Collecting a great number of active IPv6 addresses

Yudai Aratsu[1] aratsu@hongo.wide.ad.jp
Satoru Kobayashi[2] sat@nii.ac.jp
Kensuke Fukuda[2] kensuke@nii.ac.jp
Hiroshi Esaki[1] hiroshi@wide.ad.jp
1:The University of Tokyo 2: National Institute of Informatics

Background

- Scanning whole IPv4 address takes only 45 min.
- IPv6 has vast address space.
  - to scan efficiently, active IPv6 address Hitlist is needed.

Goal

- Generating IPv6 address Hitlist by employing various methods.

How to collect a large number of active IPv6 addresses?

Methods

- Using server logs
  - Web, Mail, NTP, DNS
- Crawling P2P network
  - BitTorrent, Bitcoin
- rDNS enumeration [1]

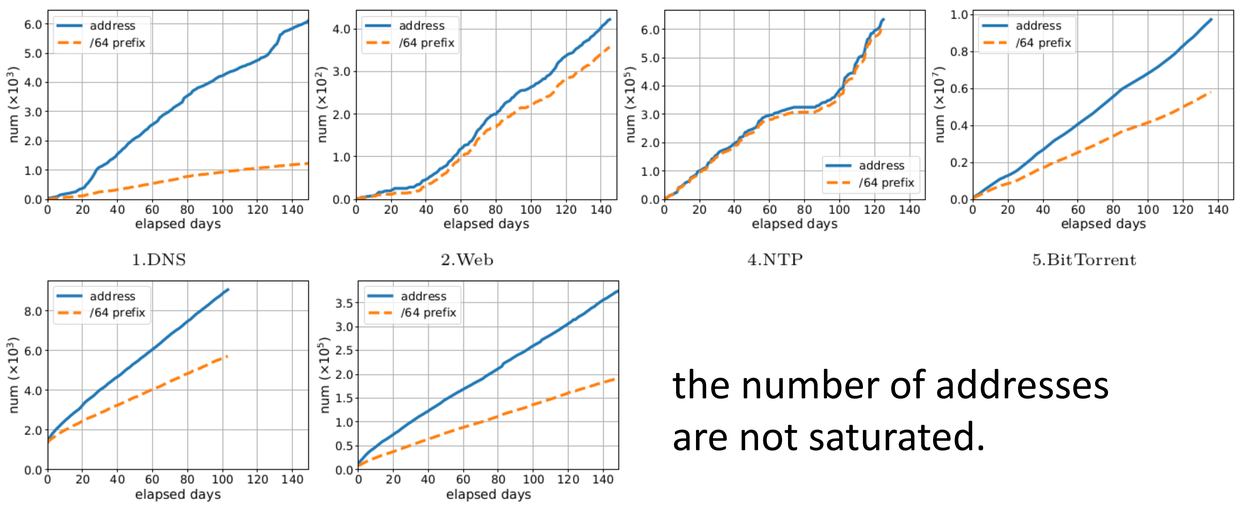
Since Jun, 2018

Result [2]

- Total numbers

		Period	#Address	#/64 prefix	#AS
1.	DNS	151 days	6,155	1,245	418
2.	Mail	146 days	1	1	1
3.	Web	149 days	422	357	101
4.	NTP	123 days	634,941	613,011	341
5.	BitTorrent	137 days	9,734,709	5,813,622	1,981
6.	Bitcoin	104 days	4,428	3,100	668
7.	rDNS enumeration	55 days	7,564,320	118,844	582
	Total	151 days	17,948,250	6,549,576	2,545
	Traffic (mawi)	151 days	377,409	193,168	4,902

- Time evolution



1.DNS 2.Web 4.NTP 5.BitTorrent 6.Bitcoin Traffic

the number of addresses are not saturated.

- IID based classification

Type	1.DNS	2.Web	4.NTP	5.BT	6.BC	7.rDNS	Traffic
"0000"	92.6%	20.6%	14.5%	13.0%	29.1%	91.3%	27.2%
"fffe"	2.3%	0.7%	3.5%	7.0%	9.8%	1.1%	9.8%
Others	5.1%	78.7%	82.0%	80.0%	63.1%	7.6%	63.1%

IP addresses of server or client?

- Response rate (icmp6)

1.DNS	2.Web	5.BT	6.BC	7.rDNS
65.8%	8.1%	0.1%	22.3%	0.2%

mostly IP addressess do not respond.
(not stable?)

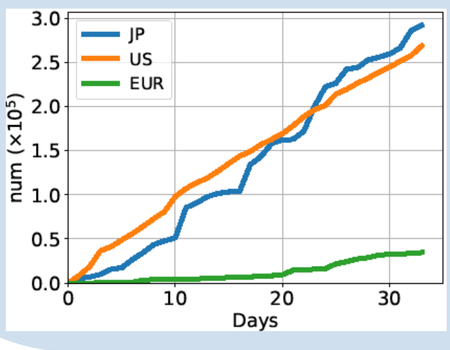
- collecting in three countries
  - Japan (Tokyo), US (California), Europe(Netherland)
  - since Sep, 2018

result of NTP server (33 days)

country based collected address classification

JP server		US server		EUR server	
IN	72,2%	US	84.5%	BR	31.4%
SA	10.1%	IN	11.1%	EU	30.0%
VN	3.2%	BR	0.7%	AT	11.3%
JP	2.5%	CN	0.6%	MX	7.8%
CN	1.5%	GT	0.4%	AR	4.9%

time evolution



Huge bias
among server location

[1] T. Fiebig, et al. “Something from nothing (There): Collecting global IPv6 datasets from DNS.” PAM’17
[2] 新津, et al. “大規模IPv6アドレス収集手法の検討” 信学技報 2018.09
[3] P. Foremski, et al. “Entropy/IP: Uncovering Structure in IPv6 Addresses” IMC’16
[4] O. Gasser, et al. “Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists” IMC’18

Discussion

How to collect more IPv6 addresses?
- multiple location
- more service
- employing machine learning [3]

How to generate “high quality” Hitlist? [4]
- collecting stable IPv6 addresses
- pseudo active space detection