DNS based Measurement and Analysis of IPv6 Domain Name Support

Feng Li¹, Huaping Cao²*, Weijun Ma³, Chao Hu⁴, Zhanfeng Wang⁵, Xiyu Lu²

¹ Nanjing NO.1 Military Representative Department of Equipment Development Ministry, Nanjing 210007, China
² National Computer Network Emergency Response Technical Team/Coordination Center of China (CNCERT/CC), Beijing 100029, China
³ Nanjing Goldenwater Shangyang Information Technology Co. Ltd, Nanjing city, 210014, China
⁴ College of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China
⁵ School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

Email of all the authors: 48810972@qq.com, caohuaping@cert.org.cn, 13981671@qq.com, huchaonj@126.com, hehengw@hotmail.com, luxiyu@cert.org.cn

* Corresponding Author: Huaping Cao; email: caohuaping@cert.org.cn; phone: 13910763332; fax: 010-82992020

Abstract: At present, the IPv4 addresses has been fully allocated, and the worldwide deployment of IPv6 addresses is accelerating due to its advantages. In order to investigate current development of IPv6, the IPv6 addresses of domain names are firstly resolved by DNS servers, and the obtained IPv6 addresses data, mainly including the geographic location statistics of IPv6, the proportion of southeast coastal cities, the comprehensive statistics of top-level domain names, etc., can be used to analyze the domain name supported by IPv6. The analysis of the measurement results has great guiding significance for further understanding the future development and application of IPv6.

1. Introduction

Both IPv4 and IPv6 play a very important role in our daily life, and the measurement and analysis on IP network also have great guiding significant for understanding the status of IP development. In order to fast scan the IPv4 cyberspace, many researchers focus on improving the measurement speed of network probes. Nowadays, Masscan [1] can scan the entire IPv4 address space within 6 minutes, while ZMAP [2] designed by Durumeric et al. can also scan the entire IPv4 address space within 45 minutes.

On the afternoon of November 26, 2019, the European Network Coordination Center (RINCC), which is responsible for the allocation of Internet resources of the UK, Europe, the Middle East and some Central Asia, announced that all 4.3 billion IPv4 addresses have been allocated, which means that no more IPv4 addresses can be assigned to Internet Service Providers(ISPs) and other large network infrastructure providers. Moreover, the lack of effective service quality and security guarantee of...
IPv4 has become increasingly prominent. As a result, the next generation Internet Protocol, i.e. IPv6, began to accelerate the promotion and deployment around the world. At the same time, the studies on IPv6 address are also becoming one of research hotspots in computer networking [3].

With the rapid development of IPv6 technology, the cyberspace surveying, topology discovery and security of IPv6 network have become very important research issues, and the most critical and basic research approach is network measurement. Compared with the $2^{32}$ address space in IPv4 network, IPv6 has $2^{128}$ addresses. Due to the huge address space of IPv6, some traditional measurement methods cannot be used in IPv6 space, such as network-wide exhaustive scanning. Under the condition of 10 Gigabit link, ZMAPv6 with the highest configuration mode needs at least hundreds of millions of years to scan the whole network space of IPv6, which is impossible to practice.

In recent years, many researchers have proposed several effective IPv6 address detection methods (Non-exhaustive scanning), which are as shown in Fig. 1. Active measurement technologies refer to the use of technical means that will involve the operation of sending data packets, mainly including the domain name system (DNS) data based approach and the seed address based approach. The most important difference between them is whether they furtherly analyze the collected addresses and generate new addresses for active measurement.

![IPv6 address measurement](image)

**Figure 1. The Approaches of IPv6 address scanning**

It is an effective way to collect IPv6 address through DNS. The DNS data based approach mainly gathers IPv6 address by sending DNS query data packets actively. Strowes proposed a method to obtain the IPv6 addresses by reverse querying the IPv4 address and then performing AAAA query [5]. In addition, Borgolte et al. proposed a method to collect IPv6 address by combining the zone file of domain name system security extensions (DNSSEC) [6].

Another measurement approach is based on seed address. The seed addresses refer to the IPv6 addresses that have been or being alive collected by researchers. Ullrich et al. presented a technology of generating new addresses from seed address by particular rules in 2015 [7]. Gasser et al. published all datasets and codes [8], which is helpful for future researchers to continue exploring the field of IPv6 scanning based on these data and methods. However, there are still many problems, such as the availability of seed addresses.

Besides active measurement, passive measurement, e.g. collecting data from the IPv6 address sources published on the network [9,10], is also a very common means in the field of network research. In the core node of IPv6 backbone network, network monitoring and capturing packets to collect IPv6 addresses is also a very effective way. However, the conditions are more stringent for most of researchers. In addition, [11] proposed a method to collect IPv6 addresses through crowdsourcing platform.

Seed address based approach often relies on other methods to provide IPv6 seed addresses as the input of the algorithm. At the same time, the method of passive collection has great limitations, and not only the number of collected addresses is limited, but also there are not enough conditions for researchers to implement passive collection in many cases. In contrast, the DNS data based approach with the auxiliary verification of curl command can mine a large number of surviving addresses in the huge IPv6 address space, which is the focus of this paper.
2. The Process of DNS based Measurement of IPv6 address

The DNS data based approach is an active measurement method, which can send and receive a large amount of data in IPv6 network space. In this paper, we adopt this approach to collect the original information of IPv6 address of websites’ domain name, and summarize the development status of IPv6 in China. Firstly, we introduce the principle of DNS resolution, and the domain name query request can be sent to the DNS server through the client, and then the returned client resolution record and the information related to the website address accessed by curl command, which can be combined to determine whether it is IPv6 address or not. Based on these confirmed IPv6 addresses, we process and analyze the data. From the perspectives of the geographical location, the proportion of southeast coastal cities, top-level domain name data and other characteristics, we analyze the current deployment of IPv6 of China in detail.

2.1. The Principal of DNS Resolution

The process of translating domain name to IP address is called domain name resolution. There are two resolution methods for domain name resolution, which are recursive query and iterative query, respectively.

When client sends a domain name query request to the DNS server, the DNS server firstly queries whether the local resource record has a domain name query record. If there is a domain name query record, the authority resolution record is returned. If the local resource record does not have a domain name query record, DNS will query the local cache to check if there is a domain name query record. If there is a domain name query record, it will return client back with a non-authoritative resolution. If there is no query domain name record in the local cache, recursive query or iterative query is performed to the remote server. Next, we query the domain name, i.e. seu.edu.cn, as an example, and this paper introduces the process of recursive query and iterative query.

Recursive query: (1) local DNS works like a client and sends query request to root domain name server; (2) after receiving domain name query from local DNS, root domain name server will forward the query request to the top-level domain name server of .CN; (3) after receiving a domain name query request from the root domain name server, the top-level domain name server of .CN will send the domain name query request to the authoritative domain name server of edu.cn; (4) when the authoritative domain name server of edu.cn receives the query request, it will return the query results to the higher level server, i.e. the top-level domain name server, and finally the results will be forwarded to the client, which will cache the results in the local DNS server.

Iterative query: (1) the local DNS server continues to send the domain name query request to the root domain name server, and the root domain name server returns the IP address of the top-level domain server of .CN; (2) the local DNS server sends the query request to the top-level domain server of .CN, and the top-level domain name returns the IP address of the authoritative server of secondary domain name edu.cn; (3) the local DNS server sends the query request to the authoritative server of edu.cn; (4) the authoritative server returns the client with query results, and the local DNS server caches the query results.

2.2. The Format of DNS Messages

DNS service mainly solves the problem of mapping domain name to IP address. The format of DNS message is defined by RFC1035. The format of DNS request and response are the same, mainly including message header and message body.

2.2.1. Message Header

The message header of DNS packets is composed of the following parts:

- Session ID: 16 bits. It is the ID identification of DNS message. For the response message of a request message, this field is the same. Different domain name request messages can be distinguished by this field.
• Flag: 2 bits. It indicates query, authorization, truncation and recursion information of message. The flag field is shown in Fig. 3.
• All of the number of questions, the number of answer resource records, the number of authorized resource records and the number of additional resource records account for 8 bits, which represent the number of query question area, answer area, authorization area and additional area, respectively.

2.2.2. Message Body
Queries area: query name and query type can be specified in the queries area of message body. The specific message format is as follows.
• Name: The content to be queried, whose length is not fixed and cannot be filled with bytes.
• Query type: The value type, notation and description.

It can be clearly found that whether the DNS service can resolve the IPv6 address is an important premise and basis for judging if the domain name supports IPv6 address. If the IPv6 address is resolved, the type field is set to 28 (AAAA). Therefore, we can use Google, telecom, Unicom, mobile and other commonly used DNS servers to detect the IPv6 address resolution of all domain names in the country, and send UDP packets in DNS protocol format to 53 port of the specified DNS server and wait for the response of the server. If a reply is received, the received message will be resolved according to the protocol format to obtain the returned IPv6 address.

2.3. CURL for Auxiliary Verification
The HTTP connection status and page information of the website can be obtained by accessing the website address through the curl command. At the same time, the curl command supports the parameter "-6", which means the domain name can be resolved to an IPv6 address for access. This approach can solve the problem of IPv4 network interference of dual stack nodes.

The curl command can be used as a supplement to the measurement method of domain name IPv6, which is one of the effective auxiliary verification way. However, because the detected IPv6 data needs to be updated continuously for a long time, and the data detected by this method is not comprehensive, the measurement method only can be used as the relevant auxiliary verification instead of the main judgment basis.

3. The Nation-wide Address Measurement and Analysis
Using the DNS servers commonly used by Google, China Telecommunications, China Unicom and China Mobile, the IPv6 address resolution measurement and analysis of all domain names (except Hong Kong, Macao and Taiwan) are carried out. Combined with the geographical location, the proportion of southeast coastal cities and top-level domain names, the deployment status and future development direction of IPv6 in China are comprehensively analyzed.

3.1. Statistics on Geographical Location
Using the DNS servers commonly used by Google, telecom, China Unicom and China Mobile, the IPv6 address resolution of all domain names in China (except Hong Kong, Macao and Taiwan) is detected. A total of 125284 IPv6 addresses were counted, while the total number of registered domain names in China was 6511042, accounting for about 1.92%. Table 1 shows the deployment of IPv6 in China's provinces.

| Province | The number of IPv6 address found by DNS based approach | The number of registered domain name | Proportion |
|----------|-------------------------------------------------------|-------------------------------------|------------|
| Guangdong | 21,026                                                 | 972,532                             | 2.16%      |
| Jiangsu  | 13,781                                                 | 692,707                             | 1.99%      |
| Beijing  | 9,807                                                  | 692,882                             | 1.42%      |

Table 1. IPv6 deployment in China's provinces
From Table 1, we can find that the distribution of IPv6 in China's provinces is skewed. The more developed regions such as Guangdong and Jiangsu have more IPv6 addresses, while correspondingly, the economically backward regions such as Xizang and Qinghai have fewer IPv6. Generally speaking, the number of IPv6 addresses in southeast coastal cities is relatively dense.

### 3.2. Proportion of Southeast Coastal Cities

Based on the detailed research and analysis of southeast coastal cities (Shanghai, Tianjin, Liaoning, Hebei, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, Guangxi, Hainan) and Beijing, we can get the proportion of IPv6 in these cities.
Figure 2. Analysis of the proportion of IPv6 in southeast coastal cities and Beijing

From Fig. 2, we can find that the total number of IPv6 in southeast coastal cities and Beijing has accounted for about 70% of the total number of detected IPv6. In other words, about 70% of the data in China are concentrated in these cities. It indicates that the distribution of IPv6 resources in China is extremely uneven. At the present stage of vigorously developing IPv6, we need to pay more attention to the further distribution of IPv6 network.

3.3. Statistics on Top-Level Domain Name

In addition to the statistical analysis of geographical location of IPv6 and the proportion of southeast coastal cities, we also focus on the nationwide top 10 domain names to implement the relevant statistics and analysis.

Table 2. Statistics of top ten IPv6 domain names in China

| Domain name | The number of domain name found by DNS resolution | The number of registered domain name | Proportion |
|-------------|-----------------------------------------------|-----------------------------------|------------|
| .com        | 80898                                         | 3779676                           | 2.14%      |
| .cn         | 34813                                         | 1843505                           | 1.89%      |
| .net        | 4646                                          | 280924                            | 1.65%      |
| .cc         | 1015                                          | 52431                             | 1.94%      |
| .org        | 963                                           | 44505                             | 2.16%      |
| .top        | 798                                           | 192074                            | 0.42%      |
| .me         | 249                                           | 9866                              | 2.52%      |
| .vip        | 202                                           | 22014                             | 0.92%      |
| .xyz        | 200                                           | 36944                             | 0.54%      |
| .hk         | 99                                            | 3716                              | 2.66%      |
| Total       | 123883                                        |                                   |            |

From Table 2, we can find that the top ten top-level domain names, and .com, .cn, .net, etc. are the most top-level domain names in China. Most of the domain names are .com and .cn, while the number of IPv6 supported by common top-level domain names such as .edu is few.

4. Conclusions

IPv6 address measurement has important reference significance for in-depth understanding the development and application status of IPv6 address. In this paper, we use DNS based active measurement method to derive the IPv6 addresses in China, and adopt curl command for auxiliary verification. The collected data are processed and analyzed, and compared and summarized from the perspectives of the
geographical location, the proportion of southeast coastal cities, top-level domain name and other characteristics. On the basis of existing research results, we think there are still some problems for DNS based IPv6 measurement, which will be studied in our future research.

(1) The number limitation of collected IPv6 addresses
The advantage of collecting IPv6 addresses through DNS is that many of them are the addresses of servers providing particular services, and these addresses are stable and fixed, but the disadvantage is that the number of IPv6 addresses found in our paper is quite limited. It can be found from the above figures that the average level of the detected data is only about 2% of the total number of registered domain names in China. The combination of active measurement and passive measurement can be used to discover more IPv6 addresses.

(2) The increase of the IPv6 address number lengthens the scanning time
Due to the exhaustion of IPv4 address resources, it is an irresistible trend to deploy large-scale IPv6 network in the world. However, there are many random cases in the last 64 bits allocation strategy of IPv6, and its address set will continue to increase, which will lengthen the scanning time. It is a valuable issue to reduce the measurement cycle in the future study.

Acknowledgements
This research was financially supported by The National Key Research and Development Program of China (No. 2018YFB1800200), CERNET Innovation Project (No. NGII20170406), Foundation for Training Postdoctoral Innovative Talents in Southeast University (No. 2242019R20024), Zhejiang public welfare Technology Application Research Project (No. LGG20F020014).

References
[1] Durumeric, Z., Bailey, M., Halderman, J.A. (2014) An internet-wide view of internet-wide scanning. In: USENIX Conference on Security, San Diego, pp. 65-78.
[2] Durumeric, Z., Wustrow, E., Halderman, J.A. (2013) ZMap: fast internet-wide scanning and its security applications. In: USENIX Conference on Security, Washington D.C., pp. 605-620.
[3] Wu, J.P., Li, X. (2012) Next generation internet. In: Publishing House of Electronics Industry, Beijing.
[4] RFC 7707. (2016) Network reconnaissance in IPv6 networks. https://tools.ietf.org/html/rfc7707.
[5] Ottow, C., Vliet, F., Boer, P. (2012) The impact of IPv6 on penetration testing. In: Information and Communication Technologies, Heidelberg, pp. 88-89.
[6] Strowes, S.D. (2017) Bootstrapping active IPv6 measurement with IPv4 and public DNS. Computing Research Repository, 2017: 1-7.
[7] Borgolte, K., Hao, S., Fiebig, T., et al. (2018) Enumerating active IPv6 hosts for large-scale security scans via DNSSEC-signed reverse zones. In: IEEE Symposium on Security and Privacy, Francisco, pp. 770-784.
[8] Ullrich, J., Kieseberg, P., Krombholz, K., et al. (2015) On reconnaissance with IPv6: a pattern-based scanning approach. In: 10th International Conference on Availability, Reliability and Security, pp. 186-192.
[9] Gasser, O., Scheitle, Q., Foremski, P., et al. (2018) Clusters in the expanse: understanding and unbiasing IPv6 hitlists. In: Internet Measurement Conference, Boston, pp. 364-378.
[10] Liu, Z.Z., Xiong, Y.Q., Xin, L., et al. (2019) 6Tree: efficient dynamic discovery of active addresses in the IPv6 address space. Computer Networks, 155: 31-46.
[11] Qasim, L., Matthew, L., Maciej, K., et al. (2018) Using crowd sourcing market places for network measurements: the case of spoofer. In: Network Traffic Measurement and Analysis Conference, Vienna, pp. 1-8.