Where Are You Taking Me? Behavioral Analysis of Open DNS Resolvers

Jeman Park†, Aminollah Khormali†, Manar Mohaisen◇, Aziz Mohaisen†
† University of Central Florida, ◇ Korea University of Technology and Education
{parkjeman, aminkhormali}@knights.ucf.edu, manar.subhi@koreatech.ac.kr, mohaisen@ucf.edu

***Abstract*—Open DNS resolvers are resolvers that perform recursive resolution on behalf of any user. They can be exploited by adversaries because they are open to the public and require no authorization to use. Therefore, it is important to understand the state of open resolvers to gauge their potentially negative impact on the security and stability of the Internet. In this study, we conducted a comprehensive probing over the entire IPv4 address space and found that more than 3 million open resolvers still exist in the wild. Moreover, we found that many of them work in a way that deviates from the standard. More importantly, we found that many open resolvers answer queries with the incorrect, even malicious, responses. Contrasting to results obtained in 2013, we found that while the number of open resolvers has decreased significantly, the number of resolvers providing incorrect responses is almost the same, while the number of open resolvers providing malicious responses has increased, highlighting the prevalence of their threat.**

***Index Terms*—Open resolver, DNS, measurement, behavioral analysis**

I. Introduction

The Domain Name System (DNS) is a hierarchical distributed naming system and is a pillar of today's Internet. The primary goal of DNS is to supply a mapping between domain names and associated IP addresses. For instance, once a user types a domain name, e.g., www.example.com, into a web browser, the domain name will be mapped, by a set of DNS servers, to the associated IP address, e.g., 1.2.3.4. Almost all Internet services depend on DNS to connect users to hosts by resolving DNS queries. However, because DNS is an open system, anyone may query publicly accessible resolvers, called open resolvers. The operation of those resolvers is required in rare cases; mainly public services such as Google DNS [1] and Open DNS [2]. However, prior studies identified millions of publicly-accessible open resolvers on the Internet [3], [4]. It is shown that open resolvers are an attractive target for attackers to launch a wide variety of attacks, such as DNS amplification [5], DNS manipulation [6], *etc*.

Open resolvers can be used as a stepping stone for many attacks. For example, a report by CloudFlare highlights a 75Gbps DNS amplification DDoS attack in 2013 [7] using open resolvers in the wild. Takano *et al.* [8] also show the potential of DNS open resolvers for attacks by investigating the software version installed on those resolvers. Moreover, several previous studies demonstrated that DNS manipulation is widely used for malicious purpose by adversaries [9], [10], censorship by governments [11], or even monetary benefits [12]. These works showed that open resolvers in the wild expose their vulnerabilities to the adversaries and users alike.

To this end, we present in this work an up-to-date view of open resolvers' threats through an in-depth analysis. Unlike the prior work that only dealt with a small subset of accessible open resolvers [11], [13], [6], [14], we attempt to investigate all open resolvers over the Internet. Moreover, we focus on the behavioral aspects of open resolvers, which provides a deeper understanding of threats posed by them. Mutual reliability is the most important factor in DNS, where domain name is queried and a response is obtained. This reliability can be guaranteed only when a role-based behavior is performed. Observing the behavior of open resolvers is a measure of their security and DNS reliability as a whole.

Contribution. Our main contributions are as follows:

- We conducted a comprehensive measurement over the entire IPv4 address space to understand the behaviors and threats of open resolvers around the world. An Internet-wide measurement allows us to have an empirical understanding of DNS open resolvers independent of arbitrary generalization. We found that there are about 3 million recursive resolvers that do not require any authorization for domain name resolution.
- Through quantitative analysis, we found that many open resolvers generate DNS responses in a way that deviates from the standard. More specifically, the responses from open resolvers marked fields in DNS response header, such as the Recursion Available bit, the Authoritative Answer bit, and the response code, improperly.
- Through measurements, we report empirical results of DNS manipulation by open resolvers. By validating the open resolvers' answers, we discovered that more than 26 thousand open resolvers redirect users to malicious destinations reported as malware, phishing, *etc*.
- For a temporal contrast, we use a dataset collected in 2013. We found that the number of open resolvers has significantly decreased, while the number of resolvers manipulating responses remains the same, and the number of open resolvers providing malicious responses has rather increased. This result shows the prevalence of open resolvers as a threat, despite their decrease in number.

While our work is the first academic work that looks into surveying open resolvers and their behavior, it is not the first operational system. For example, the open resolver

project (openresolverproject.org) is the first to survey open resolvers on the Internet. However, the project falls short in two aspects. First, it does not provide any behavioral analysis of those open resolvers. Second, it is discontinued since 2017, supposedly because of the matureness of the space and the reduced number of open resolvers. In this work, we show through behavioral analysis that the threat of open resolvers is persistent as evidenced by the increasing number of malicious open resolvers, despite the overall decrease of open resolvers.

Organization. The rest of the paper is organized as follows: In section II, we provide an overview of DNS functionality and resolution, as well as an outline of open resolvers threat. In section III we describe the methodology, followed by measurement results in section IV. In section V, we present various discussion points. In section VI, we review the related works, and draw concluding remarks in section VII.

II. Preliminaries

This section provides a brief overview of DNS operation and the sequential process of domain name resolution.

A. DNS Functionality

When a user wants to access a website on the Internet, she can do that by typing the domain name corresponding to the website, such as www.example.com, into a web browser's address bar. However, computers do not communicate using domains in a string form directly, but using numerically formed addresses, e.g., Internet Protocol (IP) address. Users, on the other hand, cannot memorize numerical addresses easily. To address this issue, Mockapetris [15] introduced the basics of DNS, which enables users to easily type natural language strings instead of numeric addresses for websites. In DNS operation, a human-readable domain name is mapped into a machine-readable address, e.g., IPv4 or IPv6 address.

Due to its convenience, scalability, and resilience, DNS has become an essential component of the Internet. Many users access websites with the help of DNS without being aware of it. Such characteristics mean that DNS components with malicious intent can be a significant threat. A well-formed DNS infrastructure can be used for malicious purposes, such as creating a command & control channel of malware that uses Domain Generation Algorithm (DGA) [16]. Moreover, a miscreant may pose as an open DNS resolver and participate in the resolution process. To clarify why it can be a threat, we briefly describe the operation of DNS in the following.

B. DNS Resolution

For efficient and stable operation, DNS was designed as a hierarchical and globally distributed system that consists of the root, Top-level Domain (TLD), and authoritative name servers. Each of these servers is partly involved in converting human-memorable domain names to machine-recognizable addresses.

The overall resolution process is shown in Fig. 1. DNS resolution begins once a user attempts to access a web service using its domain name. DNS uses cache for performance, and when a domain name mapping is not cached in the local cache

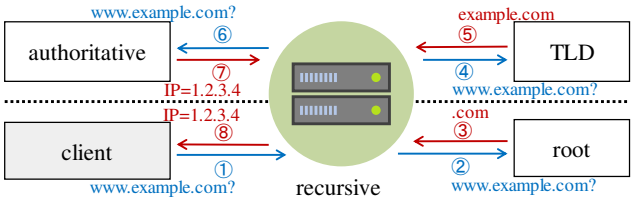

Fig. 1. Illustration of DNS resolution over recursive, root, TLD, and authoritative name server. The texts and arrows in blue correspond to DNS queries, while those in red correspond to DNS answers.

or the host table, the local resolver initiates a DNS query to the recursive resolver to retrieve the corresponding IP address to the domain name. The recursive resolver starts by asking root, then TLD, then the authoritative name servers.

Steps ② through ⑦ show the typical resolution process. The root server is the first server that receives a query from the recursive resolver, in step ②. The root servers are globally distributed and they maintain the IP addresses and location of TLD name servers. In step ③, the root name server replies to the query with the appropriate list of TLD servers for .com. In step ④ and ⑤, the recursive server sends a query for example.com and the .com TLD server responds with the IP address of the given domain's authoritative name server. In step ⑥ and ⑦, the recursive resolver communicates with the authoritative name server of example.com to find the address of www.example.com. Finally, the translated IP address of the requested domain name is forwarded to the local resolver. As a result, the browser can send a Hypertext Transfer Protocol (HTTP) request to the website to retrieve its contents.

C. Threat of Open Resolver

As described earlier, the recursive resolver is responsible for the recursive translation of domain names into IP addresses on behalf of clients. Among these recursive resolvers, open resolver is accessible by anyone on the Internet for resolution. Due to the role a typical recursive resolver plays in the resolution process, open resolvers are becoming a major threat to the security and resilience of the Internet. The rest of this section are details on how open resolver are exploited; e.g., for DNS amplification attack and DNS manipulation.

DNS Amplification Attack. The DNS amplification attack is a DDoS attack performed by exploiting the large difference between the size of a typical DNS query and the corresponding response. Originally, DNS had a packet size limited to 512 bytes. However, due to recent update [17], it is now possible to have more than 512 bytes in DNS responses.

In addition to the 'A' type query, which is commonly used to request the IP address of a webpage, there are also other types of DNS queries: 'MX' for requesting mail server information, 'CNAME' for requesting the canonical name of the server, *etc*. Moreover, 'ANY' type DNS query requests information about all domains managed by an authoritative name server including 'A', 'MX', and 'CNAME'. If the authoritative name server manages a larger number of domains, the larger DNS

response will be replied to the 'ANY' type query. Moreover, the standard DNS resolution is unauthenticated, which means it is possible for an adversary to generate a DNS query with a spoofed address as a source. Because the DNS response is returned to the source of the query, IP forgery would mean that someone who did not issue a given query may receive an overwhelming number of responses.

DNS amplification attacks use the above two features of open resolvers. 'ANY' type DNS queries with a victim's IP address as a source are sent to the open resolver, resulting in a concentration of DNS responses to the victim. An attacker can simply send hundreds of DNS queries to open resolvers to exhaust the victim's bandwidth without having to create a huge amount of packets for a direct DDoS attack. That is, in such an attack, the open resolver acts as an attack amplifier.

DNS Manipulation. Another viable threat due to open resolvers is DNS manipulation. Users typically trust the results that an open resolver provides as a result of a recursive resolution. In other words, the IP address contained in the DNS response is considered as a correct address of the given domain name. However, an attacker can exploit an open resolver to provide a manipulated result to the legitimate users. Instead of the genuine page the user wants to access, a false DNS response may mislead the user to a similar phishing page created by the attacker to distribute malicious program or to steal one's credential. Even when the attacker does not own the open resolver, he may produce the same effect by injecting the manipulated record into other existing open resolvers.

III. Methodology

The goal of this work is to answer the following questions. 1) How many open resolvers exist over the world? 2) Do open resolvers behave correctly? 3) How do such behaviors pose a threat to Internet users? To answer these questions, we analyzed DNS responses obtained using an open resolver probing system, which we describe in the following.

A. Measurement System

The overall flow of open resolver probing is shown in Fig. 2. In this figure, the flow of `Q1`, `Q2`, `R1`, and `R2` corresponds to the DNS query from the prober to the open resolver, the DNS query from the open resolver to the authoritative name server, the DNS response from the authoritative name server, and the DNS response from the open resolver to the prober, respectively. The root name server and the TLD name server are not shown in this figure because they are out of the scope of this study. The communication with both servers takes place in the time between `Q1` and `Q2` to find the address of the authoritative name server. To gather all flows in Fig. 2 during our measurements, we built and controlled two components: a prober and an authoritative name server. In the following, we elaborate on the details of each component.

1) Open Resolver Prober: A prober is responsible for sending DNS queries to the entire IPv4 address space (`Q1`) and collecting responses from open resolvers (`R2`). The `Q1`

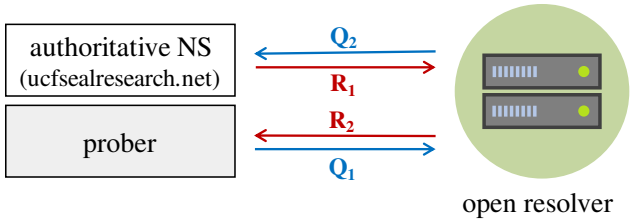

Fig. 2. The flow of DNS request and response packets among the prober, authoritative name server and open resolver. Notice that `Q1` and `R2` are captured at the prober by modified Zmap, while `Q2` and `R1` are captured at the authoritative name server by tcpdump.

TABLE I
THE EXCLUDED IP ADDRESSES FROM PROBING ACCORDING TO THE REQUEST FOR COMMENTS (RFC) DOCUMENTS. NOTICE THAT # MEANS THE NUMBER OF IPV4 ADDRESSES IN THE BLOCK.

Address Block	RFC	#
0.0.0.0/8	RFC1122	16,777,216
10.0.0.0/8	RFC1918	16,777,216
100.64.0.0/10	RFC6598	4,194,304
127.0.0.0/8	RFC1122	16,777,216
169.254.0.0/16	RFC3927	65,536
172.16.0.0/12	RFC1918	1,048,576
192.0.0.0/24	RFC6890	256
192.0.2.0/24	RFC5737	256
192.88.99.0/24	RFC3068	256
192.168.0.0/16	RFC1918	65,536
198.18.0.0/15	RFC2544	131,072
198.51.100.0/24	RFC5737	256
203.0.113.0/24	RFC5737	256
224.0.0.0/4	RFC5771	268,435,456
240.0.0.0/4	RFC1112	268,435,456
255.255.255.255/32	RFC919	1
Total	—	575,931,649

messages generated by a prober include the subdomains underneath *ucfsealresearch.net*, which is under our management.

Probing System. To perform DNS probing, we modified ZMap [18], an open-source fast Internet-wide scanner. In theory, ZMap is able to probe the entire IPv4 address space within an hour. To cope with our limited bandwidth, I/O constraints, etc., we performed a probing at 100k packets-per-second (pps). We implemented a systematic probing by combining the latest ZMap with the subdomain generation method described in section III-B.

Probing Range. In order to capture a snapshot of open resolvers on the Internet, we probed all IPv4 addresses except for some reserved areas as described in Table I. As a result, a scan of about 3.7 billion IPv4 addresses was conducted, which resulted in a comprehensive view of open resolvers.

2) Authoritative Name Server: Upon receiving a DNS query, the open resolver starts a recursive resolution. The interpretation of the domain name proceeds in the order as shown in Fig. 1. Among the components that make up the whole DNS, it is impossible to build a root name server or a TLD server by ourselves, so we built an authoritative name server to observe the behavior of open resolvers. The authoritative name server is responsible for the translation of subdomains that belonged to the Second-Level Domain

(SLD) in the DNS query. A prober generates DNS queries that include our SLD, which makes our authoritative name server participate in the recursive resolution process as a last step (⑥ and ⑦ in Fig. 1), the subdomain translation.

System Specification. We built an authoritative name server on a commercial public cloud service, Vultr [19]. The cloud instance has two 2.6 GHz virtual CPUs, with 4 GB of memory, and running CentOS 7 x64. BIND 9.9.4 was used, and the resolution of IPv6 was disabled in our configuration, since this work only focused on the IPv4 address space.

Second-Level Domain. We purchased an SLD, *ucfsealresearch.net*, from GoDaddy [20] to enable the configured authoritative name server to manage the domain name resolution of its subdomains. GoDaddy provided the option of changing responsible DNS server for the purchased SLD. We set the authoritative name server that we configured as the responsible DNS server of the purchased SLD.

B. Subdomain Generation

To understand the behavior of the open resolver, we need to keep track of `Q1, Q2, R1`, and `R2` for each open resolver. Basically, DNS matches the pair of the query and the response using the ID field in the DNS header. However, it is infeasible to assign a unique ID number to each query-response pair in our measurement. This is because the DNS ID field is only 16 bits which can represent up to 65,535 IDs, while the probing rate is about 100k pps. Therefore, we implemented and applied a subdomain generation method to deal with this issue. During the probing process, DNS queries with different subdomains (*e.g.,or000.0000000.ucfsealresearch.net*, *or000.0000001.ucfsealresearch.net,* etc.*)* are sent to different IP addresses. Using the `qname` information contained both in the DNS request and response, we were able to easily group `Q1, Q2, R1`, and `R2` for each flow.

Subdomain Cluster. Considering the limited memory resource of the authoritative name server, it cannot load about 4 billion subdomains for all IP addresses at once. In our authoritative name server, only about 5 million subdomains could be reliably loaded. Therefore, we devised a two-tiered subdomain structure for the measurement as shown in Fig. 3.

We grouped the 5 million subdomains that can be provided at once by the authoritative name server into one cluster. Five million subdomains, where each has a unique number (right 7 digits in the figure), are generated as one cluster (a zone file), and each cluster is numbered (left 3 digits in the figure). Once the predetermined number of subdomains in the cluster is exhausted, the cluster is updated with a new cluster number.

Subdomain Reuse. The application of subdomain and clustering allow us to easily match `Q1, Q2, R1`, and `R2` by comparing the `qname` field in DNS packet as well as to prevent the cached response. However, creating a cluster of 5 million subdomains also increases the probing time. To be specific, it takes about one minute to load 5 million subdomains at the authoritative name server, and the time will be very long considering that 4 billion IP addresses can make up to 800 clusters in total. Moreover, the number of open resolvers found in other projects [4], [21] was less than 10 million, which means that about 99.75% of the loaded subdomains would not be used by the open resolver. As such, we added a subdomain reuse method to improve the performance. The prober parses the response packet (`R2`) after sending the packets including subdomains within one cluster and reuses the subdomain not in the collected `R2` set, indicating that the packet was sent to the IP address which is not an open resolver. Using this approach, we could reduce the number of clusters for probing from the theoretical value of 800 to only 4.

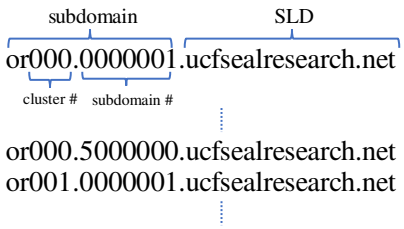

Fig. 3. The subdomain structure for open resolver probing.

IV. Measurement Results

We successfully performed an Internet-wide probing that lasted approximately 10 hours and 35 minutes, where about 3.7 billion `Q1`, 13 million of each `Q2` and `R1`, and 6.5 million `R2` packets were captured at either the prober or the authoritative name server. Compared to the number of `Q1`, the number of `Q2` and `R1` is about 0.353% and those of `R2` are only about 0.176%. Table II shows a summary of the probing results.

We compare this result with results obtained from a dataset collected in 2013. In 2013, we performed an Internet-wide measurement using a C-based system, not based on ZMap as in this study, which does not affect the settings. As shown in Table II, the probing took about 7 days for sending about 3.7 billion `Q1`. We collected about 38 million `Q2` and `R1`, and about 16.6 million `R2` packets. The percentages of `Q2` (`R1`) and `R2` to the number of `Q1` are about 1.0357 and 0.453, respectively. By observing the reduction of `Q1` and `R2` counts, we deduce that the number of open resolvers has declined over five years. In the following, we explore the change in the number of open resolvers and their behaviors in-depth.

R2 with Empty Question Field. In the sequel, we focus on `R2` to analyze the behavior of the open resolvers. However, we remark that some of the collected `R2` packets had an empty `dns_question` field. In general, the `dns_question` field is included in both the DNS query and the response [22]. As described in section III-B, `dns_question` is used to group the set of `Q1`, `Q2`, `R1`, and `R2`. Accordingly, we excluded those 494 packets without `dns_question` field from our analysis in 2018. As a result, the following analysis only covered 6,505,764 `R2` packets with `dns_question` field. However, we briefly provide a summary of those excluded packets in section IV-B4.

TABLE II
THE SUMMARY OF THE OPEN RESOLVER PROBING. NOTICE THAT THE NUMBERS IN PARENTHESES IN THE Q2 AND R2 SHOW THE PERCENTAGE OF EACH NUMBER TO THE NUMBER OF Q1.

Start time	End time	Duration	Q1	Q2, R1 (%)	R2 (%)
10/28/2013 2PM	11/04/2013 6PM	7d 5h	3,676,724,690	38,079,578 (1.0357)	16,660,123 (0.453)
04/26/2018 3PM	04/27/2018 2AM	11h	3,702,258,432	13,049,863 (0.3525)	6,506,258 (0.1757)

A. DNS Answer and Correctness

In this section, we describe a high-level analysis of the collected R2 packets, both presence and correctness. The presence of packets is simply measured by counting the number of R2 at the prober, while the correctness is measured by comparing the translated result in R2 with the ground truth.

As shown in Table III, we observed 16,660,123 R2 packets during the probing in 2013. Out of all R2 responses, 4,867,241 responses do not include dns_answer, while 11,792,882 packets contain dns_answer. Among the R2 packets which have dns_answer field, 11,671,589 packets indicate the correct IP address, but 121,293 responses include incorrect information. The rate of incorrect information is about 1.029%.

In 2018, on the other hand, we found that 2,863,655 DNS responses out of total 6,505,764 R2 collected packets had dns_answer, while the remaining 3,642,109 packets do not. Moreover, 2,752,562 of the 2,863,655 dns_answer fields contained the correct IP address result, and the other 111,093 responses had wrong results (3.879%).

From this result, we can conclude that the number of R2 packets with dns_answer has greatly decreased from about 11.8 million to 2.9 million. The reduction (≈9 million) is similar to the reduction in the R2 packets (≈10 million). Interestingly, however, the number of R2 packets providing misleading information remains similar (≈110 thousand). Consequently, the error rate has increased from about 1% to 4%. From this result, we can infer that the number of resolvers exhibiting unusual behaviors did not significantly change, despite a significant reduction in the total number of open resolvers.

We conducted a deeper analysis of the answers from open resolvers. Specifically, we looked into how open resolvers behave according to the ideal way in section IV-B, and investigated the incorrect (suspicious) answers in section IV-C.

B. Analysis of DNS Header

The operation of DNS is mainly based on RFC1034 [23] and RFC1035 [22]. These documents elaborate the standards for DNS, such as DNS packet header structure as well as the process of DNS resolution. Considering that the open resolver is a component of the whole DNS, we expect it to follow the standard when participating in the translation process.

Through this measurement, however, we found that many resolvers don't follow the standard. To be specific, when the resolvers generate DNS answer packets, many of them fill the DNS flags and the response code fields not according to the instructions described in the standard. In the following, we investigate such behaviors of open resolvers by analyzing the collected data through a comprehensive measurement.

TABLE III
THE PRESENCE AND CORRECTNESS OF dns_answer FIELD IN R2. NOTICE THAT W AND W/O CORRESPOND TO THE THE NUMBER OF R2 PACKETS WITH AND WITHOUT dns_answer, RESPECTIVELY. W_{Corr} AND W_{Incorr} CORRESPOND TO THE NUMBER OF CORRECT AND INCORRECT ANSWERS, WHICH RESULTS IN $W_{Corr} + W_{Incorr} = W$, AND Err MEANS THE PERCENTAGE OF INCORRECT ANSWERS TO THE W, SUCH THAT $Err = W_{Incorr}/W \times 100$.

Year	R2	W/O	W		Err(%)
				W_{Corr}	W_{Incorr}	
2013	16,660,123	4,867,241	11,792,882	11,671,589	121,293	1.029
2018	6,506,258	3,642,109	2,863,655	2,752,562	111,093	3.879

1) Recursion Available Flag: The Recursion Desired (RD) flag bit in the header of DNS queries sent during the probing is '1', which means that the recursive resolution is required. If the recipient of this query can perform recursive resolution (open resolver), it proceeds with recursion on behalf of the prober. Once the open resolver knows the result, it returns the translated IP address with the Recursion Available (RA) bit of '1' to the prober.

In this work, the investigation of the RA flag in R2 started to figure out how many open resolvers exist. According to the definition of open resolver, a publicly accessible and recursion-available resolver, we expected the open resolvers to answer to our query with the RA bit of 1 and a correct answer. However, by looking into the collected data we found that RA bit does not directly mean the existence of an open resolver.

Table IV shows the analysis of RA bit in R2 packets. In 2013, we can see that the RA bit of 12,270,335 R2 packets appeared as 1, which might imply that there were about 12 million open resolvers. The interesting observation we make is that there were 241,950 DNS responses with dns_answer field, even though they also have the RA bit of 0 (recursion unavailable). Moreover, 166,108 of them include the correct IP address information, which means that the senders of those 166,108 packets actually play the role of the open resolver, although they indicate they are not open resolvers. Conversely, there were 719,403 DNS responses without dns_answer field, but with the RA bit of 1 (recursion available), which means they do not perform the resolution. Such resolvers can be assumed to be either publicly inaccessible or unable to perform the recursive resolution. If the latter is the reason for the blank answer, it can be inferred that those resolvers do not

follow the standard implementation for DNS resolution.

In the result collected in 2018, 3,503,581 R2 packets have the RA bit of 0 (recursion unavailable), while 3,002,183 include RA bit of 1 (recursion available). Moreover, among the responses with RA bit of 0, 69,166 packets include dns_answer, 3,994 of correct answers and 65,172 of incorrect answers, which results in an error rate of 94%. On the other hand, 207,694 R2 responses do not contain any resolved IP address, even though they claim to have recursion available.

Regardless of the value of the RA bit, the number of packets with the dns_answer field decreased to about one quarter (from 241 thousand to 69 thousand for RA bit of 0 and from 11.5 million to 2.8 million for RA bit of 1). However, the number of incorrect answers is similar, or even larger in 2018 (from 75 thousand to 65 thousand for RA bit of 0 and from 42 thousand to 46 thousand for RA bit of 1).

In terms of the accuracy of the response, when the RA bit is 0 and the dns_answer field is included, the resolved IP address is often wrong in 2018, with 94.225% of wrong dns_answer fields. Moreover, 31.346% of responses with RA bit of 0 and dns_answer field include incorrect information in 2013. If the RA bit of 0 and dns_answer were given together, the probability that the included IP address was inaccurate increased by more than three times. When the RA bit is 1, it can be seen that about 0.393% and 1.643% of the packets containing dns_answer include the wrong result in 2013 and 2018, respectively. Considering that only less than 6% of the cases with the RA bit of 0 include dns_answer field, the ratio of incorrect responses to the total would be low. Obviously, however, an improper combination of the RA bit and dns_answer can be a clear indicator of a false result.

While investigating the RA bits in the responses, it is difficult to determine the number of open resolvers. It can be estimated that there were about 11.5 million open resolvers with the strictest criteria in 2013 (with the RA flag of 1 and correct dns_answer). Using the same criteria, it is estimated that there are about 2.74 million open resolvers in 2018. However, if we only use the RA flag as a criterion, we can also estimate that there were 12.2 and 3 million open resolvers in 2013 and 2018, respectively. On the other hand, it is also possible to conclude that there were about 11.7 and 2.75 million open resolvers by only counting the R2 packets with correct answer regardless of the RA bit.

2) Authoritative Answer Flag: Authoritative Answer (AA) means that the server responding to the DNS query is the authoritative name server for the given domain. In our probing, we sent DNS queries for all IPv4 addresses, and we were the only owner for the SLD, *ucfsealresearch.net*. Therefore, intuitively, it is appropriate to assume that the AA bit of all R2 is set to 0 except one response from our authoritative name server itself. However, as shown in Table V, 381,124 R2 packets came back with the AA bit of 1, which is about 2.29% of all responses in 2013. Among them, 231,368 responses contained dns_answer, of which 78,279 had incorrect results. The ratio of the false answers to total answers with the AA bit of 1 is about 20.539%, which is significantly high compared to 0.372% when the AA is 0.

In 2018, we received 249,193 R2 responses with AA of 1. Among them, 119,147 packets had dns_answer and 94,052 had incorrect information (79%), which is more than twice the rate of 2013, while the rate for AA bit of 0 is about 0.621%. The number of responses with AA bit of 1 is less than 4% of the total answers, while incorrect answer with AA bit of 1 account for about 84.661% of all incorrect packets.

Comparing the results of 2013 and 2018, we can see that the R2 with AA bit of 0 is greatly reduced (from 16 million to 6 million). In the case of AA bit of 1, the number of packets decreased from about 381k to 249k, which is about 61%. Nevertheless, the value of 249k is still abnormally higher than the expected value of 1. As interesting as the findings of the RA bit, however, the number of responses with inaccurate information was similar or even higher, which results in a significantly high error rate in 2018, which more than doubled from 2013.

As in the previous analysis of the RA flag, the analysis of the AA flag also shows that a large number of open resolvers do not work in a reliable manner.

3) Response Code: The response code (rcode) of the DNS response provides metadata about the outcome of the resolution. The rcode, which usually has a value from 0 to 15, is set to 0 for NoError, 1 for FormErr, 2 for ServFail, 3 for NXDomain, 4 for NotImp, 5 for Refused, 6 for YXDomain, 7 for YXRRSet, 8 for NXRRSet, and 9 for NotAuth [24]. All but 0 (NoError) indicate that the resolution was not successful.

Table VI shows the distribution of rcode in the collected R2. As expected, most responses containing dns_answer field contained an rcode of 0, while most responses had a nonzero rcode without dns_answer. Except for this general tendency, however, we found some R2 packets with abnormal combinations: 14,005 contained a nonzero (error) rcode despite having dns_answer field; 12,723 for ServFail, 10 for NXDomain, and 1,272 for Refused. Conversely, 1,198,772 R2 packets without the dns_answer field had rcode of NoError.

In 2018, we also found 2,715 R2 packets with a nonzero rcode, even with dns_answer field. In particular, 23 R2 s have rcode of 1, 2,489 have 2, 10 have 3, and 193 have 5; 377,803 responses with rcode of 0 had no dns_answer field.

In analyzing response code, we found that the number of packets with most response codes decreased (NoError, FormErr, ServFail, NXDomain, and Refused). However, we also can see that the number of responses with the rcode of 1 (NotImp) and 9 (Not Auth) significantly increased, while those with the rcode of 6 (YXDomain) and 7 (YXRRSet) remained at a similar level.

4) DNS Response with empty dns_question: We briefly describe the analysis of 494 packets without dns_question field in 2018; excluded from the previous analysis.

DNS Answer Presence. Among the 494 packets, 19 R2 packets have the dns_answer field, which is about 3.8%. However, none of the 19 packets includes the correct answer. There are 14 packets containing a private network address in dns_answer (13 for 192.168.0.0/16, 1 for 10.0.0.0/8) and

TABLE IV
THE STATISTICS OF THE dns_answer FIELD AND THE VALUE OF RA BIT IN R2. NOTICE THAT RA_0 AND RA_1 CORRESPOND TO THE VALUE OF RA FLAG BIT.

	2013					2018				
	W/O	W		Total	$Err(\%)$	W/O	W		Total	$Err(\%)$
		W_{Corr}	W_{Incorr}				W_{Corr}	W_{Incorr}		
RA_0	4,147,838	241,950		4,389,788	31.346	3,434,415	69,166		3,503,581	94.225
		166,108	75,842				3,994	65,172		
RA_1	719,403	11,550,932		12,270,335	0.393	207,694	2,794,489		3,002,183	1.643
		11,505,481	45,451				2,748,568	45,921		

TABLE V
THE STATISTICS OF THE dns_answer FIELD AND THE VALUE OF AA BIT IN R2. NOTICE THAT AA_0 AND AA_1 CORRESPOND TO THE VALUE OF AA FLAG BIT.

	2013					2018				
	W/O	W		Total	$Err(\%)$	W/O	W		Total	$Err(\%)$
		W_{Corr}	W_{Incorr}				W_{Corr}	W_{Incorr}		
AA_0	4,717,485	11,561,514		16,278,999	0.372	3,512,053	2,744,518		6,256,571	0.621
		11,518,500	43,014				2,727,477	17,041		
AA_1	149,756	231,368		381,124	20.539	130,046	119,147		249,193	78.938
		153,089	78,279				25,095	94,052		

TABLE VI
THE RCODE OF THE DNS REPONSES. Notice that each column corresponds to rcode of 0 to 7, and 9. The rcode of 8 (NXRRSet) is omitted due to the absence in our dataset.

		NoError	FormErr	ServFail	NXDomain	NotImp	Refused	YXDomain	YXRRSet	Not Auth
2013	W	11,780,575	0	12,723	10	0	1,272	0	0	0
	W/O	1,198,772	453	354,176	145,724	38	3,168,053	0	2	11
	Total	12,979,347	453	366,899	145,734	38	3,169,325	0	2	11
2018	W	2,860,940	23	2,489	10	0	193	0	0	0
	W/O	377,803	233	200,320	48,830	605	2,934,269	1	2	80,032
	Total	3,238,743	256	202,809	48,840	605	2,934,462	1	2	80,032

one with incorrect format (e.g., 0000). Moreover, 4 R2 packets had addresses which could not be found in Whois.

RA Flag. The 184 responses with RA bit of 1 were found. The 19 responses with incorrect IP address above had a RA bit of 1, and the other 165 packets did not include the resulted address even if they had an RA of 1. All the 303 packets with RA set to 0 did not contain dns_answer.

AA Flag. With AA flag, only two responses out of 494 had an AA bit of 1, and the rest did not. Only one of two responses contained dns_answer, although incorrect. The 19 R2 packets with the answer field had the AA bit set to 0.

Response Code. Among the R2 packets, 26 responses had an rcode of 0 (NoError), 1 response of 1 (FormErr), 301 responses of 2 (ServFail), 2 responses of 3 (NXDomain), and 163 responses of 5 (Refused). We can see that the failure and refusal are major reasons for the blank dns_question.

C. Incorrect DNS Answers

Here we describe further analysis on the incorrect IP addresses included in R2 packets based on the result observed in 2013 and 2018. As shown in Table III, we notice that the wrong answer was provided in 110,093 packets out of 6,506,258 R2 packets in 2018, while 121,293 packets out of 16,660,123 provided the wrond answer in 2013.

Table VII shows a summary of the incorrect answers collected through the measurement. We categorized 121,293 and 111,093 R2 packets in 2013 and 2018 with the wrong result into three types: IP address, URL, string, according to dns_answer. As a result, we found that 112,270 in 2013 and 110,790 in 2018 of the R2 packets had incorrect IP addresses, while 249 and 231 R2 packets had incorrect URLs. We also found that 10 and 72 responses include the abnormal strings such as *wild, ff, OK, 04b400000000, etc.*

Caveats. As mentioned earlier, we used a C based system in the process of collecting dataset in 2013 and stored the results in a .pcap file. While parsing the packets using the libpcap based code, we found in some packets that dns_answer was not decoded appropriately. It appears that the open resolver incorrectly filled some values in the process of creating the response packet. Among the total 16 million of R2 packets in 2013, 8,764 packets were not decoded correctly, corresponding to about 0.05% of total number.

1) Top 10 Analysis: Table VIII shows the top 10 IP addresses with the most occurrences in R2 packets in

TABLE VII
THE SUMMARY OF INCORRECT ANSWERS. NOTICE THAT $\#_{R2}$ MEANS THE NUMBER OF R2 PACKETS THAT INCLUDE THE ANSWER IN EACH FORM, AND $\#_u$ MEANS THE NUMBER OF UNIQUE VALUES APPEARING IN $\#_{R2}$.

Form	2013		2018		Example
	$\#_{R2}$	$\#_u$	$\#_{R2}$	$\#_u$	
IP	112,270	28,443	110,790	15,022	216.194.64.193
URL	249	175	231	80	u.dcoin.co
string	10	57	72	29	wild, OK, ff
N/A	8,764	-	-	-	<0x00>
Total	121,293	28,675	111,093	15,131	-

2018. The most frequently observed IP address in incorrect `dns_answer` was found in 23,692 responses, which is a domain and web hosting related company. The summed number of top 10 appearances is 50,669, which is about half the total number of incorrect `R2` responses (111,093).

By examining the top 10 addresses, we found that four of them are private networks that belonged to 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16. However, in the case of the IP addresses, we found that 74.220.199.15, 208.91.197.91, and 141.8.225.68 are located at the second, third, and fourth places in the table, with suspicious information was found from security information vendor. For IP address 208.91.197.91, for instance, Ransomware Tracker [25] states that the address is a ransomware IP, and Cymon [26] shows that malware, phishing, and botnet activities are reported for the given address as shown in Fig. 4. Therefore, 22,805 `R2` packets pointing to the IP addresses can be considered to have a deceptive `dns_answer` for malicious purpose.

For the 2013 dataset, we inspected the top-10 frequent IP addresses. The total number of `R2` packets that include those addresses is 26,514, which is almost half of the number in 2018. Specifically, in 2013, the most frequently appeared address with 9,651 `R2` packets is 74.220.199.15, the second rank in 2018, and it is the only address reported as malicious. Moreover, there are 3 private network addresses, 192.168.1.254, 192.168.2.1, and 192.168.1.1 as a second, third, and tenth places. More than 5k packets, in third place, include the address 20.20.20.20, which is owned by Microsoft, while 173.192.59.63 appeared in 995 packets (seventh rank), 221.238.203.46 in 811 packets (eighth rank), and 68.87.91.199 in 748 packets (ninth rank). As for the unusual point, 1,032 packets include 0.0.0.0.

2) Suspicious IP Addresses: Based on the possibility of malicious activities performed by open resolvers, we conducted a deeper analysis to identify the open resolvers misleading users to malicious destination. For answers of IP addresses in Table VII, we conducted an additional analysis using Cymon API [27]. From Cymon, we gathered reported information about the given addresses and judged their maliciousness. As a result, we found that there were 335 IP addresses reported as malicious. When there are multiple reports for different categories, the most frequently reported category is selected.

As shown at the right side of Table IX, the most common

TABLE VIII
TOP 10 IP ADDRESSES INCLUDED IN INCORRECT DNS RESPONSES IN 2018. 'REPORTS' IS WHETHER A SUSPICIOUS REPORT WAS FOUND WHEN QUERYING THE ADDRESS USING CYMON API.

IP address	#	Org Name	Reports
216.194.64.193	23,692	Tera-byte Dot Com	N
74.220.199.15	13,369	Unified Layer	**Y**
208.91.197.91	8,239	Confluence Network Inc	**Y**
141.8.225.68	1,197	Rook Media GmbH	**Y**
192.168.1.1	1,014	private network	N/A
192.168.2.1	741	private network	N/A
114.44.34.86	734	Chunghwa Telecom	N
172.30.1.254	607	private network	N/A
10.0.0.1	548	private network	N/A
118.166.1.6	528	Chunghwa Telecom	N
Total	50,669	-	-

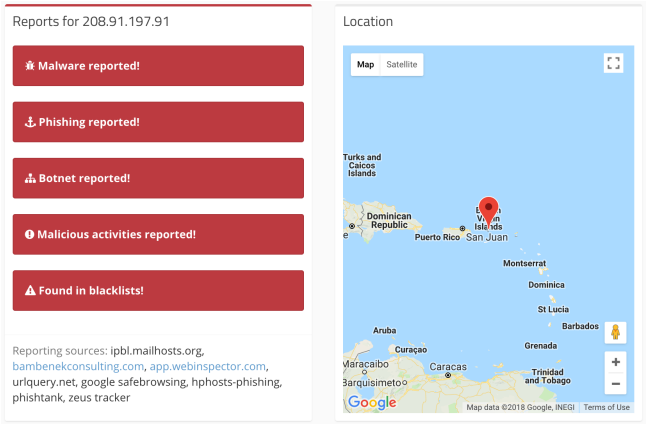

Fig. 4. The information in Cymon about the IP address of 208.91.197.91 that ranks in the third highest reference in 2018. Note the multiple reports about malware, botnet, phishing, etc. It can be assumed that the open resolvers which redirect the users to this address are exploited by the adversaries.

category for the malicious IP address is malware. The number of IP addresses related to malware is 170, accounting for over half. Moreover, the number of IP addresses related to phishing is 125, accounting for more than one third, alluding to the possibilities of DNS poisoning or manipulation. Moreover, the number of IP addresses reported as spam, SSH bruteforce, scan, and botnet is about 40. When the analysis is conducted w.r.t. the number of `R2` packets, the result is different. In `R2` packets, malware addresses account for more than 85% of the total, which means that 170 malware reported addresses are observed in `R2`, on average 136 times each. On the other hand, the 125 phishing related IP addresses are observed in 2,878 `R2` packets(≈10% of the total; 23 occurrences for each address).

To measure changes in the malicious use of open resolvers, we also conducted the same analysis on the result in 2013. In total, there were 100 unique malicious IP addresses in 12,874 responses. Among them, 65 addresses appearing in 11,149 `R2` packets were reported as malware. For addresses reported as phishing, there were 18 unique addresses that were included in 1,092 responses. In addition to the above two categories,

16 IP addresses in 633 responses were reported as Spam, SSH Bruteforce, Scan, Botnet, and Email Bruteforce.

The interesting observation we make by comparing the results of 2013 and 2018 is that the malicious behavior of open resolvers has increased from 12,874 to 26,926 in terms of the number of R2 responses. This corresponds to more than 100% of increase. From the point of view of the unique IP addresses in the R2 packets, the increase in malicious behavior is also significant: from 100 unique addresses to 335 addresses, which is more than tripled (235%).

We notice that the number of unique IP addresses reported as malware increased from 65 to 170, but the ratio to all malicious addresses decreased from 65% to 50%. The most rapid change can be found in phishing: from 19 in 2013 to 125 in 2018, which is about seven folds increase. The ratio is also doubled from about 19% to 37%, indicating that today's open resolvers are more exploitable for phishing purposes than before.

Our analysis can be considered as a lower bound of the malicious activities in that it deals only with information in Cymon. However, more malicious addresses can appear when validating using threat information from multiple vendors.

Distribution of Malicious Resolvers. To further explore malicious resolvers, we look up their geolocation and the autonomous system (AS) using ip2location [28].

As a result, we found that 12,874 malicious resolvers in 2013 were distributed over 36 countries. Specifically, 12,616 resolvers (about 98%) were the US, 91 resolvers were in TR (Turkey), 28 in VG (Virgin Islands), 24 in PL (Poland), and 18 in IR (Iran). Other 31 countries which had less than 10 malicious resolvers were (the number in the parentheses is the number of resolvers in that country): BR (9), KR, TW (8), AR (7), BG (6), ES, PT (5), AT, CA, DE, NL, VN (4), CH, RU, SA (3), AU, ID, KE, SE (2), CN, FR, GB, HK, MA, NA, NI, PR, SG, TH, VA, ZA (1). The country code by International Organization for Standardization (ISO) can be found in [29].

In 2018, there were 31 countries with malicious resolvers. 21,819 out of 26,926 (about 81%) resolvers were located in the US, 3,596 in IN (India), 714 in HK (Hong Kong), 291 in VG, 162 in AE (United Arab Emirates), and 146 in CN (China). The countries where less than 100 resolvers were located are DE (31), PL (24), RU (18), BG (16), NL (14), IE (12), AU, KY (11), CA (8), FR, GB, JP (7), CH, PT (6), IT (5), SG, TR (3), VN (2), AR, AT, ES, JO, LT, MY, and UA (1).

From the 2013 and 2018 datasets, we can see that the percentage of malicious resolvers in the US at the top rank moved from 98% to 81%, while the raw number increased from 12,616 to 21,819. Moreover, there were five countries, namely IN, HK, VG, AE, and CN, where the number of malicious resolvers has increased 10x from 2013 to 2018. It can be deduced that the wider the regional distribution of open resolvers, the more negative impact those resolvers will likely have on more people.

DNS Manipulation. The above analysis shows that DNS manipulation happens. Queries sent to each IP address were a subdomain instantaneously created, and subsequently manipulated. As mentioned earlier, one of the purposes of using subdomain is to prevent caching of results at the open resolver. In other words, the malicious IP address in the R2 packets we received does not match the information stored in the cache of the open resolver, but it is likely to be the result of an actual but illegitimate response. It is unreasonable to assume that an attacker applies a cache poisoning to the legitimate open resolver, because of the short time window, but it is more plausible to say that the open resolver itself is under the adversary's control. It can be assumed that those open resolvers will work in a way that provides the predetermined answer which includes the malicious IP address for every query they receive.

3) DNS Header in Malicious Responses: In addition to the general analysis, we also provide an analysis of R2 packets that may mislead the users to malicious IP addresses.

RA and AA Flags. Table X shows the statistics of RA and AA flags in R2 packets that contain a malicious IP address. With regard to the RA bit, more than 70% of R2 packets indicate that the senders are recursion unavailable although the responses have the dns_answer field. On the other hand, about 27% of R2 packets include the RA bit of 1, which means that the contained dns_answer fields are the result from recursive resolution. However, we already know that the IP addresses in those R2 packets are malicious and not true, which allows us to infer that RA bit is used improperly.

We also make several interesting observations from the AA bit in R2 packets. More than 70% of the responses have a AA bit of 1, which means that they are from the authoritative name server. Considering that they were not directly sent to our authoritative name server, and even they included the malicious IP address and not true result, the use of AA flag can be assumed to be a malicious attempt to allude to the credibility of the response.

Response Code. In the analysis of rcode, we found that all 26,926 R2 packets with malicious IP address have the rcode of 0 (NoError). The use of rcode can also be seen as an intention to encourage the requester to trust the response and to access the IP address by claiming a reliability of the answer.

V. Discussion

The Need for Continuous Monitoring of Open Resolvers. As show in the above analysis, open DNS resolvers still pose a threat to the Internet. The fact that the number of open resolvers has declined does not mean that their threat is going to go away anytime soon. For example, the number of open resolvers with a malicious behavior has increased, which is a clear example of the need for steady observation of those resolvers and the role they play in the DNS ecosystem.

However, and to the best of our knowledge, such a continuous and steady observation of the open resolvers on the Internet is not well performed. For example, one of the most popular open resolver-related projects is the openresolverproject.org [4], which shows the number of open resolvers distributed over the Internet and some flag values (RA bit or rcode). However, this project but does not provide

TABLE IX

MALICIOUS IP ADDRESSES IN R2 PACKETS. NOTICE THAT $\#_{IP}$ CORRESPONDS TO THE NUMBER OF IP ADDRESSES REPORTED TO CYMON IN EACH CATEGORY. WHEN THE IP ADDRESS IS REPORTED WITH MULTIPLE CATEGORIES, THE CATEGORY WITH THE MOST FREQUENCY IS SELECTED. NOTICE THAT $\#_{R2}$ MEANS THE NUMBER OF R2 PACKETS THAT INCLUDE THE IP ADDRESSES BELONGED TO EACH CATEGORY.

Report Category	2013				2018			
	$\#_{IP}$	$(\%_{IP})$	$\#_{R2}$	$(\%_{R2})$	$\#_{IP}$	$(\%_{IP})$	$\#_{R2}$	$(\%_{R2})$
Malware	65	65.0	11,149	86.6	170	50.7	23,189	86.1
Phishing	19	19.0	1,092	8.5	125	37.3	2,878	10.7
Spam	4	4.0	67	0.5	15	4.5	44	0.2
SSH Bruteforce	2	2.0	2	0	10	3.0	323	1.2
Scan	8	8.0	493	3.8	9	2.7	388	1.4
Botnet	1	1.0	70	0.5	4	1.2	102	0.4
Email Bruteforce	1	1.0	1	0	2	0.6	2	0
Total	100	-	12,874	-	335	-	26,926	-

TABLE X

RA AND AA ANALYSIS ON R2 PACKETS WITH THE MALICIOUS IP ADDRESS IN 2018. NOTICE THAT $\#_R$ AND $\#_A$ CORRESPOND TO THE NUMBER OF PACKETS WITH EACH FLAG AND VALUE. ALSO, $\%_R$ AND $\%_A$ CORRESPOND TO THE PERCENTAGE OF EACH FLAG TO THE TOTAL R2 PACKETS INCLUDING THE MALICIOUS INFORMATION (26,926).

RA	$\#_R$	$\%_R$	AA	$\#_A$	$\%_A$
RA_0	19,534	72.5	AA_0	7,472	27.8
RA_1	7,392	27.5	AA_1	19,454	72.2

any in-depth analysis of malicious IP addresses included in the responses. Moreover, and most importantly, *the project has been discontinued since January 2017*.

Another project, which is called the shadowserver dnsscan [21], provides daily information on the number and geographical distribution of open resolvers, but does not specifically analyze the behavior of each open resolver. Therefore, it is difficult to use the result of this project for understanding the threat of open resolvers in such relevant details. This is because the decrease in the number of open resolvers, as pointed by our work, does not directly mean that the associated threats are also reduced. Just as the number of open resolvers showing malicious behavior has increased, an in-depth analysis of their behavior is required for an accurate understanding of the role of open resolver on the Internet.

Censys [30] and Rapid7 [31] also provide weekly (censys) or monthly (rapid7) scan using ZMap. These raw scan datasets are more useful in that the DNS response packets can be inspected, but still have limitations. First, this raw dataset is from the measurement result only using prober, not the authoritative name server. As shown in Fig. 2, if the measurement is conducted only at the prober, we cannot catch the packet flow of R1 and Q2, which makes it difficult to investigate the behavior of open resolvers in-depth. Moreover, because both repositories use ZMap as a scanning tool, these datasets may have a blind spot that ZMap has. For example, the current ZMap can miss packets in that it only stores results for the responses from the target port of the scan (e.g., DNS responses only from port 53). This incomplete measurement can lead to the underestimation of the threat of misbehaving resolvers.

To this end, we believe a systematic and constant follow-up of the behavioral analysis in the open resolver ecosystem is a gap in the literature, and is needed for improving DNS security. For understanding the behavioral changes in open resolvers and finding countermeasures against the malicious activities such a steady observation is required.

Private Network in Incorrect Information. In Table VIII, we show that four of the top 10 IP addresses with the incorrect R2 responses in 2018 are addresses in private networks (196.168.1.1, 192.168.2.1, 172.30.1.254, and 10.0.0.1). Besides the Top 10, several private networks appeared in the incorrect responses as well.

We speculate multiple scenarios that could lead to such a behavior. For example, landing on such a private network may be a redirection to a webpage for the user's consent or form submission in a public network (e.g., Wi-Fi in airport). It is also likely to be a similar redirection in the responses with the particular company's IP address. However, in the case of a private network, and given the fact that our DNS query was sent from outside the network, this behavior is difficult to accurately understand. If it is a DNS server for users inside the network, it means that the connection is also allowed from the outside. For accurate analysis, we will conduct an in-depth analysis that focuses on these behaviors as a future work.

Open Resolver as an Existent Threat. In section II-C, we described two threats that open resolvers can bring about: DNS amplification (DDoS attacks) and DNS manipulation. In our analysis, we found that there are millions of open resolvers still exist in the wild, which allows us to deduce that these resolvers can be exploited by adversaries for launching amplification attacks. The number of open resolvers around the world can be equated to the magnitude of the potential threat as it is a threat from the functional loophole of the open resolver (no verification method for spoofed source IP address is in place). The mere existence of open resolvers and the adversary's malice are a guarantee for an attack.

However, in terms of DNS manipulation, the existence of malicious open DNS resolvers may not directly correspond to an actual threat. This is due to the passive role of open resolvers in DNS resolution. A malicious open resolver can perform its (malicious) actions only when it receives a domain

name resolution request. If no user queries the malicious open resolver, the manipulated DNS record is essentially meaningless. At this point, we need to see how malicious open resolvers are actually queried by legitimate users. Moreover, it would be further important research topic to investigate how malicious open resolvers attract legitimate users, which is our future work.

To answer these questions, we plan to conduct a follow-up analysis to investigate the actual use of malicious open resolvers with the annual Day In The Life of the Internet (DITL) collection from Domain Name System Operations Analysis and Research Center (DNS-OARC). Through the analysis combined with the DITL data, we expect to measure the actual impact of malicious open resolvers.

VI. Related Work

Open resolvers can be attacked and abused to conduct wide variety of attacks on behalf of the attackers. Therefore, researchers have done a lot of work to understand open DNS resolvers and associated threats.

Internet-Wide Scanning. Durumeric *et al.* [32] proposed ZMap, a high-speed application to run Internet-wide scans capable of surveying the IPv4 address space within 1 hour on a single machine. In addition, Open Resolver Project [4] is a project that actively investigates DNS servers world-wide since March 2013 and regularly provides open resolver statistics on the web. One can browse information of open resolvers from March 2013 until January 2017. Moreover, Shadowserver [21] is an organization that conducts surveys related to Internet security, and they also conduct active measurements of open resolvers. Takano *et al.* [8] focused on DNS server software and their distribution and performed measurements.

DNS Measurement. A large body of work exists on analyzing DNS resolvers, however most of them focused only on a small subset of resolvers. Therefore, it is unclear if the observed results can be generalized to all resolvers around the globe. For instance, Sisson [14] analyzed open resolvers based on sampled scans that repeatedly queried the same set of resolvers, thus covering only a small fraction of all open resolvers. Furthermore, Jiang *et al.* [33] analyzed the caching behavior of resolvers. The authors identified an attack vector in DNS software that allows to extend the caching of domains even after they have been removed from the upper DNS hierarchy. Schomp *et al.* [34] randomly probed the IPv4 address space to enumerate DNS resolvers and distinguish between recursive DNS resolvers and DNS proxies. Furthermore, the authors closely analyzed the caching behavior of resolvers in more detail. Gao *et al.* [35] analyzed a large set of DNS query-response pairs collected from over 600 recursive DNS resolvers. They observed that although there is a great variation in the characteristics of the DNS traffic across networks, the behavior of resolvers within an organization is very similar. In addition, Scott *et al.* [36] analyzed DNS resolutions by probing the IPv4 address space for open resolvers. Top 10,000 Alexa domain names at the identified resolvers were queried to analyze the infrastructure of the Content Delivery Networks (CDNs). By deploying the automated clustering algorithms, they detected CDN deployments in their scanning results. Hao *et al.* [37] conducted a large-scale measurement to figure out the authoritative DNS deployment patterns of modern web services and their characteristics. In addition, Thomas and Mohaisen [38] conducted a measurement of the leakage of Tor's .onion in global DNS.

DNS Manipulation and Poisoning. As the role of DNS on the Internet becomes more and more important, research on DNS poisoning or manipulation has been actively conducted. Here, we introduce a couple of representative works about DNS manipulation. Antonakakis *et al.* [6] analyzed geographically diverse set of about 300,000 open recursive DNS servers and found that attackers generally point victims to rogue IP addresses. Kuhrer *et al.* [9] tried to shed light on the negative aspect of open resolvers which can be abused by attackers. The authors measured the response authenticity of the resolvers from users' point of view and found that a large number of resolvers intentionally manipulate DNS resolutions. This work is similar to our study, but different in two points. First, we deal with more comprehensive behavioral aspects of open resolvers including how to fill the DNS header. Second, the previous work only focused on the manipulation for phishing, so they parsed and analyzed the HTTP file from the resolved address to identify the webpage for phishing purpose. In contrast, our work covers the wider scope of malicious activities including malware, phishing, botnet, *etc*. Schomp *et al.* [10] measured the vulnerability of the user-side DNS infrastructure to record injection threats and found that many open DNS resolvers, which are vulnerable to record injection attack, are being abused to attack shared DNS infrastructure. The recent work also highlighted that more than 92% of DNS resolution platforms are vulnerable to cache injection [39]. On the other hands, there have been lots of works to improve the consistency of DNS cache [40], [41], [42].

Many studies for dealing with DNS manipulation also have been proposed. DNSSEC provides the authentication and data integrity, which allows it to counter the DNS manipulation. However, DNSSEC did not yet completely replace DNS, which leaves a threat to malicious behavior on DNS. There are also studies for estimating the open resolvers supporting DNSSEC validation in real networks [43], [44]. Perdisci *et al.* [45] proposed a WSEC DNS as a solution for DNS poisoning attack. By combining the use of wildcard domain name and TXT resource, WSEC DNS can protect the recursive name server against the poisoning attack. Recently, Pearce *et al.* [11] proposed a scalable and lightweight system, Iris, to detect and measure DNS manipulation. It performs DNS queries through geographically distributed DNS resolvers and analyzes the responses.

VII. Conclusion

In this study, we conducted an up-to-date measurement of the distribution and behavior of open resolvers. Through an Internet-wide probing, we can see that about 3 million open resolvers still exist on the Internet and many of them operate

in a way that deviates from the standard. From the result, we detected two threats posed by open resolvers.

First, the presence of millions of open resolver increases the threat of a DNS amplification DDoS attack. The adversary can exploit open resolvers as an amplifier by simply sending DNS 'ANY' queries to them, which results in the concentration of large DNS answers to the victim. Moreover, we also found evidence suggesting open resolvers' abnormal behaviors. The flag bits in the DNS response from open resolvers are often inappropriately marked. More than 69k open resolvers in 2018, for example, state that they are not recursion available (RA bit of 0) although they include the result of recursive resolution. More seriously, it is also shown that over 110k open resolvers provide the incorrect IP address as a DNS response, while more than 26k open resolvers return the IP addresses reported as malware, phishing, *etc*.

A comparison between a recent scan and data obtained in 2013 shows the threat of DNS manipulation has increased, despite the decrease in the number of open resolvers. This result stresses the need of continuous observations of those resolvers to understand and mitigate their risk.

Acknowledgement. We thank our shepherd Kaustubh Joshi and the anonymous reviewers for their feedback and suggestions. This work is supported in part by NSF grant CNS-1809000 and NRF grant 2016K1A1A2912757.

REFERENCES

[1] "Public DNS - google developers." https://developers.google.com/speed/public-dns/.
[2] "Cloud delivered enterprise security by opendns." https://www.opendns.com.
[3] M. Kührer, T. Hupperich, C. Rossow, and T. Holz, "Exit from hell? reducing the impact of amplification ddos attacks," in *Proceedings of the USENIX Security Symposium*, 2014.
[4] Open Resolver Project. http://openresolverproject.org/.
[5] D. Dagon, N. Provos, C. P. Lee, and W. Lee, "Corrupted DNS resolution paths: The rise of a malicious resolution authority," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2008.
[6] M. Antonakakis, D. Dagon, X. Luo, R. Perdisci, W. Lee, and J. Bellmor, "A centralized monitoring infrastructure for improving DNS security," in *proceedings of the International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2010.
[7] CloudFlare, "The ddos that knocked spamhaus offline (and how we mitigated it)." http://blog.cloudflare.com/the-ddos-that-knocked-spamhaus-offline-and-ho, 2013.
[8] Y. Takano, R. Ando, T. Takahashi, S. Uda, and T. Inoue, "A measurement study of open resolvers and dns server version," in *Proceedings of the Internet Conference (IC)*, 2013.
[9] M. Kührer, T. Hupperich, J. Bushart, C. Rossow, and T. Holz, "Going wild: Large-scale classification of open DNS resolvers," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2015.
[10] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman, "Assessing DNS vulnerability to record injection," in *Proceedings of the International Conference on Passive and Active Measurement (PAM)*, 2014.
[11] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, "Global measurement of DNS manipulation," in *Proceedings of the USENIX Security Symposium*, 2017.
[12] N. Weaver, C. Kreibich, and V. Paxson, "Redirecting DNS for ads and profit," in *Proceedings of the USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, 2011.
[13] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig, "Comparing DNS resolvers in the wild," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2010.
[14] G. Sisson, "Dns survey: October 2010." http://dns.measurement-factory.com/surveys/201010/dns_survey_2010.pdf, 2010.
[15] P. V. Mockapetris, "Domain names: Implementation specification." IETF RFC 883, 1983.
[16] M. Thomas and A. Mohaisen, "Kindred domains: detecting and clustering botnet domains using dns traffic," in *Proceedings of the ACM International Conference on World Wide Web (WWW)*, 2014.
[17] J. Damas, M. Graff, and P. Vixie, "Extension mechanisms for DNS (EDNS(0))." IETF RFC 6891, 2013.
[18] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications.," in *Proceedings of the USENIX Security Symposium*, 2013.
[19] Vultr. https://www.vultr.com/.
[20] GoDaddy. https://www.godaddy.com/.
[21] Shadowserver. https://dnsscan.shadowserver.org/.
[22] P. V. Mockapetris, "Domain names - implementation and specification." IETF RFC 1035, 1987.
[23] P. V. Mockapetris, "Domain names - concepts and facilities." IETF RFC 1034, 1987.
[24] D. Eastlake, "Domain name system (dns) iana considerations." IETF RFC 6895, 2013.
[25] R. Tracker. https://ransomwaretracker.abuse.ch/ip/208.91.197.91/.
[26] Cymon. https://cymon.io/208.91.197.91.
[27] C. API. http://docs.cymon.io/.
[28] IP2location. https://lite.ip2location.com/.
[29] I. O. for Standardization. https://www.iso.org/obp/ui/#search.
[30] Censys. https://censys.io/data/.
[31] Rapid7. https://opendata.rapid7.com/.
[32] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications," in *Proceedings of the 22th USENIX Security Symposium*, pp. 605–620, 2013.
[33] J. Jiang, J. Liang, K. Li, J. Li, H. Duan, and J. Wu, "Ghost domain names: Revoked yet still resolvable," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2012.
[34] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman, "On measuring the client-side DNS infrastructure," in *Proceedings of the ACM Internet Measurement Conference (IMC)*, 2013.
[35] H. Gao, V. Yegneswaran, Y. Chen, P. A. Porras, S. Ghosh, J. Jiang, and H. Duan, "An empirical reexamination of global DNS behavior," in *Proceedings of the ACM SIGCOMM conference*, 2013.
[36] W. Scott, S. Berg, and A. Krishnamurth, "Satellite: Observations of the internet's star," tech. rep., University of Washington, 2015.
[37] S. Hao, H. Wang, A. Stavrou, and E. Smirni, "On the dns deployment of modern web services," in *2015 IEEE 23rd International Conference on Network Protocols (ICNP)*, pp. 100–110, IEEE, 2015.
[38] M. Thomas and A. Mohaisen, "Measuring the leakage of onion at the root: A measurement of tor's. onion pseudo-tld in the global domain name system," in *Proceedings of the ACM Workshop on Privacy in the Electronic Society (WPES)*, 2014.
[39] A. Klein, H. Shulman, and M. Waidner, "Internet-wide study of dns cache injections," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2017.
[40] X. Chen, H. Wang, S. Ren, and X. Zhang, "Maintaining strong cache consistency for the domain name system," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1057–1071, 2007.
[41] S. Hao and H. Wang, "Exploring domain name based features on the effectiveness of dns caching," *ACM SIGCOMM Computer Communication Review*, vol. 47, no. 1, pp. 36–42, 2017.
[42] X. Chen, H. Wang, and S. Ren, "Dnscup: Strong cache consistency protocol for dns," in *Distributed Computing Systems, 2006. ICDCS 2006. 26th IEEE International Conference on*, pp. 40–40, IEEE, 2006.
[43] K. Fukuda, S. Sato, and T. Mitamura, "A technique for counting dnssec validators," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2013.
[44] Y. Yu, D. Wessels, M. Larson, and L. Zhang, "Check-repeat: A new method of measuring dnssec validating resolvers," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2013.
[45] R. Perdisci, M. Antonakakis, X. Luo, and W. Lee, "WSEC DNS: Protecting recursive DNS resolvers from poisoning attacks," in *Proceedings of the IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*, 2009.