2021

Understanding the Impact of Encrypted DNS on Internet Censorship

Lin Jin

Shuai Hao
Old Dominion University

Haining Wang

Chase Cotton

Original Publication Citation

Jin, L., Hao, S., Wang, H., & Cotton, C. (2021). Understanding the impact of encrypted DNS on internet censorship. *In Proceedings of the Web Conference 2021 (WWW'21) April 19-23, 2021, Ljubljana, Slovenia*. ACM, New York, NY, USA, 484-495. https://doi.org/10.1145/3442381.3450084

Understanding the Impact of Encrypted DNS on Internet Censorship

Lin Jin
University of Delaware
Newark, Delaware, USA
linjin@udel.edu

Shuai Hao
Old Dominion University
Norfolk, Virginia, USA
shao@odu.edu

Haining Wang
Virginia Tech
Arlington, Virginia, USA
hnw@vt.edu

Chase Cotton
University of Delaware
Newark, Delaware, USA
ccotton@udel.edu

ABSTRACT

DNS traffic is transmitted in plaintext, resulting in privacy leakage. To combat this problem, secure protocols have been used to encrypt DNS messages. Existing studies have investigated the performance overhead and privacy benefits of encrypted DNS communications, yet little has been done from the perspective of censorship. In this paper, we study the impact of the encrypted DNS on Internet censorship in two aspects. On one hand, we explore the severity of DNS manipulation, which could be leveraged for Internet censorship, given the use of encrypted DNS resolvers. In particular, we perform 7.4 million DNS lookup measurements on 3,813 DoT and 75 DoH resolvers and identify that 1.66% of DoT responses and 1.42% of DoH responses undergo DNS manipulation. More importantly, we observe that more than two-thirds of the DoT and DoH resolvers manipulate DNS responses from at least one domain, indicating that the DNS manipulation is prevalent in encrypted DNS, which can be further exploited for enhancing Internet censorship. On the other hand, we evaluate the effectiveness of using encrypted DNS resolvers for censorship circumvention. Specifically, we first discover those vantage points that involve DNS manipulation through on-path devices, and then we apply encrypted DNS resolvers at these vantage points to access the censored domains. We reveal that 37% of the domains are accessible from the vantage points in China, but none of the domains is accessible from the vantage points in Iran, indicating that the censorship circumvention of using encrypted DNS resolvers varies from country to country. Moreover, for a vantage point, using a different encrypted DNS resolver does not lead to a noticeable difference in accessing the censored domains.

KEYWORDS

DNS-over-TLS, DNS-over-HTTPS, DNS Manipulation, Internet Censorship

ACM Reference Format:
Lin Jin, Shuai Hao, Haining Wang, and Chase Cotton. 2021. Understanding the Impact of Encrypted DNS on Internet Censorship. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442381.3450084

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450084

1 INTRODUCTION

The Domain Name System (DNS) provides important mappings between domain names and their numerical IP addresses to direct users to Internet services. As a fundamental component of the Internet, DNS was designed as an unencrypted protocol. However, this allows eavesdroppers to sniff the domain that a user is going to visit, raising a privacy concern. In order to mitigate this privacy issue, secure protocols, such as DNS-over-TLS (DoT) [29] and DNS-over-HTTPS (DoH) [25], have been proposed to encrypt DNS traffic, and DNS service providers, such as Google and Cloudflare, have gradually supported these protocols on their resolvers.

Considerable research efforts [12, 24, 28, 35, 47, 48] have been devoted to assessing the performance overhead and privacy benefits of using encrypted DNS, yet none of them has considered the impact of encrypted DNS on Internet censorship. Complementary to these prior studies, we provide the first comprehensive measurement study to investigate the impact of encrypted DNS on Internet censorship from two-fold: (1) the DNS manipulation occurred on the use of encrypted DNS resolvers and (2) the effectiveness of using encrypted DNS resolvers for censorship circumvention.

In this paper, we first explore the possibility of facilitating Internet censorship through DNS manipulation when adopting encrypted DNS resolvers. Concretely, we collect DoT and DoH resolvers by actively scanning the Internet and passively analyzing the public datasets. Then, we compile a list of sensitive and popular domains and resolve the domains at the DoT and DoH resolvers to identify the occurrence of DNS manipulation. Overall, we conduct 7.4 million DNS lookup measurements on 3,813 DoT and 75 DoH resolvers and identify that 1.66% of DoT responses and 1.42% of DoH responses are manipulated. In addition, we discover that more than two-thirds of the encrypted DNS resolvers manipulate at least one domain's DNS response, showing that the DNS manipulation in the encrypted DNS is even more prevalent than that in the traditional DNS [42], where only 11% of the resolvers have been identified to manipulate DNS responses. Also, the resolvers of a provider could behave very differently in terms of the number and category of the censored domains, which implies the adoption of different policies with an encrypted DNS provider.

On the other hand, encrypted DNS resolvers have the potentials to help end-users circumvent Internet censorship, since encrypted DNS traffic cannot be easily manipulated by on-path censorship devices. We evaluate the effectiveness of using encrypted DNS resolvers for censorship circumvention. To do so, we first recruit geographically distributed vantage points and detect the occurrence of DNS manipulation when using unencrypted DNS. Here, we only focus on the DNS manipulation conducted by an on-path censorship device as the DNS manipulation conducted by a resolver

can be easily bypassed by switching to another traditional DNS resolver. After that, we use encrypted DNS resolvers to perform DNS resolution on the censored domains from the corresponding vantage points and further verify the accessibility of the censored domains. In total, we identify vantage points in five countries where DNS manipulation is conducted by on-path censorship devices. By using encrypted DNS resolvers from those vantage points, we find that the effectiveness of encrypted DNS resolvers to circumvent Internet censorship varies by country. For example, with the use of encrypted DNS resolvers, all the vantage points in China are able to access approximately 37% of the censored domains, but none of the censored domains can be accessed from the vantage points in Iran. Additionally, we observe no noticeable differences in accessing the censored domains from different resolvers given a same vantage point. Furthermore, we identify that those domains remain inaccessible because they also suffer other types of censorship, such as IP-based blocking, HTTP-based blocking, and SNI-based blocking. Note that considering the ethical concerns, we carefully design these experiments to reduce potential risks on the vantage points (see details in Section 5.2).

The remainder of this paper is organized as follows. Section 2 introduces the background of encrypted DNS and Internet censorship. In Section 3, we describe our approach to evaluating the occurrence of DNS manipulation given the use of encrypted DNS and present the analysis results. The methodology and results on censorship circumvention with encrypted DNS resolvers are detailed in Section 4. A further discussion on our study is presented in Section 5. We survey the related work in Section 6, and finally, we conclude the paper in Section 7.

2 BACKGROUND

In this section, we introduce the background of encrypted DNS and censorship techniques that have been widely employed to prevent users from visiting undesired websites.

2.1 Encrypted DNS

DNS is a distributed and hierarchical database that hosts resource records of Internet services. As shown in Figure 1, to conduct a DNS resolution, a client first issues a DNS query to a recursive resolver that will then traverse the DNS hierarchical tree and return the requested resource records back to the client. DNS was originally designed as an unencrypted protocol, raising serious privacy concerns and censorship issues. More specifically, an eavesdropper who monitors the traffic on the wire can view and collect the client's browsing activities, or manipulate the DNS responses to control the information access if the requested domain is undesired.

Encrypted DNS (*e.g.*, DoT and DoH)[1] was then proposed to encrypt the DNS communications so as to address the privacy leakage and DNS manipulation. However, we notice that the current development of encrypted DNS only secures the communication channel between a client and resolvers, and the DNS messages between the resolvers and nameservers remain unencrypted. As a result, DNS manipulation between resolvers and nameservers is still possible. In addition, as a resolver itself may have a list of undesired domains, the use of encrypted DNS resolvers may also suffer from DNS manipulation. On the other hand, encrypted DNS prevents the DNS manipulation between clients and resolvers, so it conceivably improves the client's censorship-resistance.

[1]Other than DoT and DoH, encrypted DNS protocols have other proposals such as DNSCrypt [17]. In this study, we focus on DoT and DoH as they have been well-standardized [25, 29] and well-supported [15, 22] by the industry.

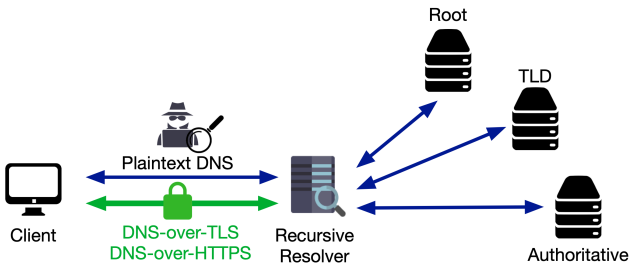

Figure 1: Illustration of DNS and Encrypted DNS. The blue lines plot unencrypted DNS traffic and the green line depicts encrypted DNS traffic.

2.2 Internet Censorship

Internet censorship has been widely witnessed for decades and steadily increasing in recent years. To restrict the access to certain information on the Internet, censor devices examine the network traffic to block undesired contents of domains that are prohibited. In doing so, censors usually check the destination IP addresses and the domain names that are normally transmitted in plaintext. Here, we briefly describe different types of censorship techniques that are widely used to block access to certain Internet resources.

2.2.1 DNS Manipulation. Conducting DNS resolution is the very first step for a client to access a domain. As we mentioned above, the plaintext DNS is vulnerable to DNS manipulation since censors on the path can accurately learn which domain a client is going to visit and block DNS queries if the requested domain is prohibited.

In our study, we mainly consider DNS manipulation in the following two scenarios. First, resolvers receive DNS queries from their customers, so they would be able to manipulate the DNS responses if the DNS queries contain their prohibited domains. However, as those resolvers can only manipulate the DNS responses replied by themselves, it is fairly easy to bypass such DNS manipulation by switching to a different DNS resolver that does not manipulate DNS responses, such as Google Public DNS. Second, an on-path censorship device can actively inspect the DNS messages on the wire and manipulate the DNS responses if needed. A traditional DNS resolver cannot avoid such a DNS manipulation, but an encrypted DNS resolver (beyond the censor's scope) is immune to the DNS manipulation as their DNS messages are encrypted.

2.2.2 HTTP-based Censorship. An HTTP request contains a `Host` header that displays the domain name of a website. The `Host` header was introduced to inform the web server which domain the client attempts to connect to so that the web server could provide the requested content when the web server hosts multiple domains. Like the DNS protocol, the HTTP protocol is also unencrypted and thus a censor can block the traffic based on the domain name in the `Host` header.

2.2.3 SNI-based Censorship. HTTPS encrypts the HTTP messages using TLS so that censors are unable to examine the `Host` header in the HTTP requests. However, the domain name of a website is still exposed during the process of the TLS handshake. In particular, a web server may host multiple domains and different domains may associate with different certificates. Thus, in order to present a correct certificate, a client needs to inform the web server which domain it connects to and the domain name is embedded in the SNI extension of the `Client Hello` message, which is sent in plaintext. As a result, a censor can block the HTTPS traffic based on the domain name in the SNI extension. To bypass the SNI-based censorship, encrypted SNI has been proposed and deployed in Cloudflare's networks [7, 14].

2.2.4 IP-based Censorship. Besides blocking traffic based on domain names described above, a censor may simply block users' packets based on IP addresses. A censor can compile a list of IP addresses that are hosting undesired content and block traffic to and from those IP addresses. However, the IP-based censorship could result in collateral damage [19, 26, 38, 52]. For example, IP addresses are shared resources in CDNs, where contents of multiple domains could be delivered from the same IP address. As a result, blocking an IP address of a CDN's edge server would also block the access to other innocent domains. Such collateral damage is significant since CDNs are widely adopted by many web services [30].

3 CENSORSHIP OF ENCRYPTED DNS

In this section, we investigate DNS manipulation given the use of encrypted DNS resolvers. The basic idea is to first collect a list of DoT/DoH resolvers and compile a list of test domains that are likely to be censored by authorities. Then, we resolve the domains with the DoT/DoH resolvers and determine whether a DNS response is valid.[2]

3.1 Discovering DoT and DoH Resolvers

A previous study [35] has already discovered more than one thousand of DoT resolvers and tens of DoH resolvers, and a public resolver list [16] is also available. However, it is unclear whether those resolvers are still in service. Besides, since the concerns of DNS privacy leakage and manipulation have attracted more attention in the industry recently, more DoT and DoH resolvers have been deployed. Thus, we first perform a large-scale scanning to discover and validate the operational DoT and DoH resolvers.

3.1.1 DoT resolvers. By default, DoT resolvers listen for TCP connections on port 853 [29]. Therefore, our first step is to find out all the IP addresses that have port 853 open. To do so, we leverage Rapid7's public dataset [44] that contains a monthly Internet-wild SYN scan on TCP port 853. As the previous study [35] shows, the churn of DoT resolvers are relatively high, we thereby select long-term active DoT resolvers by analyzing 4 snapshots of scan results conducted from Jun 2020 to September 2020 and extracting the IP addresses that have TCP port 853 open in every scan during this period. In total, we obtain 1.99 million IP addresses. After that, we validate if the hosts with those IP addresses are functional DoT resolvers by sending DoT requests to those IP addresses, resolving a domain name under our control. We then determine a host as a functional DoT resolver if it sends back the correct IP address of our domain name. In the end, we successfully validate 6,016 functional DoT resolvers.

[2]Note that the DNS manipulation could happen on the resolver itself or the on-path censorship devices placed between the resolver and the corresponding nameservers. In the paper, we do not distinguish these two scenarios. We identify DNS manipulation if through the resolver we cannot obtain a valid DNS response.

Table 1: Ethical Resolver Selection.

	DoT Resolver	DoH Resolver
Total	6,016	82
Invalid Certificate	1,880	0
Low Stability	318	7
Ethical	3,813	75

3.1.2 DoH resolvers. Unlike the DoT resolvers that operate on the exclusive port 853, DoH resolvers operate on port 443 that is used by web servers to support HTTPS connections and the URL of the resolvers is not well-defined. Therefore, it is inefficient to actively scan all the possible URLs to identify DoH resolvers. Instead, we simply search for public known DoH resolvers from online documents [1, 2]. Still, like the DoT resolvers, we validate if a DoH resolver is functional by resolving our domain name and examine DNS responses. In total, we obtain 82 functional DoH resolvers. Although the number of DoH resolvers is much less than that of the DoT resolvers, we believe that we obtained the majority of open DoH resolvers since we collected more DoH resolvers than a recent study [35], which actively scanned billions of URLs to identify DoH resolvers.

3.1.3 Ethical Resolver Selection. During our experiments, the DoT and DoH resolvers are used to resolve domains potentially being censored (Section 3.2). Therefore, it is necessary to carefully design the experiments to reduce the risks on participants. To do so, we follow an ethical principle discussed in Iris [42], i.e., we only use the resolvers that are unlikely to be associated with any individual users, as our guideline to perform experiments. To this end, we design two filtering processes to identify resolvers that are well maintained for public services. Table 1 presents the resolver selection results.

First, the previous study [35] found that some DoT resolvers installed invalid certificates, and the use of invalid certificates strongly indicates that the resolvers are not well maintained as a public service. Thus, we actively validate certificates of DoT resolvers and exclude invalid ones. In total, we identify 1,880 DoT resolvers with invalid certificates, leaving us 4,136 DoT resolvers with valid certificates. We also check the certificates of DoH resolvers, and all of them are valid.

In addition, as the stability of a resolver is critical to the public service and its users, a well-maintained resolver should not have significant downtime. Therefore, we test if the resolvers are steadily active by resolving our domain name every hour for consecutive 10 days, and exclude those that demonstrate any failed responses. In total, we discard 318 DoT and 7 DoH resolvers with unreliable availability. As a result, we obtain 3,818 ethical DoT resolvers and 75 ethical DoH resolvers for our experiments.

Table 2: Providers of Ethical DoT/DoH Resolvers

	Total Providers	Top Providers	Count
DoT Resolver	461	cleanbrowsing.org	1,074
		cloudflare-dns.com	1,044
		nextdns.io	805
		ibvpn.com	92
		perfect-privacy.com	78
DoH Resolver	47	pi-dns.com	7
		nixnet.xyz	5
		quad9.net	4
		cloudflare-dns.com	4
		rubyfish.cn	3

With the ethical DoT and DoH resolvers, we then identify the service providers of these resolvers. For a DoH resolver that is always associated with a URL, we directly use the apex domain (*i.e.*, second-level domain name) to recognize the provider of the DoH resolver. For DoT resolvers, however, we only have their IP addresses. Reverse DNS lookup is a straightforward way to obtain the domain name of a resolver, but we find that many PTR records of the IP addresses do not exist or they point to the domains of the hosting service providers. As such, to obtain the provider of a DoT resolver, we resort to the certificate carried by the resolver. Specifically, we determine the provider as the second-level domain of the Common Name (CN) in the certificate's subject field. Table 2 presents the providers of ethical DoT and DoH resolvers. Overall, our ethical resolvers are from 461 DoT providers and 47 DoH providers. Also, we observe three large DoT providers, *i.e.*, CleanBrowsing, Cloudflare, and NextDNS, which, in total, account for 77% of the resolvers and most of their IP addresses are concentrated in a few /24 subnets. For example, 255 IP addresses in 172.64.37.0/24 are DoT resolvers provided by Cloudflare.

3.2 Test Domain List

In order to obtain a list of domains that may be censored, we follow the standard approach used in prior studies [39, 42, 49], which compile a test domain list including popular domains and sensitive domains. We extract Alexa top 1,000 domains [3] as our popular domains and select the sensitive domains from the Citizen Lab's [4] global list, which contains 1,302 entries. Note that the entries provided by Citizen Lab are in the form of URLs, and we extract domains from the URLs to form our sensitive domains. After we deduplicate the popular and sensitive domains and exclude the problematic ones (*i.e.*, we cannot obtain an HTTP response with `200 OK` status code when we access the landing page of the domain or the domain installs an invalid certificate), we end up with 1,909 domains. We further determine the category of the domains with the classification services provided by FortiGuard [21].

3.3 DNS Response Validation

We resolve those test domains with the encrypted DNS resolvers and verify if the DNS responses contain valid IP addresses. If a DNS response cannot be validated, we consider that the DNS response is manipulated by the encrypted DNS resolver. However, it is very challenging to determine whether a DNS response is valid for a domain since the domain could have multiple valid IP addresses and it is impossible to enumerate all of them. To overcome this problem, we apply the following procedure to validate each DNS response.

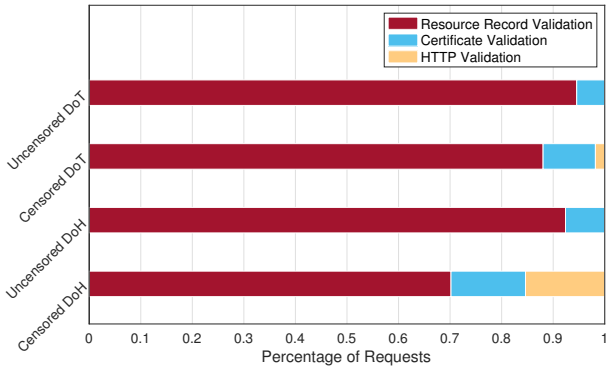

Figure 2: Validation Process to Determine the Occurrence of DNS Censorship. Each bar shows the percentages of requests that can be identified by the certain validation process.

3.3.1 Resource Record Validation. We set up a control node that does not suffer DNS manipulation to resolve the test domains, and then we compare the results with the DNS responses collected from the encrypted DNS resolvers.

First, if we obtain a failure DNS record, such as NXDOMAIN, SERVFAIL, *etc.*, from an encrypted DNS resolver or the DNS record contains a non-routable IP address, such as 127.0.0.1, but our control node obtains a public routable IP address of a domain, we consider that the DNS response of the corresponding domain is manipulated. Second, if the encrypted DNS resolver receives a public routable IP address and it is in the same Autonomous System (AS) as the IP address obtained by our control node, we consider that the encrypted DNS resolver obtains a valid IP address. This is because that DNS is widely used to balance the traffic to different web servers (e.g., DNS load-balancing in CDN [23, 31]), and therefore different resolvers may receive different IP addresses in the same time window, but those IP addresses are normally possessed by the same AS.

If the encrypted DNS resolver receives a public routable IP address but it does not belong to the same AS of the IP address obtained by the control node, we then validate the IP address with the certificate and the HTTP webpage described below.

3.3.2 Certificate Validation. In order to establish an HTTPS connection, a web server needs to present a valid, non-expired certificate, which carries valid signatures and is issued to the requested domain, *i.e.*, the domain name must appear in the Subject field or the Subject Alternative Names extension of the certificate.

As such, we validate an IP address by retrieving the certificate from that address. If the certificate can be successfully validated, we consider the IP address received by the encrypted DNS resolver is authentic, and vice versa. Note that we have already excluded

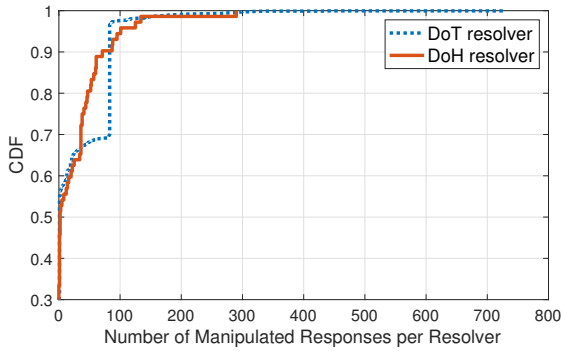

Figure 3: CDF of Manipulated Responses per Resolver.

domains with an invalid certificate in our test domains, so all the remaining domains should have valid certificates if they support HTTPS. Thus, if we identify an invalid certificate now, it could be provided by the servers controlled by the resolver providers or the on-path censorship devices, and those servers are used to provide blockpages.

Next, if a website does not support HTTPS and serves the valid content through HTTP, there is no certificate available for us to retrieve. We then validate the IP address through HTTP validation.

3.3.3 HTTP Validation. To perform HTTP validation, we first retrieve the domain's landing page using the IP address obtained by the encrypted DNS resolver and that obtained from our control node. Since we have excluded the problematic domains in our domain list, our control node only receives a `200 OK` or `3XX redirection` HTTP response. As a result, if we cannot obtain a valid HTTP response or the HTTP response does not present a correct status code, we determine that the corresponding IP address obtained by the encrypted DNS resolver is invalid.

After that, for the rest of the DNS responses, their corresponding HTTP responses, retrieved from IP addresses obtained by encrypted DNS resolver and our control node respectively, would have the same `200 OK` or `3XX redirection` status code. In such cases, if they have the same `3XX redirection` status code and their `Location` headers redirect the traffic to the same domain, we determine that the IP address is valid and vice versa. If they have the same `200 OK` status code, we will further examine their HTML content. Specifically, we first check if the two HTML webpages present the same `<title>` tag. If so, we consider the IP address is valid. Otherwise, we manually review the webpages to determine their validity. Note that the manual effort is trivial as most of the DNS responses can be determined without manual review.

3.4 Observation of DNS Manipulation

In total, we conduct 7.4 million DNS lookup measurements, resolving 1,909 domains from 3,813 DoT and 75 DoH resolvers. During the experiment, we retry a DNS query at most 3 times if we do not receive a DNS response from an encrypted DNS resolver. If all retries fail, we further attempt to resolve the domain under our control from the encrypted DNS resolver to test whether the resolver operates as we expected. We then discard 152 of DoT and 3 of DoH resolvers as they fail the test. To this end, after removing those resolvers, our dataset consists of 7.1 million DNS responses.

Table 3: Blocking Behaviors.

Blocking Behaviors	DoT Responses	DoH Responses
NXDOMAIN	93,229 (80.20%)	398 (20.46%)
SERVFAIL	1,160 (1.00%)	404 (20.77%)
REFUSED	288 (0.25%)	21 (1.08%)
Empty Record	509 (0.44%)	29 (1.49%)
Reserved IP	7,021 (6.04%)	513 (26.38%)
Forged Public IP	14,036 (12.07%)	580 (29.82%)

Overall, we identify that 116,243 of DoT responses (1.66%) and 1,945 of DoH responses (1.42%) are manipulated. Figure 2 shows the results of validation processes for determining the occurrence of DNS censorship. The validity of most DNS responses can be determined by the resource record validation, especially for those unmanipulated. Meanwhile, only a small portion of responses require HTTP validation, but most of which (93.76%) are identified as manipulated responses.

3.4.1 Manipulated DNS Responses. Figure 3 shows the CDF of the number of manipulated responses by each resolver. In total, 69.7% of the DoT resolvers and 66.7% of the DoH resolvers manipulate DNS responses for at least one domain, while 2.4% of the DoT resolvers and 5.6% of the DoH resolvers manipulate the responses for more than 100 domains. In comparison to the previous study [42] where 11% of the open resolvers manipulate DNS responses, the DNS manipulation in encrypted DNS resolvers is even more prevalent and severe. In particular, we observe that two DoT resolvers, `78.47.230.59` and `94.130.183.67` located in Germany, manipulate DNS responses of 728 domains, which are the two most among all the DoT resolvers. Meanwhile, a DoH resolver, `https://dns.alidns.com/dns-query` operated by Alibaba in China, manipulates the most domains' DNS responses among all the DoH resolvers. We believe that such DNS manipulations are actually conducted by China's Great Firewall (GFW), and we will discuss the details in Section 5.1.

We identify 1,370 and 652 domains that are censored by at least one of the DoT and DoH resolvers, respectively. The most censored domain by DoT resolvers is `use-application-dns.net`, a canary domain[3] that is used as a test to disable the DoH feature in Firefox browsers, and the domain that is censored the most by DoH resolvers is `adultfriendfinder.com`, a website serving adult information.

Table 3 shows the blocking behaviors of encrypted DNS resolvers. Most of the manipulated responses by DoT resolvers present an NXDOMAIN message, and a few show a SERVFAIL or REFUSED code. Meanwhile, for DoH resolvers, the numbers of manipulated responses with NXDOMAIN and SERVFAIL code are roughly the same, but the responses with REFUSED are much less. Furthermore, we observe that a few manipulated DoT and DoH responses have

[3] A network administrator can set a DNS failure response to such a domain as a signal for a Firefox browser to learn that the network is unsuitable or undesired for DoH resolvers configured by Firefox [36].

Table 4: Top 5 IP Addresses in the Manipulated Responses

	IP addresses	Count	Providers
DoT	0.0.0.0	6,673	el3c.de†
	116.202.104.80	2,855	kidgonet.de
	45.77.77.148	2,624	cleanbrowsing.org
	207.246.127.171	1,324	cleanbrowsing.org
	176.103.130.135	1,046	adguard.com
DoH	0.0.0.0	463	brahma.world†
	146.112.61.106	123	opendns.com
	75.2.78.201	45	cira.ca
	99.83.232.37	45	cira.ca
	127.0.0.1	45	dnsforge.de

† 202 DoT resolvers and 13 DoH resolvers return 0.0.0.0, and we show the provider of the resolver that returns the IP address the most.

a valid format but without an IP address. A significant amount of manipulated responses contain forged IP addresses that are reserved IP addresses or invalid public IP addresses. The top 5 IP addresses in the manipulated responses are listed in Table 4. Specifically, `0.0.0.0` is the most used IP address. For those forged public IP addresses listed, each of them belongs to a corresponding resolver provider, and a blockpage is hosted on each IP address. Finally, the DoT and DoH providers may configure multiple IP addresses that are associated with blockpages, *e.g.*, `cleanbrowsing.org` and `cira.ca` in Table 4.

3.4.2 DNS Manipulation by Providers. Next, we investigate the DNS manipulation policies of different providers by aggregating the observations of resolvers deployed by a same provider, as detailed in Table 5. In particular, the DoT providers censor significantly more domains than the DoH providers. For example, `alidns.com` censors the most domains among DoH providers, but the number of its censored domains is still less than that of the 5th of DoT providers. Interestingly, we observe that the number of censored domains by different resolvers from a same provider could vary significantly. For example, more than 95% of the DoT resolvers of `cleanbrowsing.org` manipulate DNS response of a similar domain set that contains around 83 domains, while a few resolvers in the address ranges of `185.228.168.2XX` and `185.228.169.2XX` manipulate the DNS responses associated with hundreds of more domains, making a total of 410 domains. We infer that such a difference could originate from the different policies at different service plans offered by the encrypted DNS provider.

Moreover, from Table 5, we can see that the top DoT providers tend to manipulate DNS responses of domains that publish news and provide online search services, while the top DoH providers tend to manipulate the DNS responses of domains that provide proxy services or serve pornography contents.

3.4.3 Commonly Censored Domains. Table 6 lists the top 5 commonly censored domains as well as the top 5 commonly censored domain categories. A category is considered as being blocked if at least one domain that falls into the category is blocked. For DoT providers, the top 3 commonly censored domains are all advertising platforms, indicating that ad-blocking services are popularly adopted by the DoT providers. With respect to domain categories, information-related categories such as information technology, business, and advertisement are ranked in high positions.

For DoH providers, the top 2 commonly censored domains are related to adult content, indicating that family protection policies are widely supported by DoH providers. Meanwhile, information technology is also ranked on the top of the censored categories, followed by two adult-related categories.

In comparison to the results in the previous study [42] showing that open resolvers incline to manipulate DNS responses of pornography, gambling, and P2P sharing domains, encrypted DNS providers are usually in favor of manipulating the DNS responses of information technology domains.

Table 5: DNS Manipulation by Providers. The Categories **column lists the top 3 categories of domains whose responses are manipulated.**

	Providers	Count	Categories†
DoT	kidgonet.de	728	NEWS, SHOP, INFO
	cleanbrowsing.org	410	PORN, PROX, SRCH
	ideame.top	332	NEWS, SRCH, PROX
	waitquietly.com	326	NEWS, SRCH, PROX
	el3c.de	315	SRCH, INFO, NEWS
DoH	alidns.com	290	NEWS, SRCH, PROX
	cleanbrowsing.org	134	PORN, PROX, ADUL
	opendns.com	125	PROX, PORN, ADUL
	rubyfish.cn	119	NEWS, PORN, PROX
	brahma.world	87	GAMB, INFO, SRCH

† The abbreviations of domain categories are specified in Table 12.

Table 6: Top 5 Commonly Censored Domains and Top 5 Commonly Censored Domain Categories by Providers. NoP**: Number of Providers.**

	Domains	NoP	Categories	NoP
DoT	www.mgid.com	270	INFO	301
	doubleclick.net	265	BUSI	284
	www.exoclick.com	264	GAMB	276
	pingomatic.com	261	ADVR	274
	use-application-dns.net	230	BLOG	270
DoH	adultfriendfinder.com	12	INFO	19
	www.livejasmin.com	11	ADUL	16
	www.exoclick.com	10	PORN	16
	use-application-dns.net	10	SRCH	16
	www.hotspotshield.com	9	BUSI	15

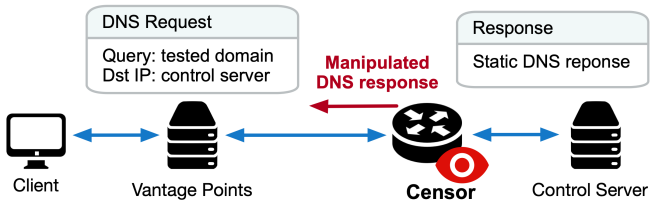

Figure 4: DNS Manipulation Detection. Our vantage points should receive a static DNS response from our control server if no DNS manipulation occurs, and vice versa.

4 CIRCUMVENTION WITH ENCRYPTED DNS

In this section, we evaluate the effectiveness of adopting DoT/DoH resolvers for circumventing Internet censorship. The basic idea is to identify the vantage points (e.g., VPN severs) that suffer DNS manipulation, from which perform DNS resolution on the censored domains via DoT/DoH resolvers, and then attempt to visit those domains for assessing the accessibility.

4.1 Vantage Points

To conduct extensive experiments, we need a pool of vantage points distributed across the world, which can issue plaintext DNS requests, DoT/DoH requests, and HTTP/HTTPS requests. One of the straightforward methods is to use crowdsourcing platforms to recruit volunteers for performing the experiments, but it may raise some ethical concerns when the volunteers attempt to access the domains that are supposed to be censored in their locations. As a result, we utilize commercial VPN servers as our vantage points, same as the strategy adopted by a well-established measurement platform, ICLab [39], which has been using VPN servers to detect censorship activities for two years and no ethical issues have been reported. Moreover, commercial VPN services have been widely used to bypass Internet censorship, and hence the VPN operators understand the risks of deploying their servers in a country.

However, it is hard to find VPN service providers that deploy physical VPN servers in China, while the GFW in China is known to perform DNS manipulation [8, 42]. Therefore, we instead launch instances in cloud services as the vantage points in China to conduct our experiments.

In total, we obtain vantage points from five commercial VPN providers that deploy more than one thousand VPN servers across over one hundred countries. As the vantage points in China, we operate four instances in cloud services launched from different locations.

4.2 DNS Manipulation Detection

Since encrypted DNS prevents the DNS manipulation between users and resolvers, it could help users circumvent Internet censorship. However, as we mentioned in Section 2.2.1, DNS manipulation can be done by an on-path censorship device. To assess the effectiveness of the censorship circumvention enabled by DoT/DoH resolvers, we need to identify, at each vantage point, the domains censored by on-path censorship devices, and those domains that will be used in the circumvention test (see Section 4.3.2).

Table 7: DoT/DoH Resolvers.

Provider	DoT Resolver	DoH Resolver
Cloudflare	1.1.1.1	https://1.1.1.1/dns-query
Google	8.8.8.8	https://8.8.8.8/dns-query
Quad9	9.9.9.10	https://9.9.9.10/dns-query
Self-built	our_IP	https://[our_IP]/dns-query

As such, we design an approach that can *accurately* identify the DNS manipulation conducted by on-path censorship devices, as illustrated in Figure 4. In particular, we first set up a control server that replies to arbitrary DNS requests with a static DNS response. Then, we issue the DNS requests from the vantage points to our control server, resolving test domains. If a DNS request triggers an on-path device, our vantage point will receive a manipulated DNS response that will be different from the static DNS response provided by the control server. Hence, it is straightforward to detect the DNS manipulation by comparing the received DNS response with the static DNS response, *i.e.*, the ground-truth. For the purpose of identifying which domains are censored at each vantage point, we use the same test domain list in Section 3.2 as the test base.

Ideally, our detection method does not introduce any false negatives or false positives since our control server provides a ground truth of resolution responses. However, ISPs may deploy DNS cache proxies as a part of network infrastructures to serve their users so that popular DNS records can be reused, improving the performance of DNS resolution [34]. These cache proxies may intercept the connections between our vantage points and control server, perform their own DNS resolution on the behalf of our vantage points, and finally return valid DNS responses to the requested domains. As such, a valid DNS response will be different from the static DNS response provided by our control server, and we will falsely identify them as the activities of DNS manipulation.

To eliminate those false positives, we conduct a cache test to exclude those vantage points that could be impacted by cache proxies. Specifically, we resolve a separate domain operated by us at our control server, which will serve as an incorrect, static DNS response. Meanwhile, an authentic DNS response of this domain is also hosted at its legitimate nameserver. If a DNS cache proxy presents, it will resolve our domain through the normal resolution path and provide us an authentic DNS response. If not, the tested DoT/DoH resolver will reach our control server and obtain the static response.

4.3 Circumvention Measurement

With the censored domains identified at each vantage point, we conduct our tests for the circumvention. From each vantage point, we resolve each censored domain through DoT and DoH resolvers and retrieve the domain's landing page with the IP address returned by those encrypted DNS resolvers. Then, we analyze the accessibility of the domain and examine the root cause if the domain is still inaccessible.

4.3.1 Resolver Selection. Although we have collected a large number of DoT/DoH resolvers (Section 3.1), it is inefficient and unnecessary to test all the resolvers from each vantage point since it will

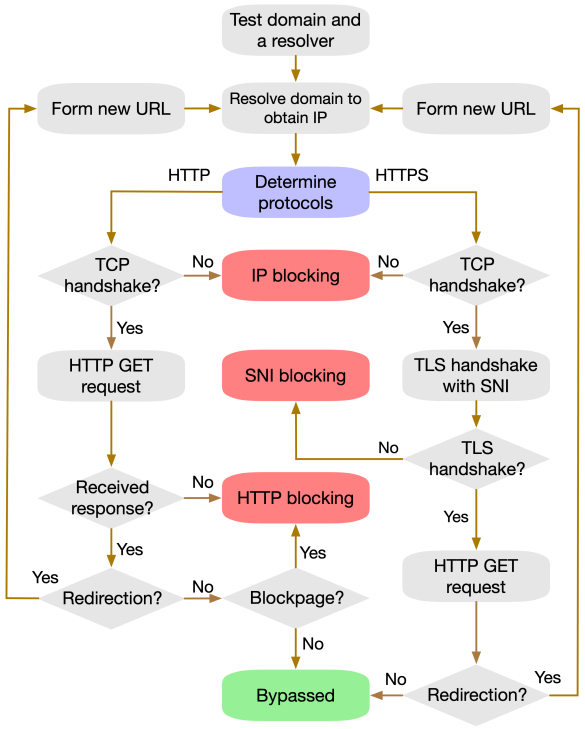

Figure 5: Flow Chart of Circumvention Test for a Pair of Resolver and Censored Domain.

generate too much traffic on the vantage point. Therefore, as shown in Table 7, we select resolvers from three well-known encrypted DNS service providers: Cloudflare, Google, and Quad9. Moreover, we also set up a new resolver to represent all other unpopular resolvers.

Note that the resolvers from the three popular providers are typically operated at multiple IP addresses, and we only evaluate the most well-known ones summarized in Table 7.[4] We confirm that there is no DNS manipulation occurred in these popular resolvers.

4.3.2 Circumvention Test. For testing the feasibility of circumvention, we must not, and cannot, directly configure a DoT or DoH resolver as the default resolver of the vantage point. Instead, we manually specify the DoT/DoH resolver when performing a DNS resolution. Also, it is possible that the resolvers are blocked in the region of our vantage points. Therefore, we first test the reachability of a DoT/DoH resolver by resolving our domain through the resolver and validating the response. If the resolver is not reachable, we exclude it from the circumvention test of the corresponding vantage point.

We then perform the circumvention test with those censored domains and reachable resolvers. We consider that an encrypted DNS resolver can be used to circumvent the censorship if the vantage point can retrieve the legitimate web content from the censored domain. In particular, we will follow any HTTP redirection if it presents. Figure 5 depicts a flow chart of the circumvention test for a pair of a censored domain and a resolver. First, we resolve the domain through the resolver. As those popular DoT/DoH resolvers do not perform DNS manipulation, we expect to obtain the authentic IP addresses. We then randomly select an IP address from the returned IP pool and determine a protocol, either HTTP or HTTPS, to visit the domain.

Here, we describe our method of determining the protocol for fetching the webpages. For a particular pair of a resolver and a domain, we first attempt to make a TLS handshake with the IP address at port 443 from a node that does not experience censorship. If the TLS handshake succeeds, we will issue HTTPS requests for this pair. Otherwise, we will use HTTP. We give priority to the HTTPS since, compared to the HTTP, it is less likely to be blocked due to the fact that (1) the majority of the websites support HTTPS connections and many of them would redirect HTTP connections to HTTPS connections [18] and (2) previous studies [49] show that the HTTP-based censorship is much more prevalent than the HTTPS (*i.e.*, SNI-based) censorship. Moreover, if we encounter an HTTP redirection, we then use the protocol indicated in the HTTP redirection response to fetch the webpages.

For the domains we use HTTPS connections, we first attempt to make a TCP handshake from the vantage point. If the connection is failed (*i.e.*, the connection timeouts or we receive an RST or FIN packet), we consider that the IP address is blocked. Otherwise, we further proceed with the TLS handshake and include the SNI extension in the `Client Hello` message. If the connection becomes failed or an invalid certificate is received, we determine that the domain experiences SNI-based censorship. If the TLS handshake succeeds, we send an HTTP `GET` request. If the HTTP response requires a redirection, we then form a new URL and rerun the test with the new URL; otherwise, we conclude that the tested domain becomes accessible with the adoption of the DoT or DoH resolver.

For the domains we determine to use HTTP connections, we first make a TCP handshake from the vantage point as well. If the connection is failed, the IP address is determined as blocked. Otherwise, we send an HTTP `GET` request. Then, if the connection fails, we determine that the domain experiences HTTP-based censorship. If the response indicates redirection, we then form a new URL and rerun the test for the new URL. Otherwise, we examine whether the web content presents a blockpage. To do so, we first obtain the webpage of the corresponding domain from a node that does not experience censorship, and we compare its `<title>` tag to that of the webpage obtained from the vantage point. If they are matched, we determine that the DoT or DoH resolver obtained a legitimate webpage. Otherwise, we manually review the webpage to determine whether it is a blockpage returned by an on-path censor.

Table 8: DNS Manipulation Observation.

Vantage Points	Censored Domains
China (4)	298
Denmark (1)	6
Iran (1)	197
Portugal (1)	3
United States (1)	6

[4]Quad9's resolvers at 9.9.9.9 do enable some censorship policies, but Quad9 also provides uncensored resolvers at 9.9.9.10 [5].

Table 9: Accessibility of Censored Domains with DoT/DoH Resolvers.

Vantage Points	Cloudflare		Google		Quad9		Self-built	
	DoT	DoH	DoT	DoH	DoT	DoH	DoT	DoH
China (Chengdu)	37.81%	37.10%	37.81%	—[†]	38.16%	37.10%	37.10%	37.81%
China (Qingdao)	35.84%	36.52%	37.54%	—	37.20%	36.51%	37.20%	37.20%
China (Shanghai)	37.33%	37.33%	—	—	36.64%	37.33%	36.30%	36.30%
China (Shenzhen)	36.39%	36.39%	—	—	36.05%	36.05%	36.05%	36.05%
Denmark	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Iran	—	—	—	—	0.00%	0.00%	0.00%	0.00%
Portugal	100%	100%	100%	100%	100%	100%	100%	100%
United States	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

[†] The "—" sign indicates that the resolver is not reachable from the vantage point.

4.4 Results of Circumvention Test

4.4.1 DNS Manipulation. We detect the DNS manipulation conducted by on-path devices from vantage points in 5 countries, listed in Table 8. More specifically, in China, all four vantage points experience DNS manipulation and they observe a very similar set of censored domains, resulting in a total of 298 censored domains. In Iran, we identify one vantage point that observes the DNS manipulation on 197 domains. Note that it is the only vantage point available for us to use in Iran during our experiments. Meanwhile, we find that only one vantage point observes DNS manipulation on a few domains[5] (6, 3, 6) in Denmark, Portugal, and the United States, respectively, indicating that the DNS manipulation in those three countries is not nation-wide censorship.

Although the previous study [42] shows that DNS manipulation is prevalent across the world, our results show that DNS manipulation by on-path devices only happens in a few countries. Thus, we infer that the majority of DNS manipulation observed in [42] should have been conducted by DNS resolvers. A similar observation has been made by ICLab [39], which only detects the DNS manipulation by on-path devices in three countries.

4.4.2 Accessibility of Censored Domains. Table 9 lists the percentage of the censored domains that can be accessed when DoT/DoH resolvers are used. In particular, we observe that the Google DoH resolver is unreachable from vantage points in China, and DoT and DoH resolvers from Cloudflare and Google are unreachable from the vantage point in Iran. On the other hand, although we observe the vantage points in Shanghai and Shenzhen can correctly resolve our domain through the Google DoT resolver at the beginning of our experiment, as the experiments progress, two-thirds of the connections to the Google DoT resolver become being blocked. Thus, we here mark the Google DoT resolver as unreachable from these two vantage points.

For a particular vantage point, we observe that different resolvers have approximately the same capability to help users access the censored domains. In other words, switching to another encrypted DNS resolver does not make a noticeable difference in accessing the censored domains. On the other hand, the effectiveness of using encrypted DNS resolvers to circumvent censorship varies from country to country. For example, when a DoT or DoH resolver is reachable, we identify that all the censored domains of the vantage point in Portugal can be accessed, and 50% of the censored domains of the vantage point in Denmark become accessible. Also, around 37% of the censored domains are accessible from all the vantage points in China, and all the sets of accessible domains are highly similar. However, none of the censored domains is accessible from the vantage point in Iran or the United States, indicating that the censors enforce very strict policies on blocking those domains.

Furthermore, we investigate the circumvention rates, with respect to the categories of censored domains. Since the DoT/DoH resolvers enable access to a similar set of censored domains for all the vantage points in China, we only present the results observed from one vantage point (Qingdao) through our self-built DoH resolver. Table 10 lists the top 5 categories of censored domains and their circumvention rates in China. We can see that the domains in three categories, *i.e.*, news and media, search engines/portals, and reference, have a relatively low circumvention rate, while for the other two categories, proxy avoidance and pornography, a circumvention rate of more than 70% can be achieved. Besides China, we observe that 50% and 100% of the censored file sharing domains become accessible for the vantage points in Denmark and Portugal, respectively.

For all the censored domains that are still not accessible when using reachable DoT and DoH resolvers, we further explore the blocking techniques that prevent the access to these domains. Again, we notice that the observations across different resolvers from the same vantage point do not have a noticeable difference, and hence we only present the results observed by our self-built DoH resolver, as listed in Table 11. More specifically, the SNI-based censorship is the most observable blocking technique for all the vantage points since the majority of connections are made through HTTPS. Also, all the vantage points in China present a similar distribution of the blocking techniques, along with a very similar set of censored domains as well.

In addition, we only observe IP-based blocking in China and Iran, and it is much less popular than the domain-name-based (*i.e.*, HTTP-based and SNI-based) blocking techniques. This is most likely due to the collateral damage of directly blocking the IP addresses (Section 2.2.4). It is worth noting that a few domains are blocked during

[5] All of the domains are related to file sharing services.

Table 10: China's Top Censored Domain Categories and their Circumvention Rates. Observed from the vantage point in Qingdao, China, and tested with our self-built DoH resolver.

Categories	Total	Circumvented
News and Media	55	18 (32.7%)
Search Engines and Portals	50	5 (10.0%)
Proxy Avoidance	28	22 (78.6%)
Reference	17	0 (0.0%)
Pornography	17	12 (70.6%)

their redirection phase (the numbers of such domains are listed in the parentheses of Table 11), and they are only observed by the vantage points in China. For example, the HTTPS requests sent from the vantage points in China to `https://torrentz.eu` are not blocked, but redirection responses are received and they indicate a new location of `https://torrentz2.is/`. Then, the HTTPS requests sent to `https://torrentz2.is/` are censored by the SNI-based blocking.

5 DISCUSSION

5.1 Importance of Deployment

As we mentioned in Section 2.2.1, DNS manipulation could be done by a resolver itself or on-path censorship devices. During our experiments, we notice that the set of censored domains that were blocked when using Alibaba's encrypted DNS resolvers (Section 3.4.2) deployed in China, is very similar to the set of censored domains observed from vantage points at different locations in China (Section 4.4.1), implying that the censored domains being observed from Alibaba's encrypted DNS resolvers are actually censored by an on-path censorship device, *i.e.*, GFW.

Given such an observation, we consider that the deployment of encrypted DNS resolvers is critical for the providers to offer reliable services. Therefore, we recommend encrypted DNS providers deploy their resolvers in the regions that do not suffer DNS manipulation by on-path censorship devices, and the existence of an on-path censorship device can be easily determined by the simple approach we proposed in Figure 4.

5.2 Ethical Considerations

Since our study does not involve the collection of personal information or human participation, it falls outside the purview of IRBs [32]. However, censorship studies often still pose ethical concerns due to active, large-scale measurement-based experiments. As a result, we carefully consider the ethical issues in our experiment design, and we highlight our key procedures to reduce the risk on the vantage points.

To study DNS manipulation at the use of encrypted DNS, we design two filtering processes to screen out the resolvers that are possibly associated with individuals. The two filtering processes include identifying the resolvers with invalid certificates and low stability, strongly indicating that the resolvers are not well maintained as a public service.

Table 11: Techniques for Blocking Censored Domains when Encrypted DNS is used. The results are observed by our self-built DoH resolver. The numbers in the parentheses are the numbers of domains being blocked in the redirection phase.

Vantage Points	Blocking Techniques		
	IP-based	HTTP-based	SNI-based
China (Chengdu)	59	5 (3)	112 (4)
China (Qingdao)	60	5 (3)	119 (6)
China (Shanghai)	59	5 (3)	122 (4)
China (Shenzhen)	60	5 (3)	123 (5)
Denmark	0	0	3
Iran	3	1	193
United States	0	1	5

To assess the effectiveness of encrypted DNS to circumvent censorship, we need to make connections to the domains that are supposed to be blocked in the region of vantage points. Therefore, to avoid potential risks on any individuals, we restrain our vantage points to VPN servers and cloud instances that are demonstrated to be ethical for censorship studies in previous research [39, 43].

In addition, we rate-limit the requests sent by each vantage point to minimize their traffic burden. Also, we attempt to avoid the request aggregation at nameservers and web servers involved in our experiments by randomizing the request sequence of the tested domains at each vantage point.

5.3 Limitation

While studying DNS manipulation at the use of DoT and DoH resolvers in Section 3, we only evaluated open DoT and DoH resolvers. However, ISPs and organizations may also provide DoT and DoH resolvers to their users but not open to the public, thereby we may miss a portion of the DoT and DoH resolvers. However, the previous study [35] showed that the deployment of local DoT resolvers is rare, and thus it would not impact the scale of our study.

Because our circumvention tests are mainly conducted at the VPN servers that are usually hosted in a cloud environment, we lack the capability to evaluate the circumvention of encrypted DNS in a residential environment. Also, VPN providers have been recognized to often lie about their servers' locations [50]. To mitigate this issue, we conduct traceroute from the VPN server to a node outside its advertised country and collect (1) the nationality of intermediate routers by looking up the geolocation database [6], and (2) the domain names of routers by issuing reverse DNS lookups to search for any geolocation hints (*e.g.*, country code, city, or airport code). If any geolocation information confirms the country of the VPN server as advertised, we consider its location to be validated.

6 RELATED WORK

6.1 Encrypted DNS

Prior studies have investigated encrypted DNS mainly from the privacy and performance perspectives. However, none of them has explored the impact of encrypted DNS on Internet censorship. Here, we briefly survey the previous research on performance overhead and privacy benefits of encrypted DNS.

Lu *et al.* [35] conducted large-scale measurements on the server availability, client reachability, and performance of DoT and DoH resolvers, demonstrating that their reachability is satisfying and the induced performance overhead is tolerable. Bottger *et al.* [12] further examined the performance overhead of DoH, and they made an observation that DoH would lead to a longer DNS resolution time but it does not significantly increase the user-perceived page load time.

Several studies have explored approaches to breaking the privacy protection provided by encrypted DNS. Concretely, Houser *et al.* [28] proposed a DoT fingerprinting method that examines the patterns of packet size to determine if a user visits a website that is interesting to adversaries, achieving a 17% of false negative rate and 0.5% of false positive rate when DNS messages are not padded. Siby *et al.* [47] developed a new feature set to effectively fingerprint the visited websites and demonstrated that the traffic analysis defenses, such as EDNS0 padding, are unable to mitigate the proposed fingerprinting attack. Hoang *et al.* [24] investigated the privacy benefits brought by encrypting domain names. They concluded that 20% of the domains cannot gain privacy benefits since their hostnames and the corresponding IP addresses have a one-to-one mapping, so the IP addresses leak a user's privacy. Trevisan *et al.* [48] leveraged the plaintext DNS traffic to infer the encrypted DNS information with a simple classification algorithm and demonstrated the efficiency of the proposed classifier.

6.2 Internet Censorship

Recent research has studied Internet censorship in some particular countries [9, 13, 37] and demonstrated that DNS manipulation is pervasively used in many countries for censorship. A group of anonymous authors [8] studied the collateral damage that affects the communication beyond the censored networks when ISPs deploy censorship tools that inject forged DNS responses to block users from accessing censored websites.

Several censorship measurement platforms have also investigated global DNS manipulation. In particular, OONI [20] is a network measurement platform that recruits volunteers to perform censorship measurements, but its DNS manipulation detection is inaccurate as it simply compares the DNS responses obtained from the volunteers with the DNS response from a control node [41], and Yadav *et al.* [51] reported that the false negative rate could be as high as 89%. Scott *et al.* [46] collected information on DNS resolution and resource availability to develop a toolchain for measuring web infrastructure deployments and detecting ISP-level DNS manipulation in the wild. Pearce *et al.* [42] proposed Iris, a scalable system for measuring and understanding the heterogeneity of global DNS manipulation. Their work focuses on investigating DNS manipulation at the use of unencrypted DNS resolvers, while our work focuses on exploring DNS manipulation at the use of encrypted DNS resolvers. More importantly, we discovered that DNS manipulation is even more prevalent in the encrypted DNS than that in the unencrypted DNS. Niaki *et al.* [39] presented ICLab, a platform that leverages commercial VPNs as vantage points to discover various censorship techniques, such as DNS manipulation, TCP packet injection, and web page blocking. They found that the majority of DNS manipulation is carried out by resolvers, rather than on-path censorship devices, and our work further confirms this observation.

Other than censorship detection, researchers have also proposed approaches to circumventing censorship. Specifically, Fifield *et al.* [19] proposed domain fronting, which runs circumvention proxies on the web services that share IP addresses with other uncensored services. Bock *et al.* [11] proposed Geneva, a genetic algorithm that automates the discovery and circumvention of packet-manipulation-based censorship techniques, and demonstrated its efficiency against censors on the Internet. Bock *et al.* [10] further applied the genetic algorithm at the server-side to bypass censorship so that the client-side does not need any modification. As ESNI has been proposed to bypass SNI-based censorship, Chai *et al.* [14] evaluated the effectiveness of using ESNI for censorship circumvention, and they concluded that ESNI makes more than 101 thousand websites more censorship-resistant.

Additionally, CDNBrowsing systems have been developed in [26, 38, 52] to bypass censorship by leveraging the collateral damage of IP blocking. Nisar *et al.* [40] implemented C-Saw, a system that conducts censorship measurements through crowdsourced users and provides adaptive censorship circumvention for them. Another line of work [27, 33] leveraged Decoy routing for bridging the connections to the censored destination through intermediate routers. However, Schuchard *et al.* [45] showed that censors can effectively block the participating routers to nullify Decoy routing.

7 CONCLUSION

In this paper, we presented the first comprehensive measurement study of investigating the impact of encrypted DNS on Internet censorship from two perspectives. On one hand, we performed a large-scale measurement to assess the severity of DNS manipulation, which can be exploited for censorship enhancement, when adopting encrypted DNS resolvers. In particular, we conducted 7.4 million DNS lookup measurements on 3,813 DoT and 75 of DoH resolvers, and identified that 1.66% of DoT responses and 1.42% DoH responses suffer DNS manipulation. More importantly, these manipulated DNS responses were originated from more than two-thirds of the DoT and DoH resolvers, showing that DNS manipulation is more prevalent in encrypted DNS than traditional DNS. On the other hand, we evaluated the effectiveness of using encrypted DNS resolvers to circumvent Internet censorship. We observed that, with the use of encrypted DNS resolvers, 37% of the censored domains are accessible from the vantage points in China, but none of the censored domains are accessible from the vantage points in Iran. Also, for a vantage point, using a different encrypted DNS resolver does not generate a non-trivial difference in censorship circumvention. Our further analysis reveals that many censored domains remain inaccessible because they are also blocked by other censorship techniques such as SNI-based blocking.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their insightful and constructive comments, which helped us improve the quality of this paper. This work was supported in part by the U.S. Army Research Office (ARO) grant W911NF-19-1-0049 and National Science Foundation (NSF) grant DGE-1821744.

REFERENCES

[1] 2020. https://dnscrypt.info/public-servers/.
[2] 2020. https://github.com/curl/curl/wiki/DNS-over-HTTPS.
[3] 2020. https://www.alexa.com/topsites.
[4] 2020. https://github.com/citizenlab/test-lists/.
[5] 2020. https://www.quad9.net/doh-quad9-dns-servers/.
[6] 2020. https://ipinfo.io/.
[7] Alessandro Ghedini. 2018. Encrypt it or lose it: how encrypted SNI works. https://blog.cloudflare.com/encrypted-sni/.
[8] Anonymous. 2012. The Collateral Damage of Internet Censorship by DNS Injection. *ACM SIGCOMM Computer Communication Review* (2012).
[9] Simurgh Aryan, Homa Aryan, and J. Alex Halderman. 2013. Internet Censorship in Iran: A First Look. In *USENIX Workshop on Free and Open Communications on the Internet.*
[10] Kevin Bock, George Hughey, Louis-Henri Merino, Tania Arya, Daniel Liscinsky, Regina Pogosian, and Dave Levin. 2020. Come as You Are: Helping Unmodified Clients Bypass Censorship with Server-side Evasion. In *ACM SIGCOMM.*
[11] Kevin Bock, George Hughey, Xiao Qiang, and Dave Levin. 2019. Geneva: Evolving Censorship Evasion Strategies. In *ACM Conference on Computer and Communications Security (CCS).*
[12] Timm Böttger, Felix Cuadrado, Gianni Antichi, Eder Leão Fernandes, Gareth Tyson, Ignacio Castro, and Steve Uhlig. 2019. An Empirical Study of the Cost of DNS-over-HTTPS. In *ACM Internet Measurement Conference (IMC).*
[13] Abdelberi Chaabane, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. 2014. Censorship in the Wild: Analyzing Internet Filtering in Syria. In *ACM Internet Measurement Conference (IMC).*
[14] Zimo Chai, Amirhossein Ghafari, and Amir Houmansadr. 2019. On the Importance of Encrypted-SNI (ESNI) to Censorship Circumvention. In *USENIX Workshop on Free and Open Communications on the Internet.*
[15] Cloudflare. 2020. https://developers.cloudflare.com/1.1.1.1/setting-up-1.1.1.1.
[16] Sara Dickinson. 2019. DNS Privacy Public Resolvers. https://dnsprivacy.org/wiki/display/DP/DNS+Privacy+Public+Resolvers.
[17] DNSCrypt. 2020. https://www.dnscrypt.org/.
[18] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. 2018. Measuring HTTPS Adoption on the Web. In *USENIX Security Symposium.*
[19] David Fifield, Chang Lan, Rod Hynes, Percy Wegmann, and Vern Paxson. 2015. Blocking-resistant Communication through Domain Fronting. In *Privacy Enhancing Technologies Symposium (PETS).*
[20] Arturo Filastò and Jacob Appelbaum. 2012. OONI: Open Observatory of Network Interference. In *USENIX Workshop on Free and Open Communications the Internet.*
[21] FortiGuard Labs. 2020. https://fortiguard.com/webfilter.
[22] Google. 2020. https://developers.google.com/speed/public-dns/docs/secure-transports.
[23] Shuai Hao, Yubao Zhang, Haining Wang, and Angelos Stavrou. 2018. End-Users Get Maneuvered: Empirical Analysis of Redirection Hijacking in Content Delivery Networks. In *USENIX Security Symposium.*
[24] Nguyen Phong Hoang, Arian Akhavan Niaki, Nikita Borisov, Phillipa Gill, and Michalis Polychronakis. 2020. Assessing the Privacy Benefits of Domain Name Encryption. In *ACM ASIA Conference on Computer and Communications Security.*
[25] P. Hoffman and P. McManus. 2018. DNS Queries over HTTPS (DoH). *IETF RFC 8484* (2018).
[26] John Holowczak and Amir Houmansadr. 2015. CacheBrowser: Bypassing Chinese Censorship Without Proxies Using Cached Content. In *ACM Conference on Computer and Communications Security (CCS).*
[27] Amir Houmansadr, Giang T. K. Nguyen, Matthew Caesar, and Nikita Borisov. 2011. Cirripede: Circumvention Infrastructure using Router Redirection with Plausible Deniability. In *ACM Conference on Computer and Communications Security (CCS).*
[28] Rebekah Houser, Zhou Li, Chase Cotton, and Haining Wang. 2019. An Investigation on Information Leakage of DNS over TLS. In *ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT).*
[29] Z. Hu, L. Zhu, J. Heidemann, A. Mankin, D. Wessels, and P. Hoffman. 2016. Specification for DNS over Transport Layer Security (TLS). *IETF RFC 7858* (2016).
[30] Lin Jin, Shuai Hao, Haining Wang, and Chase Cotton. 2018. Your Remnant Tells Secret: Residual Resolution in DDoS Protection Services. In *IEEE International Conference on Dependable Systems and Networks (DSN).*
[31] Lin Jin, Shuai Hao, Haining Wang, and Chase Cotton. 2019. Unveil the Hidden Presence: Characterizing the Backend Interface of Content Delivery Networks. In *IEEE International Conference on Network Protocols (ICNP).*
[32] Ben Jones, Roya Ensafi, Nick Feamster, Vern Paxson, and Nick Weaver. 2015. Ethical Concerns for Censorship Measurement. In *ACM SIGCOMM Workshop on Ethics in Networked Systems Research (NS Ethics).*
[33] Josh Karlin, Daniel Ellard, Alden W. Jackson, Christine E. Jones, Greg Lauer, David P. Mankins, and W. Timothy Strayer. 2011. Decoy Routing: Toward Unblockable Internet Communication. In *USENIX Workshop on Free and Open Communications the Internet.*
[34] Baojun Liu, Chaoyi Lu, Haixin Duan, Ying Liu, Zhou Li, Shuang Hao, and Min Yang. 2018. Who Is Answering My Queries: Understanding and Characterizing Interception of the DNS Resolution Path. In *USENIX Security Symposium.*
[35] Chaoyi Lu, Baojun Liu, Shuang Hao, Haixin Duan, Chunying Leng Mingming Zhang, Ying Liu, Zaifeng Zhang, and Jianping Wu. 2019. An End-to-End, Large-Scale Measurement of DNS-over-Encryption: How Far Have We Come?. In *ACM Internet Measurement Conference (IMC).*
[36] Mozilla. 2020. Configuring Networks to Disable DNS over HTTPS. https://support.mozilla.org/en-US/kb/configuring-networks-disable-dns-over-https.
[37] Zubair Nabi. 2013. The Anatomy of Web Censorship in Pakistan. In *USENIX Workshop on Free and Open Communications on the Internet.*
[38] Milad Nasr, Hadi Zolfaghari, Amir Houmansadr, and Amirhossein Ghafari. 2020. MassBrowser: Unblocking the Censored Web for the Masses, by the Masses. In *Network and Distributed System Security Symposium (NDSS).*
[39] Arian Akhavan Niaki, Shinyoung Cho, Zachary Weinberg, Nguyen Phong Hoang, Abbas Razaghpanah, Nicolas Christin, and Phillipa Gill. 2020. ICLab: A Global, Longitudinal Internet Censorship Measurement Platform. In *IEEE Symposium on Security and Privacy (S&P).*
[40] Aqib Nisar, Aqsa Kashaf, Ihsan Ayyub Qazi, and Zartash Afzal Uzmi. 2018. Incentivizing Censorship Measurements via Circumvention. In *ACM SIGCOMM.*
[41] OONI. 2020. DNS consistency. https://ooni.org/nettest/dns-consistency/.
[42] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. 2017. Global Measurement of DNS Manipulation. In *USENIX Security Symposium.*
[43] Reethika Ramesh, Ram Sundara Raman, Matthew Bernhard, Victor Ongkowijaya, Leonid Evdokimov, Anne Edmundson, Steven Sprecher, Muhammad Ikram, and Roya Ensafi. 2020. Decentralized Control: A Case Study of Russia. In *Network and Distributed System Security Symposium (NDSS).*
[44] Rapid7. 2020. https://opendata.rapid7.com/.
[45] Max Schuchard, John Geddes, Christopher Thompson, and Nicholas Hopper. 2012. Routing Around Decoys. In *ACM Conference on Computer and Communications Security (CCS).*
[46] Will Scott, Thomas Anderson, Tadayoshi Kohno, and Arvind Krishnamurthy. 2016. Satellite: Joint Analysis of CDNs and Network-level Interference. In *USENIX Annual Technical Conference (ATC).*
[47] Sandra Siby, Marc Juarez, Claudia Diaz, Narseo Vallina-Rodriguez, and Carmela Troncoso. 2020. Encrypted DNS –> Privacy? A Traffic Analysis Perspective. In *Network and Distributed System Security Symposium (NDSS).*
[48] Martino Trevisan, Francesca Soro, Marco Mellia, Idilio Drago, and Ricardo Morla. 2020. Does Domain Name Encryption Increase Users' Privacy? *ACM SIGCOMM Computer Communication Review (CCR)* (2020).
[49] Benjamin VanderSloot, Allison McDonald, Will Scott, J. Alex Halderman, and Roya Ensafi. 2018. Quack: Scalable Remote Measurement of Application-Layer Censorship. In *USENIX Security Symposium.*
[50] Zachary Weinberg, Shinyoung Cho, Nicolas Christin, Vyas Sekar, and Phillipa Gill. 2018. How to Catch When Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation. In *ACM Internet Measurement Conference (IMC).*
[51] Tarun Kumar Yadav, Akshat Sinha, Devashish Gosain, Piyush Kumar Sharma, and Sambuddho Chakravarty. 2018. Where The Light Gets In: Analyzing Web Censorship Mechanisms in India. In *ACM Internet Measurement Conference (IMC).*
[52] Hadi Zolfaghari and Amir Houmansadr. 2016. Practical Censorship Evasion Leveraging Content Delivery Networks. In *ACM Conference on Computer and Communications Security (CCS).*

A ABBREVIATION OF DOMAIN CATEGORY

Table 12: Domain Category Abbreviation.

Abbreviation	Category
ADUL	Other Adult Materials
ADVR	Advertising
BLOG	Personal Websites and Blogs
BUSI	Business
GAMB	Gambling
INFO	Information Technology
NEWS	News and Media
PORN	Pornography
PROX	Proxy Avoidance
SHOP	Shopping
SRCH	Search Engines and Portals