electronics

MDPI

Article

In-Depth Evaluation of the Impact of National-Level DNS Filtering on DNS Resolvers over Space and Time

Yanan Cheng [1], Yali Liu [1], Chao Li [1], Zhaoxin Zhang [1,*], Ning Li [1,*] and Yuejin Du [2]

[1] Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China; mrcheng0910@gmail.com (Y.C.); 20162204110@stu.kust.edu.cn (Y.L.); 20b903094@stu.hit.edu.cn (C.L.)
[2] Beijing Qihoo Technology Co., Ltd., Beijing 100015, China; du.yuejin@gmail.com
* Correspondence: zhangzhaoxin@hit.edu.cn (Z.Z.); li.ning@hit.edu.cn (N.L.)

Abstract: DNS filtering is the practice of blocking access to certain sites for a specific purpose, often content-based filtering. Unlike previous studies that focused on the behavior of national-level DNS filtering itself (e.g., location of filtering devices), we demonstrate and evaluate in depth the impact of DNS filtering on different types (public, ISP, and open) of DNS resolvers in the censored networks. In particular, we actively send DNS queries for 83 well-selected domain names to three types of DNS resolvers and keep track of the resolvers' responses changing over time and space in China. Here, we present the results of our system running for 40 days, during which we obtained a total of 1.7 billion DNS records. Using these collected data, we found that specific DNS resolvers are unaffected by DNS filtering devices and can respond with the correct IP addresses for particular blocked domains. Furthermore, we revealed that three factors should be considered to evaluate the impact of a country-level DNS filtering mechanism: DNS resolver, client location, and blocked domain. Finally, we propose and implement a system to identify the correct IP addresses of blocked domain names in censored networks based on the characteristics of country-level DNS filtering.

Keywords: DNS; DNS filtering; DNS resolvers; censored networks

check for updates

Citation: Cheng, Y.; Liu, Y.; Li, C.; Zhang, Z.; Li, N.; Du, Y. In-Depth Evaluation of the Impact of National-Level DNS Filtering on DNS Resolvers over Space and Time. *Electronics* **2022**, *11*, 1276. https://doi.org/10.3390/electronics11081276

Academic Editors: Martin Reisslein and Wojciech Mazurczyk

Received: 22 February 2022
Accepted: 15 April 2022
Published: 18 April 2022

1. Introduction

The Domain Name System (DNS) is a vital network infrastructure that translates a human-readable domain name into a numerical IP address, which affects how well the whole Internet works. However, many methods cause DNS facilities to respond to incorrect results. For example, as studied in this paper, DNS filtering, also known as DNS redirection or poisoning, causes the DNS resolvers to return incorrect domain records (e.g., IP addresses) to the clients [1–5]. This method may be used for malicious purposes such as phishing, for security and business purposes by a company [6,7] (providing parental control or antivirus filtering services), for advertising purposes of Internet service providers (ISPs), or for the governments to block access to specific domains [8–10].

At the same time, there are many different types of DNS resolvers on the Internet that provide domain name resolution services to clients. Along with the well-known public DNS (*pDNS*) and ISP DNS (*iDNS*) resolvers, there are a large number of open DNS resolvers (*oDNS*) that serve special purposes and are unfamiliar to most people. As a result, given the complexity of national-level cyberspace and the diversity of Internet entities involved (e.g., multiple DNS resolver types and diverse network locations of users), the effectiveness of country-level DNS filtering mechanisms to affect DNS resolvers within their jurisdiction is unknown. In addition, whether this mechanism can fully succeed in blocking all users from accessing the blocked domains. These issues are the driving force behind the research conducted in this paper.

In order to provide a comprehensive and in-depth analysis of the impact of DNS filtering mechanisms on DNS resolvers, we use 33 probes deployed across numerous regions and ISPs to obtain IP addresses of 83 blocked domains in different categories from

86,876 different types of DNS resolvers (13 *pDNS*, 19 *iDNS*, and 86,834 *oDNS*). Of these, 86,834 open DNS resolvers were obtained from the IP address space of Shanghai in our designed experiment (Section 3). Finally, we obtain a total of 1.7 billion DNS records between 3 December 2021 and 12 January 2022.

Based on the above data, first, we demonstrate that the DNS filtering devices occasionally fail, resulting in the client obtaining the correct IP address of the blocked domain. Still, the number of correct responses is minimal (Section 4.1). On the other hand, the pDNS and iDNS resolvers can respond more frequently to clients with the correct IP addresses of blocked domain names, even in DNS censorship networks. Therefore, we figured out that the effectiveness of the country-level DNS filtering mechanism is related to three factors: DNS resolver, client location, and blocked domain name. This indicates that the DNS filtering mechanism has a more complex and variable impact on resolvers in the censored networks, rather than having a single effect as expected, i.e., all resolvers respond to all blocked domain names with the wrong IP addresses (Sections 4.2 and 4.3).

Secondly, we also find that, because of the characteristics of open DNS resolvers (e.g., configuration flexibility and stealthiness), DNS filtering mechanisms have a more complex impact on them compared to public and ISP DNS resolvers. For example, some open DNS resolvers can correctly resolve specific blocked domains. Another example is that the open resolvers' responses to the correct IPs of blocked domains vary more over time, while public DNS resolvers are weakly correlated with time (Section 5).

Finally, we develop and implement a system for identifying the valid IP addresses of blocked domain names in the censored network based on the features of forged IPs used by DNS filtering devices. The system is able to obtain the correct IPs of the blocked domains in the DNS censored networks with 100% accuracy in a short period of time (Section 6).

The contributions of our study are summarized below:

- For the first time, we systematically and quantitatively evaluate the impact of DNS filtering mechanisms on the response of three types of DNS resolvers (public, ISP, and open DNS resolvers) to blocked domain names in censored networks.
- We reveal that many open DNS resolvers are not affected by DNS filtering due to their own characteristics (e.g., configuration flexibility and stealthiness). This indicates that using these open DNS resolvers can normally resolve blocked domains in country-level censorship networks.
- Using the behavior of DNS resolvers responding to blocked domain names in censored networks, we design and implement a novel system to identify the correct IP addresses of blocked domain names in untrusted networks.

China's national-level DNS filtering is a long-standing academic research hotspot, and researchers have done many studies on the network location of DNS filtering devices, blocking policies, lists of blocked domain names, and evasion policies [8–17]. On the other hand, we have to point out that Internet censorship, including DNS filtering, HTTP filtering, etc., has also long been controversial (e.g., restricting Internet freedom). Therefore, our work aims to provide a valuable reference for implementers of DNS filtering mechanisms or entities (e.g., domain owners and DNS resolver administrators) affected by censorship from a technical perspective.

2. Background and Related Work

2.1. Background

The DNS provides one of the most critical network services on the Internet, which translates domain names into IP addresses. The standard process of obtaining the IP address corresponding to a domain name consists of multiple facilities, including the recursive resolver, the root servers, Top Level Domain (TLD) servers, and the domain authority nameservers, as illustrated in Figure 1 (Black solid or dotted lines).

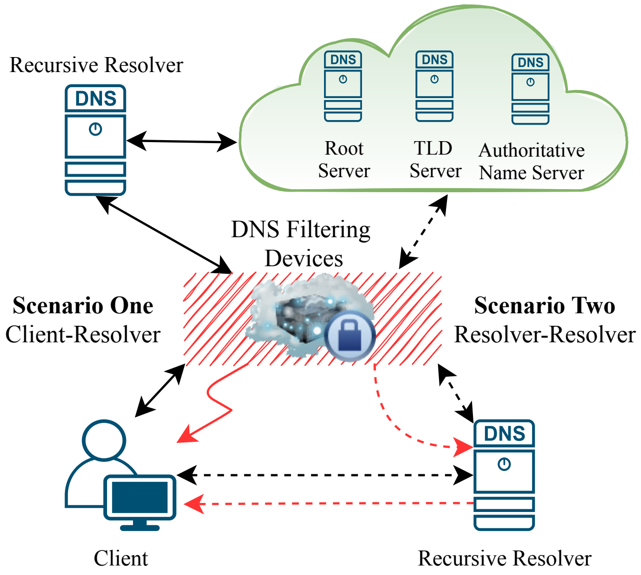


Figure 1. Standard DNS query procedure and location of DNS filtering.

Generally, DNS filtering is an on-path DNS hijacking method that can passively observe every DNS query request "passing through" its devices [4]. At the same time, it can directly return forged response packets to the DNS resolvers or clients [18]. Suppose the DNS resolver is not configured with DNSSEC (Domain Name System Security Extensions) validation. In that case, it will usually accept the forged response IP because it is earlier than the correct response IP. Therefore, access to the block-listed domain name will be blocked or redirected. From a high perspective, based on the location of the DNS filtering device, DNS filtering scenarios can be classified into two categories: Client-Resolver and Resolver-Resolver, as shown in Figure 1 (Solid and dotted red lines).

- **Client–Resolver Scenario.** The client's DNS request goes directly through the DNS filtering devices, and then the filtering devices directly return error responses. For example, in China, clients query IP addresses of the domain pornhub.com to Google's public DNS resolver 8.8.8.8.
- **Resolver–Resolver Scenario**. The DNS filtering devices are between DNS resolvers (e.g., recursive DNS and TLD servers), and the recursive resolvers obtain incorrect responses and return them to the user. For example, In China, clients query the domain pornhub.com IP address to ISP-provided DNS resolvers.

Overall, most studies have focused on the scenario Client–Server, but, in practice, since most clients interact directly with DNS resolvers, studying the scenario Server–Server provides a more comprehensive and representative assessment of the impact of DNS filtering mechanisms.

From the clients' point of view, both of these scenarios eventually affect users' access to domain names on the blocklist. Moreover, the "clean" rate of the recursive DNS resolver directly responding to user requests affects the accuracy rate of domain names in the blocklist. Therefore, the focus of this paper is the recursive DNS resolver, which we refer to as the DNS resolver for ease of description. We divide the resolvers into three categories according to the characteristics of the DNS resolver's operation and maintenance organization.

- **Public DNS Resolver (*pDNS*)**. These DNS resolvers are managed by businesses, institutions, or organizations, such as Google 8.8.8.8 and AliCloud 223.5.5.5. Public DNS resolvers are able to provide domain name resolution services to clients in different cyberspaces around the world.

- **ISP DNS Resolver (*iDNS*)**. These DNS resolvers are provided by Internet Service Providers (ISPs), and, in general, they only provide domain name resolution services to clients who use their Internet services, such as China Unicom 221.12.1.227.
- **Open DNS Resolvers (*oDNS*)**. In addition to the above two types of DNS parsers, there are a large number of open DNS that are not well known to ordinary people on the Internet [5,12]. This type of DNS resolvers may be used for special purposes, such as DNS experiments, while they may only serve a smaller scope of people. Moreover, most open DNS resolvers are not actual resolvers but DNS forwarders in home routers and other Internet devices [19].

In general, we request the IP addresses of the blocked domain names to the three types of DNS resolvers mentioned above in different cyberspaces. Based on the data collected over time, we evaluate the effectiveness of the DNS filtering mechanism on the impact of DNS resolvers.

2.2. Related Work

Many countries [12,17,18,20–22] and enterprises [6,23–25] deploy DNS filtering devices of different scales (e.g., nation-wide, ISP-wide, and Family-wide) to prevent users from accessing domain names on the blocklist from being attacked. The types of domain names on the blocklist include pornography, malware, phishing, and illegal domain names (according to relevant national laws). In addition, the methods used to implement DNS filtering at the country level are different. For example, in China, filtering devices return invalid IP addresses, but in some countries, answer that the domain name does not exist, and, in countries such as South Korea and Iran, redirect users to the blocking page to inform users about the blocked content [26]. In general, because of the complexity, effectiveness, extensiveness, and controversy of the DNS filtering mechanism in China, it has long been the research target of many scholars. Previous researchers have done a lot of research on the operation of DNS filtering mechanisms in various countries, mainly including the following:

- Many previous studies [8,9,27] focused on discovering the location of DNS filtering devices in cyberspace and collecting the set of keywords that trigger filtering. Jin et al. [3,28] proposed a novel framework that enables end-to-end measurement to detect censorship activities accurately (e.g., DNS filtering) and revealed the censor deployment without manual efforts. To discover DNS tampering, their control server will reply to the requests with a static and reserved IP address that has never been used in manipulated DNS responses by any censors in different countries. Park et al. [5] conducted a comprehensive probing over the entire IPv4 address space and found that more than 3 million open resolvers still exist in the wild. In particular, we found that many open resolvers answer queries with incorrect, even malicious, responses.
- To find deployments of DNS injection and to understand their behavior and the effect that they can have on third parties. Sparks et al. [11] analyzed the causes of the collateral damages comprehensively and measured the Internet to identify the injecting activities and their effect. They found that 26% of 43,000 measured open resolvers outside China, distributed in 109 countries, may suffer some collateral damage. Wander et al. [10] looked for large deployments of DNS injection, measured from vantage points outside of the censored networks. After testing many open resolvers outside of China, they determined that about 15,300 resolvers were potentially affected by Chinese DNS injection when querying top-level domains outside of China. Padmanabhan et al. [29] analyzed Internet censorship in Myanmar following a military coup in February 2021. They found that many ISPs in Myanmar showed evidence of confirmed DNS blocking, usually resolving to an IP address that hosted a block page or responding with NXDOMAIN to domains (e.g., facebook.com).
- Using blocked domains and open DNS resolvers to evaluate the efficiency of DNS injection, while analyzing the reasons why DNS filtering occasionally fails [12–15]. These studies are only preliminary exposures of the impact of DNS filtering mecha-

nisms on blocked domains, and do not systematically and quantitatively assess the effectiveness of censorship networks on the impact of different types of DNS resolvers.

- Some studies have focused on how to evade DNS filtering mechanisms. According to the mechanism of DNS filtering in China, that is, fake DNS response packets reach the client before valid response packets reach the client. Duan et al. [16] investigate the benefits of stub resolvers or forwarders waiting for a "Hold-On" period to allow subsequent legitimate replies to arrive. Hoang et al. [30] built a measurement system to study the accessibility of DNS-over-TLS(DoT)/DNS-over-HTTPS(DoH) and Encrypted Server Name Indication (ESNI) and to investigate whether these protocols are tampered with by network providers in censored networks. Chai et al. [31] measured the deployment prevalence of ESNI and further analyze its current and potential effectiveness in censorship circumvention. Based on their analysis, they discussed the key factors to the success of ESNI and potential problems in a post-ESNI era.

Based on previous research, we designed experiments to more systematically and quantitatively evaluate the effectiveness of country-level DNS filtering mechanisms, especially on different types of DNS resolvers. In addition, using the characteristics of DNS filtering responses to incorrect IPs, we developed a method to identify the correct IPs of blocked domain names to evade DNS censorship.

3. Methodology

This section details our research methodology for determining the effect of DNS filtering on various types of DNS resolvers. Additionally, we describe the approaches we took to address ethical issues that arose during our research.

3.1. Blocked Domain Names Collection

In addition to being DNS censored, the domains to be studied in our work also need to have high traffic and cover multiple categories. Therefore, we follow the standard approach used in prior studies [17,28,32] and carefully select blocked domains (BLDs) from Alexa Top 1M, Citizen Lab Block List (Citizen Lab. Block Test List. https://github.com/citizenlab/test-lists, accessed on 8 March 2022), and blocked domain list from Wikipedia (https://en.wikipedia.org/wiki/Websites_blocked_in_mainland_China, accessed on 8 March 2022). In the end, our blocklist consists of 83 domain names, and the domain category (categorized by FortiGuard (https://www.fortiguard.com/webfilter?q=bbc.com&version=9), accessed on 8 March 2022) covers News, Social, VPN, Porn, Video, and Search Engine, as shown in Table 1.

Table 1. Number and category of blocked domain names in China.

#	Category	# Domains
1	Search Engine	26
2	News	20
3	VPN	19
4	Porn	6
5	Video	7
6	Social	5

3.2. Public and ISP DNS Resolvers Collection

Most users or organizations use public or ISP DNS resolvers to resolve domain names. If the client does not configure a specific resolver, it is the default resolver configured by the ISP. We use search engines (e.g., Google, Baidu) to retrieve many public and ISP DNS resolvers and select DNS resolvers that can provide services for the vantage points deployed (introduced in Section 3.4.1). In addition to the Chinese public and ISP DNS resolvers, we also selected two non-Chinese public DNS (Google 8.8.8.8 and Cloudflare 1.1.1.1) as our comparison objects. In order to make our experiment more convincing, we set up two DNS recursive resolvers (using Bind 9.16.1) under our control in China and the

United States, respectively, as research objects. In the end, we obtain 32 public and ISP resolvers, as shown in Tables 2 and 3.

Table 2. The list of public DNS resolvers.

Resolver	Country	Owner	Resolver	Country	Owner
1.2.4.8	China	CNNIC	8.8.8.8	USA	Google
180.76.76.76	China	Baidu	1.1.1.1	USA	Cloudflare
223.5.5.5	China	Alibaba	1.33.184.193	Japan	NTT PC
114.114.114.114	China	Xinfeng	122.248.129.45	AU	ATT
119.29.29.29	China	DNSPod	45.63.86.214	USA	Self-build
101.226.4.6	China	DNSPai	47.93.151.157	China	Self-build
117.50.11.11	China	oneDNS	-	-	-

Table 3. The list of ISP DNS resolvers.

ISP	Resolver	Province	ISP	Resolver	Province
Mobile	211.140.13.188	Zhejiang (ZJ)	Unicom	221.12.1.227	Zhejiang
	211.140.188.188	Zhejiang (ZJ)		221.12.33.227	Zhejiang (ZJ)
	218.201.4.3	Chongqing (CQ)		210.22.84.3	Shanghai (SH)
	211.137.130.19	Shaanxi (SHX)		202.106.196.115	Beijing (BJ)
	221.131.143.69	Jiangsu (JS)		123.123.123.123	Beijing (BJ)
	211.138.106.2	Shanxi (SX)	Telecom	60.191.244.5	Zhejiang (ZJ)
	211.137.191.26	Shandong (SD)		60.191.134.196	Zhejiang (ZJ)
	211.136.17.107	Hubei (HB)		60.191.134.206	Zhejiang (ZJ)
	211.141.85.68	Jiangxi (JX)		202.98.224.69	Xizang (XZ)
	211.138.180.2	Anhui (AH)	-	-	-

3.3. Open DNS Resolvers Discovering

3.3.1. Discovering Method

As shown in Figure 2, the open DNS resolvers discovering stage is responsible for sending A and AAAA records query for two well-selected domains to the specific IPv4 address space and collecting responses from target IP addresses. These two domain names are composed of a random blocked domain and a purpose-registered domain name that we control.

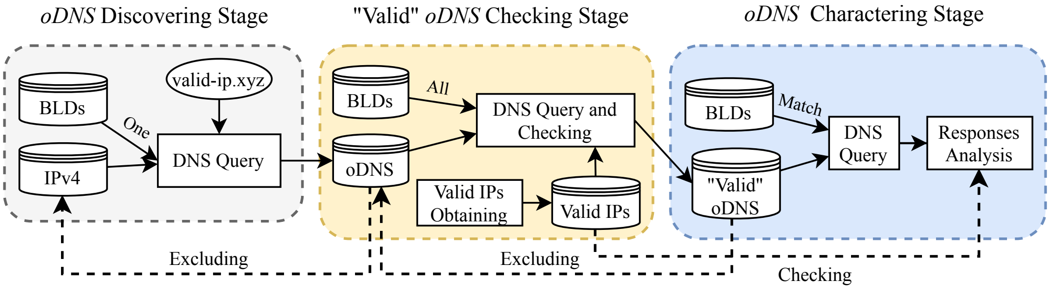


Figure 2. Processing chain of discovering and characterizing open DNS resolvers.

The purpose of using a domain name that we can control is to exclude two fake "open DNS resolver" types. One type is a resolver with its configuration error, and all of its responses are wrong. For example, a resolver only responds to the client with IP 127.0.0.1. The other type is not an actual DNS resolver, but the DNS filtering device exists between the client and the target IP. When the DNS request of the blocked domain name passes through the devices, it triggers and responds with invalid IPs. In this case, the target IP is not even a DNS resolver. Theoretically, because our vantage point and the open resolvers are both in the same city, this type of resolver does not exist. Moreover, we prove our conjecture by analyzing our experimental data.

3.3.2. Target IPv4 Space

We discovered open DNS resolvers using 11,135,744 IP addresses in Shanghai (Chacuo. http://ips.chacuo.net/, accessed on 8 March 2022). This city has a high number of IP addresses and a higher Internet penetration rate than the rest of the country. In practice, this city has a large number of resolvers available for our study.

3.3.3. Open DNS Resolvers Dataset

As of 12 January 2022, Shanghai cities have a total of 86,844 open DNS resolvers. Marc et al. [33] observed that open resolvers experience a significant churn rate over time, particularly during the first week. As a result, we determined the churn rate of open DNS resolvers to decide whether or not they continue to provide DNS service over time, as illustrated in Figure 3. Additionally, because a resolution's stability is crucial to the public service and its users, a well-maintained resolver should avoid prolonged outages [34]. As a result, we decided to concentrate our research on open resolvers capable of providing stable domain name resolution services for at least 30 days. In the end, we obtained 23,669 open DNS resolvers in Shanghai.

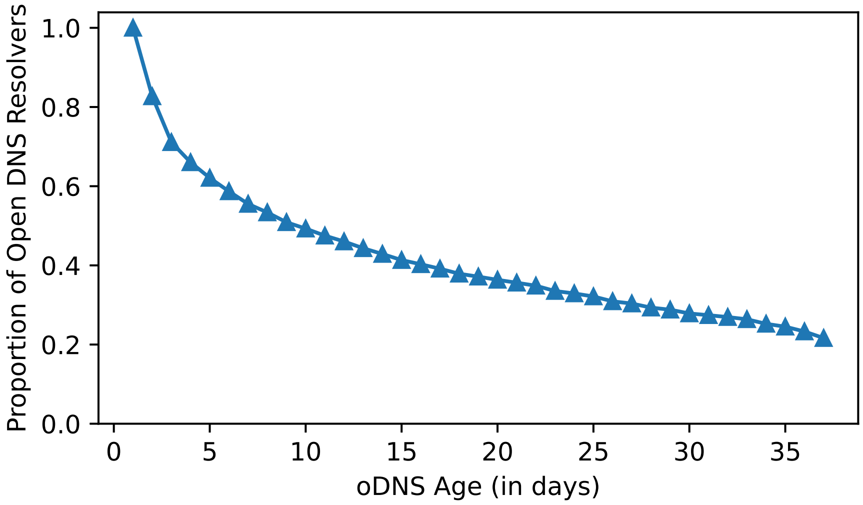


Figure 3. IP address churn of DNS resolvers for 30 days.

3.4. Performing Large-Scale DNS Queries

3.4.1. Vantage Point Selection

In order to ensure that the experiment obtains enough data about DNS resolvers' response to blocked domain names, we select the appropriate advantage points according to the features of different types of DNS resolvers. For example, the China Telecom ISP DNS resolver in Zhejiang Province corresponds to the vantage point of the Telecom ISP in Zhejiang Province. For the open DNS resolvers found in Shanghai, the probers in these two cities are used as advantage points. Finally, we leased 33 virtual private servers(VPSs) in total, covering 11 cities and four ISPs. Figure 4 shows the geographical distribution of the vantage points and all DNS resolvers used in this paper.

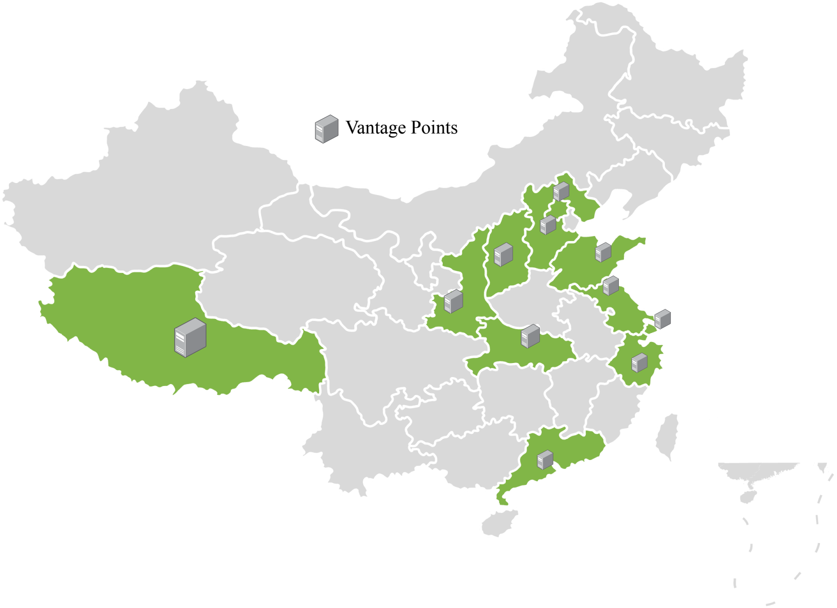


Figure 4. Geographical distribution of vantage points and DNS resolvers in China.

3.4.2. Performing DNS Queries

Theoretically, evaluating the impact of DNS filtering on DNS resolvers is straightforward. At a high level, we can determine the effect on DNS resolvers by examining the rate at which DNS resolvers correctly respond to DNS records for blocked domain names requested by clients in various locations. However, the procedure appears straightforward; implementing a system capable of scaling DNS data collection and analysis of the DNS filtering mechanism's impact on resolvers presents both ethical (described in Section 3.7) and technical hurdles.

We employ two detection strategies to collect blocked domain DNS records from the three types of DNS resolvers. On the one hand, DNS records for domains are requested at a fair rate directly from vantage points to the public and ISP DNS resolvers. On the other hand, due to a large number of open DNS resolvers, we designed an effective evaluation process for them, as illustrated in Figure 2, which we detail below.

- **oDNS Discovering Stage**. As mentioned in Section 3.3, this stage is primarily for the acquisition of open resolvers in Shanghai city.
- **"Valid" oDNS Checking Stage**. At this stage, we send DNS requests for all domains on the blocked list to the discovered open DNS resolvers. Then, depending on the correct IP addresses (to be described in Section 3.5) of the blocked domains, we discovered the "valid" open DNS resolvers that correctly responded with at least one correct request for blocked domains. Finally, we establish a set of "valid" open DNS resolvers for each domain that can properly respond to the requests. This set is passed to the following stage as input.
- **oDNS Charactering Stage.** At this stage, we request DNS queries for the matching domain name from "valid" open DNS resolvers (identified in the previous stage). Then, using the correct IP addresses of blocked domains, the response of the open DNS resolver is studied, and the open DNS resolver's features are comprehensively determined.

In summary, the three stages described above are each responsible for a distinct set of duties. Stage 1 is responsible for discovering oDNS; Stage 2 is responsible for identifying "valid" oDNS from the discovered oDNS for each blocked domain with at least one correct response; Stage 3 is responsible for obtaining sufficient response records from each oDNS in order to determine the oDNS's characteristics.

3.5. Obtaining and Verifying Correct IP Addresses

We resolve blocked domain names using three different types of DNS resolvers and check to see if the DNS answers include correct IP addresses. If the DNS response does not include the valid IP address, it is false. Therefore, first, we have to obtain and verify the correct IP addresses of the blocked domains.

The fact is that a vast number of domains, particularly popular domains, make use of content delivery network (CDN) services, which refer to a geographically distributed group of servers that work together to provide fast delivery of Internet content. This fact results in domains with multiple IP addresses, and these IPs are impossible to enumerate. The existence of a comprehensive list of valid IP addresses directly impacts our analytical results. This leads to the challenge of obtaining as many correct IP addresses of blocked domains as possible. Thus, we address this problem by applying the following procedure to obtain valid IP addresses, as shown in Figure 5.

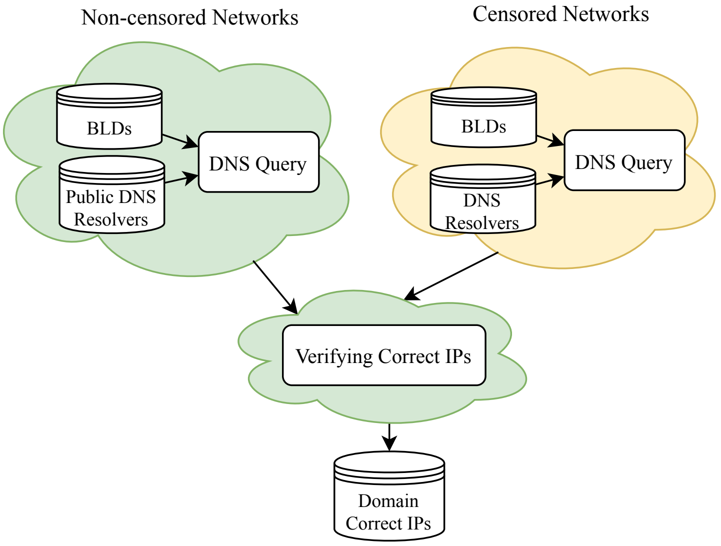


Figure 5. Processing chain to obtain and verify domain correct IP addresses.

Step 1: Obtaining IP addresses. In each of the two networks (censored and uncensored), we obtain the IP addresses of the blocked domains. In the uncensored network, we query multiple global public DNS resolvers (e.g., Google 8.8.8.8 and Cloudflare 1.1.1.1) for the IP addresses of the blocked domain names. Theoretically, the IP addresses that these DNS resolvers respond to are the correct addresses of the blocked domains. However, there is a tiny chance of an error due to DNS tampering [28]. On the other hand, we request the IP addresses of the blocked domain names from different types of DNS resolvers in the censored network.

Step 2: Verifying IP addresses. This verification takes place in uncensored networks. We send a domain HTTP or HTTPS request to an IP address obtained in Step 1 (using the command *curl-resolve*). Because the blocked domains chose all host HTTP or HTTPS services, we believe the IP address is correct if we receive the correct web content of the domain; otherwise, we believe the IP address is incorrect.

3.6. Evaluating Impact Metric

To summarize the findings of this research succinctly and accurately, we suggest one metric (as illustrated in Equation (1)) for analyzing the impact of country-level DNS filtering on DNS resolvers. Unless otherwise specified, this work refers to the DNS resolver response record as the domain name A record, which corresponds to the IPv4 address.

Absolute Correct Rate of Responses (ACR) is the proportion of correct records in the response records for a specific domain name requested from a DNS resolver at a Vantage Point (VP), during a certain time period (or number of times), as shown in Equation (1):

$$ACR_{(domain,DNS,VP)} = NCR/TNR \tag{1}$$

NCR is the **N**umber of **C**orrect **R**esponses that the DNS resolver responds to. *TNR* is the **T**otal **N**umber of **R**esponses from the DNS resolver.

3.7. Ethics and Legal Issues

First, we describe in detail the hosts used in this paper, including the rental channels, configuration, and spending, as shown in Table 4. We rented the appropriate hosting from two channels: Alibaba Cloud and Taobao. Alibaba Cloud (Alibaba Cloud. https://www.aliyun.com, accessed on 8 March 2022) (AliCloud) is a technology company that provides cloud computing services and offers a large number of customizable, configurable virtual machines. We rented 23 lower-performance virtual machines for probing and one higher-configuration host for storing the result data. Taobao (Taobao. https://www.taobao.com, accessed on 8 March 2022) is an online shopping platform in China on which you can rent hosting from different provinces and ISPs. Most of the merchants only provide fixed-configuration virtual machines, and the cost is much higher than AliCloud's. We rented 10 virtual machines distributed across different provinces and ISPs, such as Zhejiang. As shown in Table 4, we spent a total of about ¥8434 for two months renting the hosts.

Table 4. The performance and cost of the rented hosts.

Source	CPU	Memory	Storage	Price	Number	SUM
AliCloud	2	4 G	40/80 G	¥90	23	¥2070
AliCloud	4	8 G	500 G	¥447	1	¥447
Taobao	2	2 G	40 G	¥170	10	¥1700
			SUM			¥4217

For all the hosts we rent, we limit the maximum traffic rate for sending DNS queries to no more than 500 Kb/s (using Linux command *iftop* statistics). Such low host network traffic will not cause host providers or ISPs. In addition, we are careful to minimize the burden on remote hosts by limiting the rate and do not want to trigger unwanted attention from the hosting provider.

We send domain name requests to each public and ISP DNS resolver at a low rate. We did not find any denial of service by the DNS resolver during the experiment. In addition, public and ISP resolvers provide services to many clients, and they usually implement multiple technologies (such as DNS anycast) to ensure regular domain name resolution services. We believe that our experiment does not cause excessive load to these resolvers and does not affect other users.

Because the number of open DNS resolvers is large, the host's performance performing the resolution service is unknown.Therefore, during our experiments, we have to consider both efficient discovery of open resolvers and not overloading the open resolvers. As shown in Figure 2, we design the process of open resolver discovery and domain name request processing. This case can ensure that we obtain enough domain name response data within a reasonable time, reduce the number of additional domain name requests, and avoid bringing unnecessary traffic data to the Internet.

We only proceed from a technical point of view to evaluate the impact of DNS filtering technology on DNS resolvers. We do not disclose or disseminate all the essential data collected (e.g., valid IPs of the domain names, open DNS resolver). We only describe our detection results and explain why in the paper.

4. Analyzing Public and ISP DNS Resolvers

We obtained a total of 650 million DNS response records from 32 DNS resolvers (19 iDNS and 13 pDNS) after 40 days. Public and ISP DNS resolvers are well-known to individuals or organizations and are the most often used. As a result, we undertook an exhaustive investigation of the resolution of blocked domain names by these two types of DNS resolvers in order to determine the scope and magnitude of the impact of DNS filtering mechanism at the country level.

4.1. Non-Chinese Public DNS Resolvers

The ACR of the non-Chinese public DNS resolver answer is less than 1% (the red square shown in Figure 6) for blocked domain name lookup requests made through the DNS filtering devices, which is essentially the same as the results from the previous study [11,27]. As illustrated in Figure 6, DNS response results for each blocked domain name are queried from Beijing's vantage points to Google resolver 8.8.8.8. Moreover, the response results in other places (different regions and ISPs) are essentially the same. The nation-level DNS filtering mechanism affects the DNS resolver 45.63.86.214 that we built in the United States.

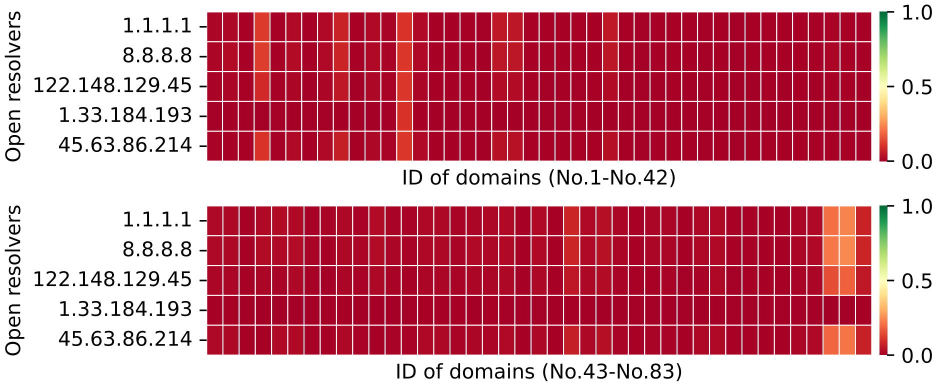


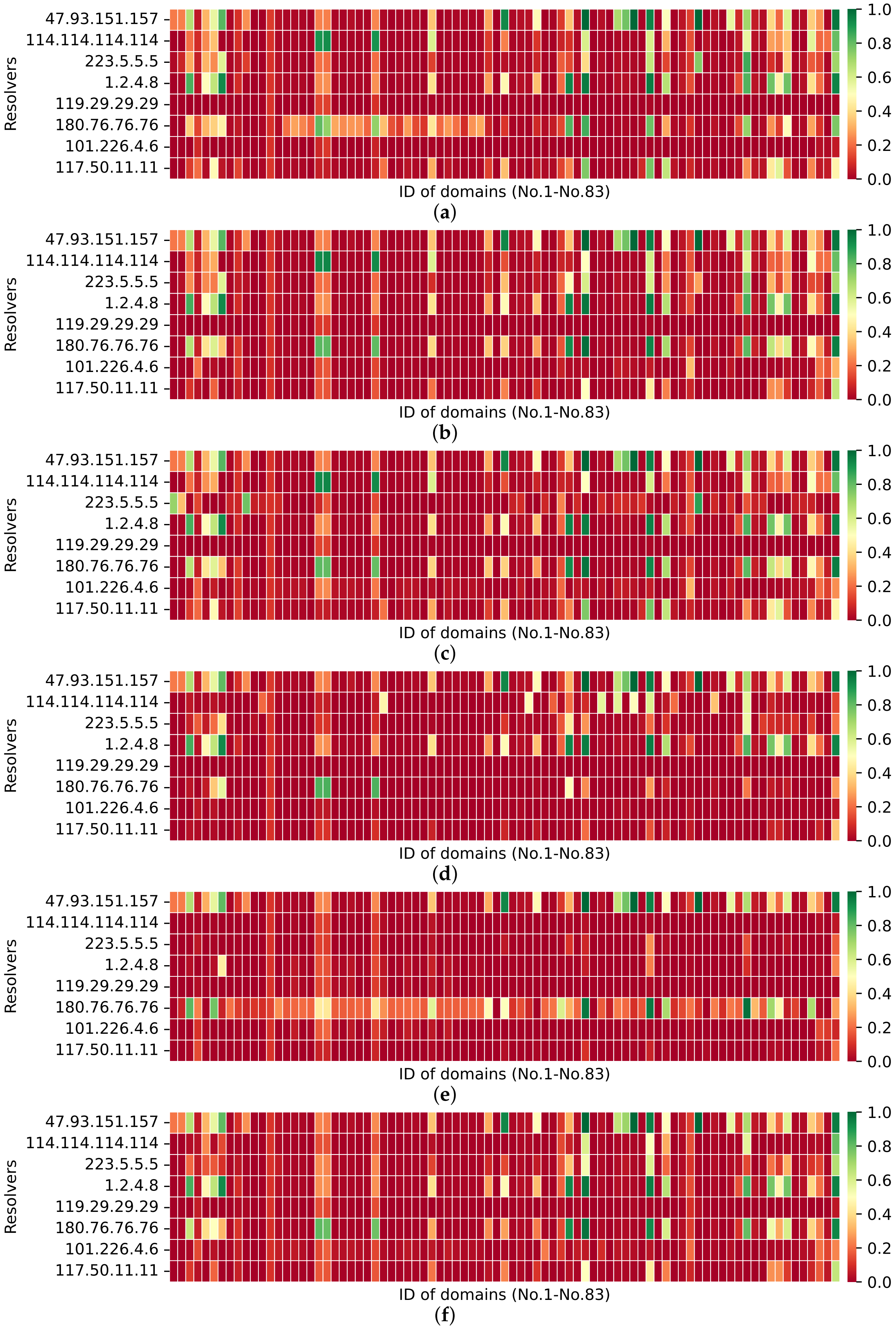
Figure 6. The ACR of each non-Chinese public resolver for blocked domains from Beijing.

In summary, these results exemplify how the DNS filtering device's centralized censorship strategy works. These results also indicate the significant effects of the country-level DNS filtering devices for DNS queries of blocked domains that pass directly through them. Clients in censored networks have a very low success rate in requesting the IP addresses of the blocked domains from DNS resolvers in non-censored networks, regardless of what cyberspace the Clients are in.

4.2. Chinese Public DNS Resolvers

As illustrated in Figure 7, the ACR of Chinese public DNS responses to blocked domains is significantly higher than the relatively low accurate response rate of non-Chinese DNS resolvers. We discovered that Chinese public DNS resolvers resolve specific domain names without being affected by DNS filtering devices (green squares in the Figure 7) or with less influence (light-colored squares in Figure 7). We summarize the impact of the country-level DNS filtering mechanism on Chinese public DNS resolvers as follows.

Public DNS resolvers can respond correctly to specific domain names. We discovered that, inside a country-level network of DNS filtering techniques, public DNS resolvers consistently replied successfully to DNS requests for certain blocked domains. This is in stark contrast to our anticipation that all DNS resolvers would fail to correctly reply to domain names within the censored network. However, in practice, some DNS resolvers do correctly or partially correctly reply to requests for blocked domains. For instance, all six DNS resolvers, including the one we built, resolve the pornographic domain xvideo.es successfully. Certain DNS resolvers answered successfully to specific blocked domains (e.g., wsj.com, vpnintouch.com) with an ACR of greater than 0.6.



(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

Figure 7. The ACR of each public resolver for blocked domains from different vantage points. (**a**) Beijing City, AliCloud; (**b**) Shanghai City, AliCloud; (**c**) Hebei Province, AliCloud; (**d**) Zhejiang Province, Mobile; (**e**) Zhejiang Province, Unicom; (**f**) Zhejiang Province, Telecom.

Inconsistent responses from the same public DNS resolver in different cyberspace. When the same public DNS resolver is queried at vantage points in different cyberspaces, we discovered that the responses are inconsistent. As illustrated in Figure 8, the four public DNS resolver respond to requests for blocked domain names (i.e., resolvers' ACR) acquired from each of the six distinct cyberspace vantage points. For example, 119.29.29.29's average accurate response rate to domain name queries from various areas is less than 1%, whereas other resolvers can respond 100% correctly to some domain name requests from various regions (as shown in the green squares in the figure). Simultaneously, 180.76.76.76's response results vary for clients from different ISPs within the same region; for example, its response results to Zhejiang Telecom and Mobile ISP are inconsistent. The explanation for such inconsistent response outcomes is inferred from the fact that public DNS resolvers maintain numerous DNS cache resolvers in distinct network regions in order to provide more stable and faster domain name resolution services. These cached DNS resolver services serve customers in several areas, and the cache servers are impacted by a variety of DNS filtering techniques, resulting in inconsistent response returns [35].

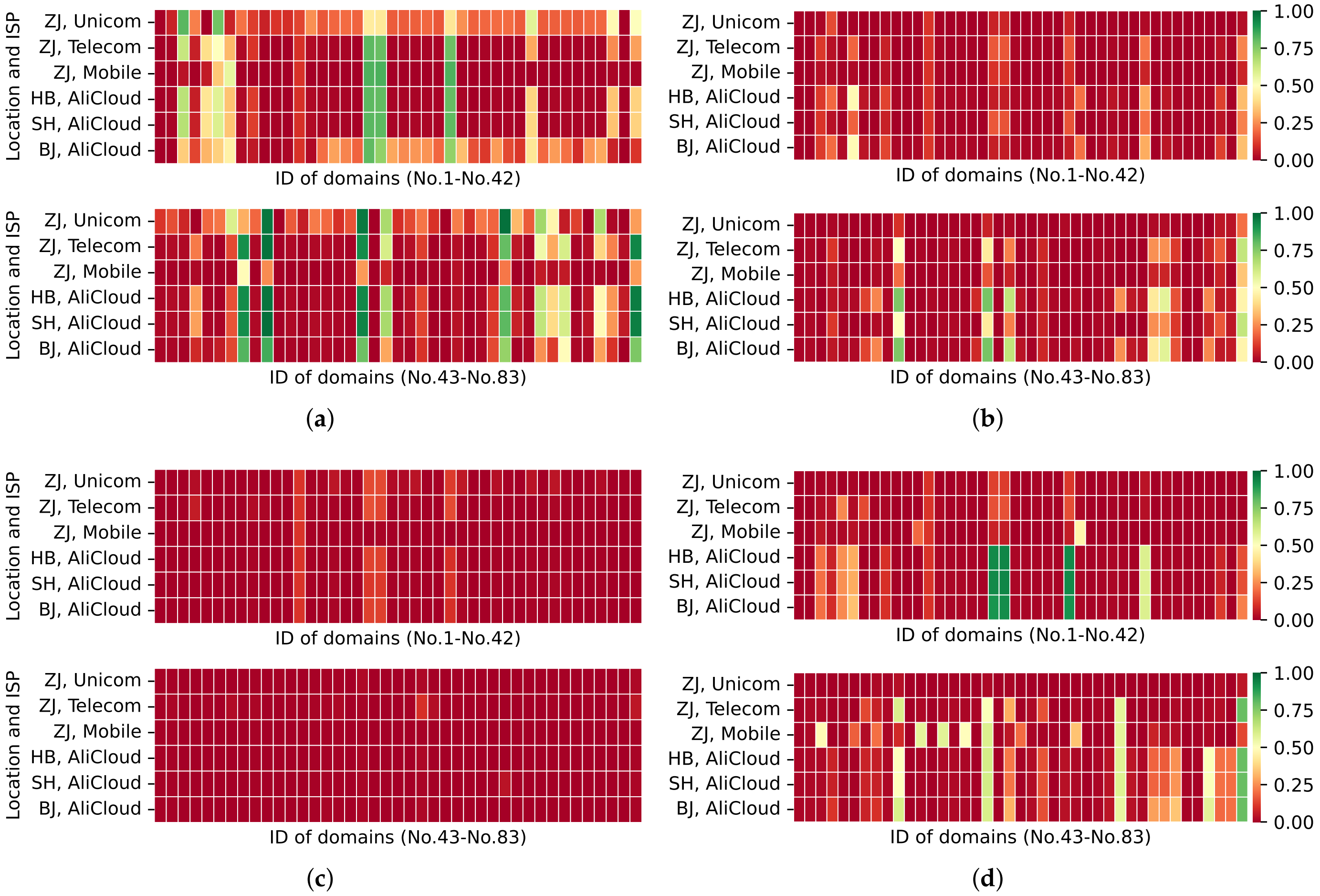
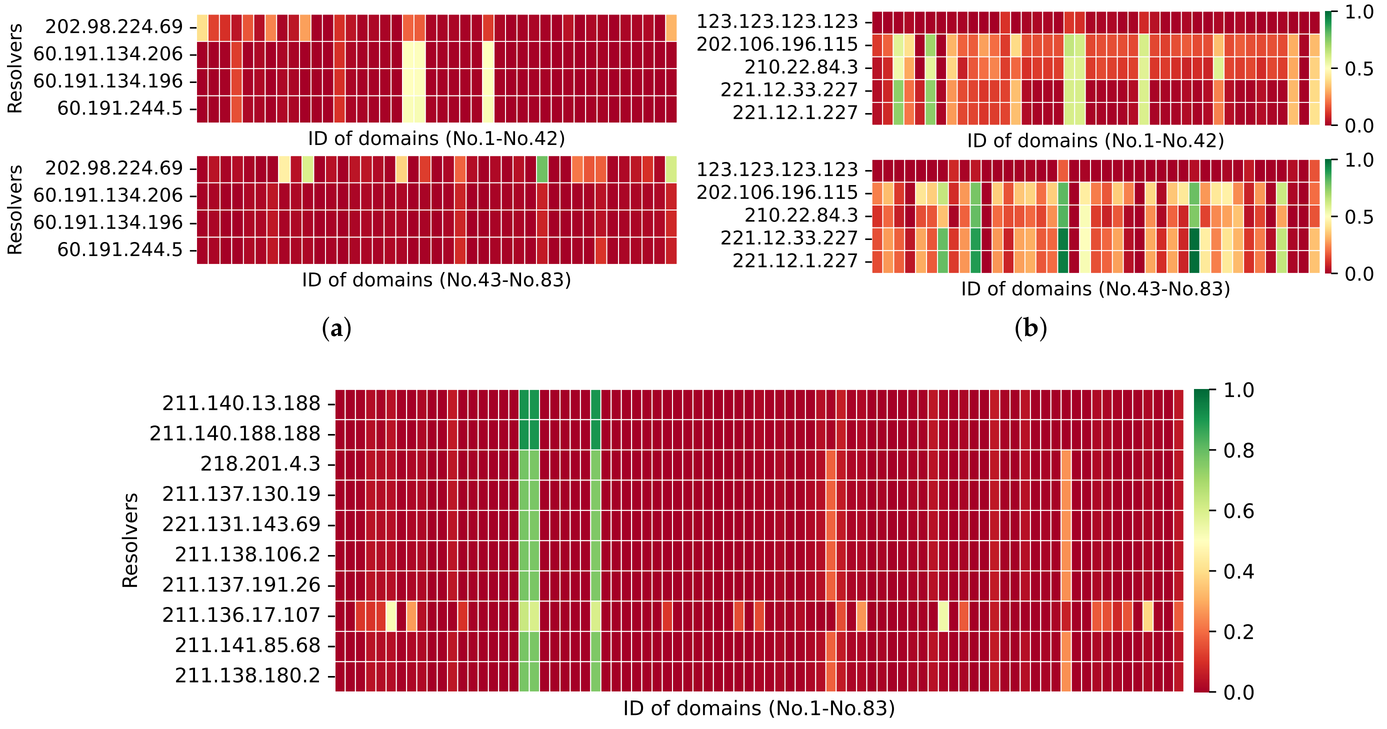


Figure 8. Clients in different cyberspace lookup blocked domain names to the public DNS Resolvers. (**a**) 180.76.76.76; (**b**) 117.50.11.11; (**c**) 119.29.29.29; (**d**) 114.114.114.114.

4.3. Chinese ISP DNS Resolvers

As shown in Figure 9, ISP DNS resolvers have some of the same characteristics as public DNS resolvers for resolving blocked domain names (e.g., they can resolve blocked domain names correctly), in addition to their own distinctive features, which are described in detail below.

DNS filtering has varying effects on different ISPs. As illustrated in Figure 9a,b, ISP Telecom and Mobile are the most impacted by the DNS filtering mechanism, with the majority of resolvers failing to retrieve the right IP address for the domain name and an ACR value of less than 0.2. The ISP DNS resolver responds appropriately to three domain names with an ACR greater than 0.8. As shown in Figure 9b, the ISP Unicom is distinct from the other two in that their resolver correctly resolves a large number of domain names with an ACR greater than 0.4.



(**a**) (**b**)

(**c**)

Figure 9. The correct rate of DNS resolvers responding to blocked domains for different ISP. (**a**) ISP Telecom; (**b**) ISP Unicom; (**c**) ISP Mobile.

DNS filtering may affect resolvers from the same ISP in various countries inconsistently. As illustrated in Figure 9a, the ISP Telecom DNS resolver 202.98.224.69 in Tibet can correctly respond to the domain names windscribe.com and xvideos.es (with ACR values more than 0.7), while the Telecom DNS resolver in other regions replies to the domain names with a lower accurate rate. Similarly, the ISP Unicom DNS resolver in Beijing, 123.123.123.123, resolves all domains with an ACR value of less than 0.1, which is notably different from the other Unicom DNS resolvers (as seen in Figure 9b).

Consistent responses from the same ISP and Location DNS resolvers. We discovered that, when we analyzed the responses of DNS resolvers from the same ISP in the same area, their accurate rates are nearly identical. For instance, Figure 10 depicts the correct response rates of seven ISP DNS resolvers (Telecom, Unicom, and Mobile in Zhejiang Province) to blocked domain names. This case means that clients in a region who use any DNS resolver provided by their ISP are similarly affected by DNS filtering. Additionally, it demonstrates once again that DNS is affected differentially by DNS filtering at the national level for different ISPs.

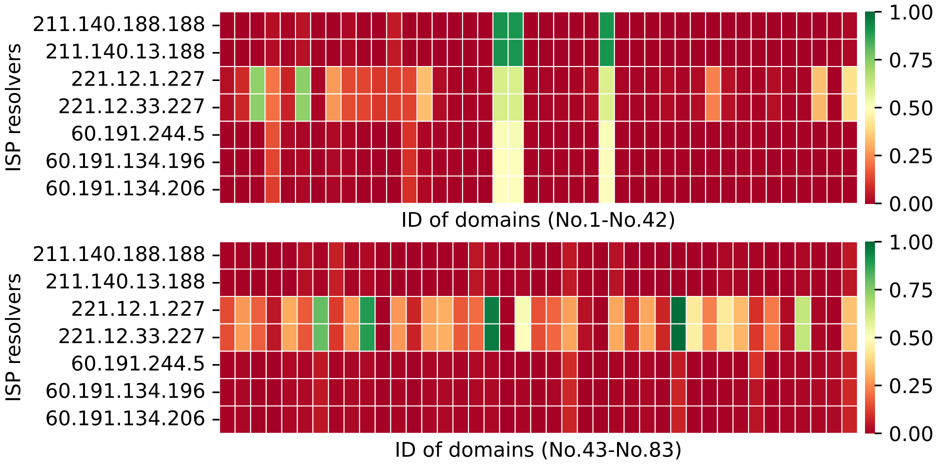


Figure 10. The correct response rate of resolvers belonging to the same ISP is basically the same.

Correct responses are only loosely related to time. To determine whether the rate of correct responses by DNS resolvers to domain name requests is time-dependent, as illustrated in Figure 11, we take hourly intervals and calculate the average accurate rate of DNS resolvers responding to each domain name throughout this time period. The average number of requests per domain name per hour is approximately 120, which means that the IP addresses for domain names are requested twice every minute from a single resolver. The following illustrates the effect of DNS filtering on DNS resolvers:

- According to the trend of the correct rate at which the six DNS resolvers reply to domain names each hour, the mean value of the correct rate throughout different time periods is not constant, and there are oscillations. Additionally, the proper rate fluctuation trend is nearly the same for all six resolvers. For instance, the correct rate is lower in the early morning hours until 6:00 a.m. The correct rate steadily climbs beginning at 6:00 a.m. We believe that this condition exists because the DNS filtering device is unable to process all domain name requests in a timely way when network traffic is heavy.
- Although the correct rate of DNS resolver response to a domain name varies over time, the variance is relatively tiny, with a maximum standard deviation of 0.03. At the national level, the effect of the DNS filtering mechanism is rather constant over time.
- By examining the hourly correct response rates for the six DNS resolvers (2 non-Chinese pDNS, 2 Chinese pDNS, and 2 iDNS), it appears that the more scatter in the graph, the greater the correct response rate. Once again, the low correctness rate of domain name requests that pass through the DNS filtering device (e.g., requests to Google 8.8.8.8) is illustrated, while the internal Chinese pDNS and iDNS have a relatively high correct rate.

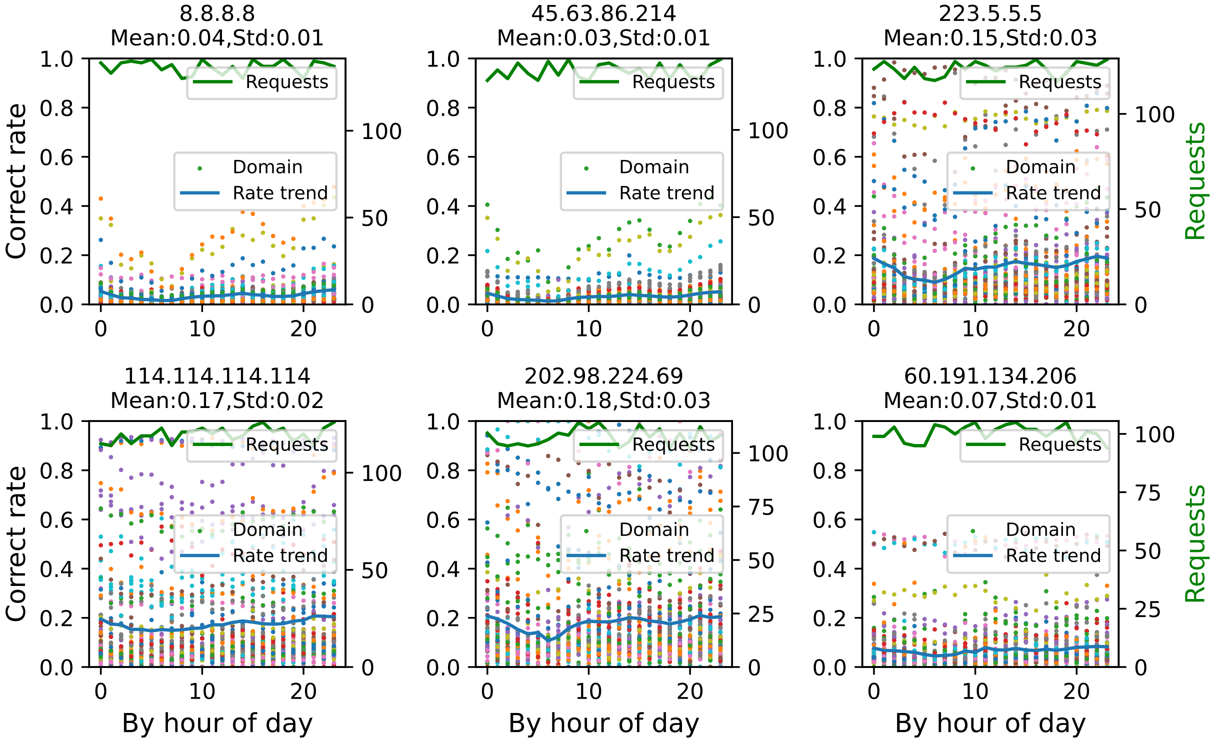


Figure 11. Correct response rate of DNS resolvers answering domain from 00:00 to 24:00.

5. Analyzing Open DNS Resolvers

In this section, we examine the effectiveness of nation-level DNS filtering mechanisms on the more numerous, flexibly configurable, and cryptic open DNS resolvers. We analyze open DNS resolvers with a lifecycle of more than 30 days in Shanghai.

5.1. Overview

Finally, based on the method described in Section 3.3, we obtained 23,669 open DNS resolvers that normally provide resolution services for more than 30 days. However, 5470 of these resolvers only responded to domains under our control and did not respond to any blocked domains, so we ended up analyzing 18,199 open DNS resolvers. In censored networks, open DNS resolvers work similarly to public or ISP DNS resolvers in resolving blocked domains, and their correct rate may vary from domain to domain. However, due to the vast number of open DNS resolvers, there are many blocked domains that can be resolved correctly, as illustrated in Figure 12. From the perspective of the domain name, we outline the following features.

Immunity from DNS filtering. There are a few open DNS resolvers that are immune to country-level DNS filtering methods. We discovered a total of 136 resolvers in Shanghai that could resolve any blocked domain name normally, unaffected by DNS filtering.

DNS filtering impact varies based on domain name. The results show that, similarly to public and ISP DNS resolvers, the extent to which country-level DNS filtering schemes affect open DNS resolvers is substantially related to domain names. Additionally, this phenomenon is more evident when the number of open resolvers is greater (as shown in Figure 12). For example, the correct rate of open DNS resolvers responding to the domain xvideos.es is concentrated between 0.2 and 0.4, whereas the correct rate for xnxx.com is mainly less than 0.2 and the correct rate for google.bj is mostly greater than 0.6.

Furthermore, as illustrated in Figure 13, we counted the number of open DNS resolvers responding to each domain using multiple accurate rate intervals (ACR higher than or equal to 0.5). The results indicate that each domain name has an average of more than

436 open resolvers with an accuracy rate more than 0.9, with a maximum of approximately

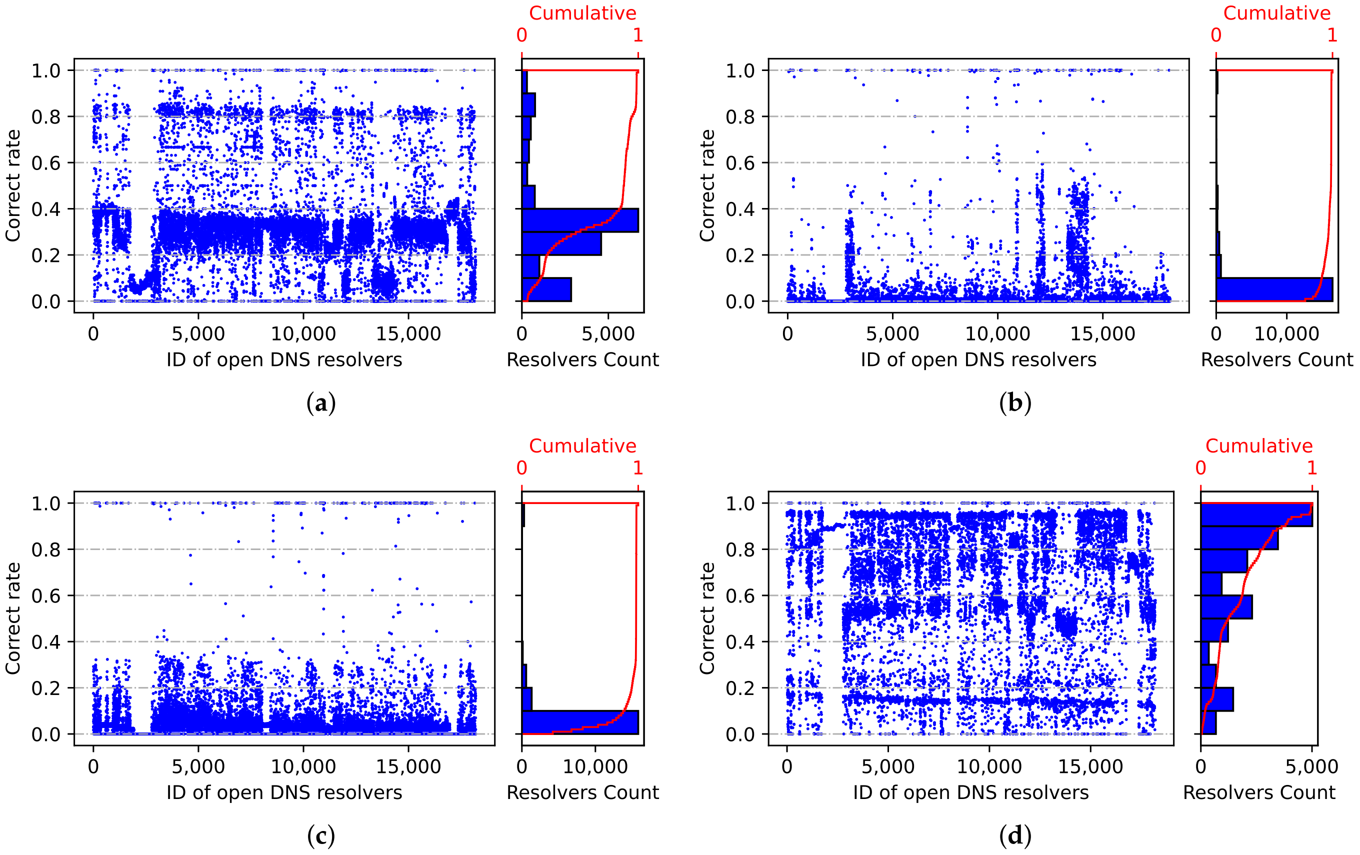


Figure 12. Distribution of correct rate of open DNS resolvers responding to four domains. (**a**) xvideos.es; (**b**) express-vpn.com; (**c**) xnxx.com; (**d**) google.bj.

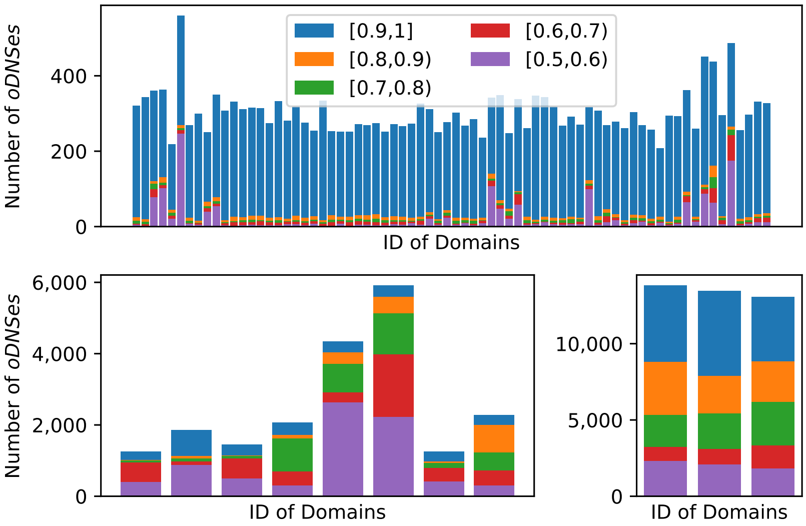


Figure 13. Distribution of the number of Open DNS resolvers responding to each blocked domain name (based on correct response rate).

5.2. Cluster Analysis of Open DNS Resolvers

In analyzing the resolution correct rate of open DNS resolvers, we have found that the resolvers have “cluster” features, for example, the correct response rate of Open DNS resolvers for the domain xvideos.es is concentrated between 0.2 and 0.4. Therefore, we use the algorithm t-SNE (t-Distributed Stochastic Neighbor Embedding) to measure the association between open DNS resolvers. t-SNE is a nonlinear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions [36].

We take the correct rate of each resolver response to 83 sensitive domain names as each resolver's feature vector. Therefore, we ran the t-SNE algorithm to reduce the dimensionality of the 83-dimensional data to two-dimensional, and the clustering of the resolvers is shown in Figure 14. The different colors indicate the value of the maximum correct rate for the open DNS resolver to respond to a particular blocked domain.

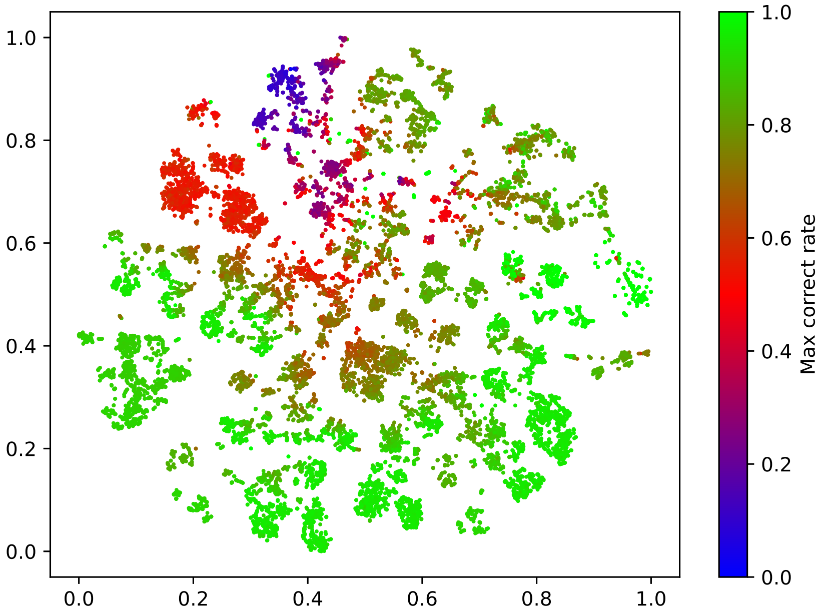


Figure 14. Clustering of open DNS resolvers in Shanghai.

The findings indicate that a large number of open DNS resolvers have comparable accurate rates in response to blocked domain names, resulting in the resolver clustering phenomena seen in Figure 14. This means that the DNS filtering technique has a similar effect on open resolvers with similar behavior (e.g., the same ASN). This feature provides us with an idea to discover the set of open DNS resolvers with specific behaviors (e.g., resolving a specific domain name with a correct rate above 0.9). This feature gives us ideas for quickly discovering oDNS sets with a specified proper rate (e.g., resolving a specific domain name with a correct rate of more than 0.9).

5.3. AS-Level Feature

We examine the ownership and management of a sample of open DNS resolvers. The selection criteria specified that DNS resolvers must be capable of responding to at least ten blocked domains with a correct rate greater than 0.8. Finally, 344 DNS resolvers were discovered. From the WHOIS information for the IP addresses, we derive their ASNs and ASN names. We found that these DNS resolvers are mainly distributed in six ASes (as shown in Table 5). Moreover, the managers of these ASes include ISPs or Internet exchange centers. For example, 80% of the open resolvers come from AS4821, which belongs to the network access points of China Net and China Telecom.

Table 5. ASN and ASN name of the open DNS resolvers.

ASN	# Resolvers	ASN Name
4812	276	CHINANET-SH-AP China Telecom Group, CN
17621	35	CNCGROUP-SH China Unicom Shanghai network, CN
9929	22	CUII CHINA UNICOM Industrial Internet Backbone, CN
24427	6	CNNIC-FREENET Freecomm Corporation, CN
17775	4	STN-CN shanghai science and technology network communication limited company, CN
45061	1	CNNIC-SIN-AP Shanghai Information Network Co., Ltd., CN

5.4. Inconsistent Response Correct Rate over Time

We discovered that, when we analyzed the results of open DNS responses to blocked domains, the correct response rate of various open DNS resolvers became inconsistent over time due to the influence of DNS filtering systems. This is a significant distinction from pDNS and iDNS (described in Section 4.3). As illustrated in Figure 15, between 4 December and 12 January, two open DNS resolvers (123.49.245.50 and 122.144.144.103) exhibit significant variation in their daily response correct rate for the domain porndude.com. This indicates that country-wide DNS filtering strategies impacting open DNS resolvers are more complicated, owing to the invisibility and flexibility of open DNS resolvers, which are more likely to elude filtering effects.

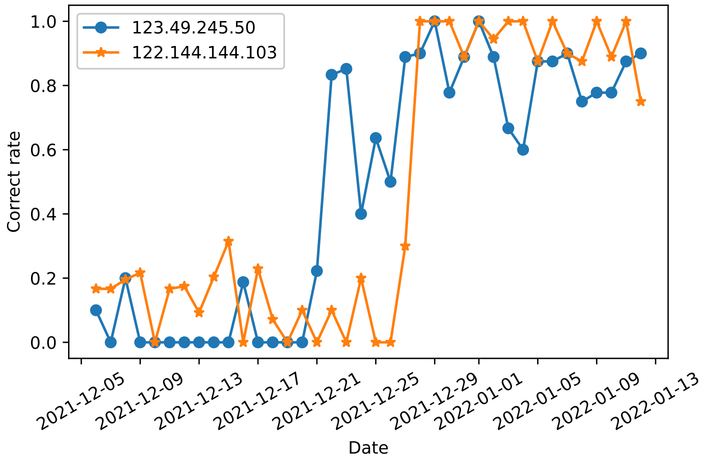


Figure 15. The correct rate of DNS resolvers responding to domain theporndude.com varies over time.

6. Identifying Valid IPs of Blocked Domain

It is well-known that, when DNS requests for blocked domains are sent through DNS filtering devices, the devices respond with forged IP addresses. As a result, clients are unable to effectively access the blocked domain names. In this section, we examine the features of the bogus IP addresses that DNS filtering devices respond to. We then design and implement a system for identifying the valid IP addresses of blocked domains in the filtered network based on these features.

6.1. Changing Forged IP Addresses over Time and Space

First, we extract the forged IP addresses of DNS responses for 83 blocked domain names that pass through the DNS filtering device. That is, we extract the faked IP addresses of responses from four non-Chinese public DNS resolvers at vantage points in Beijing, Shanghai, and Zhangjiakou, respectively, and present the results in Table 6. There are 788 of these bogus IP addresses in all.

Table 6. The number of forged IP addresses contained in DNS resolver responses.

	1.1.1.1	**8.8.8.8**	**122.148.129.45**	**45.63.86.214**	**Total**
Beijing	753	752	761	746	774
Shanghai	756	747	761	746	772
Hebei	752	748	759	750	771
Total	757	753	763	750	788

The results show that the forged IP addresses obtained from responses by various DNS resolvers at different vantage points are essentially the same but contain a few unique IP addresses. As a result, this indicates that obtaining all forged IP addresses returned by the DNS filtering device is challenging, even with additional vantage points and DNS resolvers. Moreover, as prior research [26] has demonstrated, the quantity of forged IP addresses varies, with different DNS filtering devices employing separate forged IP addresses at various periods.

6.2. Shared Forged IP Addresses

In addition, we find that we obtain the same forged IP address of each blocked domain, i.e., the IP addresses are shared by the blocked domains, as shown in Figure 16. Seventy-two domain names shared 60 fake IP addresses. Over 600 bogus IP addresses have been discovered in 71 domains. Additionally, a few IP addresses were shared across a few domains. This could be because the DNS filtering devices recently utilized these forged IP addresses to tamper with the correct IP addresses of the blocked domains. As a result, we can determine the valid IP address for blocked domains in the censorship network by utilizing this feature of the DNS filtering mechanism.

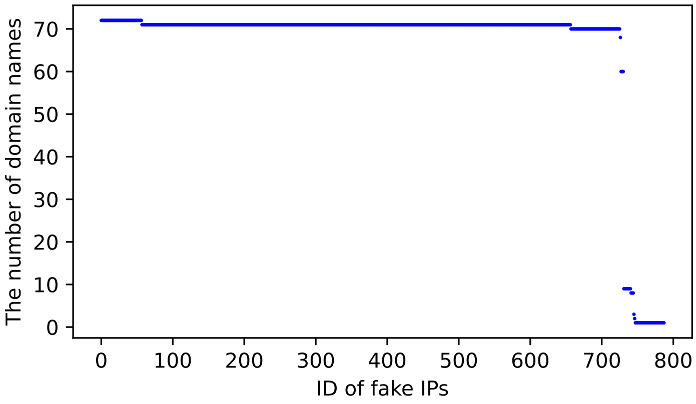


Figure 16. DNS filtering devices answer the same forged IP to different blocked domain names.

6.3. Low Frequency of Forged IP Addresses

We learned in the previous sections that most DNS resolvers reply to blocked domains with a correct rate of less than 0.1 and that a domain may contain numerous fake IP addresses. However, for a domain name in a DNS filtering network, is there any pattern to the frequency with which each forged IP and correct IP emerge during a specific period? As a result, in this subsection, we compare the frequency with which the valid IP address and each forged IP address appear in a given period.

We randomly selected 10,000 response records for requesting the domain theporndude.com (ACR value below 0.1) from five non-Chinese public DNS resolvers at the Beijing vantage point, and counted the frequency of correct IPs and incorrect IPs, respectively, as shown in Figure 17.

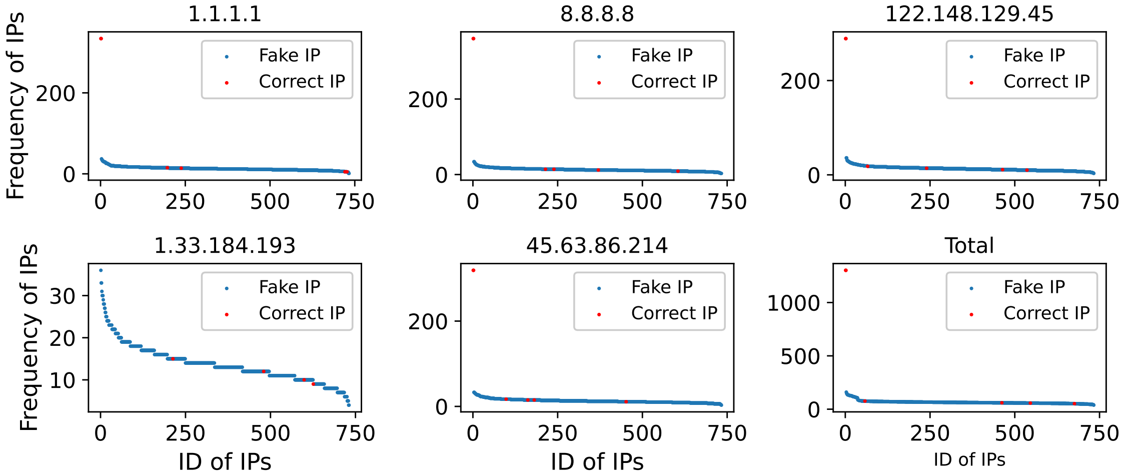


Figure 17. The frequency with which non-Chinese pDNS respond to correct and forged IPs for the domain theporndude.com.

Except for the DNS resolver 1.33.184.193, the findings indicate that the most commonly occurring IP address is the correct one, which appears 100 times more frequently than the other forged IP addresses. We believe this is due to the DNS device failing sometimes, resulting in proper replies to domain requests. Given that the number of correct IPs for a domain name is significantly smaller than the number of forged IPs (almost 800), the frequency of valid IPs is considerably larger than the frequency of forged IPs.

Then, we examined the frequency with which public and ISP DNS resolvers responded to the correct IPs and forged IPs of blocked domain names (as shown in Figures 18–20), and discovered that the situation is comparable to that of non-Chinese public DNS resolvers.

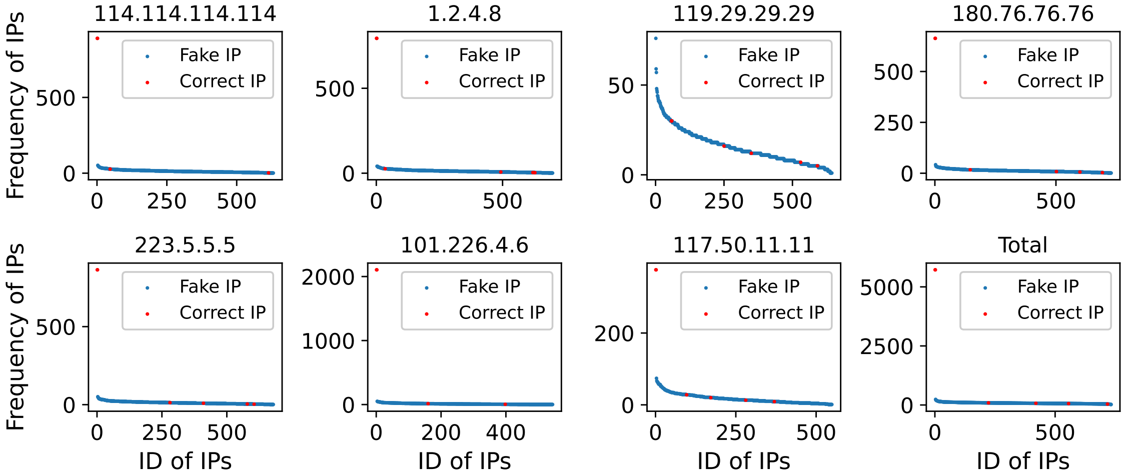


Figure 18. The frequency with which Chinese pDNS respond to correct and forged IPs for the domain theporndude.com.

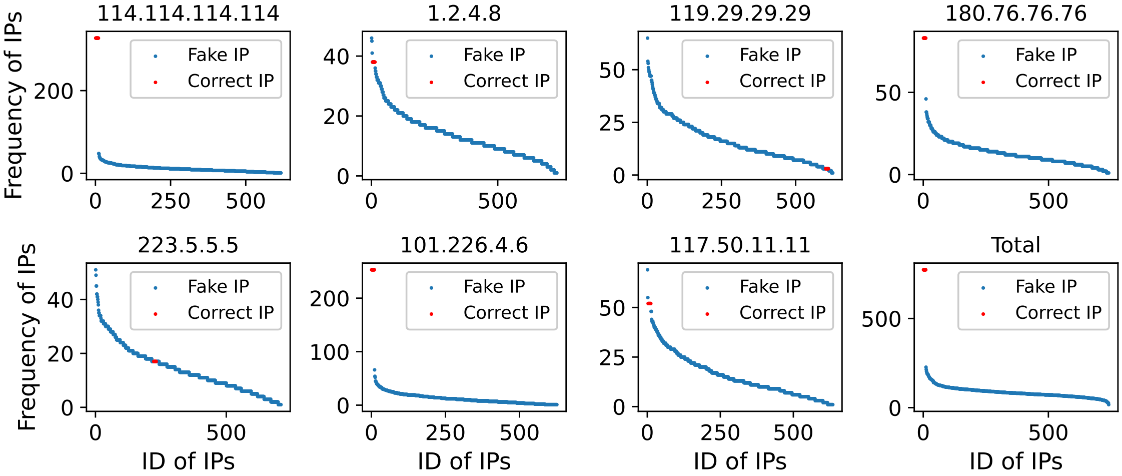


Figure 19. The frequency with which Chinese pDNS respond to correct and forged IPs for the domain xvideos.com.

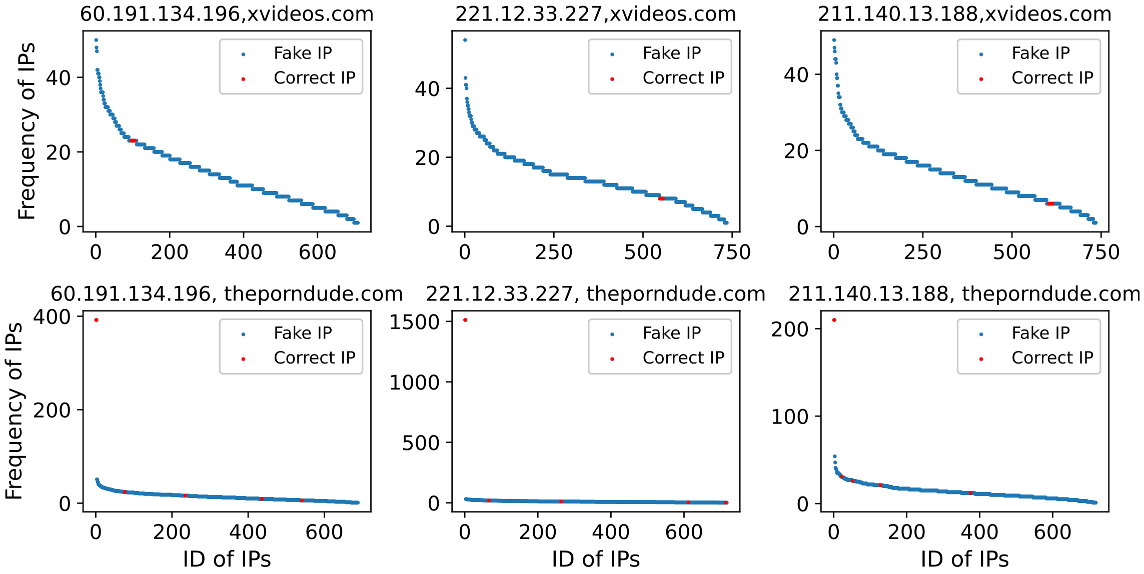


Figure 20. The frequency with which iDNS respond to correct and forged IPs for the domain theporndude.com and xvideos.com.

We estimate that DNS resolvers that do not fit this pattern result from the deployment of DNS filtering devices. Although the DNS filtering system is administered centrally, the DNS devices are dispersed across multiple network paths. Specific network pathways feature more capable DNS devices that are less likely to provide correct IP addresses. When a client or a recursive DNS server sends a DNS query to the device, the device responds with a fake response.

6.4. Implementation and Evaluation of a System for Identifying Domain Valid IP

We develop and implement a system for identifying the valid IP address for blocked domain names in the nation-level censored networks based on the features of faked IPs used by DNS filtering devices in responses.

6.4.1. Principles and Advantages

The principle of the identification system is based on the behavioral characteristics of the fake IPs used for response by DNS filtering and our practical experience in the process of analyzing the data.

Principle 1: Multiple domains frequently share forged IPs in the response results of DNS resolvers affected by DNS filtering devices, which is one of the primary features, because, in general, valid IP addresses are usually shared and used by very few domains concurrently, especially the blocked domains with high traffic, whose IP addresses are not shared with other domains in order to ensure the quality of service. Of course, we should consider the exceptions. For example, the Google search engine globally has many different top-level domain names (such as google.com, google.co.jp, and google.es), whose IP addresses can be used by more than one domain at the same time.

Principle 2: As previously stated, some DNS resolvers respond to blocked domain names with a significantly higher frequency of valid IP addresses than forged IPs. Thus, the frequency of legitimate IP addresses for a domain name appears to be the "anomaly" data among all IP frequencies. Thus, the valid IP address may be discovered using the anomaly detection technique. While we can only identify the IP address with the highest frequency as the correct one, this does not provide a high recall rate of legitimate IP addresses for domain names. Therefore, this paper adopts the IQR (Inter-Quartile Range) method of outlier detection to obtain as many correct IP addresses as possible.

Principle 3: During our analysis of the data, we find that many of the valid IP addresses of many domains belong to the same IP segment. This is different from other fake IPs that appear in the DNS records of domains at the same time. These forged IPs may belong to the same company [26] (e.g., Facebook or Twitter) but are not in the same segment.

Finally, we design and implement a valid IP identification scheme for blocked domain names based on the three principles outlined above, as illustrated in Figure 21. Additionally, the system circumvents the dynamic nature of the DNS filtering mechanism, relying solely on the behavioral characteristics of forged IPs to obtain correct IP addresses for blocked domain names in the long term.

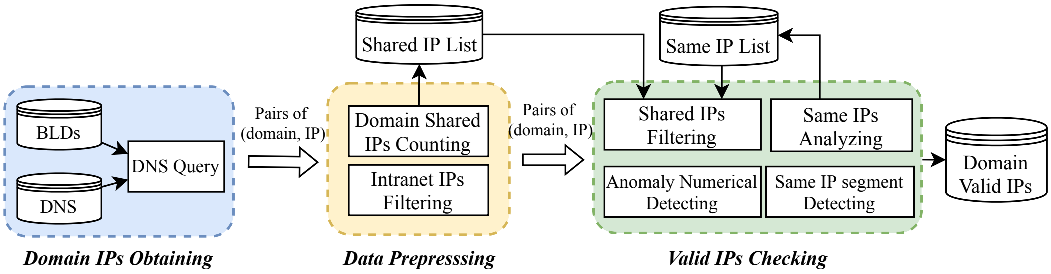


Figure 21. The system to identify valid IPs for blocked domain names.

6.4.2. Design and Implementation

As shown in Figure 21, the correct IP identification system consists of three main stages: *Domain IPs Obtaining*, *Data Prepressing*, and *IPs Checking*.

- **Domain IP Obtaining Stage.** This stage's primary objective is to obtain the blocked domains' IP addresses from a large number of DNS resolvers. The interval between queries for a specific domain name to a specific DNS resolver must be greater than the domain name's TTL (Time To Live). Additionally, the number of domains (e.g., 83 domains are used in this paper) in the blocked list should be sufficient to ensure the system's high accuracy rate. Finally, as previously stated, Chinese public DNS resolvers have a higher possibility of replying to the correct IP address, and the more resolvers there are, the more IP addresses can be retrieved.
- **Data Prepressing Stage.** This stage preprocesses the set of domain IP addresses received in the previous stage. To begin, we filter the response records for intranet IP

addresses, such as 127.0.0.1. Second, we generate a shared IP list of domains based on each domain's answer IP addresses for the subsequent stage of faked IP filtering.

- **Valid IPs Checking Stage.** Based on the three principles introduced above, this stage establishes forged IP filtering and detection rules to discover the correct IP address of the domain name. The *Same IPs Analyzing* module highlighted then configures domains that do have the same IP, and works with the *Shared IPs Filtering* module to filter forged IPs together.

6.4.3. Evaluation

We evaluate our system's performance in finding valid IP addresses for blocked domain names using two metrics: CORRECTNESS and COVERAGE. The term "CORRECTNESS" refers to the proportion of valid IP addresses to all obtained IP addresses. Simultaneously, the system must obtain as many valid IP addresses for the domain name as possible while maintaining a high rate of correctness. The "COVERAGE" refers to how many domains the system can support to obtain the correct IP addresses.

Finally, we assess our method using the DNS records obtained in this paper for the blocked domains. That is, the system utilizes 14 DNS resolvers (four non-Chinese pDNS, seven Chinese pDNS, and three iDNS) to retrieve valid IP addresses for 83 blocked domains. We randomly picked a predetermined number of DNS records for each domain name obtained, repeated the selection ten times, and calculated the average CORRECTNESS and COVERAGE values to evaluate the system's performance, as shown in Figure 22.

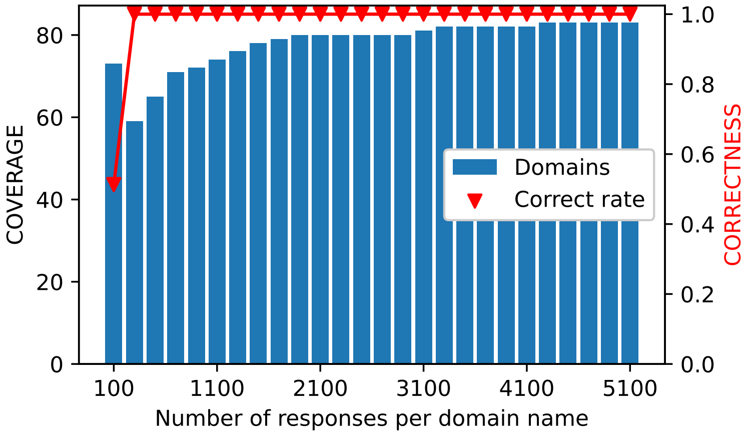


Figure 22. CORRECTNESS and COVERAGE of the valid IP's identification systems.

As illustrated in Figure 22, when fewer DNS records for domain names are received, the system's COVERAGE rate increases, but its CORRECTNESS rate decreases. As the quantity of DNS records grows, the system's CORRECTNESS rate improves, while the COVERAGE rate gradually climbs after falling. When the system obtains 4300 records for each domain, it achieves 100 percent coverage and accuracy. This result indicates that our system has a good performance in identifying blocked domain valid IPs in censored networks.

7. Future Work

Like other uses of Internet measurements to analyze data, our system has a few inherent limitations. Because there are over 31 provinces and 650 cities in China, even though our system covers a large number of different types of DNS resolvers, multiple cities, and ISPs in China, our vantage points are not granular enough to measure the effectiveness of DNS filtering mechanisms for each region—in particular, whether the DNS resolvers provided by the ISPs in each region correctly or occasionally correctly resolve blocked domain names.

Therefore, based on the above limitations, the future work concerns scaling up open resolvers, blocked domains, and regions. Firstly, we should scale up the open resolver to the global level and study how they resolve different types of domains (e.g., blocked domains, malicious domains, and popular domains). In particular, we are trying to measure as complete a list of blocked domains in China as possible to provide a reference for domain owners and researchers. Secondly, in China, we would like to increase the number of vantage points to cover more regions and ISPs, especially cellular networks. Finally, we would like to conduct an in-depth analysis of other Internet censorship techniques such as TCP reset and IP blocking. This allows us to have a more comprehensive view of Internet censorship.

8. Discussion and Conclusions

For the first time, we quantitatively evaluate how country-level DNS filtering mechanisms affect the response of three types of DNS resolvers (public, ISP, and open DNS resolvers) to blocked domain names over space and time. Based on the 1.7 billion data acquired in 40 days, we discovered that specific DNS resolvers are unaffected by DNS filtering devices and can respond with the correct IPs of blocked domains to clients. In summary, the effectiveness of national-level DNS filtering systems is directly tied to three factors: the DNS resolver, client location, and blocked domain. This indicates that the DNS filtering mechanism has a more complex and variable impact on DNS resolvers in censored networks, rather than having only a single effect as expected, i.e., all resolvers respond to all blocked domain names with the wrong IP addresses.

We expose features of open DNS resolvers not found in previous studies in the censored networks. In particular, for many open DNS resolvers on the Internet, their flexible configuration and stealthy nature lead to their more complex influence by DNS filtering mechanisms. For example, some open resolvers can completely evade the effects of DNS filtering and can correctly resolve all blocked domains. On the other hand, some resolvers can only correctly resolve specific domain names, such as pornhub.com. When analyzing the correct rate of resolvers responding to each blocked domain name, we found that these resolvers have a “cluster” feature, which indicates that a large number of open resolvers would affect the effectiveness of the DNS filtering mechanism.

We discover a weakness in the national-level DNS filtering mechanism, which one can exploit to identify the correct IP addresses of blocked domains in the censored networks. Based on the features (e.g., changing, shared, and low frequency) of the forged IP addresses used by DNS filtering devices to respond, we then propose and implement a system for identifying the correct IP addresses of blocked domains in censored networks. This system circumvents the dynamic nature of the DNS filtering mechanism, relying solely on the behavioral characteristics of forged IPs to obtain correct IP addresses for blocked domain names quickly and accurately.

It is emphatically stated that Internet censorship (e.g., DNS filtering, HTTP censorship) is a long-standing and controversial topic, and the research in this paper only seeks to analyze the entities involved in DNS filtering from a technical perspective. We can only hope that the research in this paper can give some advice to DNS filtering administrators and blocked domain owners in a neutral way.

- **DNS filtering administrators.** With the national-level DNS filtering mechanism in China, some DNS resolvers still correctly resolve blocked domains, especially with the large number of open DNS resolvers that exist on the Internet. In addition, the distinctive features of forged IP addresses used by DNS filtering devices to respond lead to the fact that we can use these features to obtain the correct IP address of blocked domains in the censored networks. As a result, if administrators want to block censored domains effectively, they need to change the response characteristics of the filtering mechanism.
- **Blocked domain name owners.** Domain owners, especially those with blocked domains, can realize that, even in the DNS censorship networks, there are still some

DNS resolvers that can resolve their domains correctly. Alternatively, domain owners can build specific functional DNS resolvers to resolve their domain names in the censored networks.

Author Contributions: Conceptualization, Y.C. and Z.Z.; methodology, Y.C. and Y.L.; software, Y.C.; validation, C.L., Y.C. and Y.L.; formal analysis, N.L.; investigation, Y.C.; resources, C.L.; data curation, C.L.; writing—original draft preparation, Y.C. and Y.L.; writing—review and editing, Y.C. and N.L.; visualization, Y.C. and N.L.; project administration, Z.Z. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province (Grant No. ZR2020KF009) and the Young Teacher Development Fund of Harbin Institute of Technology (Grant No. IDGA10002081).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DNS	Domain name system
ISP	Internet service provider
oDNS	Open DNS resolvers
iDNS	ISP DNS resolvers
pDNS	Public DNS resolvers
BLDs	Blocklist domains
VP	Vantage point
ACR	Absolute correct rate of responses
NCR	Number of correct responses
TNR	Total number of responses
TLS	Transport layer security
DoT	DNS-over-TLS
DHT	DNS-over-HTTPS
ESNI	Encrypted server name indication
IQR	Inter-quartile range
t-SNE	t-distributed stochastic neighbor embedding
TTL	Time to live

References

1. Niaki, A.A.; Hoang, N.P.; Gill, P.; Houmansadr, A. Triplet Censors: Demystifying Great Firewall's DNS Censorship Behavior. In Proceedings of the 10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20), USENIX Association, Boston, MA, USA, 12–14 August 2020.
2. Ensafi, R.; Winter, P.; Mueen, A.; Crandall, J.R. Analyzing the Great Firewall of China Over Space and Time. *Proc. Priv. Enhancing Technol.* **2015**, *2015*, 61–76. [CrossRef]
3. Jin, L.; Hao, S.; Wang, H.; Cotton, C. Understanding the Practices of Global Censorship through Accurate, End-to-End Measurements. *Proc. Acm Meas. Anal. Comput. Syst.* **2021**, *5*, 1–25. [CrossRef]
4. Liu, B.; Lu, C.; Duan, H.; Liu, Y.; Li, Z.; Hao, S.; Yang, M. Who Is Answering My Queries: Understanding and Characterizing Interception of the DNS Resolution Path. In Proceedings of the 27th USENIX Security Symposium (USENIX Security 18), USENIX Association, Baltimore, MD, USA, 15–17 August 2018; pp. 1113–1128.
5. Park, J.; Khormali, A.; Mohaisen, M.; Mohaisen, A. Where Are You Taking Me? Behavioral Analysis of Open DNS Resolvers. In Proceedings of the 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Portland, OR, USA, 24–27 June 2019; pp. 493–504. [CrossRef]
6. What Is DNS & How Does DNS Filtering Work? Available online: https://www.webroot.com/us/en/resources/glossary/what-is-dns-filtering (accessed on 8 March 2022).
7. Top 8 DNS/Content Filtering for Home to Protect Family and Kids from Cyber Threats. Available online: https://geekflare.com/dns-content-filtering-software (accessed on 8 March 2022).
8. Crandall, J.R.; Zinn, D.; Byrd, M.; Barr, E.T.; East, R. ConceptDoppler: A Weather Tracker for Internet Censorship. In Proceedings of the 14th ACM Conference on Computer and Communications Security-CCS 07, Alexandria, VA, USA, 29 October–2 November 2007; pp. 352–365. [CrossRef]

9. Xu, X.; Mao, Z.M.; Halderman, J.A. Internet Censorship in China: Where Does the Filtering Occur? In *Passive and Active Measurement*; Spring, N., Riley, G.F., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 133–142._14. [CrossRef]
10. Wander, M.; Boelmann, C.; Schwittmann, L.; Weis, T. Measurement of Globally Visible DNS Injection. *IEEE Access* **2014**, *2*, 526–536. [CrossRef]
11. Anonymous. The collateral damage of internet censorship by DNS injection. *ACM SIGCOMM Comput. Commun. Rev.* **2012**, *42*, 21–27. [CrossRef]
12. Hoang, N.P.; Doreen, S.; Polychronakis, M. Measuring I2P Censorship at a Global Scale. In Proceedings of the 9th USENIX Workshop on Free and Open Communications on the Internet (FOCI 19), Santa Clara, CA, USA, 13 August 2019.
13. Lowe, G.; Winters, P.; Marcus, M.L. *The Great DNS Wall of China*; Technical Report; New York University: New York, NY, USA, 2007.
14. Ensafi, R.; Winter, P.; Mueen, A.; Crandall, J.R. Large-scale Spatiotemporal Characterization of Inconsistencies in the World's Largest Firewall. *arXiv* **2014**, arXiv:1410.0735.
15. Wright, J. Regional variation in Chinese internet filtering. *Inf. Commun. Soc.* **2013**, *17*, 121–141. [CrossRef]
16. Duan, H.; Weaver, N.; Zhao, Z.; Hu, M.; Liang, J.; Jiang, J.; Li, K.; Paxson, V. Hold-on: Protecting against on-path DNS poisoning. In Proceedings of the Securing and Trusting Internet Names (SATIN), Teddington, London, UK, 22–23 March 2012.
17. Niaki, A.A.; Cho, S.; Weinberg, Z.; Hoang, N.P.; Razaghpanah, A.; Christin, N.; Gill, P. A Global, Longitudinal Internet Censorship Measurement Platform. In Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP), Virtual, 21 May 2020. [CrossRef]
18. Verkamp, J.P.; Gupta, M. Inferring Mechanics of Web Censorship Around the World. In Proceedings of the 2nd USENIX Workshop on Free and Open Communications on the Internet (FOCI 12), Bellevue, WA, USA, 6 August 2012.
19. Schomp, K.; Callahan, T.; Rabinovich, M.; Allman, M. On measuring the client-side DNS infrastructure. In Proceedings of the 2013 Conference on Internet Measurement Conference, Barcelona, Spain, 23–25 October 2013; pp. 77–90. [CrossRef]
20. Gebhart, G.; Kohno, T. Internet Censorship in Thailand: User Practices and Potential Threats. In Proceedings of the 2017 IEEE European Symposium on Security and Privacy (EuroS&P), Paris, France, 26–28 April 2017; pp. 417–432. [CrossRef]
21. Jones, B.; Lee, T.W.; Feamster, N.; Gill, P. Automated Detection and Fingerprinting of Censorship Block Pages. In Proceedings of the 2014 Conference on Internet Measurement Conference, Vancouver, BC, Canada, 5–7 November 2014; pp. 299–304. [CrossRef]
22. Raman, R.S.; Stoll, A.; Dalek, J.; Ramesh, R.; Scott, W.; Ensafi, R. Measuring the Deployment of Network Censorship Filters at Global Scale. In Proceedings of the 2020 Network and Distributed System Security Symposium, San Diego, CA, USA, 23–26 February 2020. [CrossRef]
23. What Is DNS Filtering? Available online: https://www.cloudflare.com/zh-cn/learning/access-management/what-is-dns-filtering (accessed on 8 March 2022).
24. DNSFilter. Powered DNS Threat Protection & Content Filtering. Available online: https://www.dnsfilter.com (accessed on 8 March 2022).
25. OpenDNS. Home Internet Security. Available online: https://www.opendns.com/home-internet-security (accessed on 8 March 2022).
26. Hoang, N.P.; Niaki, A.A.; Dalek, J.; Knockel, J.; Lin, P.; Marczak, B.; Crete-Nishihata, M.; Gill, P.; Polychronakis, M. How Great is the Great Firewall? Measuring China's DNS Censorship. In Proceedings of the 30th USENIX Security Symposium (USENIX Security 21), Virtual Event, 11–13 August 2021; pp. 3381–3398.
27. A Comprehensive Picture of the Great Firewall's DNS Censorship. In Proceedings of the 4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14), San Diego, CA, USA, 18 August 2014.
28. Pearce, P.; Jones, B.; Li, F.; Ensafi, R.; Feamster, N.; Weaver, N.; Paxson, V. Global Measurement of DNS Manipulation. In Proceedings of the 26th USENIX Security Symposium (USENIX Security 17), Vancouver, BC, USA, 16–18 August 2017; pp. 307–323.
29. Padmanabhan, R.; Filastò, A.; Xynou, M.; Raman, R.S.; Middleton, K.; Zhang, M.; Dainotti, A. A multi-perspective view of Internet censorship in Myanmar. In Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet, Virtual Event, 23–27 August 2021.
30. Hoang, N.P.; Polychronakis, M.; Gill, P. Measuring the Accessibility of Domain Name Encryption and Its Impact on Internet Filtering. In *International Conference on Passive and Active Network Measurement*; Springer: Cham, Switzerland, 2022; pp. 518–536.
31. Chai, Z.; Ghafari, A.; Houmansadr, A. On the Importance of Encrypted-SNI(ESNI) to Censorship Circumvention. In Proceedings of the 9th USENIX Workshop on Free and Open Communications on the Internet (FOCI 19), Santa Clara, CA, USA, 13 August 2019.
32. VanderSloot, B.; McDonald, A.; Scott, W.; Halderman, J.A.; Ensafi, R. Quack: Scalable Remote Measurement of Application-Layer Censorship. In Proceedings of the 27th USENIX Security Symposium (USENIX Security 18), Baltimore, MD, USA, 15–17 August 2018; pp. 187–202.
33. Kührer, M.; Hupperich, T.; Bushart, J.; Rossow, C.; Holz, T. Going Wild. In Proceedings of the 2015 Internet Measurement Conference, Tokyo, Japan, 28–30 October 2015; pp. 355–368. [CrossRef]
34. Jin, L.; Hao, S.; Wang, H.; Cotton, C. Understanding the Impact of Encrypted DNS on Internet Censorship. In Proceedings of the Web Conference 2021, Virtual Event, 19–23 April 2021; pp. 484–495. [CrossRef]
35. Randall, A.; Liu, E.; Akiwate, G.; Padmanabhan, R.; Voelker, G.M.; Savage, S.; Schulman, A. Trufflehunter. In Proceedings of the ACM Internet Measurement Conference, Virtual Event, 27–29 October 2020; pp. 50–64. [CrossRef]
36. Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.