2019 IEEE 5th International Conference on Computer and Communications

Characterizing DNS Malicious Traffic in Big Data

Wenqi Sun, Songyang Wu, Tao Zhang
The Third Research Institute of
Ministry of Public Security
Shanghai, China
e-mail: sunwenqi1989@foxmail.com

***Abstract*—DNS is a quite critical network service while it is a critical attack vector. DNS vulnerabilities and attacks have been studied for many years. However, what parameters are efficient to identify specific type of DNS attacks? This question becomes more important in order to ensure DNS security in this big data era. This paper firstly characterizes main types of DNS attacks, and then gives some analytic methods to detect malicious traffic in big network data. Based on a recent dataset in a competition, analysis is carried out to verify our methods. We try to obtain some hints along with DNS malicious traffic, which might be helpful to carry out quick and efficient attack detection in huge data.**

Keywords-DNS; attacks; big data; security

I. INTRODUCTION

In current TCP/IP network world, DNS system is much more critical and important than we can say. DNS in fact is a distributed database system that translates human-friendly URLs [1], for example google.com into machine-friendly IP addresses, for example 216.58.200.238. When a user enters the Google domain name into the address bar of a browser, the user's client will look for the IP address for the domain name. The client firstly checks whether there is a record in the local cache, e.g. in host file. If there is not any record, then it sends DNS query to DNS server configured as the default DNS server. The DNS server also checks if there is a record in local cache that can answer the client. If there is not a record for the requested domain name, the DNS server will directly send a DNS query packet to the authoritative server if it knows the corresponding authoritative server for that domain name. But if the DNS server does not know which server is authoritative for the domain name, it has to resolve the name. DNS resolution can be divided into recursive or iterative resolution. Usually, the recursive resolution happens between the client the DNS server that if performing resolution for the client, which means that the DNS server will ask other servers and return an ultimate answer to the client. Iteration resolution is used among DNS servers, which means that the DNS server firstly ask the root DNS server, and the root DNS server returns the proper top-level DNS server, and then the DNS server will ask the top-level DNS server. This process goes on iteratively until the DNS server finally find the authoritative server for the queried domain name. The following figure 1 depicts general DNS working flow.

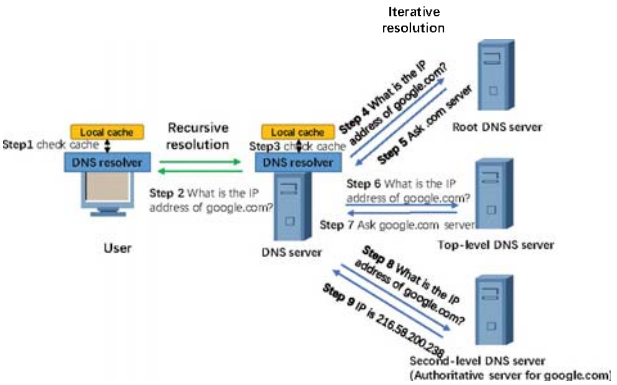

Figure 1. DNS working flow.

Although DNS is quite critical, it was designed without much considerations on security. DNS attack has always been a popular attack for these years. Exploitation of DNS can easily redirect users' traffic from a legitimate website to a fake website so that malicious events like pharming and phishing might happen. Motivated by great profits, attackers have been performing malicious behaviors towards DNS or relies on DNS for many years. There are traditional and common attacks such as cache poisoning, DDoS and amplification [2, 3, 4]. At the same time, new techniques used for attacks are continuously developed as DNS evolves to a more secure system [5]. Attacking and defensing are twin brothers. Correspondingly, there are many academic and industrial work focusing on detecting DNS malicious behaviors. [6][7] detects and prevents DNS cache poisoning, [8][9] detects DNS amplification attack and DDoS attack, [10] does researches on DNS tunneling and zone transfer. A recent work [11] analyzes and illustrates the DNS hijacking.

Different from previous related work, this paper gives an overview of how different types of DNS attacks are carried out. Based on the attacking work flows, we try to characterize DNS malicious traffic and perform analysis from a big data view. In addition, a recent dataset (a security big data analysis competition in 2019) is used to find malicious traffics and their instinct traits. How to utilize huge volume network traffic data and logs in order to find and trace malicious behaviors becomes more and more important.

978-1-7281-4743-7/19/$31.00 ©2019 IEEE

II. AN OVERVIEW OF DNS ATTACKS

A. *Cache Poisoning*

Cache poisoning means to change or add records in the resolver caches, either on the client or the server. Cache poisoning consists of traditional cache poisoning and Kaminsky cache poisoning.

(1) Traditional cache poisoning

At the beginning stage of network, DNS is designed with little security consideration. There is no verification for DNS replies. When the resolver receives a reply, it only checks the DNS ID, destination IP address and port number. When a DNS resolver receives a forged response, it accepts and caches the records without checking whether the reply comes from a legitimate source. In earlier years, DNS software like BIND always used port 53. Currently, some DNS software implementation is not so well designed that source port number randomness is not enough. Therefore, the attacker only has to guess a 16-bit pseudorandom ID in order to be successful (averagely 256 times trying by birthday attack). Detail process of traditional cache poisoning is shown in following figure 2.

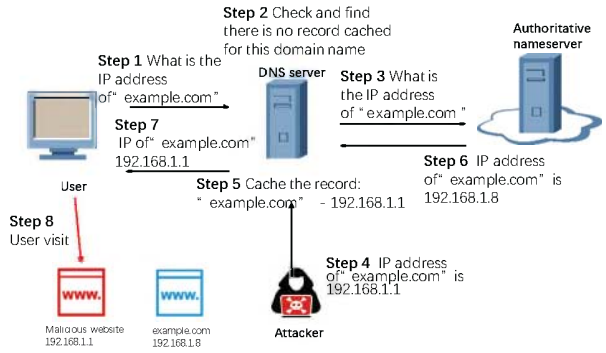

Figure 2. Traditional cache poisoning.

Traditional cache poisoning has some drawbacks. Firstly, it can only pollute domain names that are not cached. Attacks against a cached domain name have to wait until the record's TTL expired. Secondly, it poisons a single record, and it's harder to hack multiple hostnames at the same time. In addition, traditional cache poisoning is easily to be detected since it will generate a large number of forged packets destined to the resolver.

(2) Kaminsky cache poisoning

In 2008, Security Researcher Dan Kaminsky discovered a serious DNS vulnerability which can be exploited to carry out cache poisoning much more effectively. Details are shown in figure 3. In first step, the attacker requests a random name that is unlikely to be cached in the nameserver, for example, www12345678.google.com. The nameserver will request recursively. The attacker will send forged responses back which points the caching server to a fake nameserver for the domain the attacker wants to compromise. If the response packet's DNS query ID and port number matches with the information of the request packet, the nameserver will accept the packet and caches records. Unfortunately, since additional region of the DNS response packet is used to indicate a domain's authoritative server, the victim believes that attacker's server is authoritative for the entire domain: google.com. Kaminsky cache poisoning bypasses TTL limitation and gives attackers more time to forge packets and guess the query ID.

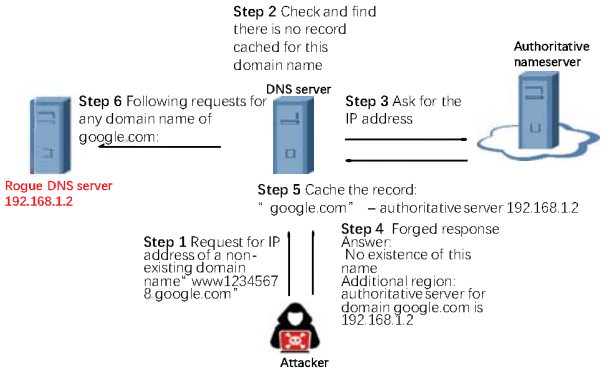

Figure 3. Kaminsky cache poisoning.

Cache poisoning attack can be alleviated by randomizing DNS source port which will increase the attacker's computation resource to guess 16-bit ID and 11-bit port correctly. DNSSEC which introduces cryptographic signature is an ultimate way to solve cache poisoning attack, but it still needs some time for worldwide deployment.

B. *DNS flood attack*

DNS flooding is a common type of Denial of Service (DoS) attack. The attacker sends a large number of requests to a DNS server to affect resolution service negatively. It becomes a Distributed Denial of Service (DDoS) attack if many malicious hosts are involved. NXDOMAIN attack is an example, the attacker floods the DNS server with requests for invalid or nonexistent records. Then the DNS server spends resources for searching non-existing things instead of serving legitimate requests. DoS and DDoS attack against a DNS server is illustrated in figure 4.

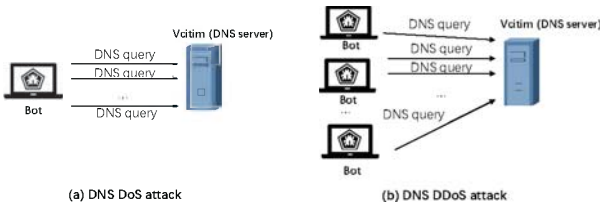

Figure 4. DoS and DDoS attack against DNS server.

C. *Distributed Reflection Denial of Service (DRDoS)*

DRDoS is also called as DNS amplification attack similar to SMURF attack [12]. DNS servers are utilized during the attack instead of being as targets. The following figure 5 depicts DRDoS process. The attacker sends DNS requests to many servers with spoofed source IP address which is the IP address of the intended victim. DNS is capable of generating a much larger response than query.

Many DNS server might send responses to the victim which results that the networking and computing resources of the victim are drastically consumed.

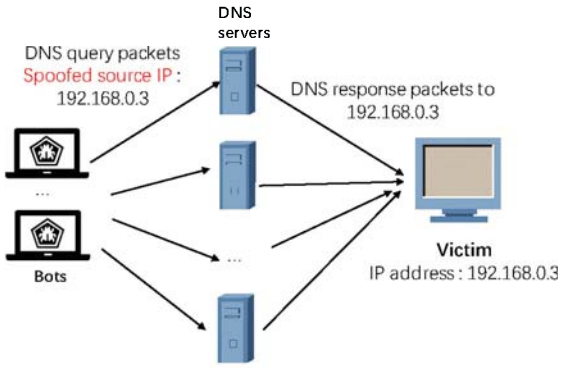

Figure 5. DRDoS utilizing DNS servers.

D. *DNS Tunneling*

Because DNS is not used for data transfer, firewalls usually do not monitor DNS traffic for finding malicious activity. Attackers use DNS tunneling for data exfiltration or other purposes. DNS tunneling relies on the DNS protocol to tunnel malware and other data through a client-server model. DNS tunneling is shown in figure 6. Attackers controls a domain and a server that can act as an authoritative server in order to execute the server-side tunneling. In addition, an internal host is infected. The infected computer is allowed to send a query to the DNS resolver. The DNS resolver relays requests to the authoritative server controlled by the attacker. Then a connection is now established between the victim and the attacker's command-and-control server through the DNS resolver. The compromised host acting as DNS tunneling client reads the data to be exfiltrated line by line. And it slices the data into small chunks and performs base64 encoding on each line. The encoded data is sent as subdomain labels suffixed with the attacker's domain name in a DNS query. The attacker's DNS server sends DNS responses back to the client with new commands encoded into Resource Record (RR) back, e.g. CNAME record. These DNS responses have low TTL values to avoid being cached. DNS tunneling is quite robust and hard to be detected.

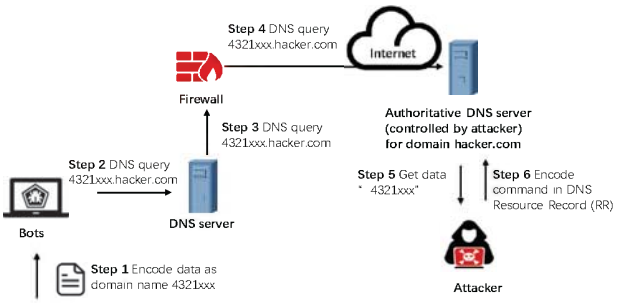

Figure 6. DNS tunneling.

E. *DNS Hijacking*

DNS hijacking or DNS redirection is to subvert the resolution of DNS queries to a malicious DNS server instead of the legitimate one, which will eventually redirect traffic to a fake website. DNS hijacking is used for a variety of reasons such as phishing, pharming and ISP's own purpose. The techniques for doing this fall into one of three categories: host-based, network-based and server-based techniques.

F. *(1) Host-Based*

The attacker simply changes the local DNS settings of the victim's computer to point to the rogue DNS server. This kind of methods needs to get access to the target or rely on malware.

(2) Network-based

This kind of method does not have to control client and server. The adversary is in your network. For example, an attacker may use ARP table cache poisoning to redirect DNS traffic to his own server. ARP table cache poisoning can be effective when the victim, legitimate DNS server and rogue DNS server are in the same LAN. Or the attacker could hack a router (acting as a NAT proxy with valid DNS settings for all clients).

(3) Server-based

A DNS server is told that the rouge server is the authoritative one for domains being hijacked.

G. *Zone Transfer*

A DNS zone is an administrative subdivision of the DNS namespace. Zone transfer is a simple mechanism to replicate DNS zone records across DNS servers. The DNS query type for zone transfer is AXFR. Any client can use AXFR to ask a DNS server for a copy of the entire zone without authentication. Therefore, DNS servers should be configured to only allow zone transfers from trusted IP addresses.

H. *Subdomain Enumeration*

When zone transfer cannot be exploited, a subdomain might be useful. Subdomain enumeration is to find valid subdomains for a domain, which is an essential part of the reconnaissance phase in the cyber kill chain. The common way is using a dictionary of common subdomain names and trying to resolve them.

I. *DNSSEC Related Attack*

DNSSEC (Domain Name System Security Extensions) adds a secure layer on top of DNS by adding cryptographic signatures to DNS records thus providing authentication. A DNS server can validate whether a DNS record comes from the authoritative name server. Ideally, it requires all DNS servers to support DNSSEC. The trust chain is built based on father and son relationship, for instance, root server keeps the hashed value of a top-level domain server's public keys to help verify the top-level domain server's public key.

NSEC and NSEC3 record types are used for authenticated denial of existence. Therefore, attackers cannot just spoof that response so that users couldn't get to the site.

But a problem is if you request for a non-existing domain name, NSEC records returns the next existing record alphabetically. NSEC walking attack is to walk through the zone and gather every single record without knowing which ones they're looking for. NSEC3 record adds a hashing mechanism so that the zone cannot be simply walked. NSEC3's drawbacks include greater cryptographic overhead for recursive validators, and more complicated DNS configuration

III. DATASET

In this paper, a recent dataset of a competition is used to performing DNS malicious traffic analysis. 2019 DataCon is a security analysis competition organized by Tsinghua University and Qihoo 360. One of the tasks is DNS malicious traffic identification. The dataset consists of 10,000,000 DNS packets. The task is to find 5 types of DNS attacks in the unlabeled data. It means that competitors have to understand the traffic characteristics of different DNS attacks.

In order to improve the analyzing efficiency, we transform pcap file into json data and put the json data on 3 HDFS nodes. Also 3 spark nodes are deployed to perform the analysis task. Spark SQL is quite efficient to perform computation on json file

IV. ANALYSIS

Firstly, we try to summarize characteristics of different types of DNS attacks. Then based on the assumed characteristics, we do analysis on the dataset to find out DNS malicious traffic.

A. *Summary of DNS Attack Characteristics*

In this part, we present some characteristics of different types of DNS malicious behavior mentioned in previous section.

Cache poisoning: if there are a large number of DNS response packets destined to one DNS server and DNS query ID in the response packets varies, it might be a cache poisoning. This method can only be used to detect a cache poisoning aiming at DNS servers. Those cache poisoning carried out by malware aiming at user host can be detected by some software installed on the host for security.

DoS or DDoS: if there are a large number of DNS query packets destined to one DNS server, DoS or DDoS might happen. And the

DRDoS: From the query's point of view, there are a large number of DNS query packets with the same source IP address and the query packets are destined to multiple DNS servers. In addition, if the query type is "ANY" (query.type is 255), there is a high probability that DRDoS happens. Because DNS response packet for this kind of requests is much larger.

DNS hijacking: detection of DNS hijacking happened at host or DNS server relies on security software installed by host or server. Detection DNS hijacking happened in network depends on the hijacking method, for instance, we may need to detect ARP cache poisoning detection to find DNS hijacking.

DNS tunneling: DNS tunneling is difficult to find because it looks like normal DNS traffic. A long and strange subdomain label might reveal some traits of this kind of attack. And the TTL of resource record is DNS response is relatively short.

Zone transfer: If DNS query packet's query type is AXFR (252) and the source IP address does not belong to an DNS server allowed to perform zone transfer, it may be a zone transfer attack.

Subdomain Enumeration: There are a large number of DNS query packets towards a DNS server, and the respective DNS response packets answer with "no such domain" (DNS flags rcode is 3). In addition, the queried domain name of those packets is the same one.

DNSSEC NEC walking: There are a large number of DNS query packets towards a DNS server and DNS response packets are NSEC type response.

B. *Analysis on Dataset*

Based on our analysis, this is a man-made dataset. Some basic information is shown in table 1.

TABLE I. BASIC INFORMATION OF THE DATASET

Time span	*Number of source IP*	*Number of requested domain name*
2019-01-25 20:29:22 ~ 2019-01-26 12:42:22	143141	174010

(1) DNSSEC

There are 71 DNS response packets with query type of 43 which means that they are DNSSEC NSEC packets. All the 71 packets come from one source IP address, and the destination is one domain name. Therefore, we can identify that the IP is a malicious source.

(2) Zone transfer

There are 5305 DNS query packets with query type of 255 (zone transfer). Among the packets 5302 are from one source IP and the number of destination IP is 1286. We can infer that this source IP is performing malicious zone transfer test towards many DNS servers.

(3) Unauthorized dynamic update

We found 5505 DNS query packets with DNS.opcode of 5 which indicates these are dynamic zone update packets. All the packets come from one IP and destined for 1597 DNS servers. It is unauthorized dynamic update.

(4) DRDoS

There are 51875 DNS query packets with query type of 255 (ANY) which have a possibility to be DRDoS. But in fact, only 33200 packets are DRDoS. Because most DRDoS packets are distributed in a short period. Malicious IPs have an abnormal QPS (query per second), so we pick out 3 abnormal source IP with 33200 DRDoS packets.

(5) Subdomain Enumeration

The source IP with the highest QPS (on average 284) generates 34194 packets requesting for 17178 domain names at only 4 servers. These are apparent subdomain enumeration malicious packets.

It is a quite clean dataset because several IP performs all the attacks. Thus, the most important thing we need to focus is to identify abnormal source IPs. Two parameters: QPS and DNS query type are quite meaningful. QPS distribution is shown in figure 7 from which we can find abnormal IP obviously.

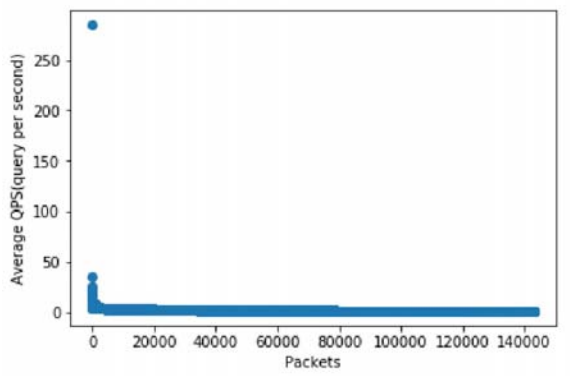

Figure 7. QPS of source IP.

V. CONCLUSION

In this paper, firstly we give a comparatively full picture of different DNS attacks. Then according to attacking methods, we try to give some characteristics of different DNS malicious traffic in order to identify DNS malicious traffic in a large volume of network data. Finally, based on a recent dataset, we try to find DNS attacks according to traffic characteristics. Although DNS malicious traffic detection has been studied for a long time, but as the data volume becomes huge, how to utilize big data and analytic methods still require more research.

ACKNOWLEDGMENT

The paper is supported by Chinese National Key Research and Development Project (2017YFC0803805); and also supported by the project of Shanghai science and technology talent plan (17XD1401300).

REFERENCES

[1] IETF RFC 1035, DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION

[2] S. Son and V. Shmatikov. The hitchhiker's guide to dns cache poisoning. In Proc. of the 6th Iternational ICST Conference on Security and Privacy in Communication Networks (SecureComm'10), Singapore, volume 50 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 466–483. Springer Berlin Heidelberg, September 2010.

[3] Ishibashi K, Toyono T, Matsuoka H, et al. Measurement of DNS Traffic Caused by DDoS Attacks[C]// Symposium on Applications & the Internet Workshops. 2005.

[4] United States Computer Emergency Readiness Team. Dns amplification attacks. Alert (TA13-088A), July 2013. https://www.us-cert.gov/ncas/alerts/TA13-088A.

[5] Hao Y, Osterweil E, Dan M, et al. Deploying Cryptography in Internet-Scale Systems: A Case Study on DNSSEC[J]. IEEE Transactions on Dependable & Secure Computing, 2011, 8(5):656-669.

[6] Yong W J, Song K H, Lee E J, et al. "Cache Poisoning Detection Method for Improving Security of Recursive DNS", International Conference on Advanced Communication Technology. 2007.

[7] Musashi Y, Kumagai M, Kubota S, et al. "Detection of Kaminsky DNS Cache Poisoning Attack", International Conference on Intelligent Networks & Intelligent Systems. 2011.

[8] Kambourakis, Georgios, et al. "Detecting DNS amplification attacks." International Workshop on Critical Information Infrastructures Security. Springer, Berlin, Heidelberg, 2007.

[9] LG Chen, YD Zhang, and Q Zhao, et al., "Detection of DNS DDoS Attacks with Random Forest Algorithm on Spark", the 13th International Conference on Future Networks and Communications, FNC-2018 and the 15th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2018.

[10] Adam Ali.Zare Hudaib & Esra'a Ali Zare Hudaib, "DNS Advanced Attacks and Analysis", International Journal of Computer Science and Security (IJCSS), Volume (8) : Issue (2) : 2014

[11] VISSERS, T., BARRON, T., VAN GOETHEM, T., JOOSEN, W., AND NIKIFORAKIS, N. The wolf of name street: Hijacking domains through their nameservers. In Proceedings of the 2017 ACM SIGSAC Con ference on Computer and Communications Security (2017), ACM, pp. 957–970.

[12] Kumar S. "Smurf-based Distributed Denial of Service (DDoS) Attack Amplification in Internet", International Conference on Internet Monitoring & Protection. 2007.