Measurement and Analysis of Adult Websites in IPv6 Networks

Shize Zhang*, Hui Zhang*, Jiahai Yang*, Guanglei Song*, Jianping Wu*
*Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing, China
*Beijing National Research Center for Information Science and Technology
Email: zsz16@mails.tsinghua.edu.cn, hzhang@cernet.edu.cn, yang@cernet.edu.cn
The corresponding author: Jiahai Yang

***Abstract*—The Internet is in the transition from IPv4 to IPv6. At present, researches on IPv6 networks mainly focus on architectural issues, such as routing, addressing, and security; there are few studies on the operational issues of IPv6 networks. Our preliminary observation shows that there are a large amount adult websites and traffic in IPv6 networks. Adult websites can damage health of teenagers and bring operational issues in IPv6 networks. This paper conducts a comprehensive measurement and analysis of the adult websites and traffic in IPv6 networks to help solve these operational issues. The data used in this paper is the raw packet traffic from CNGI-CERNET2 which is a pure IPv6 academic network in China. The duration of the data is from July 2017 to January 2018 and the total amount is 40+ terabytes. We detected about 3000 adult websites in the global IPv6 network. This paper analyzes these adult websites and traffic from the perspectives of websites, users and ISPs respectively. We find that adult websites are still in the developing stage in IPv6 networks and only 30% adult websites with full resources can be accessed in IPv6-only networks. But due to the IPv6-first policy in RFC 4038, adult traffic will continue to migrate to IPv6 networks from IPv4 networks. On the other hand, we find that CDN vendors promote the development of adult websites in IPv6 networks and many adult website owners use muti-domain policies to escape ISPs restricting. Our findings may help ISPs effectively understand adult websites and enhance the restriction of adult content in IPv6 networks.**

***Index Terms*—measurement, network management, pornographic websites, IPv6 networks**

I. Introduction

At present, the Internet is in the transition from IPv4 to IPv6. According to the latest statistics released by Google, about 19.62% of users access Google website with IPv6 in July 2017; while the proportion is 12.53% in July 2016, showing 56% growth in one year period [1]. This rapid growth has led to many studies on IPv6 networks, including protocol security in IPv6 [2]; IPv6 address structure analysis [3], [4]; IPv6 deployment measurements [5]. However, there are few studies on the application in IPv6 networks, especially on pornographic websites and traffic in IPv6 networks. Pornographic websites, as a kind of special network application, can damage the mental and physical health of teenagers. In some countries, such as China, pornographic content is illegal. So operators need to detect and restrict pornographic websites and traffic according to law. In addition, pornographic video traffic consumes a large proportions of critical link bandwidth(such as 18% of Internet traffic in some areas is pornographic traffic [4]). Therefore, a comprehensive measurement and analysis of pornographic websites and traffic in IPv6 networks is very important and significant to network operation and management.

This paper conducts a comprehensive measurement and analysis of pornographic websites and traffic in a nationwide pure IPv6 network. The data used in this paper is the raw packet traffic from one of the international links of CNGI-CERNET2 which is a pure IPv6 academic network in China. The duration of the data is from July 2017 to January 2018 and the total amount is 40+ terabytes. In order to detect pornographic websites and traffic from the raw traffic, we implemented a pornographic website detection system based on Naïve Bayes algorithm. And we detect about 3000 pornographic websites by running our implemented detection system.

We conduct a comprehensive measurement and analysis of these pornographic websites and traffic from three different perspectives: websites, users and ISPs. We find that adult websites are still in the developing stage in IPv6 networks and only 30% of all adult websites with full resources can be accessed in IPv6-only networks. However, the adult websites with full resources more than 95% can be accessed in IPv4 networks. But due to the IPv6-first policy in RFC 4038 [10], adult traffic will continue to migrate to IPv6 networks from IPv4 networks. On the other hand, we find that the support for IPv4/IPv6 dual stack by CDN vendors promotes the development of adult websites in IPv6 networks and more than 30% of adult website owners use multi-domain policies to escape ISPs restricting. In addition, We analyze the habit of users visiting adult websites.

The main contributions of this paper are as follows:

- To the best of our knowledge, this paper is the first to conduct a comprehensive measurement and analysis of pornographic websites and traffic in a nationwide pure IPv6 academic network in China. The duration of our data sets is from July 2017 to January 2018 and the total amount is 40+ terabytes.
- By running the pornographic website detection system we implemented, we detect about 3000 adult websites in the global IPv6 network. The paper provides a com-

prehensive analysis of these pornographic websites and traffic from the three different perspectives: websites, users,and ISPs.
- We find that adult websites are still in the developing stage in IPv6 networks and will continue to increase due to the IPv6-first policy in RFC 4038. Besides, we find the promotional role by CDN vendors and the muti-domain policies used by adult website owners. Our findings may help ISPs effectively understand adult websites and enhance the restriction of adult content in IPv6 networks.

The rest of this paper is organized as follows: Section II describes the related works about IPv6 network, pornographic websites and pornographic traffic; We provide a detailed description of our datasets in section III ; Section IV introduces the pornographic websites detection system based on Naïve Bayes algorithm; We discuss our measurement and analysis of pornographic websites and traffic in section V; Section VI is the summary and future work.

II. RELATED WORK

There are a lot of researches on IPv6 networks at present, but most of which are aimed at the architecture issues of IPv6 networks. For example, Plonka D et al. conducted a measurement analysis of the IPv6 active addresses in temporal and spatial dimensions [3]. Czyz J et al. studied protocol security issues in IPv6 networks which shows that protocols used in IPv6 networks, such as ICMP, SSH, Telnet, etc., may have vulnerabilities that do not exist in IPv4 networks [2].

However, these efforts are more about the research of IPv6 networks architecture, and very few are about the measurement and analysis of the applications in IPv6 networks, especially the pornographic websites and traffic. we think the main reason is the lack of IPv6 traffic data, especially the packet level data. While our research tries to bridge the gap. Pornographic websites, as a special kind of application, have an important influence on the operation and management of the networks. Therefore, there are some previous research works on pornographic websites, but most of which focus on some popular pornographic websites rather than a wide range of measurement in IPv6 networks.

Tyson G et al. conducted a measurement analysis of YouPorn, the most popular Porn 2.0 website [8]. Porn2.0 website is similar to popular video websites such as YouTube, allowing users to upload, view, comment for free. In this study, the authors crawled the contents of the YouPorn website and conducted a measurement analysis. The analysis results show that YouPorn contains 183k videos, with a total of over 60 billion views, and the average number of pageviews for YouPorn is about 7 times more than for YouTube. So we can see that videos from the YouPorn website is really popular.

Ahmed F et al. conducted a measurement study of online adult traffic by using HTTP logs collected from a major CDN vendor [18]. The HTTP logs include traffic from dozens of major adult websites and account for approximately 323 terabytes worth of traffic from 80 million users. The authors selected five adult websites to conduct measurement and analysis. Their analysis of the aggregate showed that nearly 99% of pornographic traffic is consisted of videos and images, and most of the users access adult websites by using desktop. Further, their analysis of adult website contents shows that images in adult websites are typically less than 1 megabyte and video is typically on the order of dozens of megabytes. In addition, the user analysis of adult websites shows that users have a certain degree of loyalty to the browsing of adult websites. These analyses are very important for CDN vendors and ISPs to provide efficient services. However, all of these analyses are only based on few popular adult websites. In contrast, we conduct a comprehensive measurement and analysis of a large number of adult websites and traffic in IPv6 networks. Besides the measurement and analysis, there are also many studies on the detection (e.g. [19]) and cybersecurity issues of adult websites(e.g. [20]).

Note that adult websites can damage the mental and physical health of teenagers and are illegal in China. ISPs have the right to detect and restrict pornographic websites and pornographic traffic under Chinese law. Therefore, the purpose of this paper is to help ISPs understand adult websites and enhance the restriction of adult content in IPv6 networks, rather than to promote the development of adult websites in IPv6 networks.

III. DATA

CNGI-CERNET2 is an important part of China's next generation Internet demonstration project [14]. The backbone of CNGI-CERNET2, put into operation since 2004, is the world's largest pure IPv6 Internet backbone. CNGI-6IX, its interconnection center, is currently the only IPv6 international exporter in China. CNGI-CERNET2 includes 25 core nodes distributed in 20 cities in China and the bandwidths between core nodes is 2.5G-10Gbps. As of December 2016, the total traffic of CNGI-CERNET2 backbone reaches 50.16Gbps. There are 19 32-bit prefix IPv6 address blocks in CNGI-CERNET2 [15]. The data set used in our research came from one of the international links of CNGI-CERNET2. The duration of the data is from July 2017 to January 2018 and we store the raw traffic of the 10Gbit/s link with a 100: 1 sampling ratio. The total amount is 40+ terabytes, including more than 30 billion packets and covering about 5 million users.

IV. PORNOGRAPHIC WEBSITE DETECTION SYSTEM

In order to analyze the pornographic traffic in IPv6 networks, we design and implement a pornographic website detection system to filter the pornographic traffic from raw traffic. We use the Naïve Bayes algorithm in the system to classify the content of websites to filter the pornographic traffic from the huge amount of raw traffic. Note that this paper focuses on the measurement and analysis of pornographic websites, not on the detection algorithms. Therefore, in this section, we briefly introduce our pornographic website detection system to explain how we detect pornographic websites and pornographic traffic from the raw traffic in IPv6

networks. The design of the detection algorithm will not be introduced in depth.

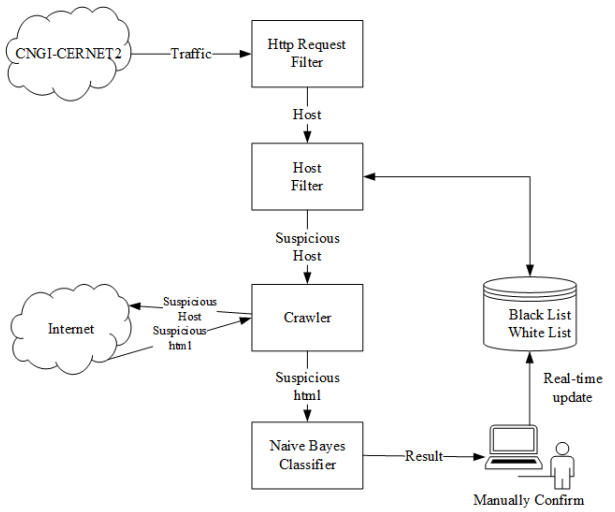

Fig. 1. The framework of pornographic websites detection system

The framework of our detection system is depicted in Figure 1. We use the raw traffic collected from CNGI-CERNET2 as input to the system. *Http Requester Filter* is responsible for filtering the http request from the raw traffic, extracting the host field from http header, and passing the host to the *Host Filter*. Then, the host is detected in the *Host Fiter* by using existing blacklist and whitelist stored in the database. If the host is in the list, it will be sent out detection system for statistical analysis. And if not, it is considered as a suspicious host and passed to the *Crawler* component for in-depth analysis. The *Crawler* is responsible for crawling the home page of the suspect host and passing the suspect html page to the *Naïve Bayes Classifier* for further analysis [16]. The *Naïve Bayes Classifier* uses a trained classifier to classify the suspect html page as a pornographic website or non-pornographic website and record the result logs. After the network operators manually confirms the result of detection, the result will be updated to the database in real time.

V. Measurement And Analysis

We use the pornographic websites detection system described in section 4 to detect the raw traffic from CNGI-CERNET2. we detected about 3,000 pornographic websites. In this section, we conduct comprehensive measurement and analysis of these pornographic websites and pornographic traffic from three different perspectives: websites, users and ISPs. The results of these analyses will help network operators understand and restrict pornographic websites and traffic in IPv6 networks.

A. Website Perspective

The analysis from website perspective includes IPv4/ IPv6 reachability, content similarity ,consistencies between domain names and IPs, geographical distribution of adult websites, and user distribution of adult websites.

1) IPv4/IPv6 reachability: When users request to a website, resources are accessed not only from the target domain, but also from other websites by automatically following the embedded hyperlinks. Therefore, whether the $\langle link \rangle$ in pornographic webpage supports IPv6 networks may affect users to obtain full resources from pornographic websites in IPv6-only network. We measure and analyze the IPv4/IPv6 reachability of all resource links in about 3000 pornographic websites detected, and the analysis results are shown in Figure 2. From Figure 2, we can see that although these adult websites are detected in IPv6 networks, but only 30% of all adult websites with full resources can be accessed in IPv6-only networks. On the other hand, the adult websites with full resources more than 95% can be accessed in IPv4 networks. Therefore, it is more important to conduct continuous measurement and analysis of pornographic websites that are still in the developing stage in IPv6 networks. In particular, some regions that need to restrict adult websites like in China, need to pay attention to the arising of adult websites in IPv6 networks.

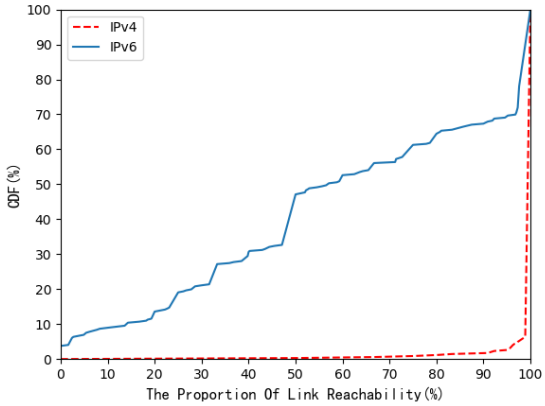

Fig. 2. The CDF of IPv4/IPv6 reachability

2) Content similarity: By further analysis of the detected pornographic websites, we find that some websites do have similar contents, although they have different domain names. In order to reveal this phenomenon, we introduce an algorithm to calculate the similarity between pages. We define the string content of different pornographic web page sets as

$$H = \{h_1, h_2, ..., h_m\} \tag{1}$$

. To measure the dissimilarity between pages, we use the "edit distance" algorithm [21]. The "edit distance" between h_x and h_y is to find the alignment between h_x and h_y that maximizes the number of matches. The resulting number of mismatches between h_x and h_y is called their "edit distance", denoted as $\mathbf{ld}\,(h_x, h_y)$. For example,

$$String\ s : S\ \mathbf{a}\ \mathbf{t}\ u\ \mathbf{r}\ d\ a\ y$$

$$String\ t : S\ \_\ \_\ u\ \mathbf{n}\ d\ a\ y$$

$$\mathbf{ld}\,(s, t) = 3 \tag{2}$$

We denote $\mathbf{LD}(\mathbf{H})$ as the "edit distance" matrix, which is defined as follows:

$$\mathbf{LD}(\mathbf{H}) = \begin{Bmatrix} \mathbf{ld}(h_1,h_1) & \mathbf{ld}(h_1,h_2) & \dots & \mathbf{ld}(h_1,h_m) \\ \mathbf{ld}(h_2,h_1) & \mathbf{ld}(h_2,h_2) & \dots & \mathbf{ld}(h_2,h_m) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{ld}(h_m,h_1) & \mathbf{ld}(h_m,h_2) & \dots & \mathbf{ld}(h_m,h_m) \end{Bmatrix} \tag{3}$$

Due to the different length of different pages, we normalize the "edit distance" matrix to get the web similarity matrix $\mathbf{S}(\mathbf{H})$, which is defined as follows:

$$\mathbf{S}(\mathbf{H}) = \begin{Bmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22} & \dots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \dots & s_{mm} \end{Bmatrix} \tag{4}$$

where $s_{xy} = \frac{\mathbf{ld}(h_x,h_y)}{\max(h_x.length, h_y.length)}$.

Then we choose $s_{xy} < 0.1$ as the criterion of similarity, and classify similar pages into different similar sets $SR = \{set_1, set_2, \dots, set_n\}$, where $set_m = \{h_j, h_k, \dots, h_l | for\ any\ h_x\ and\ h_y\ in\ set_m, s_{xy} \leq 0.1\}$. We construct a distribution of the number of pages contained in each set in the SR as shown in Figure 3. We find that there are more than 1000 adult websites with similar contents in a total of about 3000 adult websites. As can be seen from Figure 3, the most set having 2 similar web pages with different domain names, accounts for 12.1% of the total number of web sites. In addition, the largest subset with similar contents contains 65 different domain names. Some of these websites containing similar contents are shown in Table I. As can be seen from Table I, most websites with similar contents also have similar domain names, and we find that these similar domain names have the following similarities respectively: the second level domain name is the same but the top level domain name is different; or the second level domain name is basically same with one or more characters different; or the second level domain name is same but third-level domain name is different; Our preliminary speculation may be that pornographic websites registered a large number of domain names in order to escape the blockage of pornographic websites in some areas.

3) Geographic position analysis: Using the published IPv6 address database [17], we analyze the geographical distribution of adult websites and the results are shown in Figure 4. As can be seen from Figure 4, 85% of the adult websites have IPv6 addresses from a CDN vendor in Hong Kong, while Anguilla and the United States came in second and third place respectively, with less than 7%. We speculate this is mainly due to the short distance between CDN nodes in Hong Kong and the networks in mainland China. In addition, it can also be seen that CDN vendors have played a significant role in promoting the development of pornographic websites in the IPv6 networks. And since pornographic websites are illegal in mainland China, no visits to mainland Chinese are detected in the actual networks. Thus, it can be seen that in some areas where pornographic websites are illegal, the management and control of pornographic websites and traffic in IPv6 networks should be strengthen to restrict foreign pornographic websites.

TABLE I
THE DOMAIN NAME AND IPV6 ADDRESS OF ADULT WEBSITE WITH SIMILAR CONTENT.

number	domain name	IPv6 address
2	51luoben.com	2400:cb00:2048:1:0:0:681f:4adb
	www.52luoben.xyz	2400:cb00:2048:1:0:0:6818:7525
2	www.feijibt.com	2400:cb00:2048:1:0:0:681b:9099
	feijibt.com	2400:cb00:2048:1:0:0:681b:9099
3	3ba.info	2400:cb00:2048:1:0:0:681c:b15
	city9.org	2400:cb00:2048:1:0:0:681f:471b
	city9.info	2400:cb00:2048:1:0:0:681f:5384
3	stream3.tuboff.com	2607:5300:60:1885:0:0:0:1
	stream5.tuboff.com	2607:5300:60:793a:0:0:0:1
	www.tuboff.com	2400:cb00:2048:1:0:0:6818:7c44
65	www.bamumu.com	2400:cb00:2048:1:0:0:6818:7978
	m.nanahuai.com	400:cb00:2048:1:0:0:681f:59bc
	www.cecexx.in	2400:cb00:2048:1:0:0:6812:27dc
	nvnvkan.org	2400:cb00:2048:1:0:0:681b:a7fa
	cecexx.in	2400:cb00:2048:1:0:0:6812:27dc
	huaikuku.co	2400:cb00:2048:1:0:0:6818:7587
	cecepa.com	2400:cb00:2048:1:0:0:681c:c12
	...	...

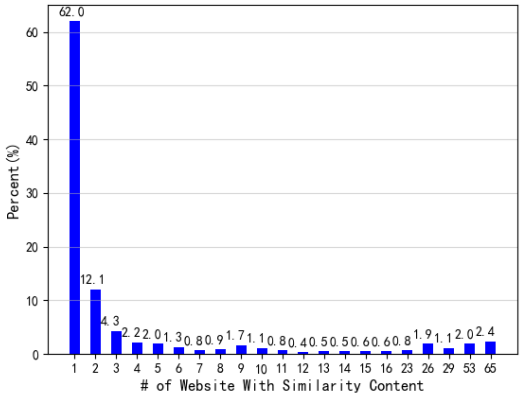

Fig. 3. The distribution of content similarity

B. User Perspective

1) Objective reasons for accessing in IPv6 networks: Some access networks of CNGI-CERNET2 are IPv4 / IPv6 dual stack networks and we analyzed the objective reasons why users use IPv6 network access instead of IPv4 networks. First of all, we use a host with dual IPv6 / IPv4 stack in an access network of CNGI-CERNET2 and configure IPv6 and IPv4 local DNS server. Then, we visit the pornographic websites that have been detected to see the usage of IPv6 and IPv4 in different browsers. The results are shown in Table II. As you can see in Table II, most browsers give priority to IPv6 DNS resolution server and use the IPv6 addresses returned by the DNS resolution server to launch HTTP requests. In order to understand the behavior of the browser, we review

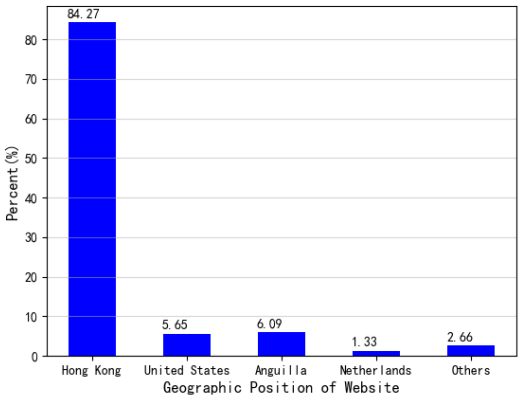

Fig. 4. The distribution of pornographic website location

TABLE II
THE POLICY OF IPV6/IPV4.

Browsers	DNS Query	HTTP Request
Chrome	IPv6	IPv6
Firefox	IPv4	IPv6
IE	IPv6	IPv6
UC	IPv6	IPv6
QQ	IPv6	IPv6
Qpera	IPv6	IPv6
Sogou	IPv6	IPv6
360	IPv6	IPv6

the relevant RFC documents and find that RFC 6724 [9] and RFC 4038 [10] set the standard for IPv6 / IPv4 priority. RFC 6724 defined the default priority policy table for socket link ordering in the operating system, as shown in Table III. As you can see in Table III, most IPv6 addresses correspond to :: 1/128 with a higher priority than IPv4 addresses :: ffff: 0: 0/96 (except for few IPv6 addresses such as, 6to4, etc.). So when the application gets the socket, the IPv6 address is in front of the IPv4 address. And RFC 4038 recommends that the application select the IP address from the socket link defaulting to the IPv6-first policy in the IPv6 / IPv4 dual stack case. We see that the current operating system and most browsers follow the standards of RFC 6724 and RFC 4038, and select IPv6 networks by default in dual-stack case. As a result, it can be inferred that pornographic traffic will continue to migrate to IPv6 networks in the future as adult websites continue to develop their support for IPv6 networks. Therefore, network operators need to pay attention to the management and supervision of pornographic websites in IPv6 networks.

C. ISP Perspective

1) Analysis of pornographic websites over time: In order to observe the long-term changes in pornographic websites in IPv6 networks, we analyze the number of pornographic websites accessing from 0:00 to 1:00 every night from August 24, 2017 to December 31, 2017. The results are shown in Figure 5. As can be seen from Figure 5, the number of pornographic

TABLE III
THE DEFAULT POLICY TABLE.

Prefix	Precedence	Label
::1/128	50	0
::/0	40	1
::ffff:0:0/96	35	4
2002::/16	30	2
2001::/32	5	5
fc00::/7	3	13
::96/16	1	3
fec0::/10	1	11
3ffe::/16	1	12

websites visited shows an overall upward trend over time, which shows that pornographic websites are in a developing state in IPv6 networks. In addition, it can be seen that the number of websites visited has a period of 3-4 days. This is mainly due to the high loyalty of pornographic websites users. Users do not usually visit pornographic websites every day and the visiting period is 3-4 days.

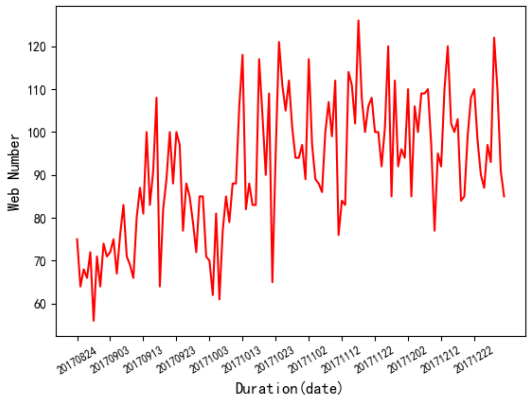

Fig. 5. The distribution of adult website access number

2) Analysis of pornographic traffic within One day: In order to study the distribution of the pornographic traffic, we analyze the distribution of normal traffic and pornographic traffic within one day, the results are shown in Figure 6. In Figure 6, we can see that the peak time of the pornographic traffic is different from the normal traffic. The peak time for pornographic traffic is 0:00 am to 2:00 am, while the normal traffic peak is between 20:00 and 22:00. In addition, there are two hours peak of pornographic traffic in one day, and two clear troughs at night and noon. While there is only a trough at night of normal traffic, and the distribution of the traffic size is uniform during the rest time of one day. It can be seen that the monitoring time of pornographic traffic should be different from normal traffic and the monitoring of pornographic traffic should focus on the daily 0:00 am to 2:00 am .

VI. SUMMARY AND FUTURE WORK

To the best of our knowledge, this paper is the first to conduct a large number of pornographic websites detection,

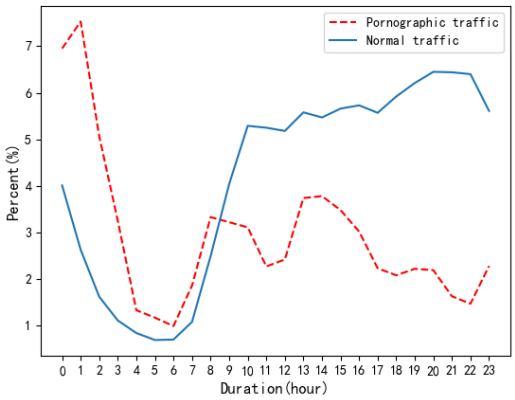

Fig. 6. The distribution of pornographic traffic within oneday

measurement and analysis in IPv6 networks. This paper implements a pornographic website detection system based on Naive Bayes algorithm, and we use the system to detect the continuous six months of traffic in CNGI-CERNET2 which is a pure IPv6 academic network, detecting a total of about 3000 pornographic websites. We conduct a comprehensive measurement and analysis of these pornographic websites and traffic from three different perspectives: websites, users and ISPs. Our analysis of the reachability of IPv6 / IPv4 adult websites shows that at present, adult websites are still in the early stage in IPv6 networks, and the proportion of available resources is still less than that in IPv4 networks. Through our analysis of the geographical distribution of pornographic websites, it can be seen that the CDN vendors played a significant role in the development of pornographic websites in IPv6 networks. In addition, we find that some pornographic websites have the same content corresponding to multiple domains and multiple IPs. This is mainly the strategy adopted by adult website owners to circumvent the blocking of some ISPs. Due to the huge IPv6 address range, this strategy will bring great challenges to ISPs in IPv6 networks. Through our analysis of user access policies, we can see that in the case of IPv4 / IPv6 dual stack, users will adopt the IPv6 network by default. As pornographic sites continue to support IPv6 networks, a large number of pornographic traffic will migrate from IPv4 networks to IPv6 networks. Finally, our measurement and analysis work is great helpful to the management and monitoring of IPv6 networks, especially in some areas where pornographic websites and traffic are illegal. And our study is important for network administrators to effectively understand adult websites and enhance the restriction of adult content in IPv6 networks..

As pornographic websites are in the developing stage in IPv6 networks, there are many future reseach works on pornographic traffic management in IPv6 networks. For example, through our analysis of the content similarity of pornographic websites, some pornographic websites adopt multi-domain name and multi-IP strategies to circumvent ISPs' blocking in IPv6 networks. Therefore, how to effectively solve this problem in IPv6 networks is one of our future work. In addition, from the point of view of network security, we find that there are some malicious js code and some "phishing" pages in adult websites, so the security analysis of these adult websites is also one of our future work.

VII. Acknowledgment

We thank the network operations center (NOC) of CERNET2 for providing the traffic data.This work is supported by the National Key Research and Development Program of China under Grant No.2017YFB0803004

References

[1] The statistics about IPv6 adoption. https://www.google.com/intl/en/ipv6/statistics.html
[2] Czyz J, Luckie M J, Allman M, et al. Don't Forget to Lock the Back Door! A Characterization of IPv6 Network Security Policy[C]//NDSS. 2016.
[3] Plonka D, Berger A. Temporal and spatial classification of active IPv6 addresses[C]//Proceedings of the 2015 ACM Conference on Internet Measurement Conference. ACM, 2015: 509-522.
[4] Foremski P, Plonka D, Berger A. Entropy/ip: Uncovering structure in ipv6 addresses[C]//Proceedings of the 2016 ACM on Internet Measurement Conference. ACM, 2016: 167-181.
[5] Czyz J, Allman M, Zhang J, et al. Measuring ipv6 adoption[C]//ACM SIGCOMM Computer Communication Review. ACM, 2014, 44(4): 87-98.
[6] Schulze, Hendrik, and Klaus Mochalski. "Internet Study 2008/2009." Ipoque Report 37 (2009): 351-362.
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[8] Tyson G, Elkhatib Y, Sastry N, et al. Demystifying porn 2.0: A look into a major adult video streaming website[C]//Proceedings of the 2013 conference on Internet measurement conference. ACM, 2013: 417-426.
[9] Chown T, Thaler D, Draves R, et al. RFC 6724: default address selection for Internet Protocol Version 6 (IPv6)[J]. Internet Engineering Task Force (IETF) RFC, 2012.
[10] Shin M K, Hong Y G, Hagino J, et al. RFC 4038: Application Aspects of IPv6 Transition[J].
[11] The market share of browers and operator system. http://gs.statcounter.com/
[12] Alexa Top 500 Global Site. http://www.alexa.com/topsites.
[13] Tang G, Wu K, Brunner R. Rethinking CDN design with distributee time-varying traffic demands[C]//INFOCOM 2017-IEEE Conference on Computer Communications, IEEE. IEEE, 2017: 1-9.
[14] Li C, Li X, Wu J, et al. IPv6 development in China[C]//Applications and the Internet Workshops, 2003. Proceedings. 2003 Symposium on. IEEE, 2003: 153-156.
[15] CNGI-CERNET2 Introduction. http://www.insc.tsinghua.edu.cn/publish/wyy/10809/index.html
[16] Murphy K P. Naive Bayes classifiers[J]. University of British Columbia, 2006, 18.
[17] IPv6 Address Database. http://ip.zxin c.org/index.htm
[18] Ahmed F, Shafiq M Z, Liu A X. The internet is for porn: Measurement and analysis of online adult traffic[C]//Distributed Computing Systems (ICDCS), 2016 IEEE 36th International Conference on. IEEE, 2016: 88-97.
[19] Staddon J, Golle P, Zimny B. Web-Based Inference Detection[C]//USENIX Security Symposium. 2007.
[20] Zhang J, Xie Y, Yu F, et al. Intention and Origination: An Inside Look at Large-Scale Bot Queries[C]//NDSS. 2013.
[21] Ristad E S, Yianilos P N. Learning string-edit distance[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(5): 522-532.
[22] Kiremire A R. The application of the Pareto principle in software engineering[J]. Consult. January, 2011, 13: 2016.