清華大學
Tsinghua University

Comprehensive measurement of IPv6 address interface identifier pattern in current IPv6 deployment

Wei Zhang, Gang Ren, Xia Yin, Lin He

Tsinghua University

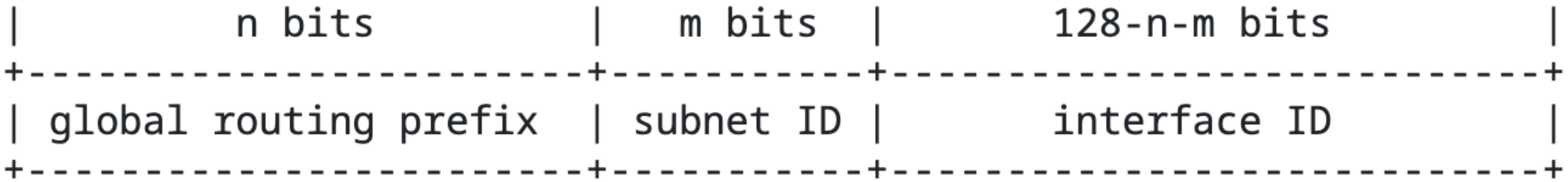
9/3/2024

CONTENTS

- Background
- Motivation
- Method
  - Data Collection
  - Pattern Analysis
- Result
  - Patterns of Servers & Routers & Clients
  - Trends of Mail Servers & Clients
- Conclusion

Background

- What is IPv6 Address Interface Identifier (IID)?
  - Component of IPv6 Address
  - Used to identify interface on a link
  - 64 bits in most cases



```
|          n bits          |   m bits   |       128-n-m bits         |
+--------------------------+------------+----------------------------+
| global routing prefix    | subnet ID  |       interface ID         |
+--------------------------+------------+----------------------------+
```

Architecture of Global Unicast Addresses

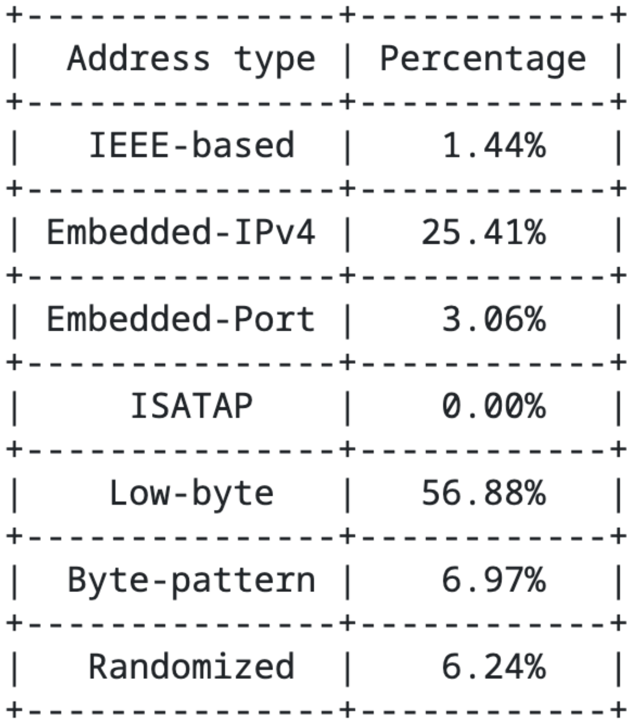
Background

- How are IIDs assigned?

Mechanism	RFC	Pattern		Scanning Difficulty	Privacy Issue
Modified EUI-64	4291	IEEE-based	****:**ff:fe**:****	Medium	Yes
Temperary Address	8981	Randomized	****:****:****:****	High	No
Stable Address	7217				
Mannually	/	Low-byte	0000:0000:00**:****	Low	No
		Embedded-port	IPv4 address in IID	Low	No
		Emdedded-IPv4	0192:0168:0001:0001	Medium	No
		Byte-pattern	zero bytes > 2	Medium	No
ISATAP	5214	ISATAP	0200:5efe:****:**** 0000:5efe:****:****	Medium	No
Teredo	4380	Teredo	IPv6 prefix 2001:0000::/32	Medium	No

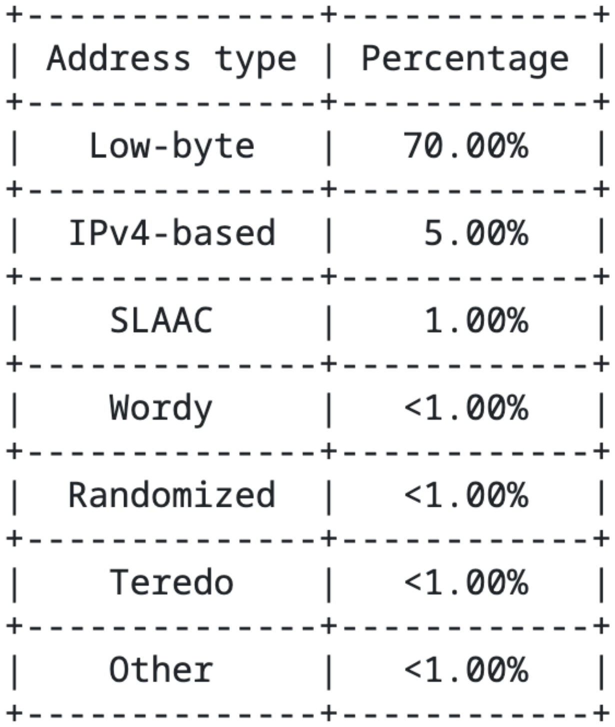
Background

- Previous work: RFC 7707



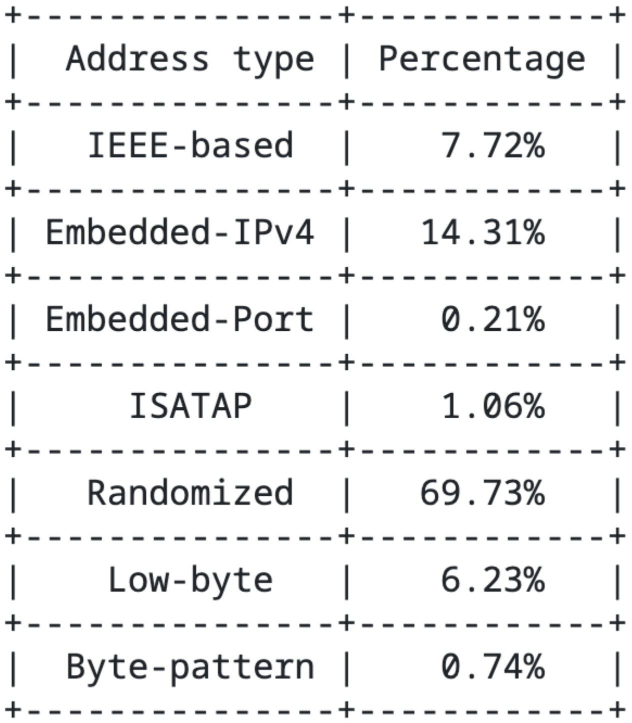
Address type	Percentage
IEEE-based	1.44%
Embedded-IPv4	25.41%
Embedded-Port	3.06%
ISATAP	0.00%
Low-byte	56.88%
Byte-pattern	6.97%
Randomized	6.24%

Figure 1: Measured Web Server Addresses



Address type	Percentage
Low-byte	70.00%
IPv4-based	5.00%
SLAAC	1.00%
Wordy	<1.00%
Randomized	<1.00%
Teredo	<1.00%
Other	<1.00%

Figure 4: Measured Router Addresses



Address type	Percentage
IEEE-based	7.72%
Embedded-IPv4	14.31%
Embedded-Port	0.21%
ISATAP	1.06%
Randomized	69.73%
Low-byte	6.23%
Byte-pattern	0.74%

Figure 5: Measured Client Addresses

Motivation

- No comprehensive measurement of IID patterns after RFC 7707
- Low accuracy for identifying random IIDs
  - Random addresses cannot be scanned practically

Motivation

- How to recognize Random IID?
  - Probability-based[1]
    - must have between 27 and 35 set bits
    - the first 32 bits must have between 9 and 21 set bits
    - the last 32 bits must have between 10 and 22 set bits
    - must not have two or more ‘words’ in it
  - Rule-based[2]
    - If an IID does not match any rule of pattern (IEEE-based, Low-byte, etc.), then it is a Randomized IID

$$\frac{1}{2^{63}} \sum_{\substack{9 \le i \le 21, 10 \le j \le 22 \\ 27 \le i+j \le 35}} \binom{31}{i} \binom{32}{j} \approx 0.7335.$$

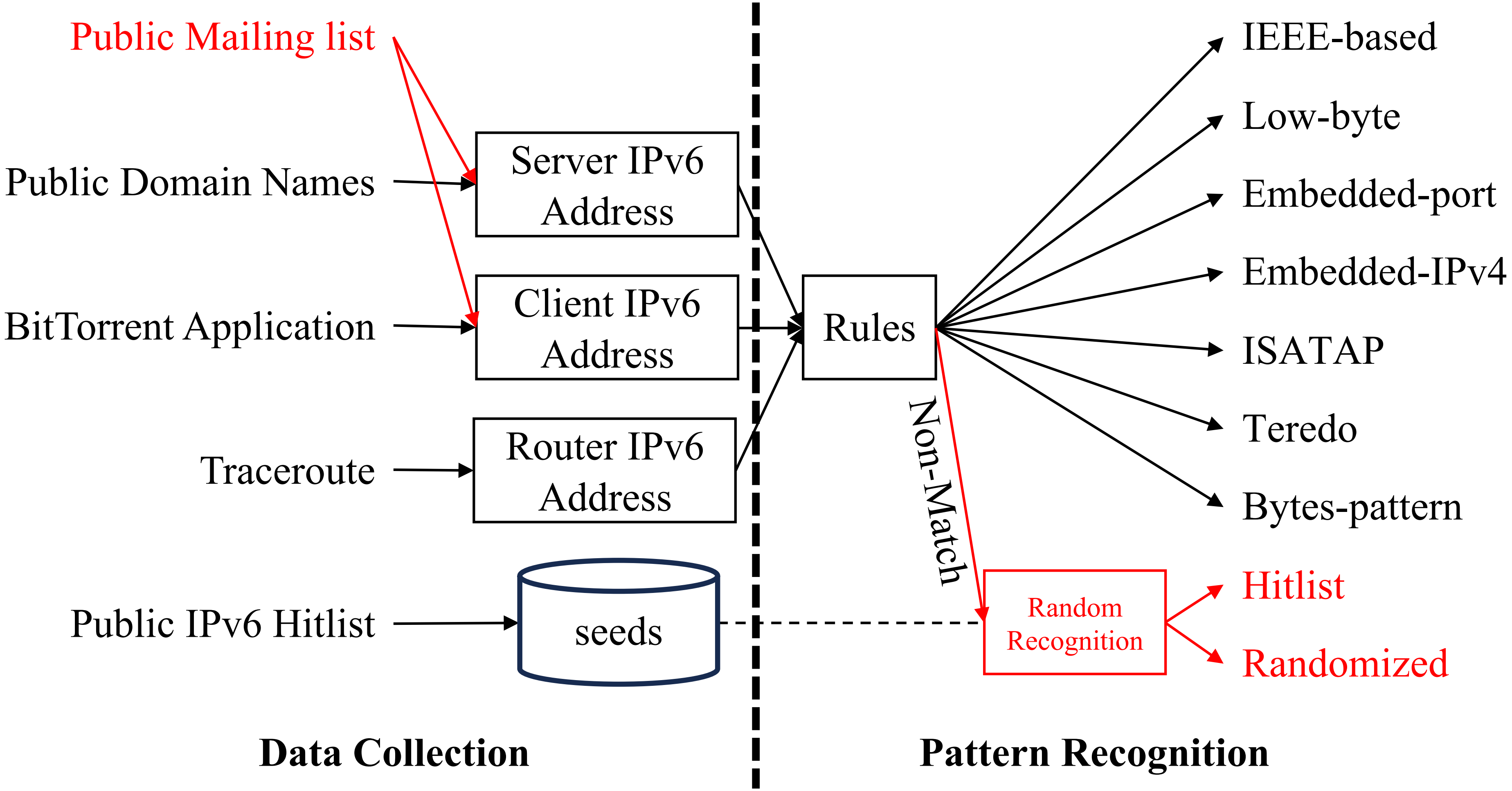
only capable of identifying approximately three-quarters of random IIDs

Identify FFFF:FFFF:FFFF:FFFF as a Randomized IID

[1] David Malone. 2008. Observations of IPv6 Addresses. In Passive and Active Network Measurement

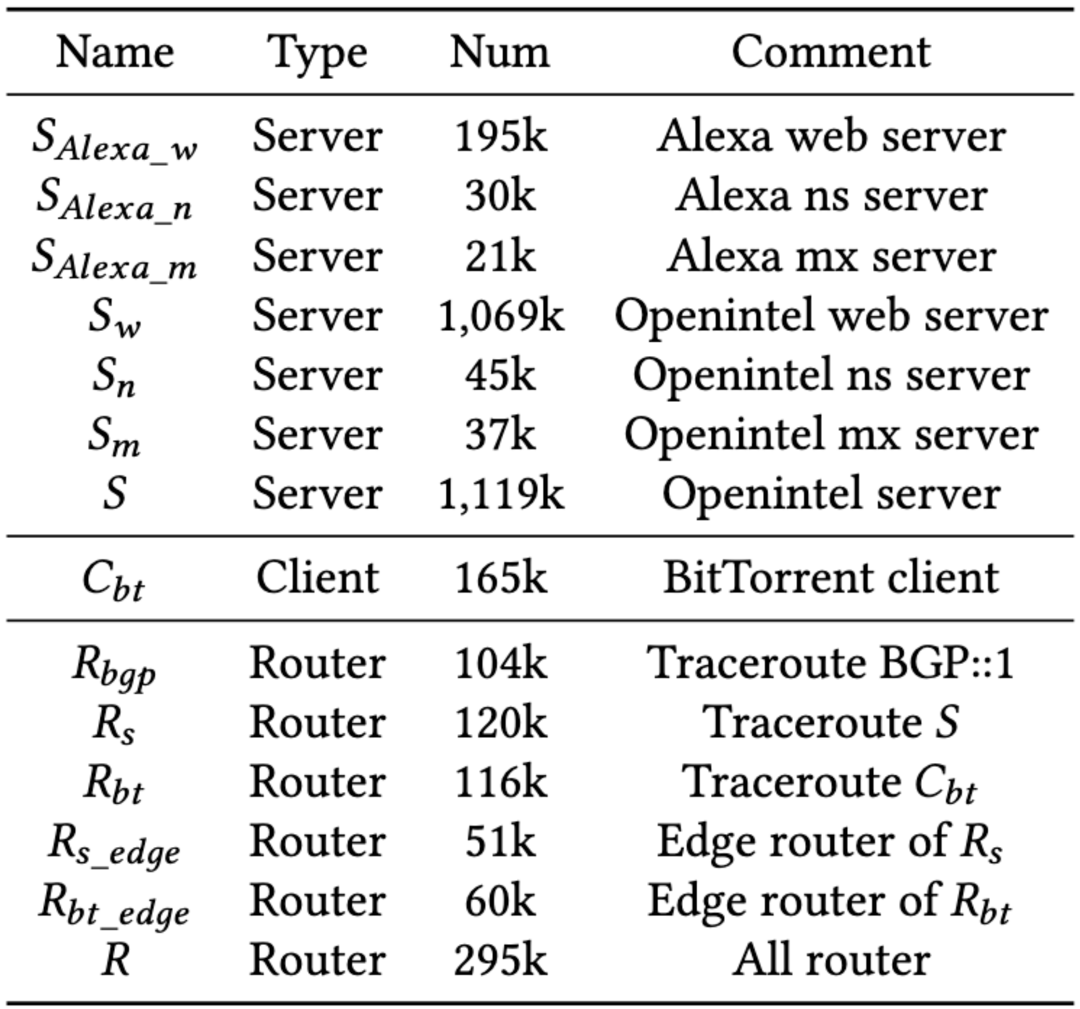
[2] Fernando Gont. IPv6 Toolkit. urlhttps://github.com/fgont/ipv6toolkit/addr6.

Methodology - Overview



Methodology - Data Collection

- Public Domain Names
  - OpenIntel[1]
- BitTorrent Application
  - Download 2000+ seeds with a BT client
- Traceroute
  - scamper

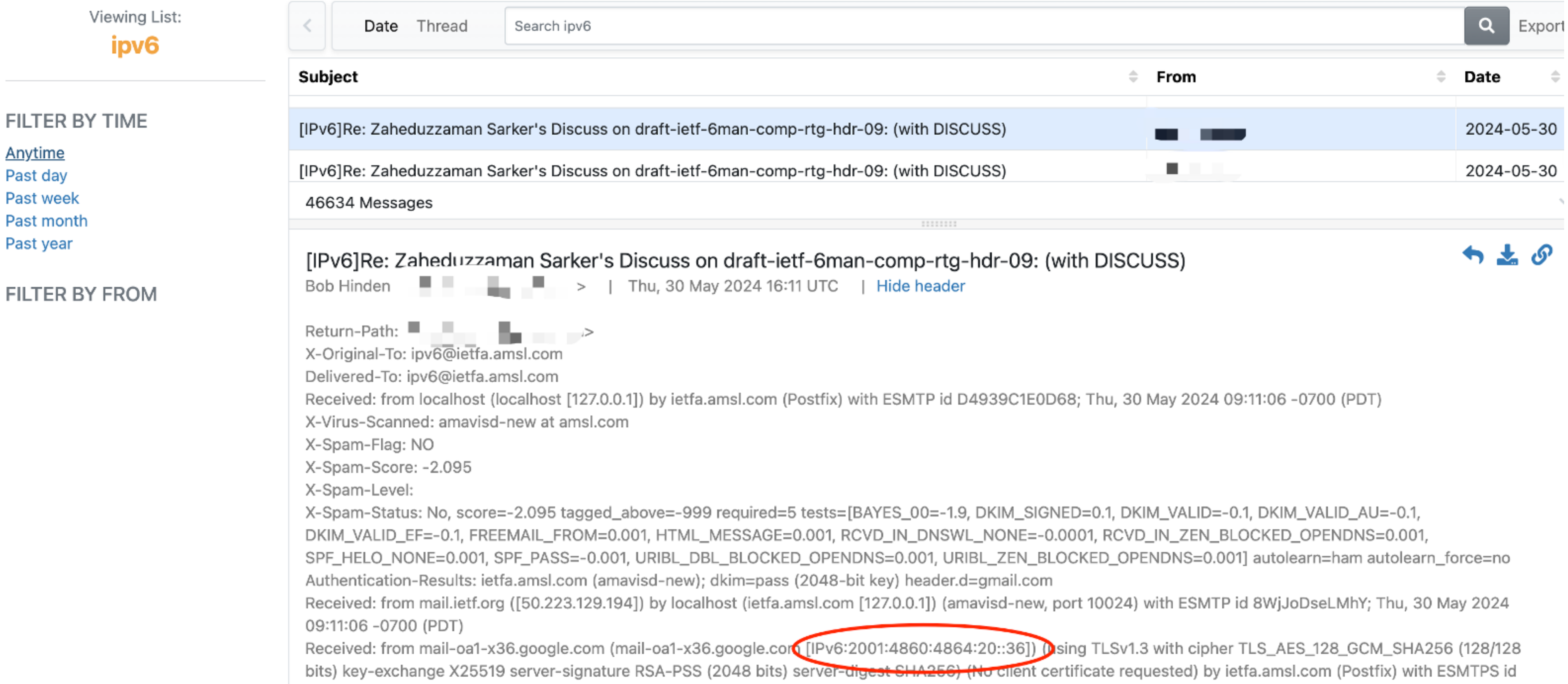


Name	Type	Num	Comment
S_{Alexa_w}	Server	195k	Alexa web server
S_{Alexa_n}	Server	30k	Alexa ns server
S_{Alexa_m}	Server	21k	Alexa mx server
S_w	Server	1,069k	Openintel web server
S_n	Server	45k	Openintel ns server
S_m	Server	37k	Openintel mx server
S	Server	1,119k	Openintel server
C_{bt}	Client	165k	BitTorrent client
R_{bgp}	Router	104k	Traceroute BGP::1
R_s	Router	120k	Traceroute S
R_{bt}	Router	116k	Traceroute C_{bt}
R_{s_edge}	Router	51k	Edge router of R_s
R_{bt_edge}	Router	60k	Edge router of R_{bt}
R	Router	295k	All router

[1] OpenINTEL: Active DNS Measurement Project. https://www. openintel.nl/

Methodology - Data Collection

- Public Mailing Lists



Viewing List:
ipv6

FILTER BY TIME
Anytime
Past day
Past week
Past month
Past year

FILTER BY FROM

Subject	From	Date
[IPv6]Re: Zaheduzzaman Sarker's Discuss on draft-ietf-6man-comp-rtg-hdr-09: (with DISCUSS)	[redacted]	2024-05-30
[IPv6]Re: Zaheduzzaman Sarker's Discuss on draft-ietf-6man-comp-rtg-hdr-09: (with DISCUSS)	[redacted]	2024-05-30

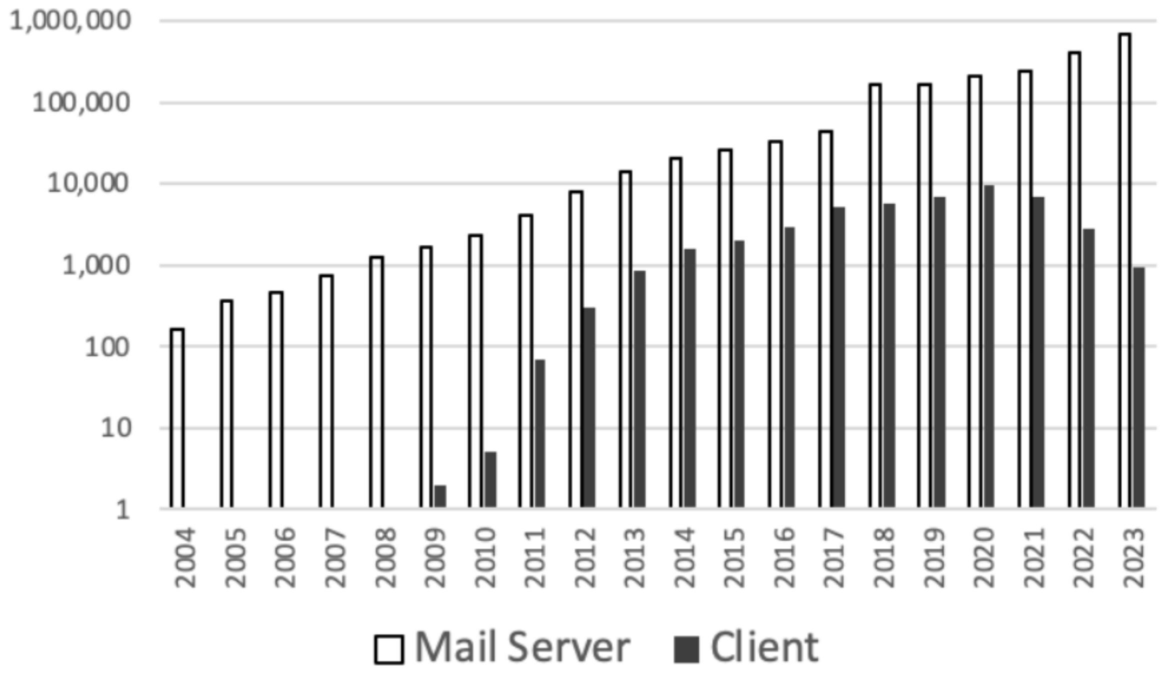
46634 Messages

[IPv6]Re: Zaheduzzaman Sarker's Discuss on draft-ietf-6man-comp-rtg-hdr-09: (with DISCUSS)
Bob Hinden [redacted] > | Thu, 30 May 2024 16:11 UTC | Hide header

```
Return-Path: [redacted]>
X-Original-To: ipv6@ietfa.amsl.com
Delivered-To: ipv6@ietfa.amsl.com
Received: from localhost (localhost [127.0.0.1]) by ietfa.amsl.com (Postfix) with ESMTP id D4939C1E0D68; Thu, 30 May 2024 09:11:06 -0700 (PDT)
X-Virus-Scanned: amavisd-new at amsl.com
X-Spam-Flag: NO
X-Spam-Score: -2.095
X-Spam-Level:
X-Spam-Status: No, score=-2.095 tagged_above=-999 required=5 tests=[BAYES_00=-1.9, DKIM_SIGNED=0.1, DKIM_VALID=-0.1, DKIM_VALID_AU=-0.1,
DKIM_VALID_EF=-0.1, FREEMAIL_FROM=0.001, HTML_MESSAGE=0.001, RCVD_IN_DNSWL_NONE=-0.0001, RCVD_IN_ZEN_BLOCKED_OPENDNS=0.001,
SPF_HELO_NONE=0.001, SPF_PASS=-0.001, URIBL_DBL_BLOCKED_OPENDNS=0.001, URIBL_ZEN_BLOCKED_OPENDNS=0.001] autolearn=ham autolearn_force=no
Authentication-Results: ietfa.amsl.com (amavisd-new); dkim=pass (2048-bit key) header.d=gmail.com
Received: from mail.ietf.org ([50.223.129.194]) by localhost (ietfa.amsl.com [127.0.0.1]) (amavisd-new, port 10024) with ESMTP id 8WjJoDseLMhY; Thu, 30 May 2024
09:11:06 -0700 (PDT)
Received: from mail-oa1-x36.google.com (mail-oa1-x36.google.com [IPv6:2001:4860:4864:20::36]) using TLSv1.3 with cipher TLS_AES_128_GCM_SHA256 (128/128
bits) key-exchange X25519 server-signature RSA-PSS (2048 bits) server-digest SHA256) (No client certificate requested) by ietfa.amsl.com (Postfix) with ESMTPS id
```

Methodology - Data Collection

- Public Mailing Lists
  - news.gmane.io
    - Public Mailing List: 30k
    - From 2004 to 2023
    - Client IPv6 Address: 43k
    - Mail Server IPv6 Address: 1,563k
      - S_{ml_2023}: 0.26%
      - S_m: 50%

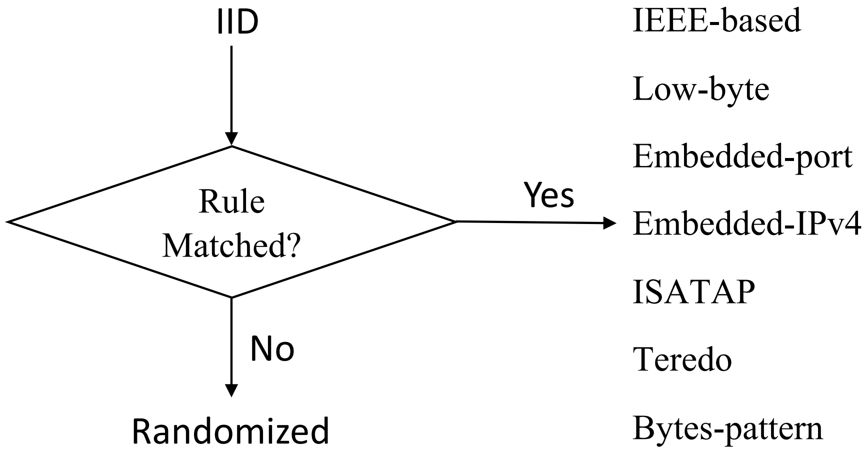


Methodology - Pattern Recognition

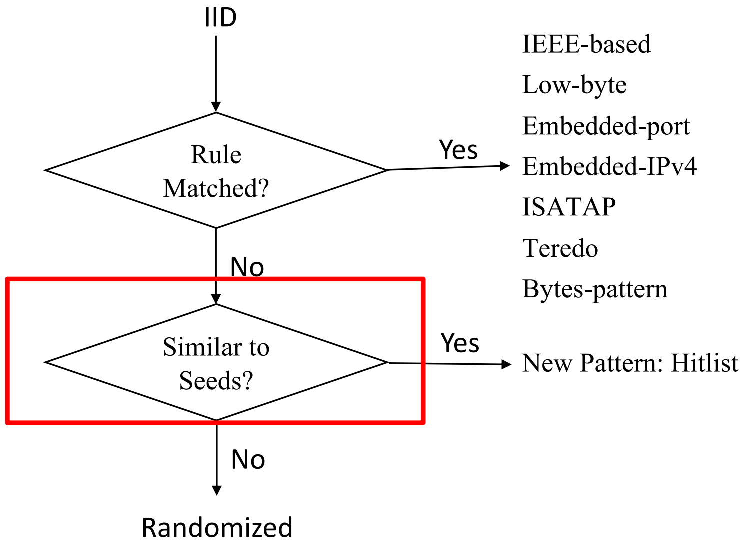
Seed-based Random IID Recognition

- If an IID does not match any rule of pattern (IEEE-based, Low-byte, etc.) and it does not similar to any IID in a list of IPv6 address (seeds), then it is a Randomized IID
- Hitlist pattern: a special type of manually configured pattern
- Seeds: IPv6 Hitlist (https://ipv6hitlist.github.io/, 9M addresses)
- https://github.com/will-zhang/iidpattern

addr6

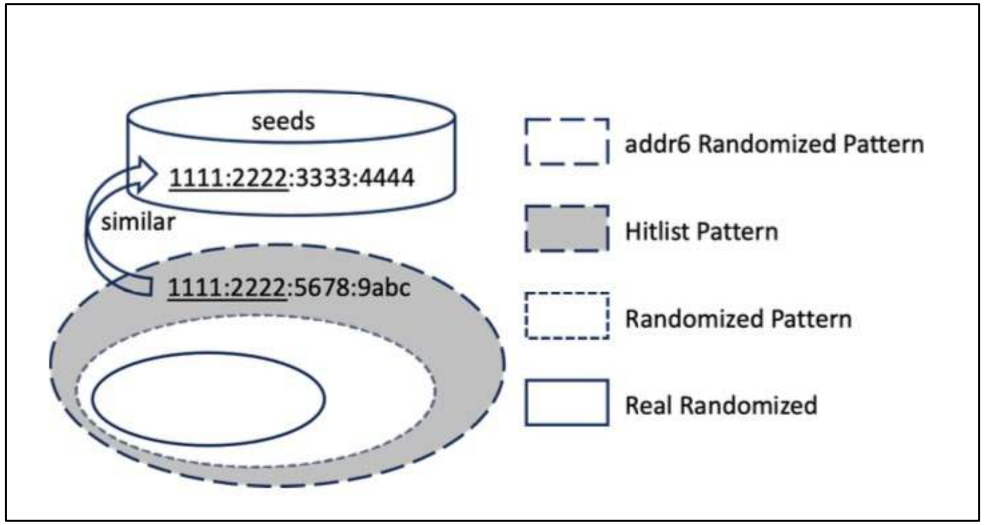


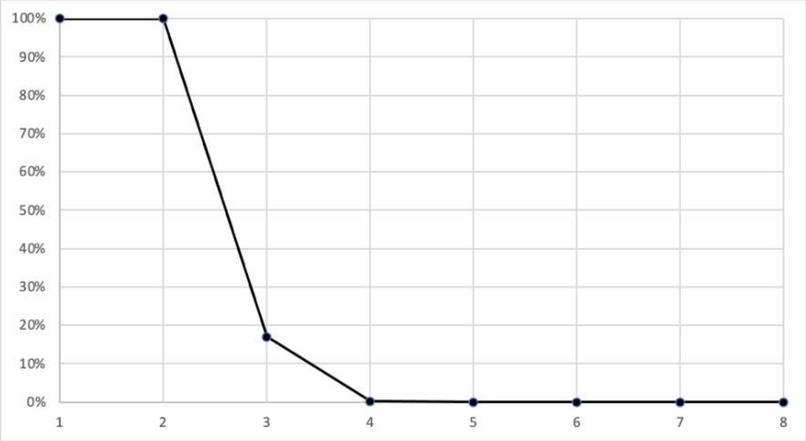
Our approch



Method - Pattern Recognition

- Seed-based Random IID Recognition
  - If the first 4 bytes or the last 4 bytes of two IIDs are the same, then the two IIDs are considered similar
  - false negative rate: 0.17%
    - Generate 10 million random IIDs, then test how many IIDs are Hitlist pattern(false negative)





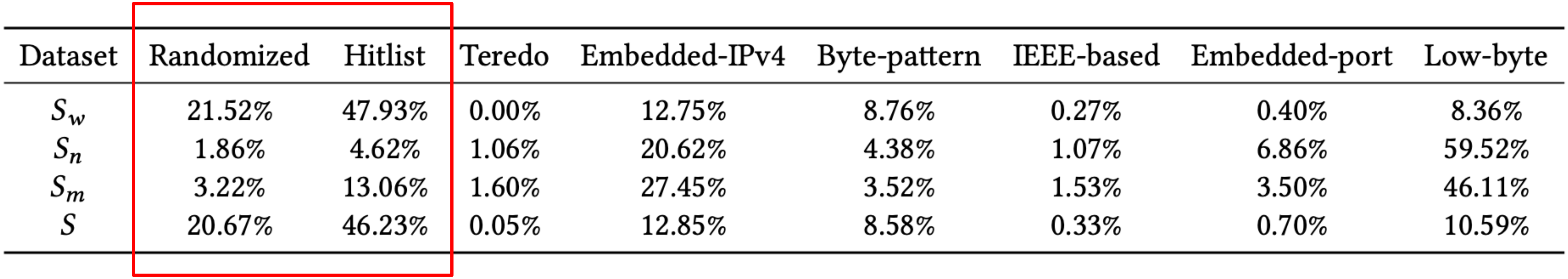
false negative rate for different length

Results

- The measurement was conducted in January 2024

Results - Server IID Patterns

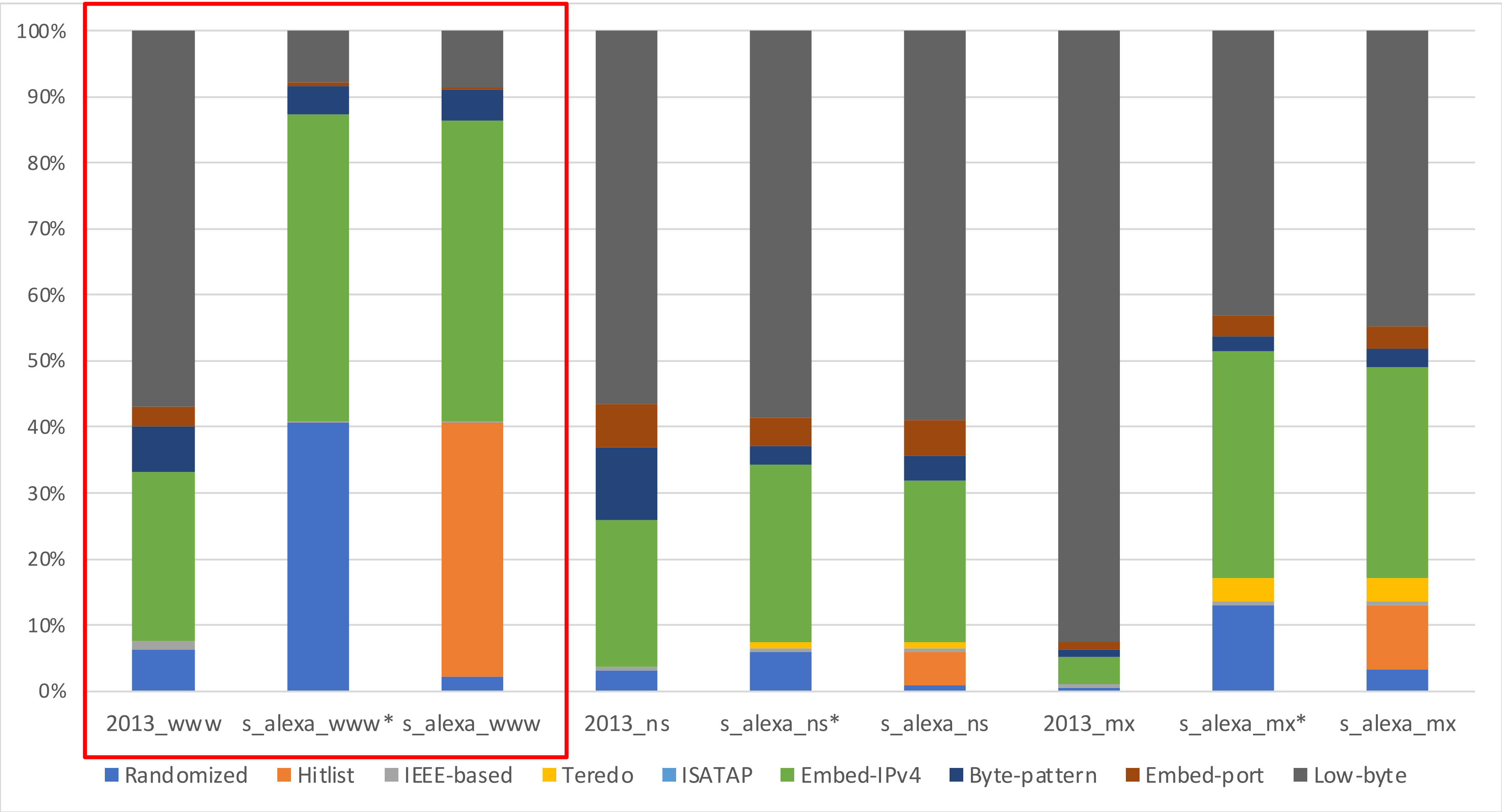
- Randomized pattern is severely overestimated
  - addr6: 67%
  - Our method: 21%



Dataset	Randomized	Hitlist	Teredo	Embedded-IPv4	Byte-pattern	IEEE-based	Embedded-port	Low-byte
S_w	21.52%	47.93%	0.00%	12.75%	8.76%	0.27%	0.40%	8.36%
S_n	1.86%	4.62%	1.06%	20.62%	4.38%	1.07%	6.86%	59.52%
S_m	3.22%	13.06%	1.60%	27.45%	3.52%	1.53%	3.50%	46.11%
S	20.67%	46.23%	0.05%	12.85%	8.58%	0.33%	0.70%	10.59%

Results - Server IID Patterns

- Increased IPv6 address scanning difficulty

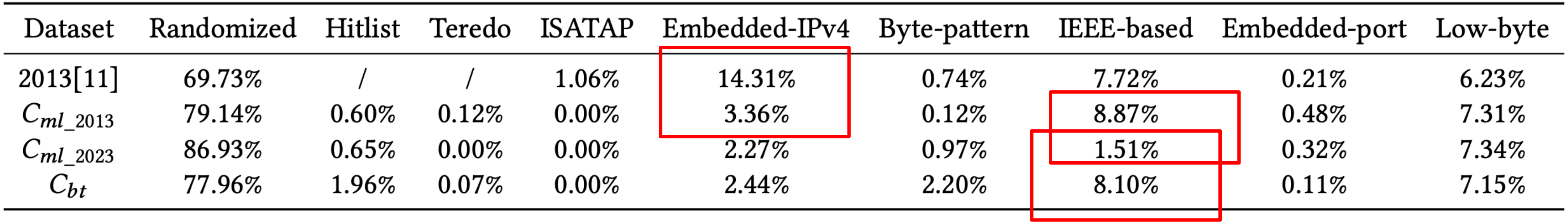


1.The dataset used in RFC 7707 is closely related to S_{Alexa}

2. * denotes results derived using addr6

Results - Client IID Patterns

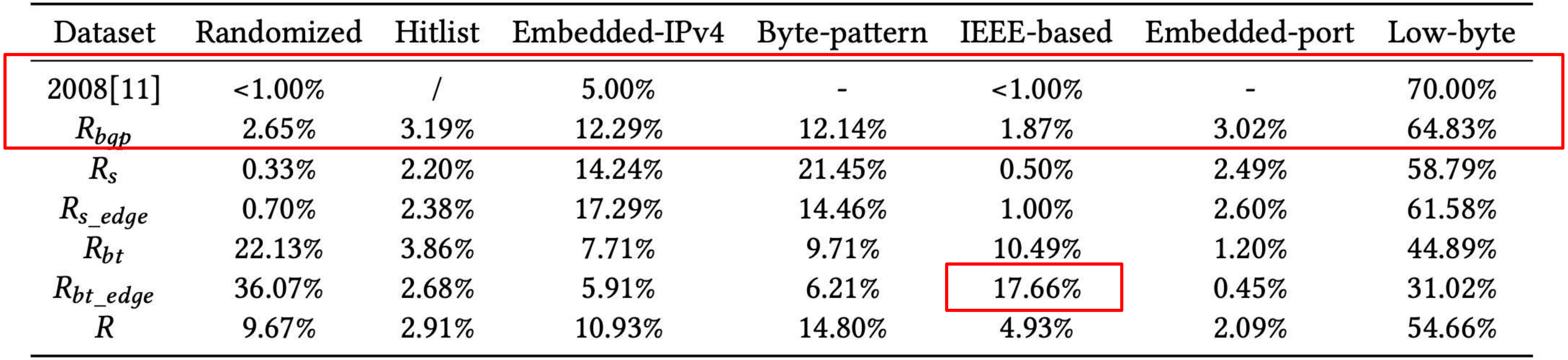
- C_{bt} VS C_{ml_2023}
- C_{ml_2013} VS RFC 7707
- Reduced IPv6 address privacy risk



Dataset	Randomized	Hitlist	Teredo	ISATAP	Embedded-IPv4	Byte-pattern	IEEE-based	Embedded-port	Low-byte
2013[11]	69.73%	/	/	1.06%	14.31%	0.74%	7.72%	0.21%	6.23%
C_{ml_2013}	79.14%	0.60%	0.12%	0.00%	3.36%	0.12%	8.87%	0.48%	7.31%
C_{ml_2023}	86.93%	0.65%	0.00%	0.00%	2.27%	0.97%	1.51%	0.32%	7.34%
C_{bt}	77.96%	1.96%	0.07%	0.00%	2.44%	2.20%	8.10%	0.11%	7.15%

Results - Router IID Patterns

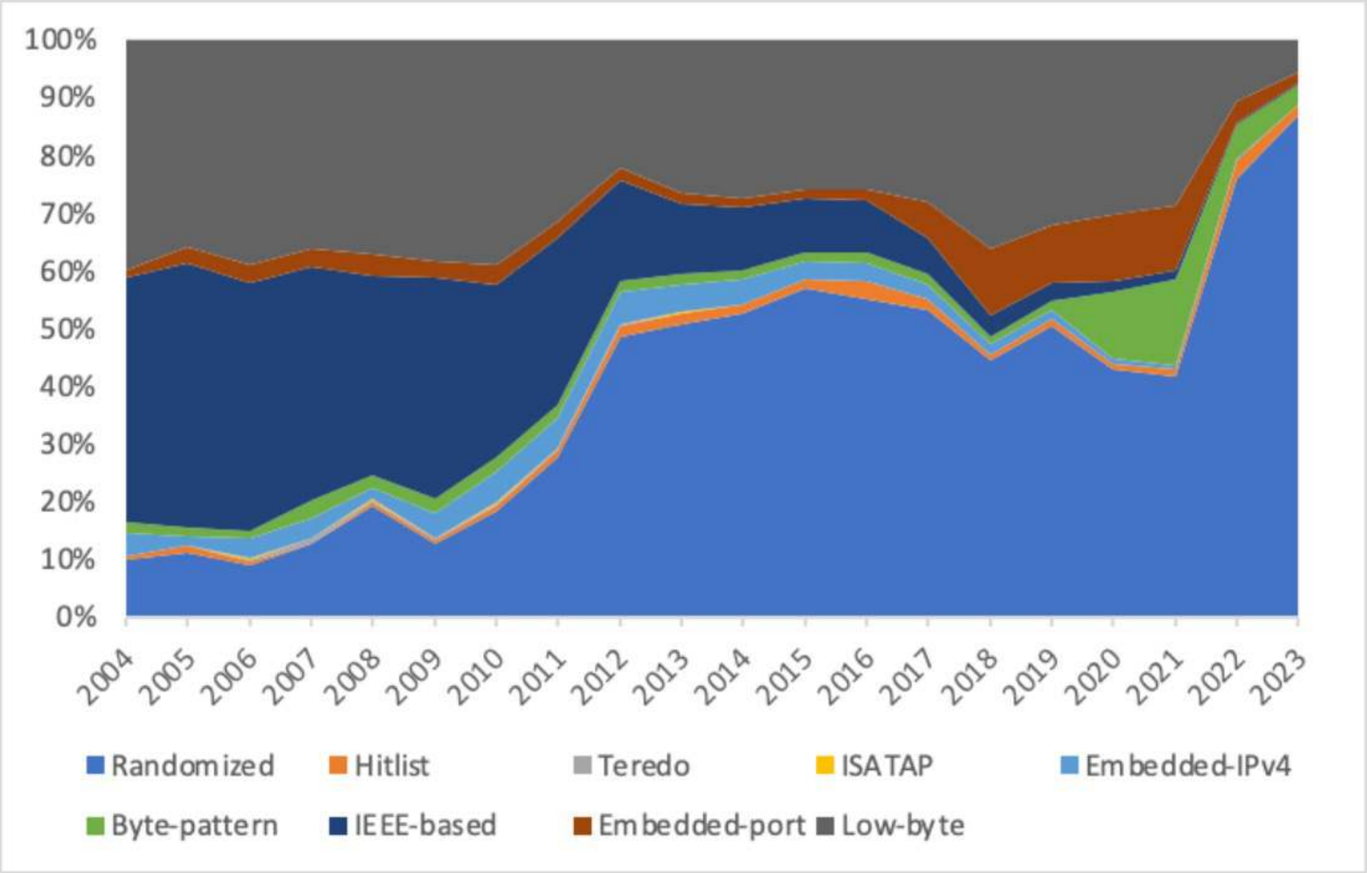
- High privacy risk for client edge routers
- Increased IPv6 address scanning difficulty



Dataset	Randomized	Hitlist	Embedded-IPv4	Byte-pattern	IEEE-based	Embedded-port	Low-byte
2008[11]	<1.00%	/	5.00%	-	<1.00%	-	70.00%
R_{bgp}	2.65%	3.19%	12.29%	12.14%	1.87%	3.02%	64.83%
R_s	0.33%	2.20%	14.24%	21.45%	0.50%	2.49%	58.79%
R_{s_edge}	0.70%	2.38%	17.29%	14.46%	1.00%	2.60%	61.58%
R_{bt}	22.13%	3.86%	7.71%	9.71%	10.49%	1.20%	44.89%
R_{bt_edge}	36.07%	2.68%	5.91%	6.21%	17.66%	0.45%	31.02%
R	9.67%	2.91%	10.93%	14.80%	4.93%	2.09%	54.66%

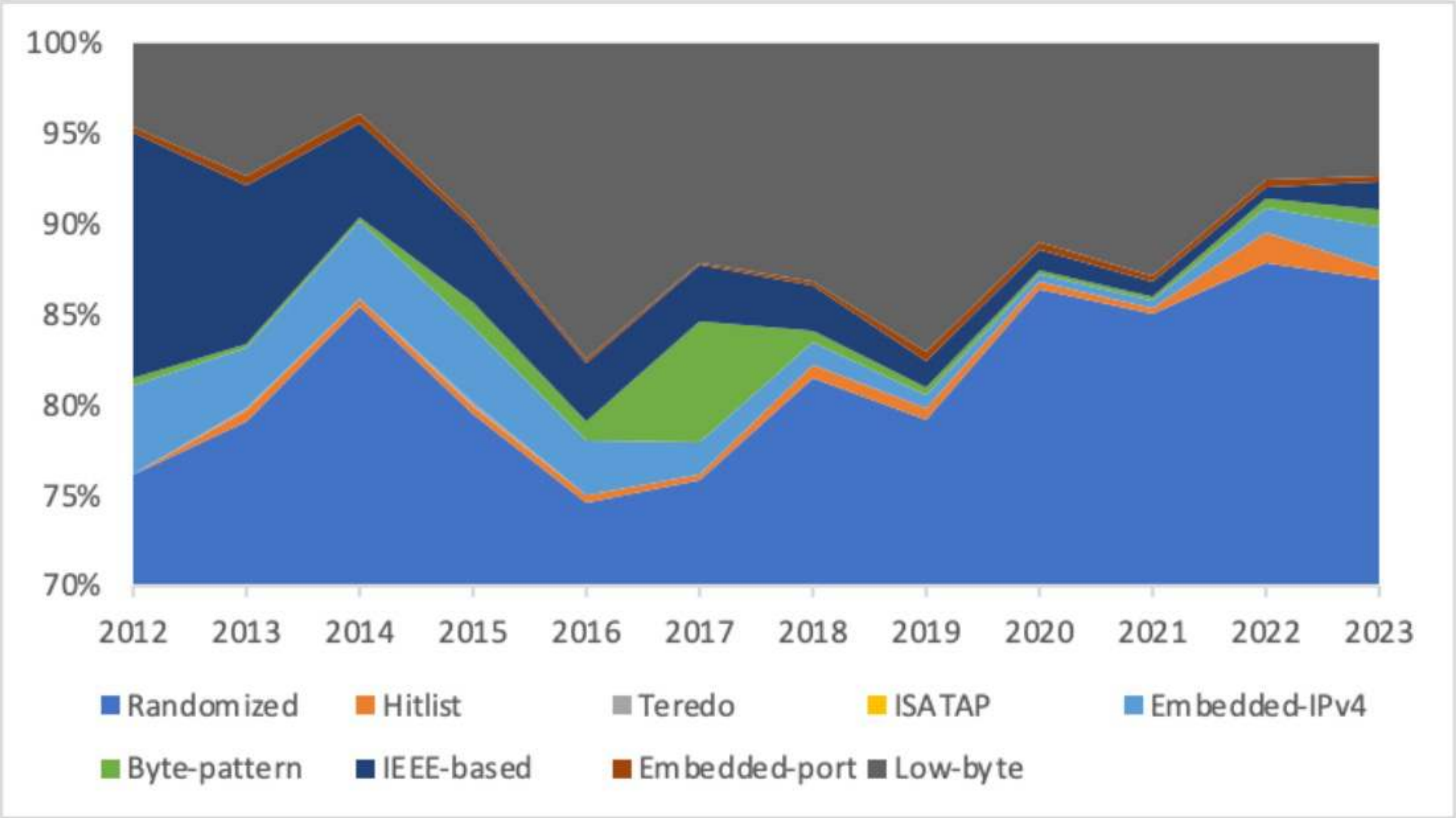
Results - IID Pattern Trend

- Mail Server



Results - IID Pattern Trend

- Client



Conclusion

- **The scanning of IPv6 addresses has become significantly more challenging for servers and routers**
  - Increased use of Randomized addresses
  - Decreased use of Low-byte addresses
- **Server Randomized pattern is severely overestimated with current method**
  - High rate of false positive for existing tools to recognize random addresses
- **The risk of privacy breaches for clients has been further reduced**
  - Decreased use of IEEE-based addresses
- **The privacy risks caused by client edge routers is a concern**
  - 18% of IEEE-based address
- **Public mailing list is an alternative source for obtaining IPv6 addresses**

Future work

- **More data sources**
  - Server logs
  - Network traffic
  - …
- **Public mailing lists**
  - IPv6 deployment rates in different countries
  - market share among different hardware manufacturers
  - …

Q&A

Wei Zhang: zhang-w22@emails.tsinghua.edu.cn

Gang Ren: rengang@cernet.edu.cn