清華大學
Tsinghua University

Towards the Construction of Global IPv6 Hitlist and Efficient Probing of IPv6 Address Space

Guanglei Song, Lin He, Zhiliang Wang,

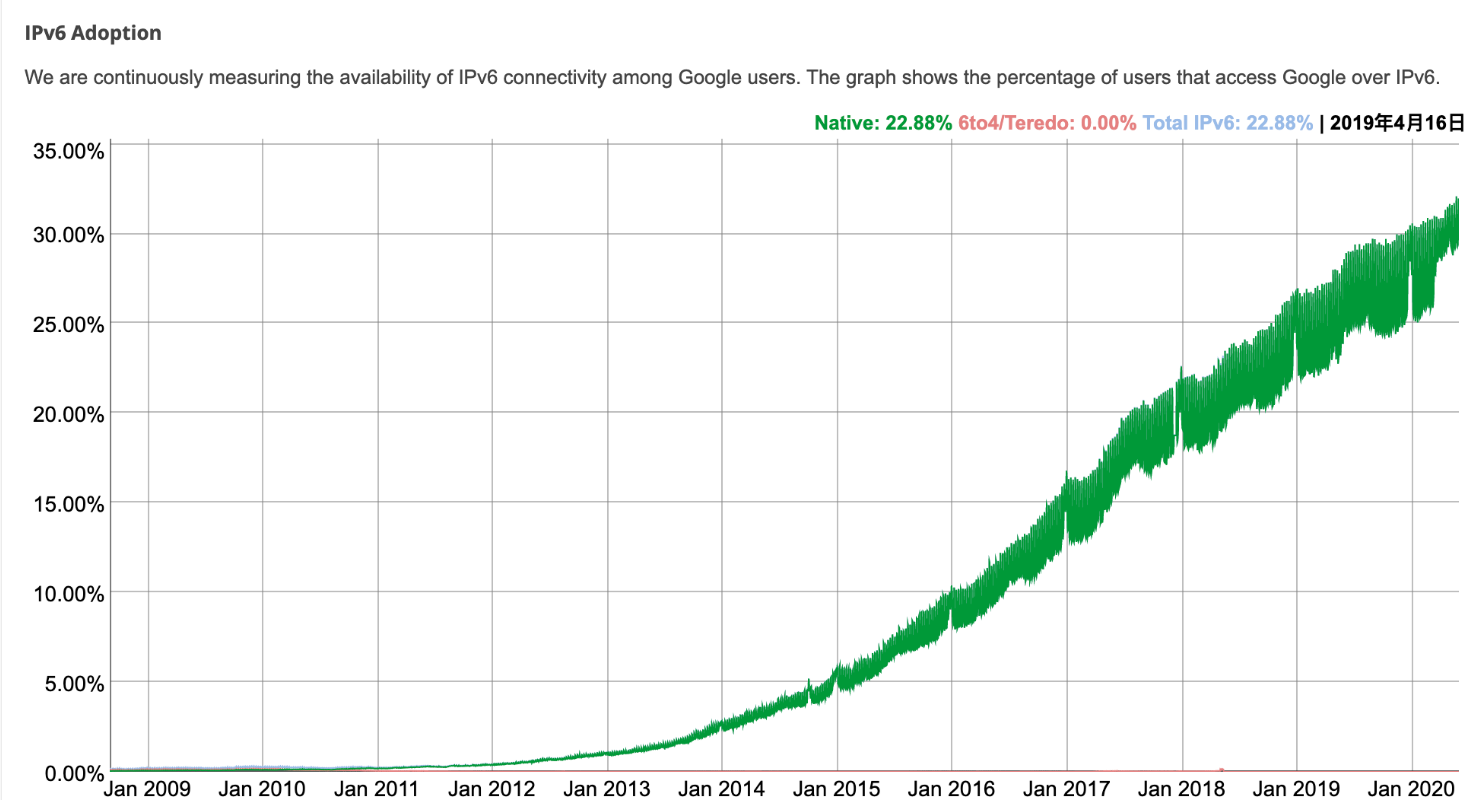
Jiahai Yang, Tao Jin, Jieling Liu, Guo Li

Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing National Research Center for Information Science and Technology

清華大學
Tsinghua University

■ Motivations

With the growing address exhaustion of IPv4 , IPv6 is being deployed increasingly commonly around the world, and this trend will accelerate.



Source: https://www.google.com/intl/en/ipv6/statistics.html

■ Motivations

- **Various IPv6 address configuration methods**
- **Vast IPv6 space**
- **Low address usage**

Scanning all IPv4 address space only needs tens of minutes

Scanning all IPv6 address space needs more than **1 million years**

Brute-force scanning of all IPv6 space is infeasible

How to quickly find active IPv6 addresses in limited probe resources?

■ Motivations

Learn structure of seed addresses

Looking for a needle in the haystack

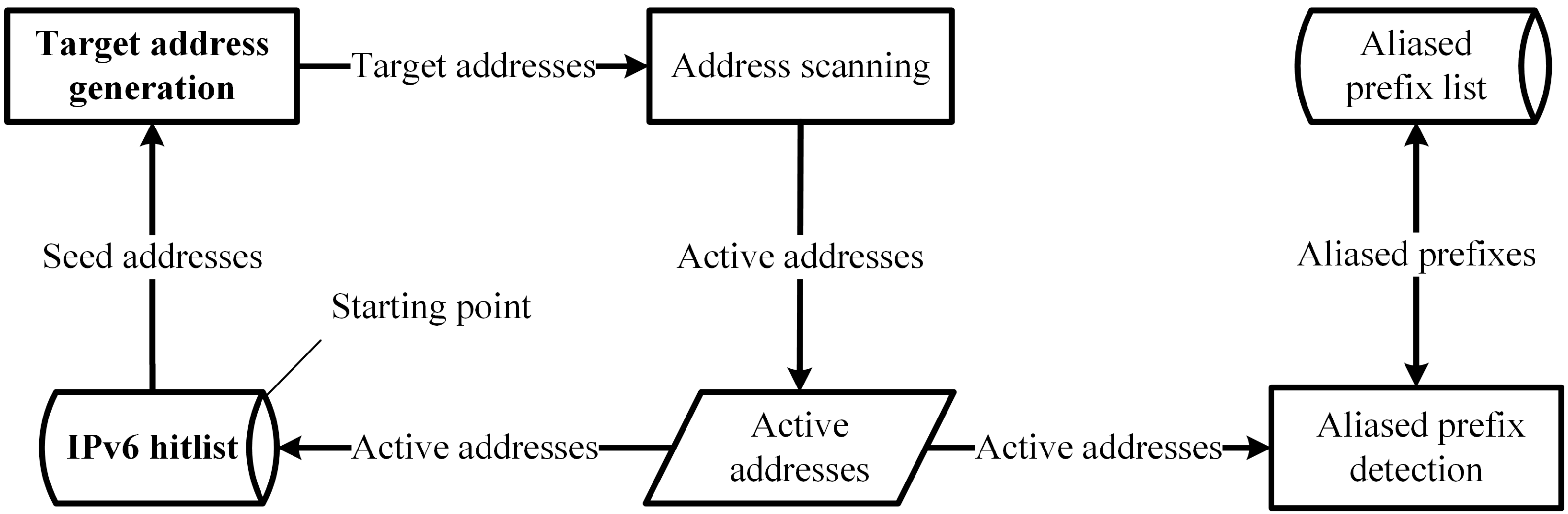
Looking for a needle in the pond

solution:

learn structure of seed address space to reduce the probing scope and generate candidate addresses that may have a higher probability to be active.

清華大學 Tsinghua University

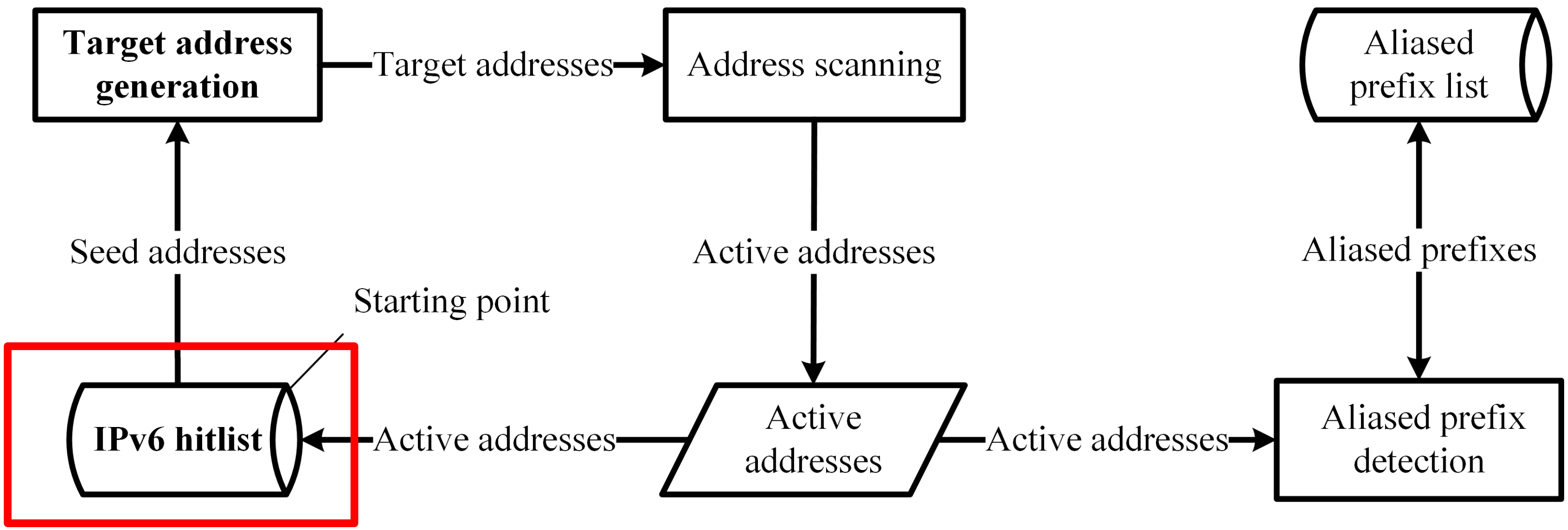
■ Motivations



Framework of IPv6 address probing

Hitlist : IPv6 address list extracted from multiple data sources
Seed addresses: Active address as input of address generation algorithms
Target address: Possible active address generated by a address generation algorithm

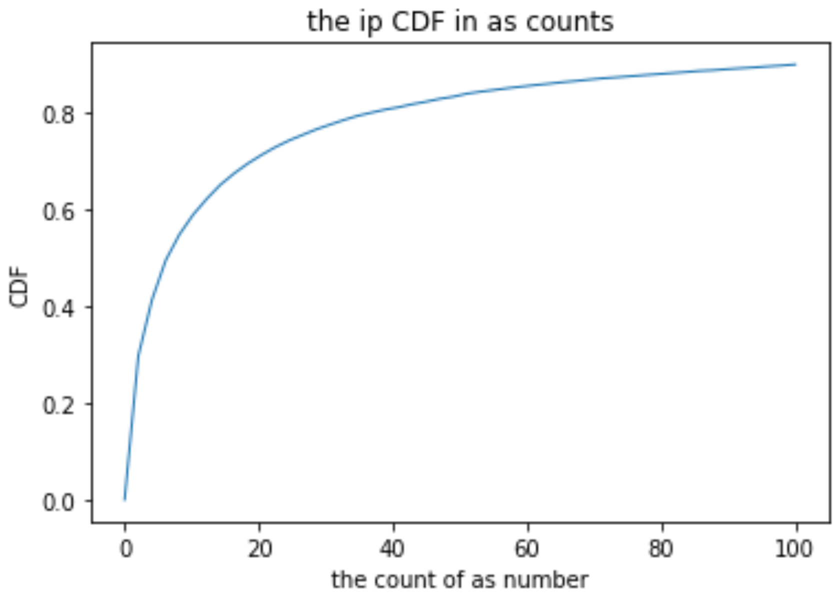
■ Motivations



Framework of IPv6 address probing

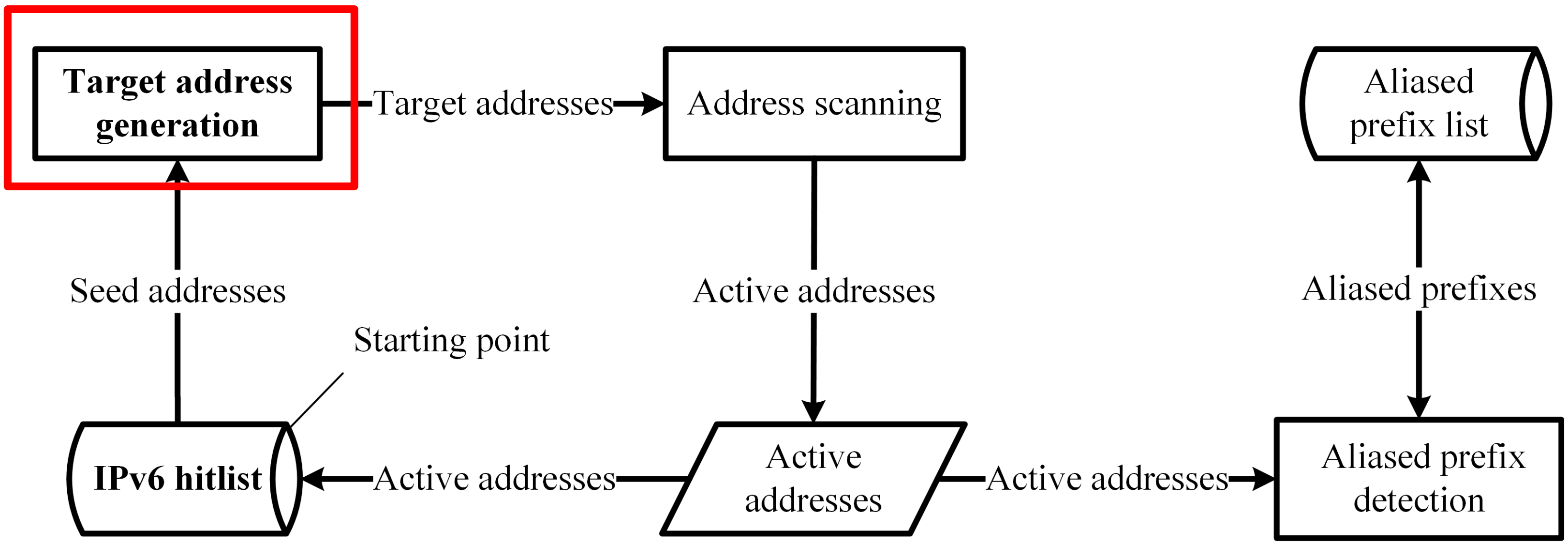
■ Motivations

Challenge 1： Uneven distribution and small quantity of public IPv6 hitlist.



Can we construct greater quantity and higher quality of IPv6 hitlist ?

Motivations



Framework of IPv6 address probing

清華大學
Tsinghua University

■ Motivations

Challenge 2: Inefficient the state-of-the-art algorithms of address generation

method	cluster-time	hit-rate[1]	limitation
Entropy/IP	slow	Low(<1%)	small detection scale
6Gen	slow ($O(n^3)$)	Low(<15%)	small detection scale
6Tree	first (linear time)	Low(<17%)	limit the scope under same budget

hit-rate[1] : Generate 50M candidate addresses

Can we design address generation algorithm with faster time and higher hit-rate ?

Talk Outline

- **IPv6 hitlist**
- **DET: IPv6 Address Generation Algorithm**
- **Evaluation About IPv6 hitlist And DET**
- **Summary**

Talk Outline

- **IPv6 hitlist**
- **DET: IPv6 Address Generation Algorithm**
- **Evaluation About IPv6 hitlist And DET**
- **Summary**

■ IPv6 Hitlist

Building greater quantity and higher quality of IPv6 hitlist from multiple sources.

1. **Public:** extract IPv6 addresses based on the data sources from December 4th, 2019 to February 10th 2020.
2. **Active:** obtain IPv6 addresses in each BGP prefix by Random-Bytes from October 1st, 2019 to January 31st, 2020.
3. **Passive:** mirror NetFlow traffic from the CERNET2 and extract IPv6 addresses from August 1st, 2019 to December 31st, 2019.
4. **TraceIP:** use scamper to perform traceroute measurement on the IPv6 addresses.

■ IPv6 Hitlist

OVERVIEW OF OUR IPV6 HITLIST ON FEBRUARY 16, 2020

Name	Nature	IPs	#ASes	#PFXes	#PFXes[1]
Public	Mixed	61M	12.4k	26.3k	25.1k
Active	Mixed	870M	14.1k	35.3k	34.1k
Passive	Mixed	340M	4.2k	8.3k	7.6k
TraceIP	Routes	49M	8.5k	12.6k	11.7k
Total	**Mixed**	**1.3B**	**16.5k**	**45.2k**	**44.1k**

#PFXes[1] : Removing aliased prefixes using aliased prefix detection
Mixed : Address types include client, server, router, etc

Talk Outline

- IPv6 hitlist
- **DET: IPv6 Address Generation Algorithm**
- Evaluation About IPv6 hitlist And DET
- Summary

清華大學 Tsinghua University

DET: IPv6 Address Generation Algorithm

- **Characteristics**

  Hierarchical characteristic (6Tree) + Distribution characteristic (6Gen)

- **Assumption**

  High-density regions of seeds are associated with high-density regions of active addresses.

- **Address vectorization**

2605:bb00:8100:0100:0001:0010:0000:0010 $\xrightarrow{\text{eg. hex}}$ 2605bb00810001000001001000000010

How to quickly discover high-density regions of addresses ?

清華大學
Tsinghua University

Discover high-density regions

- **Density Space Tree**

  root node: represents the entire seed address space
  leaf node: represents a high-density region

  Storage address hierarchy and distribution characteristics

- **Cluster**

  top-down divisive hierarchical clustering (DHC)

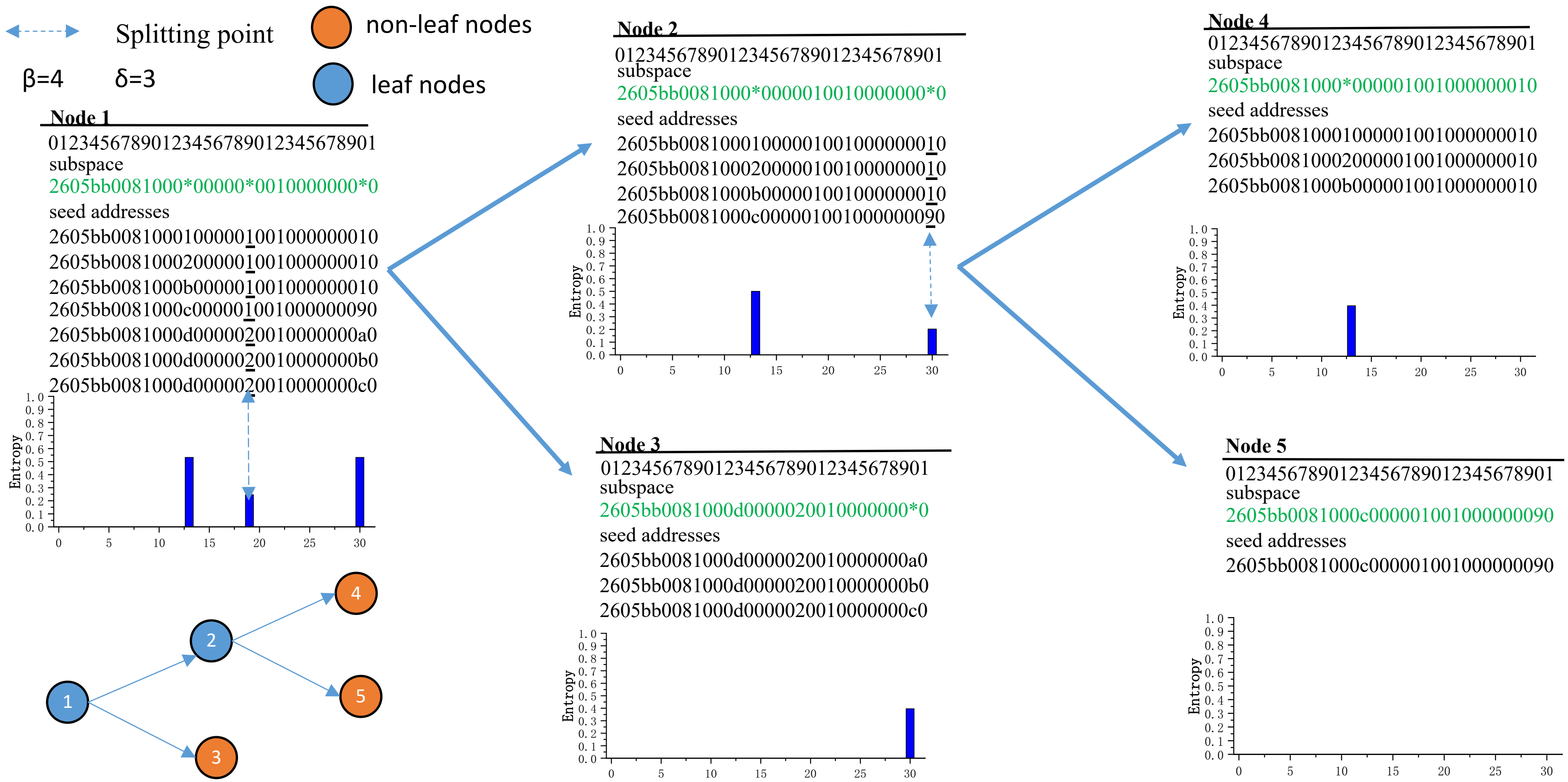
- **Splitting Indicator: The Smallest Entropy**

  Node split at the position with the smallest entropy value. Theoretical proof in paper

Ensure that high-density regions are at the leaf

Tsinghua University

Density Space Tree Generation



Public code : https://github.com/sixiangdeweicao/DET

Evaluation About IPv6 Hitlist

COMPARISON OUR HITLIST AND GASSER'S HITLIST IN ADDRESS DISTRIBUTION

Hitlist	IPs	#ASes	#PFXes	#PFXes[1]
Oliver's hitlist	61M	12.4k	26.3k	25.1k
Our hitlist	**1.3B**	**16.5k**	**45.2k**	**44.1k**

#PFXes[1] : Removing aliased prefixes using aliased prefix detection

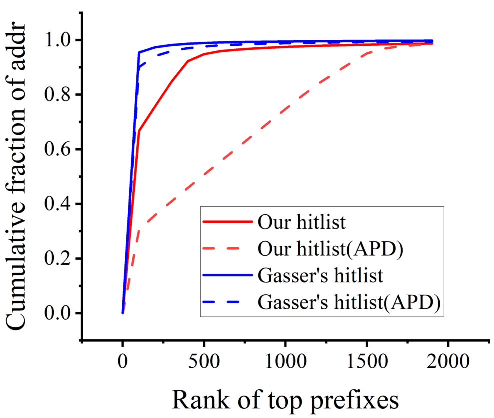


Fig. 2. Prefix distribution of two hitlists

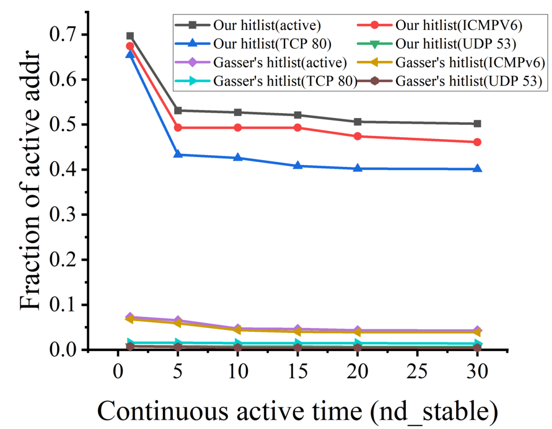


Fig. 3. Active address stability of two hitlists

nd_stable: address is continuously active for n days

Talk Outline

- IPv6 hitlist
- DET: IPv6 Address Generation Algorithm
- **Evaluation About IPv6 hitlist And DET**
- Summary

清華大學
Tsinghua University

■ Evaluation About DET

Target Generation Efficiency

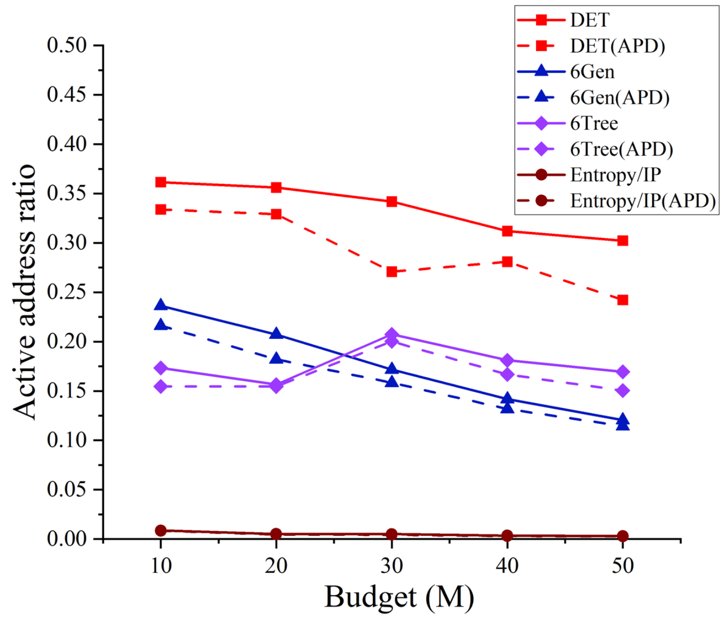


Fig. 4. Target generation in announced prefixes

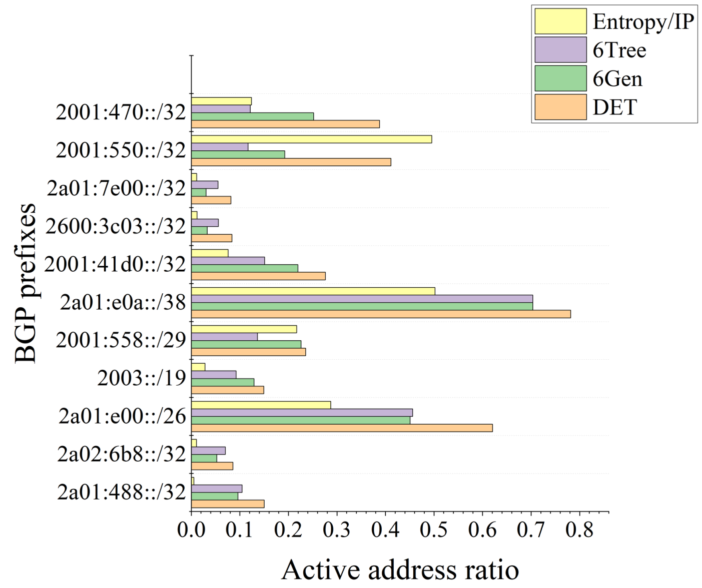


Fig. 5. Target generation in announced prefixes

Budget: the probing number of target addresses

■ Evaluation About DET

DHC Time Performance

DET learns the distribution characteristics in linear time

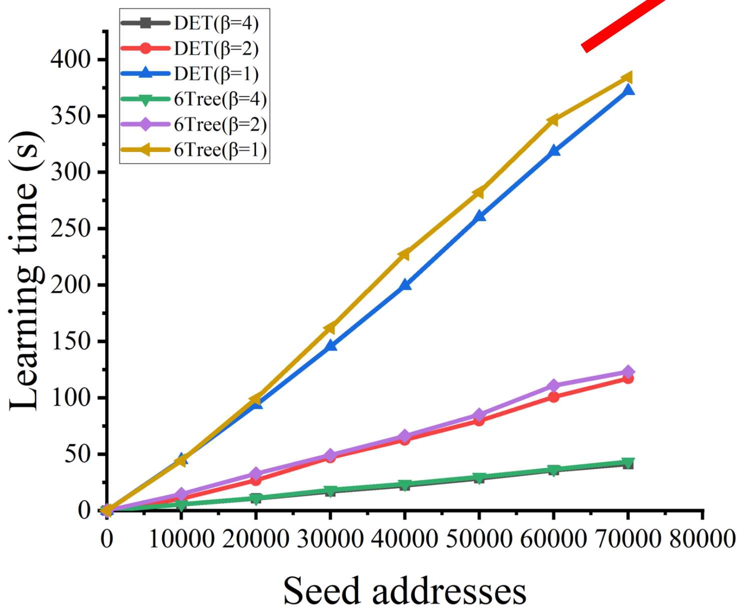


Fig. 6. DET VS 6Tree (DHC)

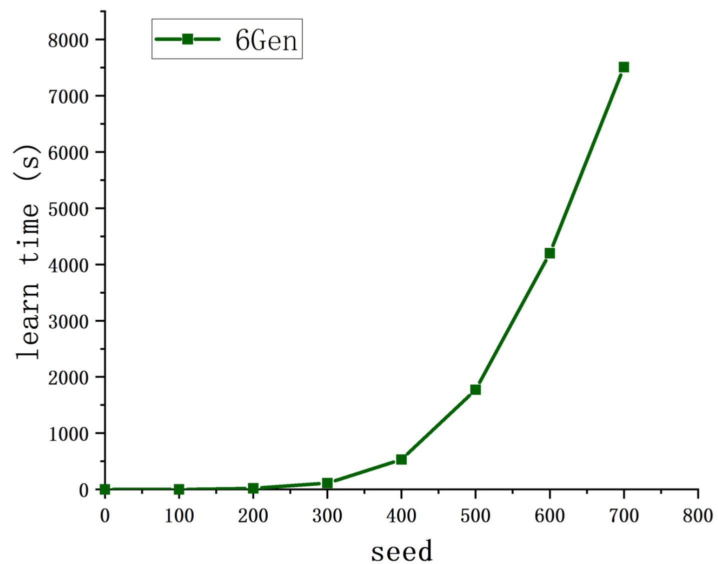


Fig. 7. 6Gen density clustering (AHC)

Talk Outline

- IPv6 hitlist
- DET: IPv6 Address Generation Algorithm
- Evaluation About IPv6 hitlist And DET
- **Summary**

Summary

- **Hitlist**

  1. Greater quantity and higher quality of IPv6 hitlist
  2. Maintain address fingerprint information

- **DET**

  1. Faster time performance of address cluster
  2. Higher IPv6 active address hit rate

Support IPv6 network measurement like IPv4

Further open the door to explore security issues on IPv6

清華大學
Tsinghua University

Thanks for your attention!

Q&A

Email: sgl18@mails.Tsinghua.edu.cn