Amazon Cloud, Microsoft Cloud, Akamai, Tencent (i.e., DNSPod), and Alibaba. The second part comprises 193 NS domains, randomly selected from smaller public DNS hosting providers, including 4.cn, SAKURA, Bodis, and Gname. This dataset is randomly selected and manually labeled by us. To the best of our knowledge, no publicly available dataset currently exists for this problem.

We used the ASN database from MaxMind's free GeoLite2 database [19] (updated on 2024-09-23) to implement the ASN method for testing. WHOIS method, the WHOIS + certificates method and our method all use WHOIS and certificates data from the same source and version to ensure consistency.

We conducted comparative tests of the three methods on the dataset, with results further detailed for Parts 1 and 2. We performed data statistics at both the NS domain level and the provider level, as shown in Table II and Table III.

Overall, from the perspective of NS, our proposed method successfully identified 732 out of 773 NS domains, achieving an identification rate of 94.70%. The WHOIS + certificates method identified 680 NS domains, with an 87.97% success rate, while the WHOIS method correctly identified 658 NS domains with an 85.12% accuracy. The ASN method correctly identified 539 NS domains, yielding an accuracy of 69.73%. Although all three tools demonstrated relatively high identification rates overall, a detailed breakdown reveals notable differences. In Part 1, the identification rates were 97.41%, 91.72%, 88.63%, and 84.14% for our method, WHOIS + certificates, WHOIS, and ASN, respectively—showing a relatively small performance gap. However, in Part 2, despite some misclassifications, our tool achieved a 86.53% identification rate, whereas the WHOIS + certificates method dropped to 76.68%, the WHOIS method dropped to 74.61%, and the ASN method only reached 26.42%. As for the perspective of provider, the differences are even more pronounced.

We analyze the dataset in an attempt to identify the reasons for the poor performance of the other three methods.

1) ASN method: The ASN tool performed well only with providers that own their server networks and are willing to deploy hosting services on their own infrastructure, such as Cloudflare, Amazon, Microsoft, and Google. For smaller providers that lack the capability to operate such a network and must rely on other cloud service providers, or for providers (such as Baidu and GoDaddy) that, for various reasons, do not deploy hosting services on their own servers, the ASN tool tends to incorrectly identify these providers as the outsourced server providers. Moreover, we observe that similar smaller providers, like those in part 2, are the majority in the real world. This suggests that the ASN tool's correct identification rate will further decrease when faced with this data.

2) WHOIS & WHOIS+Certificate method: We observe inherent limitations, such as certain providers neglecting to properly fill out WHOIS data or deploy certificates for their NS domains and NS base domains. Furthermore, with the implementation of the *General Data Protection Regulation* (GDPR) [25] by the European Union, an increasing number of domain registrars are opting to conceal information about domain holders in WHOIS and certificates. This trend indicates that methods solely reliant on WHOIS and certificates will

TABLE II: Result of evaluation (NS)

		Our method	WHOIS + certificates	WHOIS	ASN
All (773)	Correct	731 (94.57%)	680 (87.97%)	658 (85.12%)	539(69.73%)
	Wrong	42 (5.43%)	2 (0.26%)	2 (0.26%)	234 (30.27%)
	Unidentified	0 (1.29%)	91 (11.77%)	113 (14.62%)	0(0.00%)
Part1 (580)	Correct	564 (97.24%)	532 (91.72%)	514 (88.63%)	488 (84.14%)
	Wrong	16 (2.76%)	2 (0.34%)	2 (0.34%)	92 (15.86%)
	Unidentified	0 (0.00%)	46 (7.94%)	64 (11.03%)	0 (0.00%)
Part2 (193)	Correct	167 (86.53%)	148 (76.68%)	144 (74.61%)	51 (26.42%)
	Wrong	26 (13.47%)	0 (0.00%)	0 (0.00%)	142 (73.58%)
	Unidentified	0 (0.00%)	45 (23.32%)	49 (25.39%)	0 (0.00%)

TABLE III: Result of evaluation (provider)

		Our method	WHOIS + certificates	WHOIS	ASN
All (57)	Fully correct	48 (84.21%)	39 (68.42%)	37 (64.91%)	17 (29.82%)
	Partially correct	3 (5.26%)	6 (10.53%)	6 (10.53%)	9 (15.79%)
	No correct	6 (10.53%)	12 (21.05%)	14 (24.56%)	31 (54.39%)
Part1 (16)	Fully correct	13 (81.25%)	12 (75.00%)	11 (68.75%)	10 (62.50%)
	Partially correct	3 (18.75%)	3 (18.75%)	4 (25.00%)	3 (18.75%)
	No correct	0 (0.00%)	1 (6.25%)	1 (6.25%)	3 (18.75%)
Part2 (41)	Fully correct	35 (85.37%)	27 (65.85%)	26 (63.41%)	7 (17.07%)
	Partially correct	0 (0.00%)	3 (7.32%)	2 (4.88%)	6 (14.63%)
	No correct	6 (14.63%)	11 (26.83%)	13 (31.71%)	28 (68.30%)

identify fewer domains over time. In fact, when examining the intermediate processes of this method, we discover that most of the WHOIS data it relies on are outdated—often several years old. As registrars increasingly obscure the latest WHOIS data due to GDPR, this information becomes unusable. Errors in identification by this method are largely due to this issue: after the last update, domain ownership may have changed due to acquisitions or sales, but with current data concealed, the method defaults to relying on outdated WHOIS records containing information that was once accurate. This reliance on obsolete data often leads to incorrect identification.

TABLE IV: Top 10 hosting provider

Provider	Hosting domains	Proportion of sum
GoDaddy	47,498,201	26.65%
Cloudflare	11,797,644	6.62%
Google	7,819,371	4.39%
Namecheap	5,706,569	3.20%
Wixpress	5,583,020	3.13%
NameBright	4,339,109	2.43%
IONOS	3,811,401	2.14%
Alibaba	3,636,313	2.04%
Newfold	3,244,439	1.82%
Share-DNS	3,169,556	1.78%

V. Centralization Measurement

To enhance the generalizability of our results, we select the dataset, which combined with the `.com` , `.net` and `.org` zone file from September 14, 2024, all of these zone files are sourced from CZDS [26]. Our dataset includes 178,252,051 domains and 2,155,818 NS domains.

We applied the method to measure the centralization information in the dataset and obtained data on the NS domains of the providers, and we identify 530,111 hosting providers. After sorting the providers by the number of domains they host, we plotted the proportion of hosting domains, the proportion of NS domains held, and the proportion of NS base domains held for the all providers, as shown in Figure 4. We also list the top 10 hosting providers in Table IV. Our measurement results show that the top 10 providers, hold 4.05% of the NS domains in the dataset, totaling 87,333 NS domains. More concerning is that they host 54.19% of the user domains in the dataset, amounting to 96,598,224 domains. As for the top 100 providers, they hold 10.01% NS domains while they host over 82.99% domains. The results shown in the figure indicate that the reason top providers host more domains is not due to their control of a significantly larger number of NS domains. In fact, the distribution of NS domains across providers is relatively uniform. This result indicates that DNS at the hosting provider level indeed exhibits a significant centralization issue.

We acknowledge that our identification results may include users who self-host their websites, as we currently lack the ability to distinguish between true self-hosting and third-party hosting providers. However, even under a lenient criterion for identifying self-hosted users—such as classifying providers hosting 10 or fewer domains as self-hosted—this category comprises 439,294 entities. Nevertheless, the websites they host account for only 0.48% of the total, an insignificant proportion that has minimal impact on our conclusions. Even if we further relax the threshold to include providers hosting 100 or fewer domains, this would involve 514,394 providers. However, the domains they host account for only 1.69% of the total, which still does not affect our analysis results.

Beyond the centralization observed at the provider level, we also measure the centralization of domains at the AS level. We resolve NS domains to IP addresses whenever possible and use the GeoLite2 [27] database to identify the ASes associated with IP addresses. This allows us to map the hosting relationships between domains and NSes into *'hosting'* relationships between domains and ASes. We extract 356,501 IP addresses that can be mapped to their respective autonomous

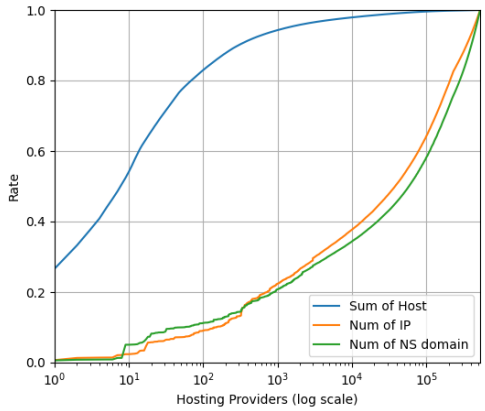

Fig. 4: The proportion of the sum of hosting domains, NS base domains and NS domains of all providers

TABLE V: Top 10 ASes

ASN	AS provider	Hosting domains	Proportion of sum
44273	Host Europe	49,471,645	28.11%
13335	Cloudflare	29,824,532	16.95%
15169	Google	13,414,589	7.62%
16509	Amazon	10,019,881	5.69%
14618	Amazon	7,375,020	4.19%
397220	Security Services	6,759,912	3.84%
19905	Security Services	6,036,960	3.43%
8560	IONOS	5,797,674	3.30%
62597	NSone	4,247,624	2.41%
37963	Alibaba	3,841,344	2.18%

systems, which correspond to 675,961 NS domains, hosting 175,980,269 domains. These IP addresses are categorized into 18,005 ASes. We present the cumulative distribution of ASes with respect to hosted domains, NS domains, and IP addresses, as shown in Figure 5. Additionally, we list the top 10 ASes hosting the most domains in Table V. It's evident that a significant level of centralization also exists at the AS level. However, this centralization is more pronounced compared to the centralization observed at the provider level. Additionally, the top 10 ASes and providers don't correspond to each other. A detailed analysis of the relationship between providers and ASes will be presented in SectionVII.

VI. Base domain ≠ provider

In our measurement results, we observe some intriguing findings. For example, domains like `squarespacedns.com` and `gannett-dns.com` are categorized under NSone. Upon further investigation of the measurement process, we determine that these results are more likely due to the complex relationships among providers—such as acquisitions and commercial collaborations—rather than misclassification. Specifically, in these two cases, we find that nearly all domains hosted by these NS domains are further delegated to nsone.net, suggesting that Squarespace and Gannett purchase hosting services from NSone.

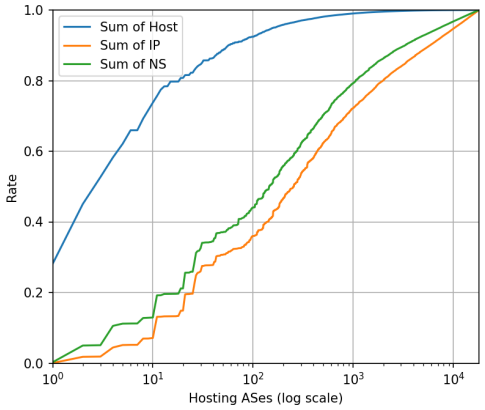

Fig. 5: The proportion of the sum of hosting domains, NS domains and IPs of all ASes

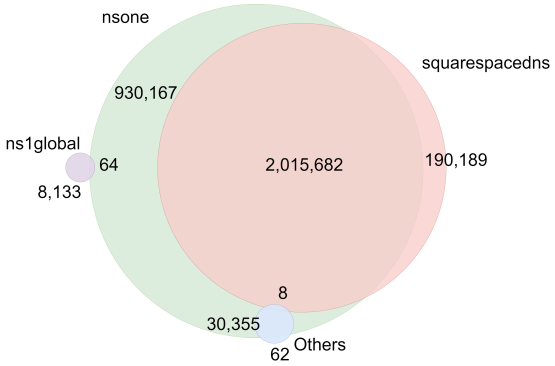

Fig. 6: Venn diagram of the hosting distribution of NSone's base domains

Using this observation as a starting point, we analyze the NS base domains within NSone, examining their co-hosting relationships and finding a high degree of consistency in the hosted content among NS domains with significantly different base domains. We further analyze the variety of base domains owned by top providers. Finally, we examine the domains in the dataset that are hosted by multiple NS base domains to determine the proportion of those that are genuinely hosted by two or more providers.

A. Base domains of NSone: the large scale of 'independent' providers within NSone

We analyze the distribution of the NS domain for the provider NSone. NSone hosts a total of 3,144,233 domains and encompasses 596 NS domains, which can be grouped into 152 unique NS base domains. The two most prominent NS base domains, `nsone` and `squarespacedns`, host 2,945,849 and 2,205,871 domains, respectively. The remaining NS domains collectively host 38,622 domains. The hosting distribution of these categories is illustrated in a Venn diagram, as shown in Figure 6.

It can be observed that the `nsone` and `squarespacedns` jointly host 2,015,682 domains, accounting for 68.42% of the domains hosted by `nsone` and 91.38% of those hosted

by squarespacedns. This strongly indicates an extremely close commercial collaboration between NSone and Squarespace in the domain hosting business, to the extent that it can be assumed Squarespace has entirely outsourced its domain hosting requirements to NSone. ns1global is a base domain owned by NSone, which can be verified through domain registration information provided by WHOIS. The various NS domains under ns1global may host different services, resulting in a lower number of co-hosted domains with NSone.

Among the domains hosted by other base domains, 99.80% are also co-hosted by NSone. We analyze specific examples among them, such as kenvuedns, jnjdns, and gannett-dns, which might otherwise be considered independent providers or self-hosted servers. The domains which are hosted by them are all co-hosted by NSone.

This specific case illustrates the complexity at the provider level. NS domains previously considered to belong to independent providers or self-hosted servers are entirely co-hosted by a large provider, revealing that they inherently belong to the same provider. The wide variety of base domains hosted by large providers further suggests that this phenomenon is not an isolated case. Moreover, the analysis of NSone exemplifies the accuracy of our provider identification method.

B. Base domains owned by top providers

In addition to analyzing NSone, we conduct a statistical analysis of the number of base domain types for the top 10 providers with the largest number of hosted domains in our identification results, where base domains with the same content, excluding symbols and numbers, are considered the same. The results are shown in Figure VI. We find that an individual provider may manage a large number of NS domains across various base domains. In the past, the distinction between different base domains was one of the most important criteria for identifying different providers. However, our results challenge this principle. Some of these base domain patterns were previously considered to belong to independent providers, while others were believed to be self-hosted by users to manage their own domains. Our findings suggest that these base domain patterns are highly correlated with the domains hosted by top providers, and thus should be considered part of these providers' portfolios.

On the other hand, we can observe that there is still a significant disparity in the number of base domains even among the top providers, which may indirectly reflect differences in the providers' cooperation and service strategies. We find that some providers offer special domain hosting services that allow VIP clients to customize the base domain field of the NS domains they use. Although this creates surface-level differences in NS domains, the underlying hosting services are still provided by the provider. Moreover, because these NS domains are often hosted alongside other NS domains from the same provider, they can be merged into a single provider through shared hosting relationships. This unique service offering may also be one of the reasons for the wide variety of NS domains owned by providers.

C. Hosting status of domains: who is actually hosted by different providers?

Building on the previous analysis, we track the hosting status of all domains in three regions. First, we count the domains that appear to use two or more providers, which we refer to as positives. Since users may lack detailed prior knowledge, they might simply assume different base domains correspond to different hosting providers. Therefore, we categorize these domains based on NS records involving multiple base domains. Furthermore, we track the domains that genuinely use multiple providers for hosting services, which we refer to as true positives. Only these domains can effectively avoid having their DNS resolution affected when one provider experiences a failure.

We analyze the number of domains in our dataset under these two scenarios and found that only 5,755,687 domains are hosted by NS domains with multiple types of base domains. Among these, only 1,027,290 NS domains are genuinely co-hosted by two or more providers, accounting for just 0.58% of all NS domains in the dataset. This means that only 0.58% of domains in our dataset can still be resolved normally if one provider fails (assuming the issue lies solely with the provider's service and not with the infrastructure). For the 5,755,687 domains identified as positives, their users might assume that hosting by multiple providers enhances the reliability of their domains. However, in reality, less than 20% of these domains are truly co-hosted by multiple providers. For the remaining 80% or more of domains, the redundant deployment that fails to achieve the intended effect may result in unnecessary resource waste.

VII. The complex relationship between infrastructure and providers.

In our study, infrastructure refers to AS. An AS is a network or group of networks under a single administrative entity. An ASN is a unique identifier assigned to an AS for BGP (Border Gateway Protocol) routing [28], [29].

A. Providers and ASes: many-to-many

We analyzed the deployment patterns between providers and ASes, revealing the complex many-to-many relationship between infrastructure and providers. We will now discuss this relationship from both the provider perspective and the AS' perspective.

From the perspective of providers, we categorize the deployment patterns of providers across ASes into three types:

1) Own: The provider has its own server network and deploys nearly all of its hosting services on this network. For example, Cloudflare's domain hosting services are almost entirely deployed within its own AS(13335), while Google's domain hosting services are almost completely deployed within its AS(15169).

2) Single external: The provider selects one or more external ASes from a fixed service provider to deploy managed services. For example, Wixpress mostly deploys its services within the Google network, AS 15169, while Hostdirekt and

TABLE VI: The kinds and examples of main NS base domains owned by top 10 providers

Provider	Kinds of base	Example	
		Base domain	NS
Godaddy	4,007	domaincontrol	ns70.domaincontrol.com
		afternic	ns4.afternic.com
		namefind	ns3.namefind.com
Cloudflare	641	cloudflare	jobs.ns.cloudflare.com
Google	332	googledomains	ns-cloud-e3.googledomains.com
Namecheap	34	registrar-servers	rs45.registrar-servers.com
		namecheaphosting	ns108.namecheaphosting.com
Wixpress	166	wixdns	ns1.wixdns.set
NameBright	24	namebrightdns	ns89.namebrightdns.com
IONOS	4,063	ui-dns	ns1077.ui-dns.com
		1and1	ns28.1and1.com
Alibaba	87	hichina	dns3.hichina.com
		alidns	ns5.alidns.com
Newfold	545	worldnic	ns92.worldnic.com
Share-DNS	1	share-dns	a.share-dns.com

TABLE VII: Examples of complex in the top 50 providers

Provider	AS number	Proportion
Hostgator	(AS 19871, Network Solutions)	48.78%
	(AS 46606, Unified Layer)	48.78%
Dnspod	(AS 45090, Tencent)	99.82%
	(AS 134756, China Net)	99.46%
	(AS 56046, China Mobile)	99.21%
	(AS 4837, China Unicom)	99.16%
Rzone	(AS 15418, Fasthosts)	99.97%
	(AS 8560, IONOS)	99.97%
Xserver	(AS 131965, Xserver)	97.87%
	(AS 16509, Amazon)	97.77%
Sedo	(AS 47846, SEDO GmbH)	99.98%
	(AS 16509, AMAZON-02)	96.38%

DreamHost deploy almost all of their services within the Cloudflare network.

3) Complex: A provider may choose multiple ASes from different service providers (including their own network) to deliver services. The quantitative criterion for this model is the provider deploys hosting services in at least two ASes, hosting over 10% of user domains. Table VII below shows examples of providers using the complex model for deployment in the top 40 providers. The percentages in the table represent the proportion of user domains hosted by the provider on each AS. Since a name server may resolve to multiple IPs across different ASes, the total sum may exceed 100%. For clarity, only the most significant ASes are displayed.

We segment the statistics for the number of top providers that host over 100,000 domains using three deployment types in the dataset, as shown in Table VIII. The results indicate that as the provider's scale decreases, they are less likely to use their own server networks, or they may not have their own networks at all, and instead turn to other service providers' networks to deliver services. Furthermore, we can observe that a considerable number of providers opt for the third deployment model, and the frequent occurrence of this hybrid deployment mode adds complexity to the relationship between providers and infrastructure.

From the perspective of infrastructure, a similar situation arises where a single infrastructure supports services from

TABLE VIII: Statistics on the deployment models of top providers in the dataset.

Hosting	Num	Own	Single external	Complex
Over 1,000,000	23	34.783%	43.478%	21.739%
500,000-1,000,000	20	25.000%	45.000%	30.000%
100,000-500,000	74	18.919%	37.838%	43.243%

multiple providers, resulting in a many-to-many relationship between infrastructure and providers. As indicated by previous results, smaller providers, likely due to the difficulty of operating their own infrastructure, increasingly tend to rely on external infrastructure to deploy their services. The high service quality and stability offered by well-known large infrastructure providers make them the preferred choice. This results in networks operated by major service providers hosting services from a range of different hosting providers. We analyze the deployment of hosting services on the top five ASes in our dataset, as shown in Table IX. These ASes host a significant portion of the total hosting services. In our dataset, we identify 18,005 ASes, which together host services from 122,236 hosting providers, covering 176,375,125 domains. On average, each AS supports hosting services from 6.789 providers, providing hosting for 9,795,897 domains. Notably, the top 10 ASes alone host services from 19,547 hosting providers, covering 136,789,181 domains. These figures clearly demonstrate the complexity of the relationship between infrastructure and providers from an infrastructure perspective.

Both perspectives reveal the complex relationship between hosting providers and infrastructure. Under such complexity, it is challenging for researchers to infer information from one perspective by using data from the other, so the two perspectives should be treated independently. Finally, we present the deployment of top providers in our dataset using a Sankey diagram to illustrate the intricate relationships between hosting providers and infrastructure visually. We select the top providers which host over 1,000,000 domains and investigate their infrastructure deployment. We focus on those infrastructures where more than 10% of their hosting services were deployed, meaning that the NS domains of a provider,

TABLE IX: The deployment of providers on the top 5 ASes

AS number	Hosting provider	Hosting	Proportion
AS 44273 Host Europe	GoDaddy	49,354,062	99.762%
	Others(35 providers)	117,583	0.238%
AS 13335 Cloudflare	cloudflare	11,733,212	39.341%
	Share-DNS	3,169,553	10.627%
	Hostdirekt	3,092,850	10.370%
	Newfold	2,852,879	9.566%
	Bluehost	1,930,272	6.472%
	Others(4251 providers)	7,045,766	23.624%
AS 15169 Google	Google	7,815,554	58.262%
	Wixpress	5,582,965	41.618%
	Others(224 providers)	16,070	0.120%
AS 16509 Amazon	Amazon	2,478,349	24.734%
	Siteground	1,190,371	11.880%
	xserver	701,247	6.999%
	Sedo	637,789	6.365%
	Bodis	589,999	5.888%
	Others(9087 providers)	4,422,126	44.134%
AS 14618 Amazon	NameBright	4,339,026	58.834%
	Buy Domains	691,071	9.370%
	HugeDomains	563,604	7.642%
	Others(2072 providers)	1,653,852	22.424%

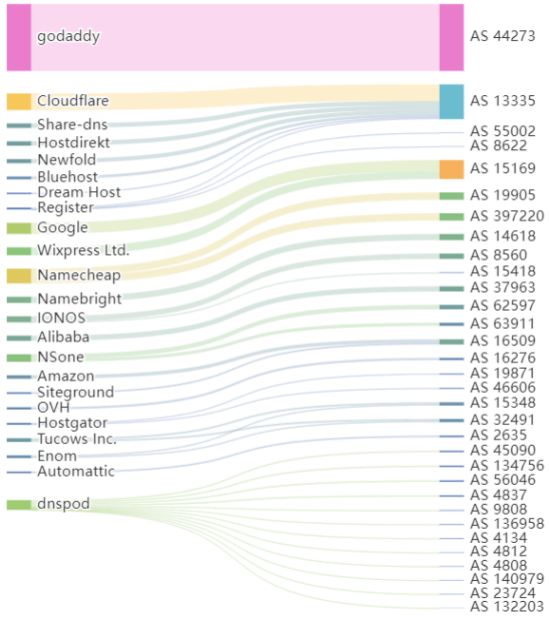

Fig. 7: The figure of top providers' usage of ASes

corresponding to a certain AS's IP, host more than 10% of the provider's total hosting domains. The correspondence between providers and infrastructure is shown in Figure 7.

B. Providers' influence on services

In addition to the complex relationship between hosting providers and infrastructure in service deployment, we also find evidence that hosting providers can independently influence the basic attributes of services, as demonstrated by measurements of basic security attributes conducted on the same infrastructure.

We conducted random sampling measurements on Amazon's AS numbered 16509, which hosts the largest variety of hosting providers in our dataset. We identified some examples

TABLE X: Attributes among providers on AS 16509

Provider	AXFR:OFF	EDNS:ON	Cookie:ON
Amazon	100.000%	100.000%	0.000%
EZOIC	100.000%	90.937%	0.000%
ServiceMaster	100.000%	100.000%	100.000%
NKsupport	0.000%	80.000%	80.000%
Megazone	100.000%	17.293%	16.541%
NFB	100.000%	100.000%	100.000%

where the security attributes of services varied depending on the provider, as shown in Table X. This provides preliminary evidence for our main argument, namely that hosting providers influence the quality and security attributes of services. Combined with the previous measurements of centralization, this indirectly highlights that the failure of security strategies by major hosting providers can have a significant negative impact on a large number of domains across the network.

C. Domains hosted by multiple infrastructures

In the dataset, we measure the infrastructure behind the NS records of all 178,252,051 domains. The IP addresses of the extracted NS domains host a total of 177,584,957 domains. Among these, 36,301,251 domains use multiple infrastructures for hosting services, meaning that about 20.442% of domains rely on two or more infrastructures for their hosting services.

We performed more fine-grained segmented statistics on the results, are shown in Table XI. The proportion of domains hosted by multiple infrastructures varies, starting low, increasing in the middle segments, and then decreasing again. We believe this phenomenon may be attributed to the following factors: top providers have sufficient trust in their own infrastructure and rely on it exclusively; mid-level providers lack their own infrastructure or do not trust it but have the capability to utilize multiple external infrastructures to deliver services; and smaller providers, due to financial constraints or other limitations, can only afford to deploy services on a single external infrastructure.

Although top providers' infrastructure is generally trustworthy, offering high security and low failure rates, relying exclusively on a single infrastructure without seeking additional backups could introduce potential single points of failure. This is particularly concerning for the top three providers—GoDaddy, Cloudflare, and Google—which together host 67,111,514 domains, accounting for 37.650% of the dataset. However, only 85,986 of these domains are hosted using multiple infrastructures. If present infrastructure of any of these three providers fails, because they lack proper backups on other infrastructures, it would inevitably have a significant negative impact on Internet availability. Furthermore, if centralization continues to intensify, the potential security risks will increase as the number of domains hosted by these three providers grows. In the event of infrastructure failures affecting these providers, the impact on Internet availability will become even more pronounced.

TABLE XI: Segmented results of multi-infrastructures hosting

Host num	Provider	Multi-ASes	Sum	Rate
Over 1,000,000	23	14,414,059	122,484,397	11.768%
500,000-1,000,000	20	6,073,781	12,941,577	46.932%
100,000-500,000	74	6,727,827	14,820,056	45.397%
10,000-100,000	454	5,634,134	14,329,847	39.317%
1,000-10,000	2124	2,278,958	6,294,772	36.204%
100-1,000	12,291	939,155	3,319,617	28.291%
Less than 100	527,416	470,329	2,615,691	17.981%

VIII. Discussion

A. Limitations

Our study has certain limitations. First, we only measure centralization within the `.com`, `.net`, and `.org` zones. While these are among the largest TLDs, we acknowledge that they may not fully reflect the overall centralization of the DNS. Second, our measurement is based on a single snapshot of these three TLDs on a specific day, without long-term observations. As a result, we do not capture trends in centralization over time.

B. Ethics considerations

We don't collect any private user information. As for the network connection, we use zgrab2 to establish TLS connections and collect certificates, applying rate limits to zgrab2 in the process to avoid impacting network availability. As for three TLD zone files we used, we got them on ICANN's Centralized Zone Data Service (CZDS) [26]. For the security measurements in Section 6, we similarly limited the packet transmission frequency to ensure network availability remained unaffected. Additionally, we did not employ potentially harmful or intrusive measurement methods, nor did we attempt to access extra private information (e.g., obtaining a DNS zone file via an open AXFR).

C. Reflections on Centralization

1) Benefits: Centralization is not entirely harmful. A centralized system can facilitate the adoption of new standards and security measures. For example, if the top 10 providers implement a new security measure, 58% of domains across the Internet would be protected, significantly lowering the deployment barrier.

2) Disadvantages and harm: **Single point of failure.** If domain hosting becomes overly dependent on a few large providers, failures in these providers could have a substantial negative impact on the overall availability of the network. For instance, the Akamai outage in 2021 and the Azure outage in 2022 both led to widespread service and website unavailability [30], [31].

Greater impact of attack. Centralization amplifies the impact of security risks associated with major providers, increasing the potential scope of their consequences. Reported attacks, particularly those related to domain hijacking and takeover [8]–[10], [32]–[43], primarily exploit mismanagement or weak security policies of certain providers.

Internet censorship. When a large number of domains are hosted by a single provider, this effectively grants the provider greater control over content censorship. The provider may reduce service quality or even halt resolution services to target content they disapprove of, which could pose further risks of influencing public opinion [44].

3) Recommendations: We urge the community and providers to recognize the challenges of centralization. Providers should ensure proper management and isolation of their hosting services, with particular attention to the security of their infrastructure. Any newly identified security risks should be promptly addressed and mitigated.

IX. Related Work

Relevant prior work includes studies of DNS centralization and Entity identification.

DNS centralization. Several researchers have made efforts on this topic, measuring DNS centralization from various perspectives. Their work has contributed to the motivation for this study. Xu et al. [11], Huston et al. [7], and Zembruzki et al. [2] used the autonomous systems corresponding to the IP addresses of NS domains as proxies for hosting service providers to measure the degree of centralization in the DNS system. These studies effectively measure the centralization of DNS hosting services at the infrastructure level, which differs from the centralization observed at the provider level. Zembruzki et al. [45] analyzed the degree of centralization of hosting services across multiple TLD zone files, but their analysis was still based on Autonomous Systems rather than hosting service providers. Moura et al. [1] analyzed DNS hosting services using traffic from B-root and two ccTLDs, examined the traffic share of five major providers, revealing that DNS traffic is also concentrated among major providers. This study does measure the market share of five major providers in the hosting service market. However, it is unlikely to become a widely applicable method due to the difficulty of obtaining traffic data. Doan et al. [46] investigated the popularity and performance of public DNS resolution services in the context of DNS centralization. Wang et al. [44] analyzed the DNS and web hosting dependencies of the top 10,000 popular websites. They discussed how a small number of large organizations control internet infrastructure and the potential risks this poses to network resilience, security, and content control. Kashaf et al. [47] investigated the dependence of modern web services on third-party infrastructure, such as DNS, CDN, and CA services. By comparing data from 2016 and 2020, their analysis revealed an increasing reliance on major DNS service providers, which contributed to greater centralization and exacerbated the potential risks of service disruptions and attacks. Radu et al. [4] analyzed changes in the recursive DNS service market from 2016 to 2019 to study trends in market share centralization, revealing how large internet companies leverage privacy and security-enhancing technologies to consolidate market dominance. They also explored the potential risks this centralization poses to the internet ecosystem. Hao et al. [48] analyze the deployment

patterns of authoritative servers of websites in Alexa's list of the top 1 million websites in June 2014. They find that most websites rely on third-party hosting providers and identify the top 10 most popular DNS hosting domains.

Entity identification. Several researchers have proposed various methods for identifying the controlling or owning entities of different objects, which have inspired the methodology of this study. Sebastián et al. [49] presented a comprehensive method including WHOIS, passive DNS, certificates, NER, etc., to identify the organization associated with a website. Liu et al. [50] proposed a method to identify email service providers behind MX domains, their analysis also reveals that email services are concentrated among a small number of major providers. Ma et al. [51] identify the controlling entities of CA certificates by analyzing certificate fingerprints, network infrastructure, and audit records. Although these studies focus on different subjects, their methods are highly transferable and provide valuable insights for this study. Some studies on domain takeover and domain hijacking also address methods for identifying the providers of NS domains. Alowaisheq et al. [8] combine the content of NS base domains with domain registration data obtained via the WHOIS protocol to determine the provider associated with NS domains. Squarcina et al. [43] relies on manual identification of third-party providers and doesn't implement an automated method. Zhang et al. [9] compile a comprehensive list of providers and the patterns of NS domains they use. These precise NS domain patterns serve as a crucial starting point for this study's in-depth examination of relationships among NS domains belonging to the same provider.

X. Conclusion

We propose a method for identifying the hosting provider of NS domains based on textual similarity and co-hosting relationships. Compared to traditional identification methods, such as ASN, WHOIS, and WHOIS+certificate approaches, our method offers higher accuracy. Using this method, we measured a dataset composed of zone files from .com, .net, and .org, quantifying the degree of centralization. Our results reveal a high degree of centralization, with the top 10 hosting providers hosting over 54% of the domains while the top 100 providers host over 82.99% domains. Based on our identification results, we also observed complex relationships at the provider level, where many entities previously regarded as independent providers or self-hosted services exhibit a high degree of co-hosting with top hosting providers. We further analyzed the relationship between infrastructure and hosting providers, revealing a complex many-to-many relationship. We identified cases where services provided by different hosting providers on the same infrastructure exhibited varying quality, indicating that these two levels cannot be conflated. Additionally, our analysis of the infrastructure behind domain hosting services revealed that only about 20% of domains use hosting services deployed across multiple infrastructures, with this percentage being even lower for top providers.

Acknowledgment

This work is in part supported by the National Key Research and Development Program of China (No. 2023YFB3105600), and Deng Feng Fund of Beijing National Research Center for Information Science and Technology.

Appendix A

TABLE XII: The accuracy and the number of iterations of methods under different thresholds.

	0.6	0.7	0.8	0.9	0.99
Correct	693	698	722	731	710
Accuracy	89.65%	90.30%	93.40%	94.57%	91.85%
Iterations	6	6	6	5	5

We record the number of iterations of our method on the dataset under different thresholds and their results on the dataset in table XII. It can be observed that a threshold of 0.9 provides a good balance between accuracy and efficiency.

References

[1] G. C. M. Moura, S. Castro, W. Hardaker, M. Wullink, and C. Hesselman, "Clouding up the internet: how centralized is DNS traffic becoming?," in *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*, pp. 42–49, ACM, 2020.

[2] L. Zembruzki, A. S. Jacobs, G. S. Landtreter, L. Z. Granville, and G. C. M. Moura, "dnstracker: Measuring centralization of DNS infrastructure in the wild," in *Advanced Information Networking and Applications - Proceedings of the 34th International Conference on Advanced Information Networking and Applications, AINA-2020, Caserta, Italy, 15-17 April* (L. Barolli, F. Amato, F. Moscato, T. Enokido, and M. Takizawa, eds.), vol. 1151 of *Advances in Intelligent Systems and Computing*, pp. 871–882, Springer, 2020.

[3] M. Allman, "Comments on DNS robustness," in *Proceedings of the Internet Measurement Conference 2018, IMC 2018, Boston, MA, USA, October 31 - November 02, 2018*, pp. 84–90, ACM, 2018.

[4] R. Radu and M. Hausding, "Consolidation in the dns resolver market – how much, how fast, how dangerous?," *Journal of Cyber Policy*, vol. 5, no. 1, pp. 46–64, 2020.

[5] G. Huston, "Dns resolver centrality." https://blog.apnic.net/2019/09/23/dns-resolver-centrality/, 2019.

[6] G. Huston, "Dns resolver centrality." https://labs.apnic.net/presentations/store/2021-09-10-hknog-resolver-centrality.pdf, 2021.

[7] G. Huston and J. Damas, "Measuring the centralization of dns resolution." https://labs.apnic.net/presentations/store/2023-02-16-OARC-resolver-concentration.pdf , 2023.

[8] E. Alowaisheq, S. Tang, Z. Wang, F. Alharbi, X. Liao, and X. Wang, "Zombie awakening: Stealthy hijacking of active domains through DNS hosting referral," in *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020* (J. Ligatti, X. Ou, J. Katz, and G. Vigna, eds.), pp. 1307–1322, ACM, 2020.

[9] Y. Zhang, M. Zhang, B. Liu, Z. Liu, J. Zhang, H. Duan, M. Zhang, F. Shi, and C. Xu, "Cross the zone: Toward a covert domain hijacking via shared DNS infrastructure," in *33rd USENIX Security Symposium (USENIX Security 24)*, (Philadelphia, PA), pp. 5751–5768, USENIX Association, aug 2024.

[10] F. Zhang, B. Liu, E. Alowaisheq, J. Chen, C. Lu, L. Song, Y. Ma, Y. Liu, H. Duan, and M. Yang, "Silence is not golden: Disrupting the load balancing of authoritative dns servers," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, CCS '23, (New York, NY, USA), p. 296–310, Association for Computing Machinery, 2023.

[11] C. Xu, Y. Zhang, F. Shi, H. Shan, B. Guo, Y. Li, and P. Xue, "Measuring the centrality of dns infrastructure in the wild," *Applied Sciences*, vol. 13, no. 9, 2023.

[12] P. V. Mockapetris, "Domain names - concepts and facilities," *RFC*, vol. 1034, pp. 1–55, 1987.

[13] P. V. Mockapetris, "Domain names - implementation and specification," *RFC*, vol. 1035, pp. 1–55, 1987.

[14] E. P. Lewis and A. Hoenes, "DNS zone transfer protocol (AXFR)," *RFC*, vol. 5936, pp. 1–29, 2010.

[15] P. Vixie, "Extension mechanisms for DNS (EDNS0)," *RFC*, vol. 2671, pp. 1–7, 1999.

[16] J. Damas, M. Graff, and P. Vixie, "Extension mechanisms for DNS (EDNS(0))," *RFC*, vol. 6891, pp. 1–16, 2013.

[17] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "DNS security introduction and requirements," *RFC*, vol. 4033, pp. 1–21, 2005.

[18] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "Resource records for the DNS security extensions," *RFC*, vol. 4034, pp. 1–29, 2005.

[19] R. Arends, R. Austein, M. Larson, D. Massey, and S. Rose, "Protocol modifications for the DNS security extensions," *RFC*, vol. 4035, pp. 1–53, 2005.

[20] D. E. E. III and M. P. Andrews, "Domain name system (DNS) cookies," *RFC*, vol. 7873, pp. 1–25, 2016.

[21] John.Kurkowski, "tldextract-5.1.2." https://pypi.org/project/tldextract/description, 2024.

[22] L. Daigle, "WHOIS protocol specification," *RFC*, vol. 3912, pp. 1–4, 2004.

[23] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. T. Polk, "Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile," *RFC*, vol. 5280, pp. 1–151, 2008.

[24] "ZGrab2." https://github.com/zmap/zgrab2, 2017.

[25] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)."

[26] ICANN, "Centralized zone data service." https://czds.icann.org/home, 2024.

[27] "GeoLite2." https://dev.maxmind.com/geoip/geolite2-free-geolocation-data, 2024.

[28] J. Hawkinson and T. Bates, "Guidelines for creation, selection, and registration of an autonomous system (AS)," *RFC*, vol. 1930, pp. 1–10, 1996.

[29] Y. Rekhter, T. Li, and S. Hares, "A border gateway protocol 4 (BGP-4)," *RFC*, vol. 4271, pp. 1–104, 2006.

[30] M. Sundaram, "Akamai summarizes service disruption (resolved)." https://www.akamai.com/blog/news/akamai-summarizes-service-disruption-resolved, 2021.

[31] M. Ruffell, "Reflecting on the Azure DNS outage - a post incident analysis." https://ruffell.nz/programming/writeups/2022/12/09/reflecting-on-the-azure-dns-outage-a-post-incident-analysis.html, 2022.

[32] X. Li, W. Xu, B. Liu, M. Zhang, Z. Li, J. Zhang, D. Chang, X. Zheng, C. Wang, J. Chen, H. Duan, and Q. Li, "Tudoor attack: Systematically exploring and exploiting logic vulnerabilities in dns response pre-processing with malformed packets," in *2024 IEEE Symposium on Security and Privacy (SP)*, pp. 4459–4477, 2024.

[33] X. Li, C. Lu, B. Liu, Q. Zhang, Z. Li, H. Duan, and Q. Li, "The maginot line: Attacking the boundary of DNS caching protection," in *32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9-11, 2023* (J. A. Calandrino and C. Troncoso, eds.), pp. 3153–3170, USENIX Association, 2023.

[34] W. Xu, X. Li, C. Lu, B. Liu, H. Duan, J. Zhang, J. Chen, and T. Wan, "Tsuking: Coordinating DNS resolvers and queries into potent dos amplifiers," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, CCS 2023, Copenhagen, Denmark, November 26-30, 2023* (W. Meng, C. D. Jensen, C. Cremers, and E. Kirda, eds.), pp. 311–325, ACM, 2023.

[35] D. Liu, S. Hao, and H. Wang, "All your DNS records point to us: Understanding the security threats of dangling DNS records," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016* (E. R. Weippl, S. Katzenbeisser, C. Kruegel, A. C. Myers, and S. Halevi, eds.), pp. 1414–1425, ACM, 2016.

[36] G. Akiwate, M. Jonker, R. Sommese, I. D. Foster, G. M. Voelker, S. Savage, and kc claffy, "Unresolved issues: Prevalence, persistence, and perils of lame delegations," in *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*, pp. 281–294, ACM, 2020.

[37] K. Borgolte, T. Fiebig, S. Hao, C. Kruegel, and G. Vigna, "Cloud strife: Mitigating the security risks of domain-validated certificates," in *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*, The Internet Society, 2018.

[38] D. Liu, Z. Li, K. Du, H. Wang, B. Liu, and H. Duan, "Don't let one rotten apple spoil the whole barrel: Towards automated detection of shadowed domains," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017* (B. Thuraisingham, D. Evans, T. Malkin, and D. Xu, eds.), pp. 537–552, ACM, 2017.

[39] M. Zhang, X. Li, B. Liu, J. Lu, Y. Zhang, J. Chen, H. Duan, S. Hao, and X. Zheng, "Detecting and measuring security risks of hosting-based dangling domains," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 7, no. 1, pp. 9:1–9:28, 2023.

[40] G. Akiwate, R. Sommese, M. Jonker, Z. Durumeric, kc claffy, G. M. Voelker, and S. Savage, "Retroactive identification of targeted DNS infrastructure hijacking," 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.), pp. 14–32, ACM, 2022.

[41] Y. Zhang, B. Liu, H. Duan, M. Zhang, X. Li, F. Shi, C. Xu, and E. Alowaisheq, "Rethinking the security threats of stale DNS glue records," in *33rd USENIX Security Symposium, USENIX Security 2024, Philadelphia, PA, USA, August 14-16, 2024* (D. Balzarotti and W. Xu, eds.), USENIX Association, 2024.

[42] T. Dai, P. Jeitner, H. Schulmann, and M. Waidner, "The hijackers guide to the galaxy: Off-path taking over internet resources," in *30th USENIX Security Symposium, USENIX Security 2021, August 11-13, 2021* (M. D. Bailey and R. Greenstadt, eds.), pp. 3147–3164, USENIX Association, 2021.

[43] M. Squarcina, M. Tempesta, L. Veronese, S. Calzavara, and M. Maffei, "Can I take your subdomain? exploring same-site attacks in the modern web," in *30th USENIX Security Symposium, USENIX Security 2021, August 11-13, 2021* (M. Bailey and R. Greenstadt, eds.), pp. 2917–2934, USENIX Association, 2021.

[44] S. Wang, K. MacMillan, B. Schaffner, N. Feamster, and M. Chetty, "A first look at the consolidation of DNS and web hosting providers," *CoRR*, vol. abs/2110.15345, 2021.

[45] L. Zembruzki, R. Sommese, L. Z. Granville, A. S. Jacobs, M. Jonker, and G. C. M. Moura, "Hosting industry centralization and consolidation," in *2022 IEEE/IFIP Network Operations and Management Symposium, NOMS 2022, Budapest, Hungary, April 25-29, 2022*, pp. 1–9, IEEE, 2022.

[46] T. V. Doan, J. Fries, and V. Bajpai, "Evaluating public DNS services in the wake of increasing centralization of DNS," in *IFIP Networking Conference, IFIP Networking 2021, Espoo and Helsinki, Finland, June 21-24, 2021* (Z. Yan, G. Tyson, and D. Koutsonikolas, eds.), pp. 1–9, IEEE, 2021.

[47] A. Kashaf, V. Sekar, and Y. Agarwal, "Analyzing third party service dependencies in modern web services: Have we learned from the mirai-dyn incident?," in *IMC '20: ACM Internet Measurement Conference, Virtual Event, USA, October 27-29, 2020*, pp. 634–647, ACM, 2020.

[48] S. Hao, H. Wang, A. Stavrou, and E. Smirni, "On the DNS deployment of modern web services," in *23rd IEEE International Conference on Network Protocols, ICNP 2015, San Francisco, CA, USA, November 10-13, 2015*, pp. 100–110, IEEE Computer Society, 2015.

[49] S. Sebastián, R. Diugan, J. Caballero, I. Sánchez-Rola, and L. Bilge, "Domain and website attribution beyond WHOIS," in *Annual Computer Security Applications Conference, ACSAC 2023, Austin, TX, USA, December 4-8, 2023*, pp. 124–137, ACM, 2023.

[50] 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.), pp. 122–136, ACM, 2021.

[51] Z. Ma, J. Mason, M. Antonakakis, Z. Durumeric, and M. D. Bailey, "What's in a name? exploring CA certificate control," in *30th USENIX Security Symposium, USENIX Security 2021, August 11-13, 2021* (M. D. Bailey and R. Greenstadt, eds.), pp. 4383–4400, USENIX Association, 2021.