Discovering Path MTU Black Holes With RIPE Atlas

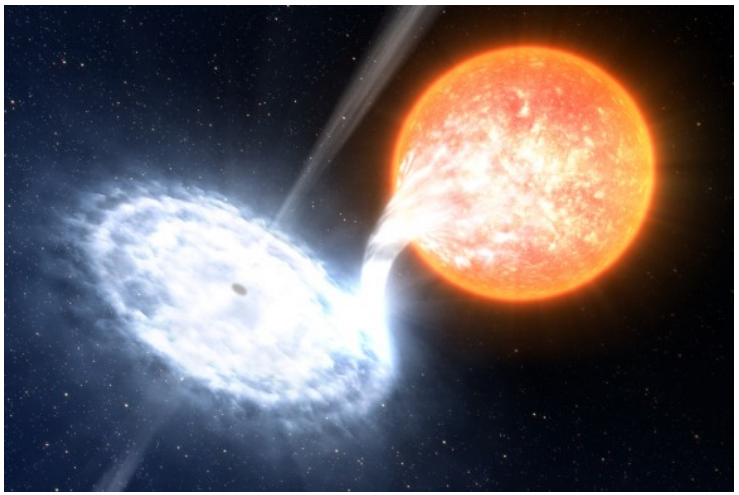
Maikel de Boer
Jeffrey Bosma
Benno Overeinder
Willem Toorop

29 July 2012

Universiteit van Amsterdam
NLnet Labs

Introduction

- Black holes
  - *"A sphere of influence into which or from which communication or similar activity is precluded."*
    ~ Wiktionary.org
  - In layman's terms: what goes in is forever lost



  - The Internet is full of black holes
    - many possible causes
      - e.g., misconfiguration, bugs in software, etc.
    - we focus on Path MTU black holes

Research Questions

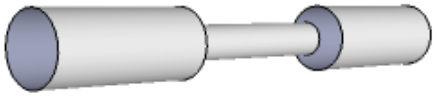
Where on the Internet do Path MTU black holes occur?

Do Path MTU black holes occur more often with IPv6 compared to IPv4?

BACKGROUND

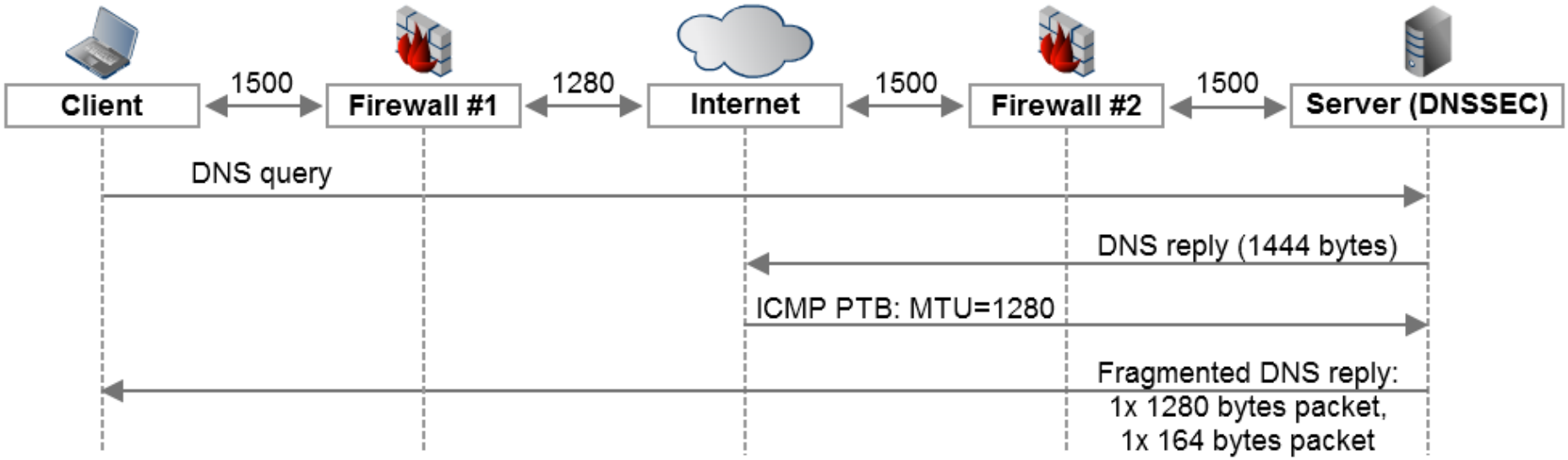
Concepts

- Maximum Transmission Unit (MTUs) on network interface
  - determines the maximum size of packets
- Path MTU (RFC1191)
  - highest possible MTU for entire path
    - determined by link with smallest MTU

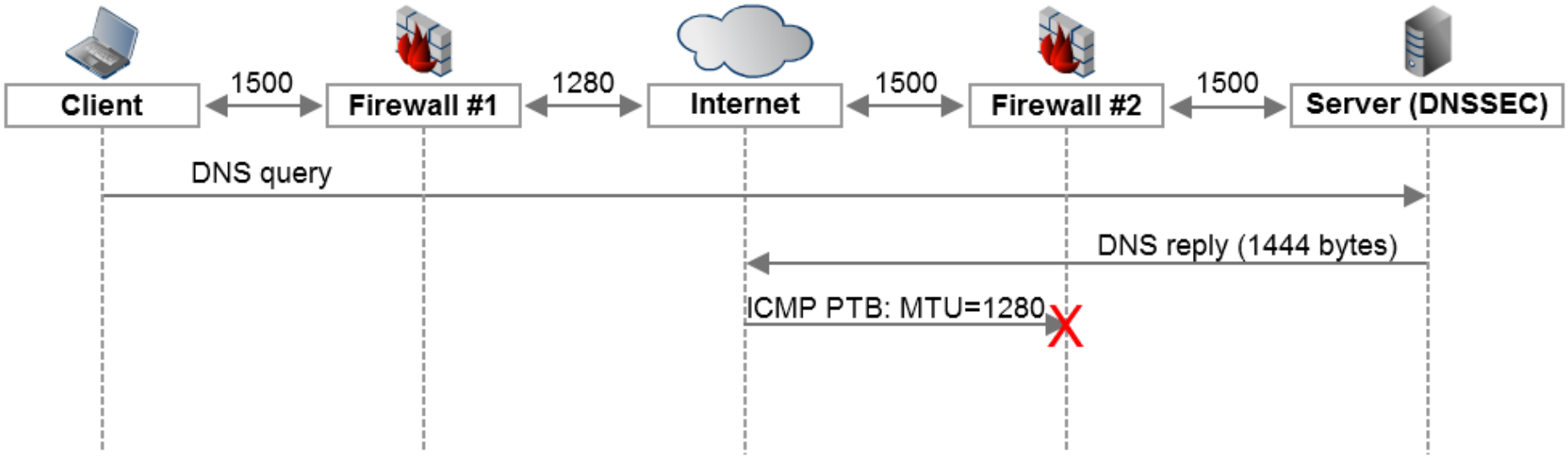


- Internet Path MTU is commonly 1500 bytes
  - not always the case
  - requires Path MTU detection mechanism

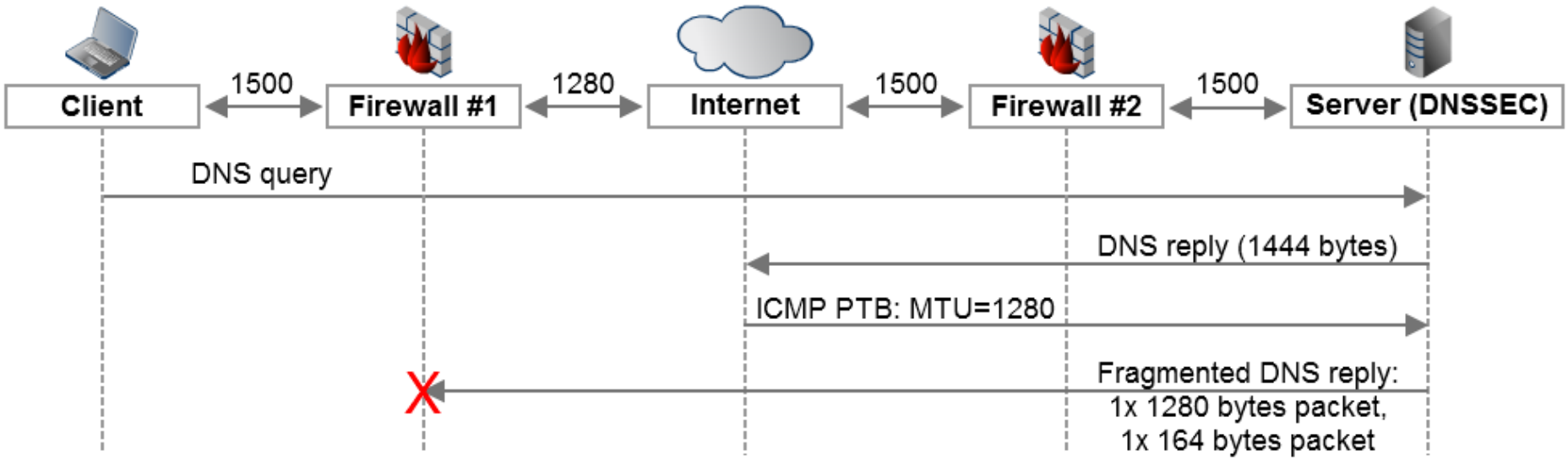
Path MTU Discovery (PMTUD)



Problem #1: ICMP PTB Filtering

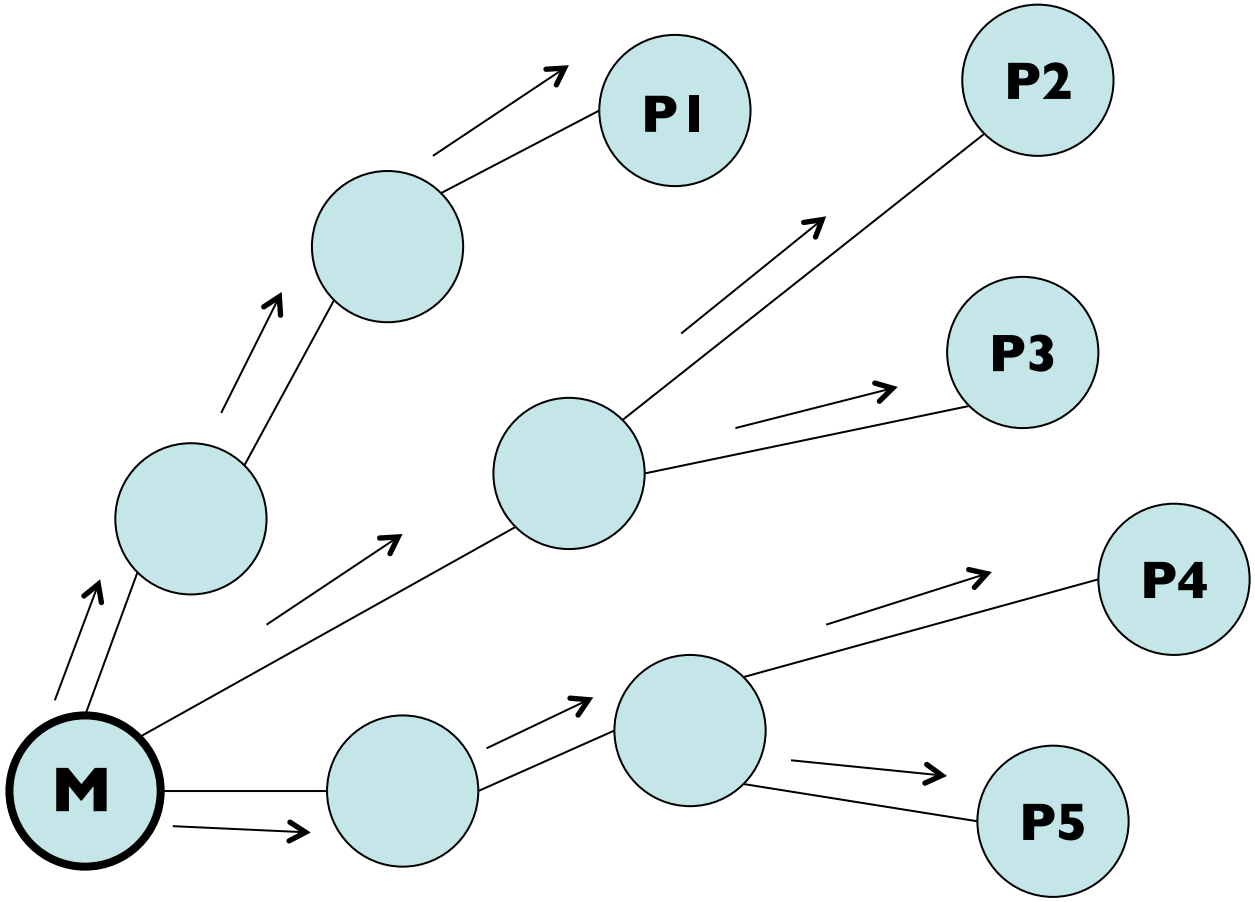


Problem #2: Fragment Filtering



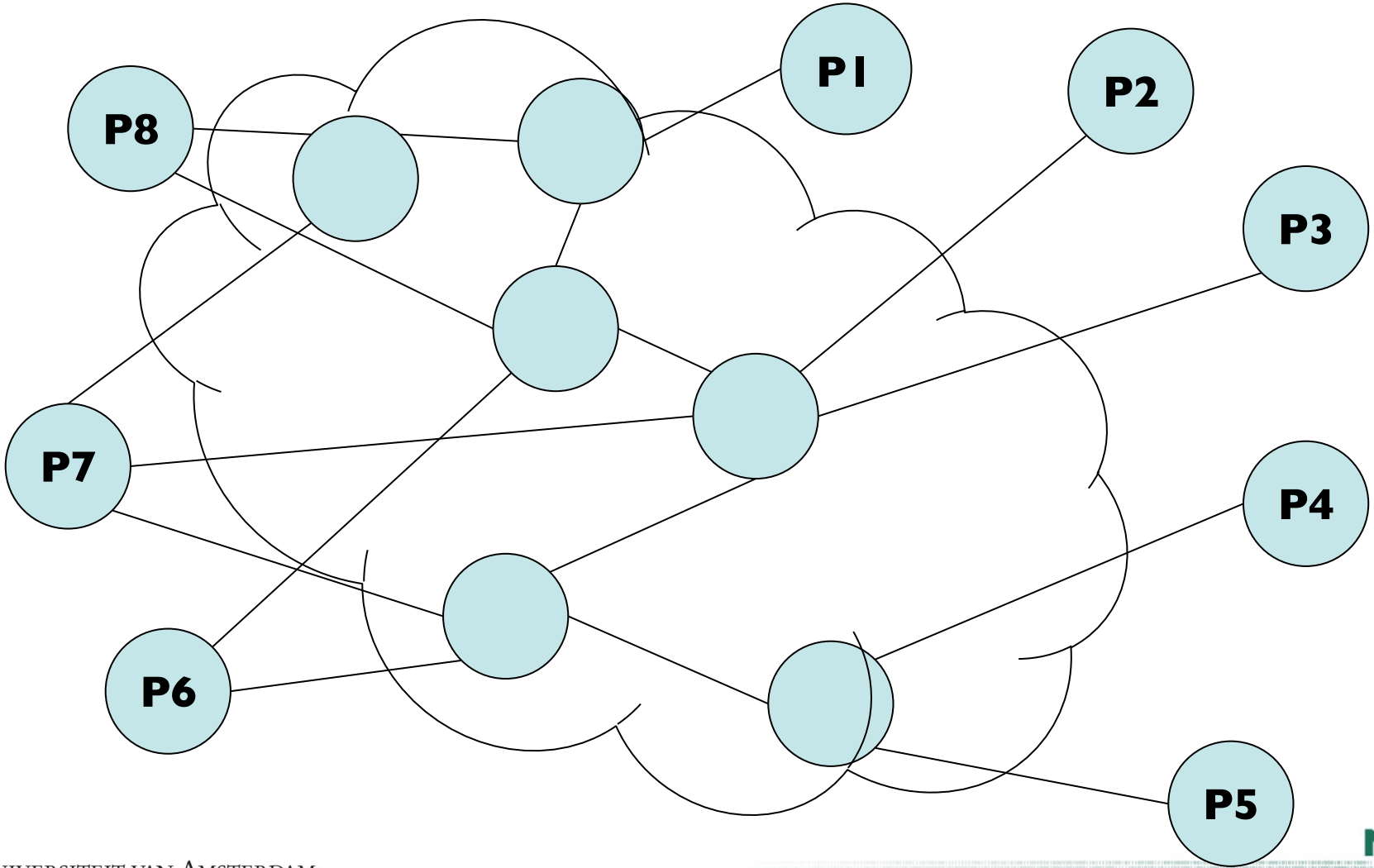
SET-UP OF MEASUREMENTS

Measurement Setup: Alternative I

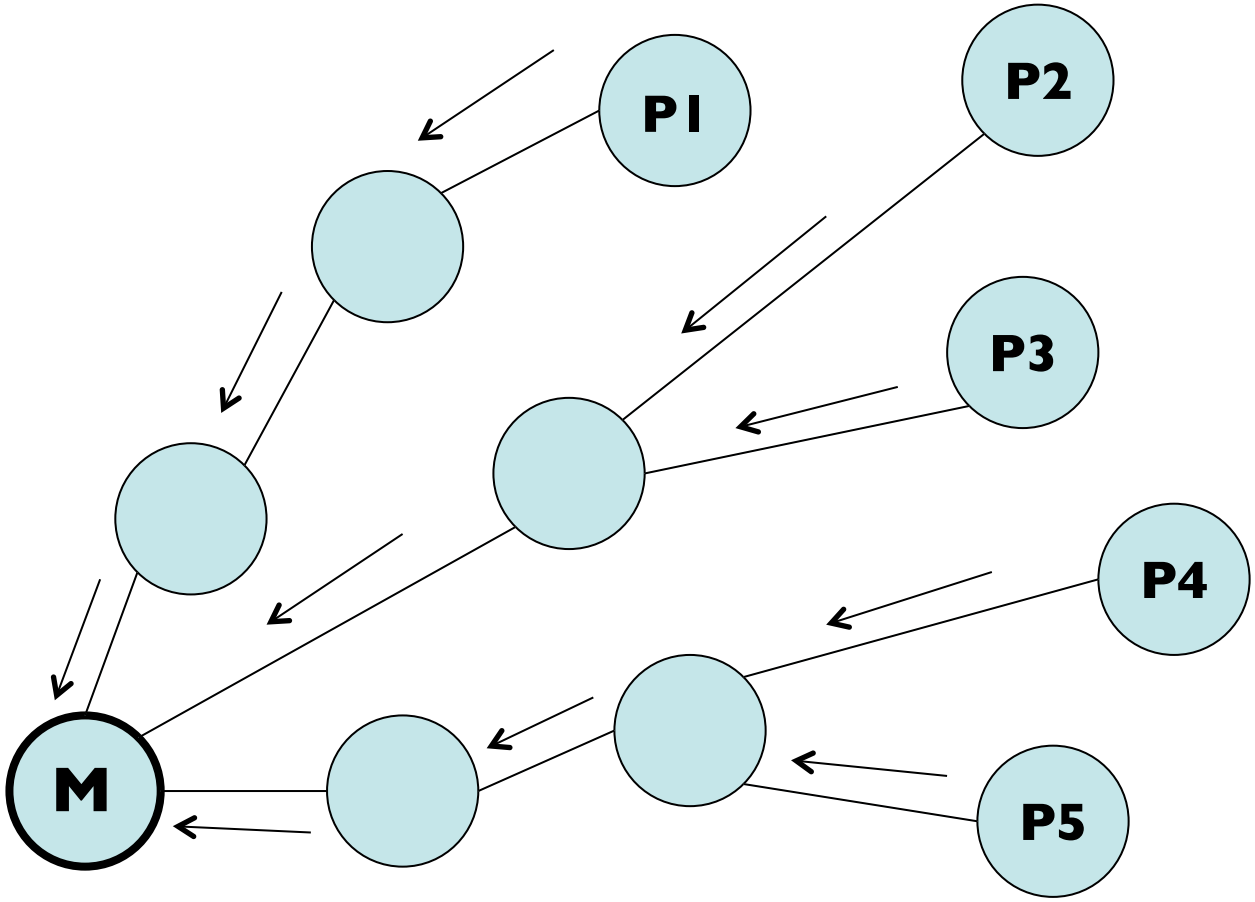


Like PMTUD measurements
by M. Luckie and B. Stasiewicz

Measurement Setup: Alternative 2—Atlas Grid



Measurement Setup: Alternative 3



Measurement Setup

- Combination of Alternative 1 and Alternative 3
  - no triangulation as possible with Alternative 2
  - capture packets on the interface of measurement server
- Use RIPE Atlas infrastructure probes to actively measure ICMP PTB and fragment filtering

RIPE Atlas

- Internet measurement system
- Driven by probes
  - USB-powered embedded devices
- Default measurement functionality:
  - *Ping*
  - *traceroute*
  - low volume, non-intrusive measurements

- Currently around 1700 probes up and running
  - located primarily in the RIPE NCC service region
  - … but also other regions around the globe

RIPE Atlas

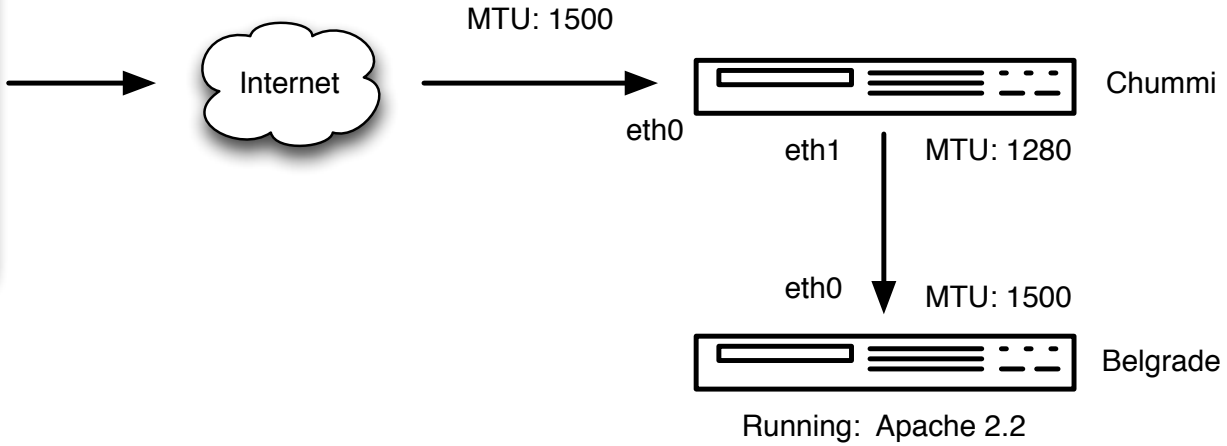
Worldwide network of probes

EXPERIMENTS

Experimental setup

ICMP PTB filtering

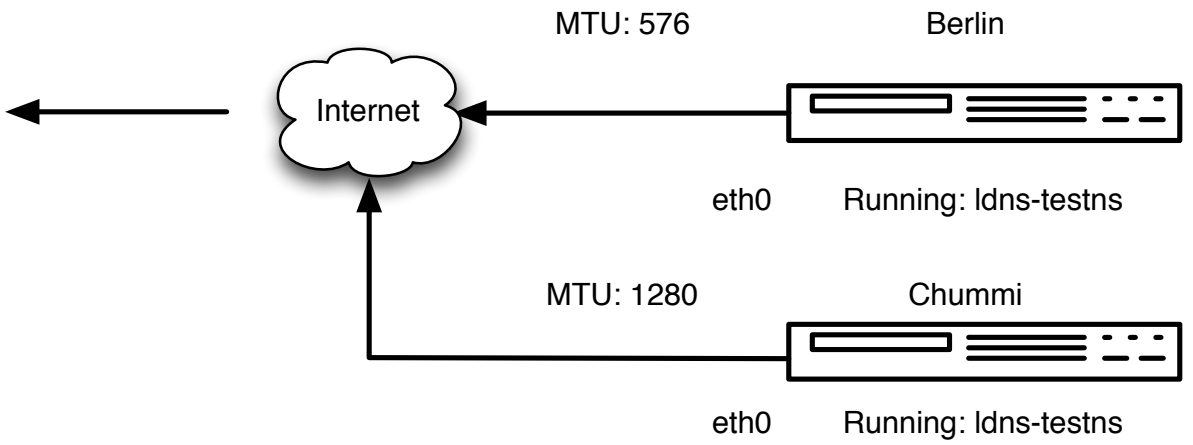
```
POST / HTTP/1.1
Host: httppost6.uranus.nlnetlabs.nl
Connection: close
User-Agent: httpget for
atlas.ripe.net
Content-Type: application/x-www-
form-urlencoded
Content-Length: 65528
```



Experimental setup

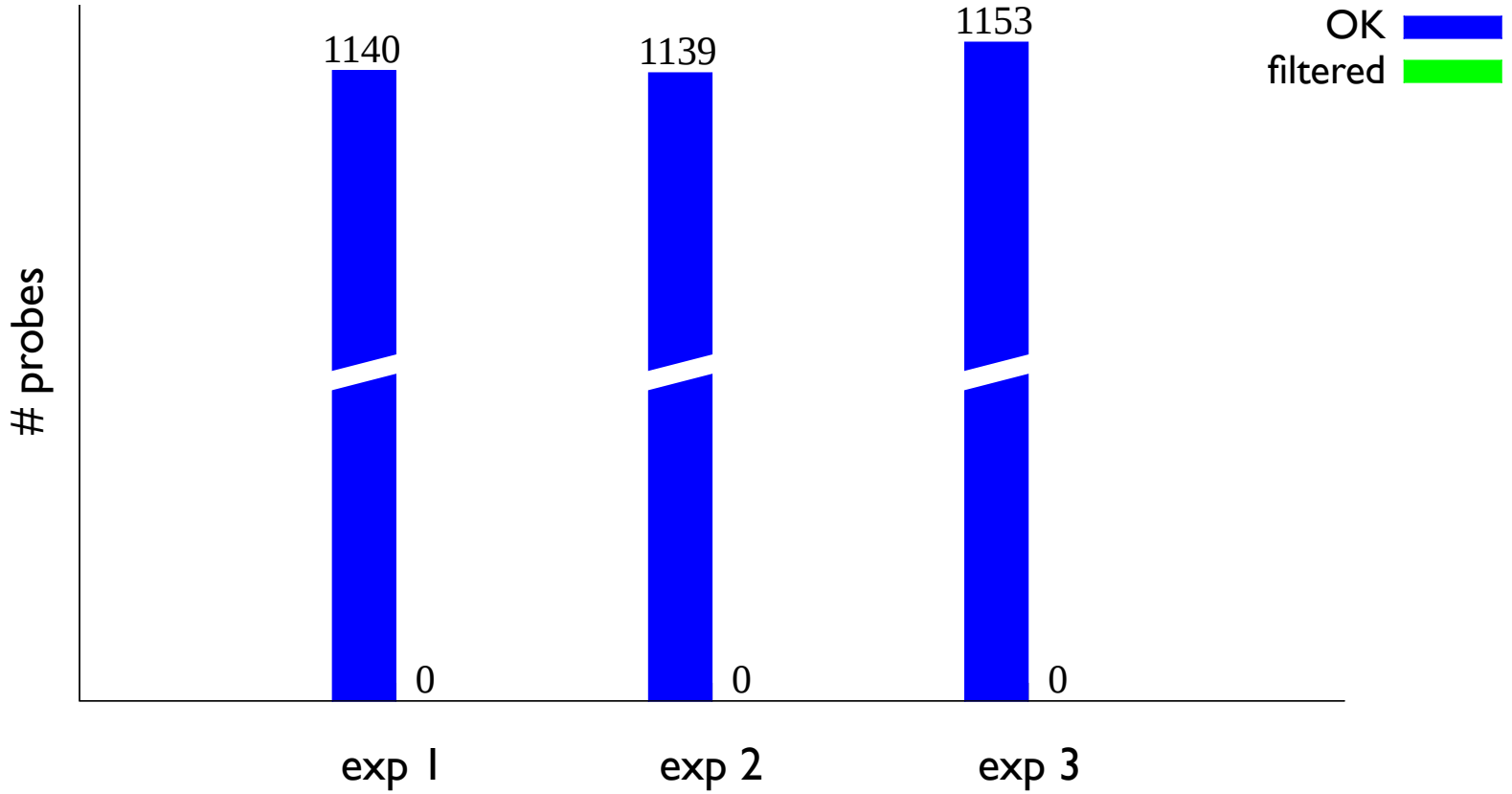
Fragment filtering

```
version.bind. 60 CH TXT

1,002,003,004,005,006,007,008,009,
010,011,012,013,014,015,016,017,01
8,019,020,021,022,023,024,025,

[truncated]

347,348,349,350,351,352,353,354,35
5,356,357,358,359,360,361,362,363,
364,365,366,367,368,369,370,371,37
2,373,374,375,376,377,378,379,380,
381,382,383

MSG SIZE snd: 1590
```



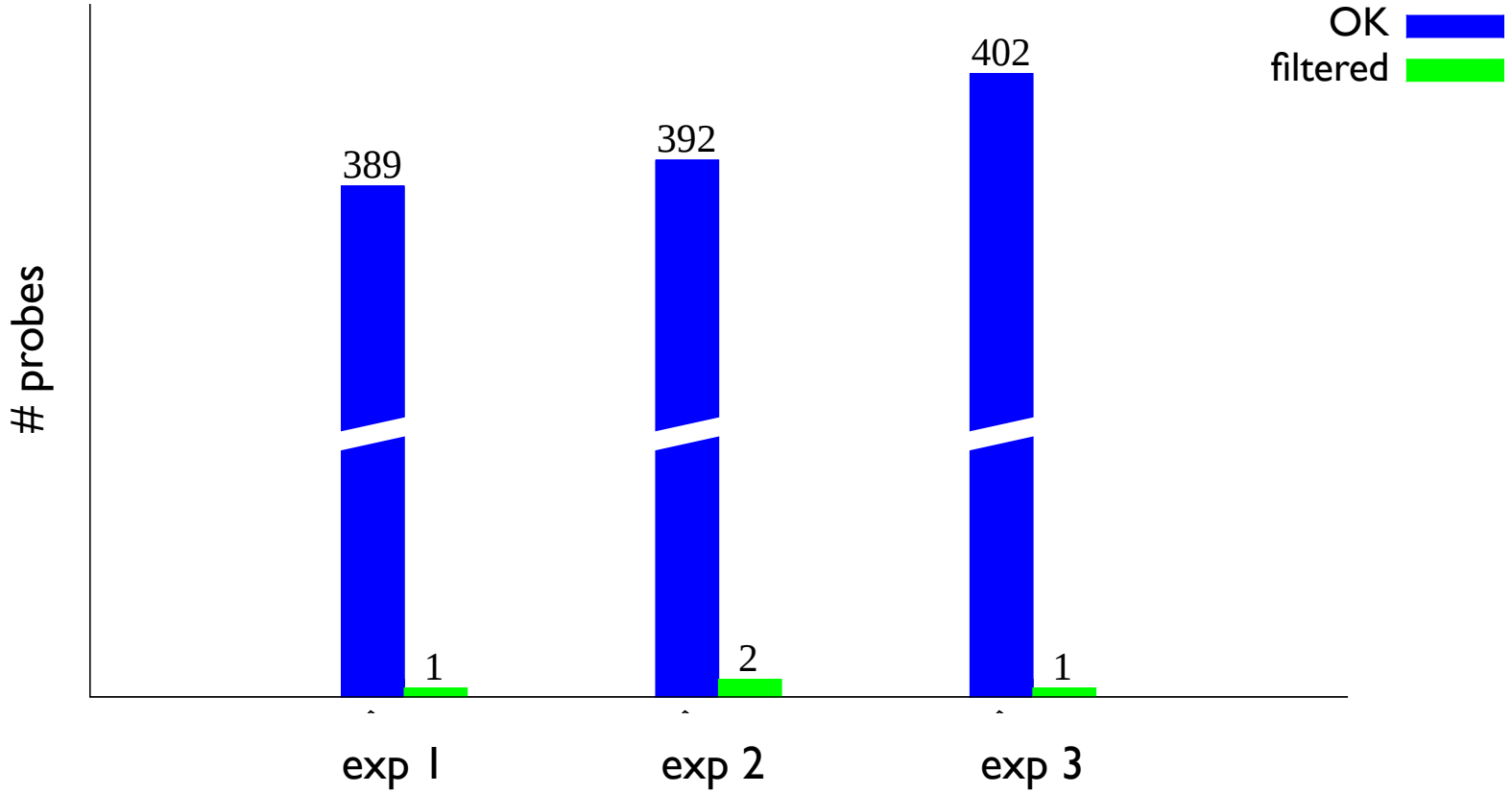
Results

ICMP PTB filtering IPv4 – MTU 1500



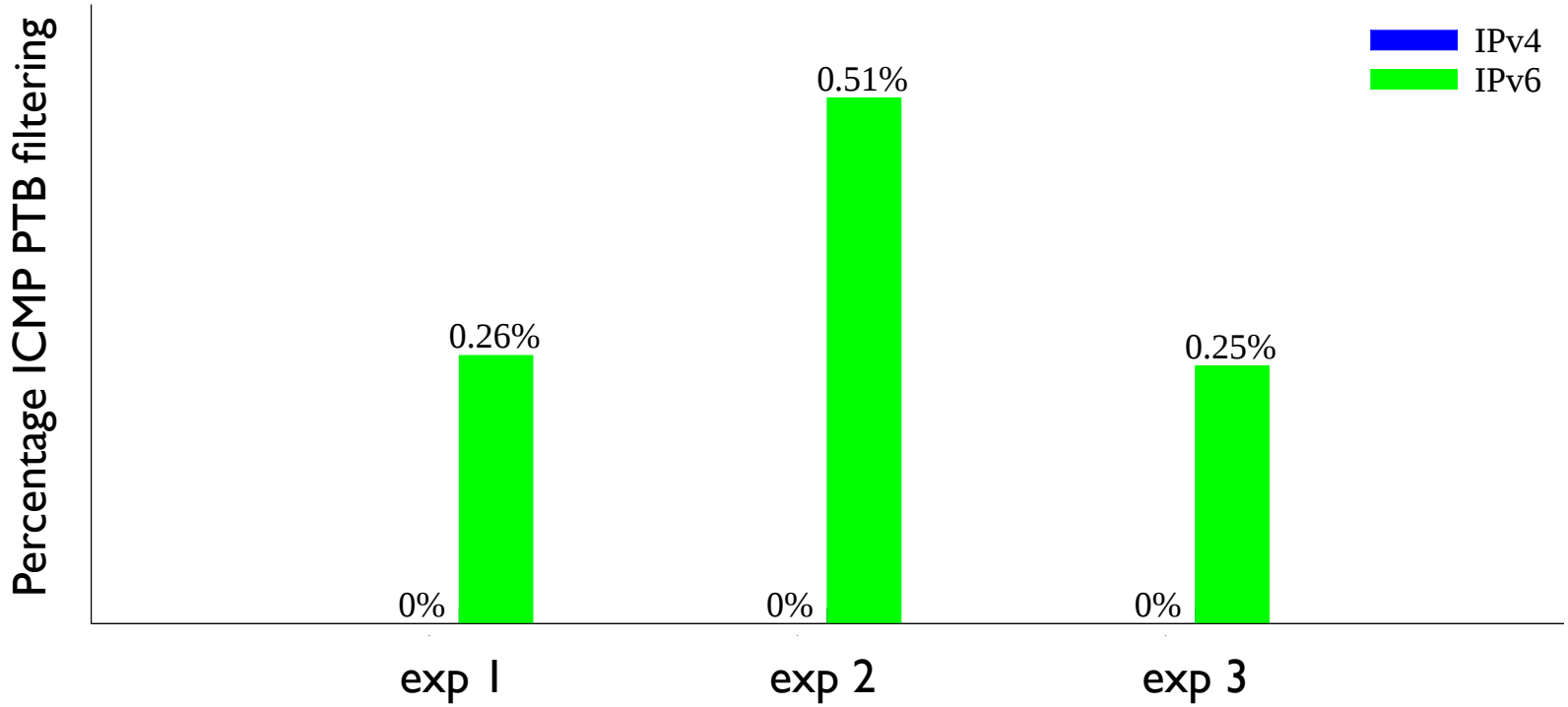
Results

ICMP PTB filtering IPv6 – MTU 1500



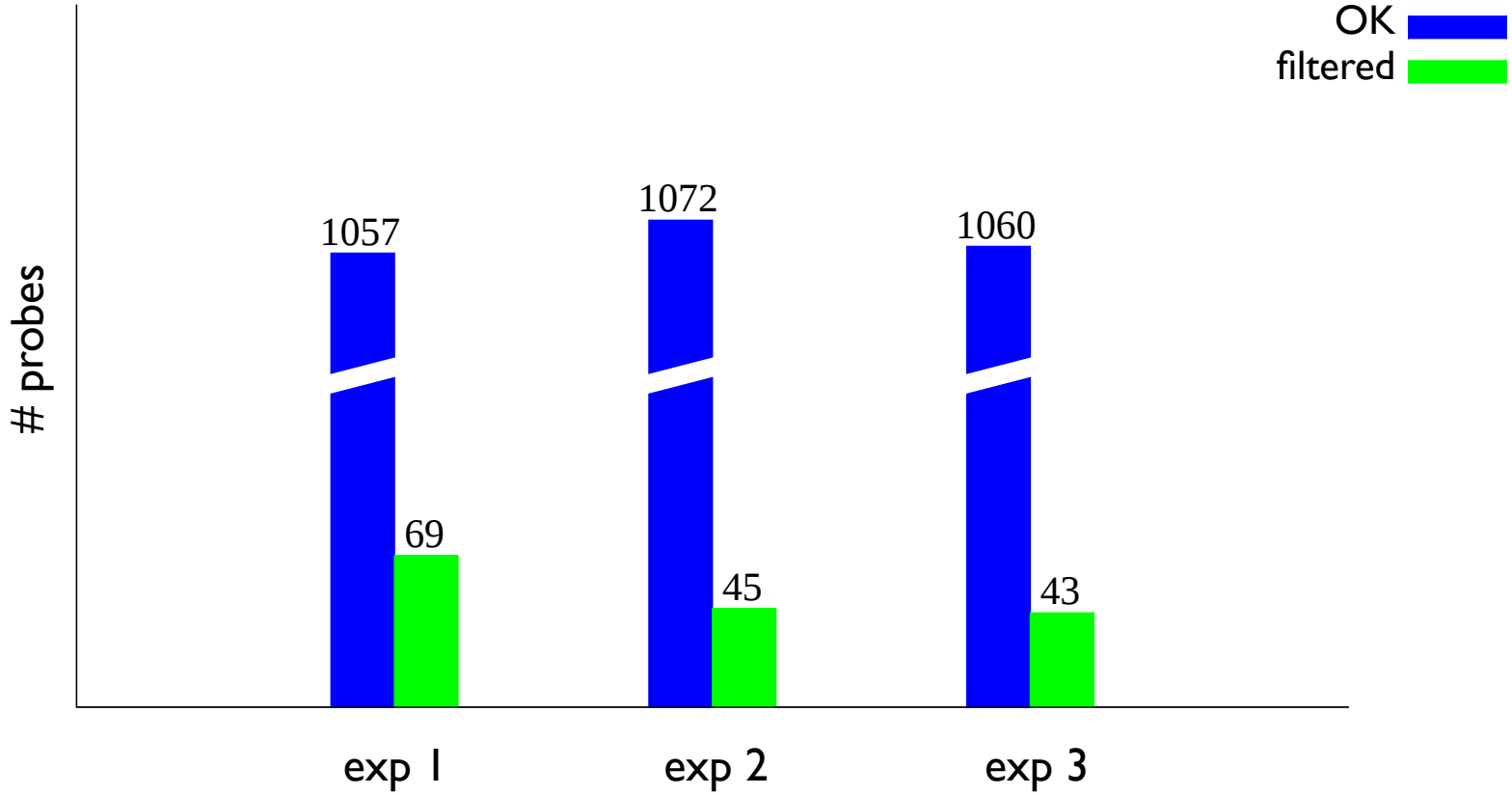
Results

ICMP PTB filtering percentages MTU 1500



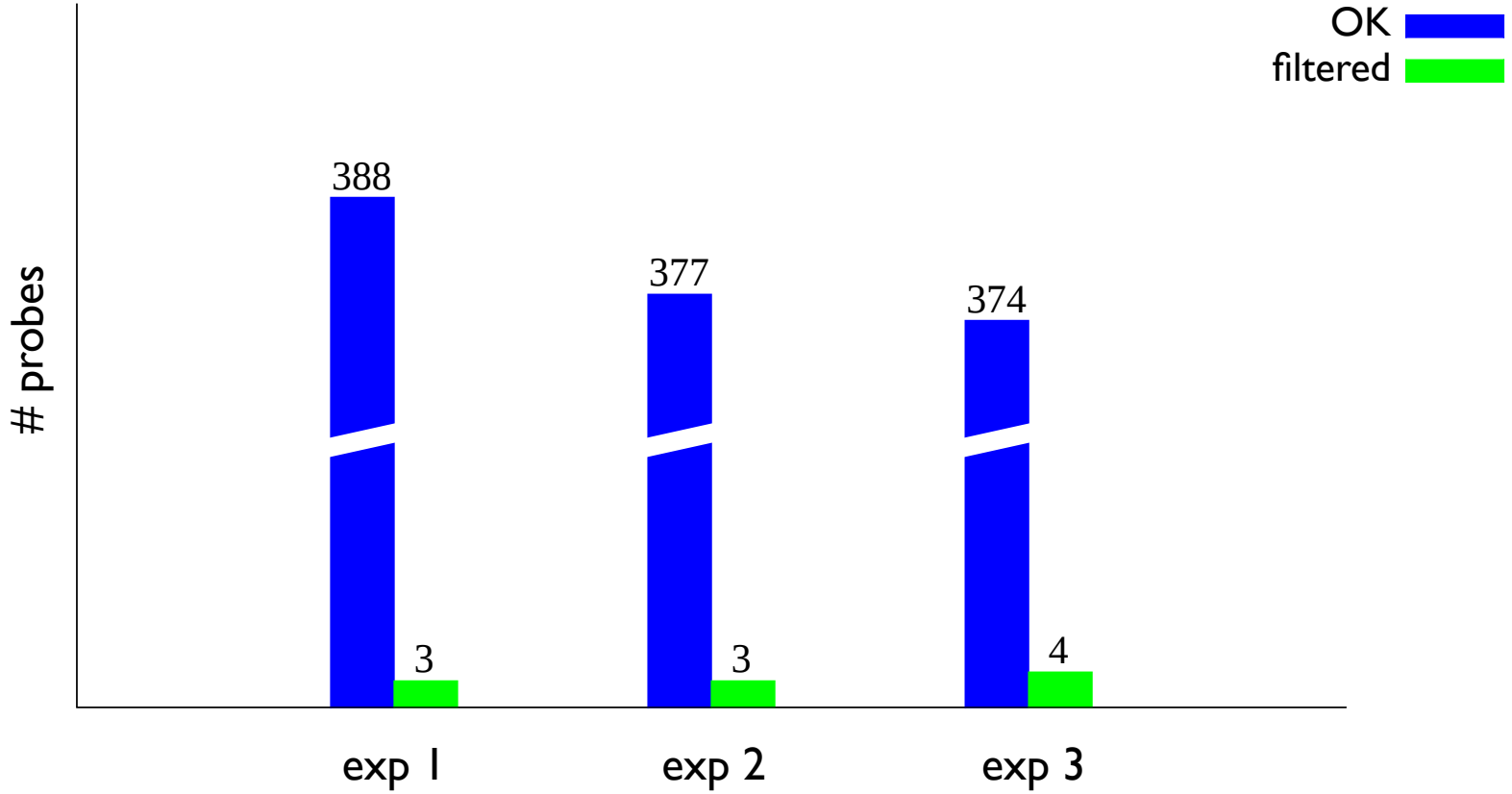
Results

ICMP PTB filtering IPv4 – MTU 1280



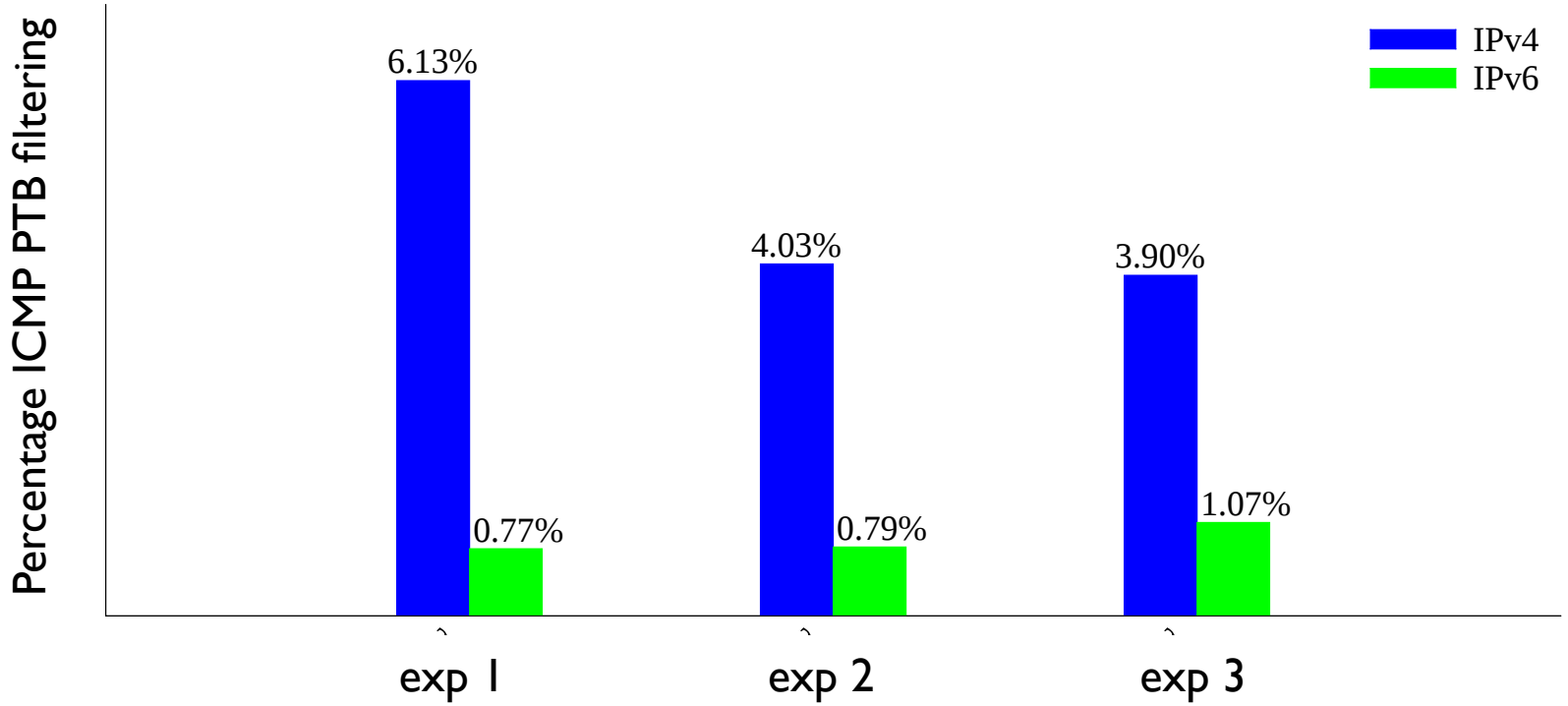
Results

ICMP PTB filtering IPv6 – MTU 1280



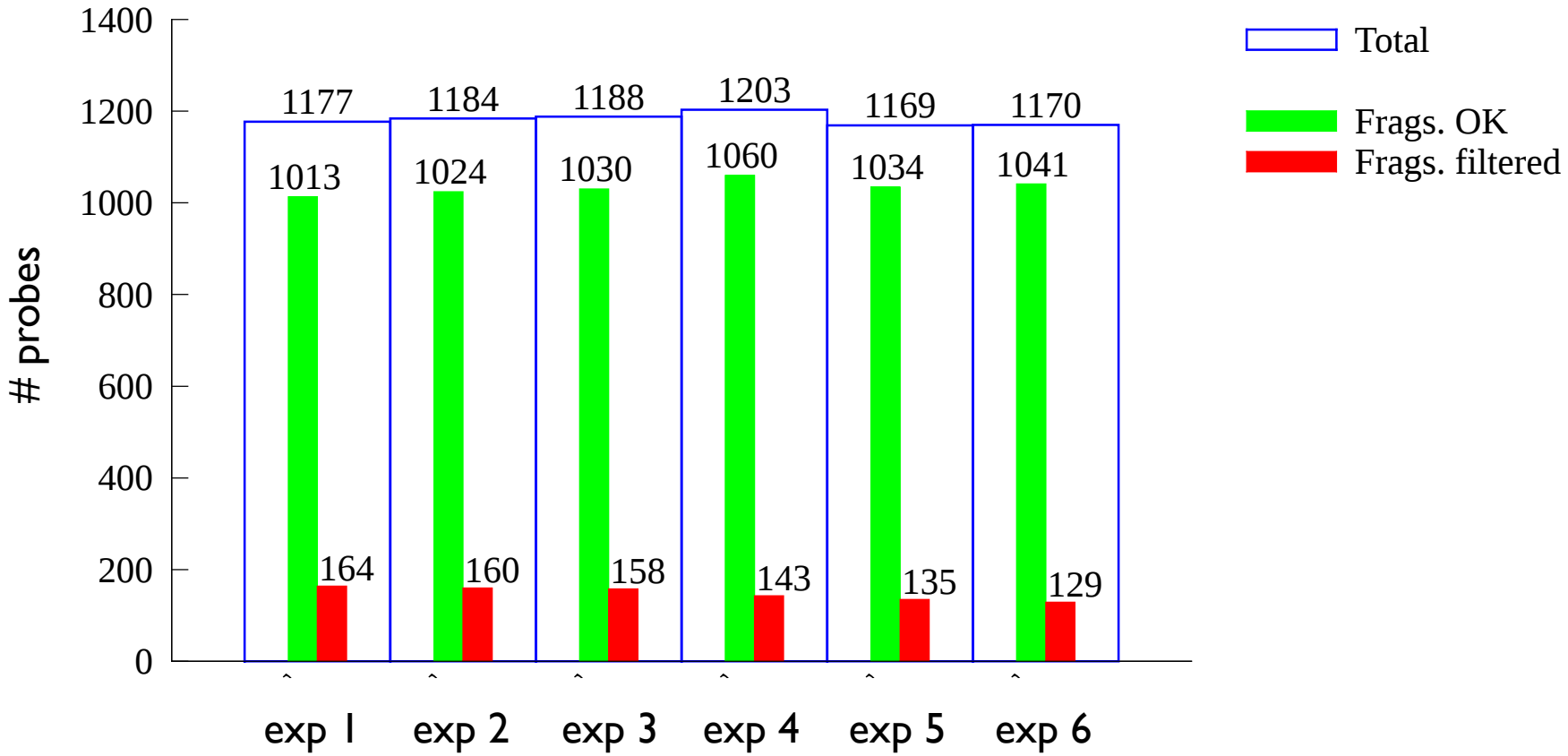
Results

ICMP PTB filtering percentages MTU 1280



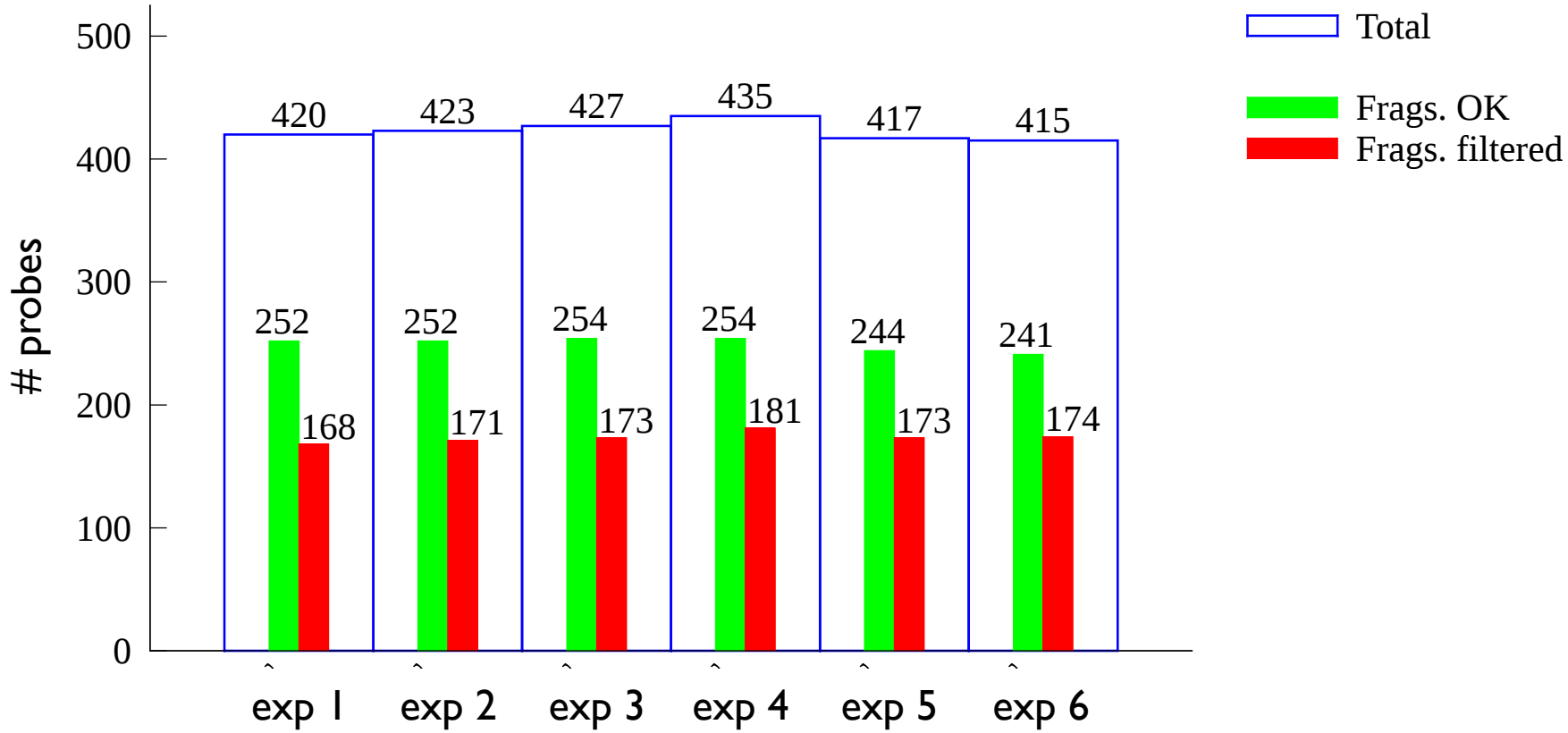
Results

Fragment filtering IPv4 – MTU 1500



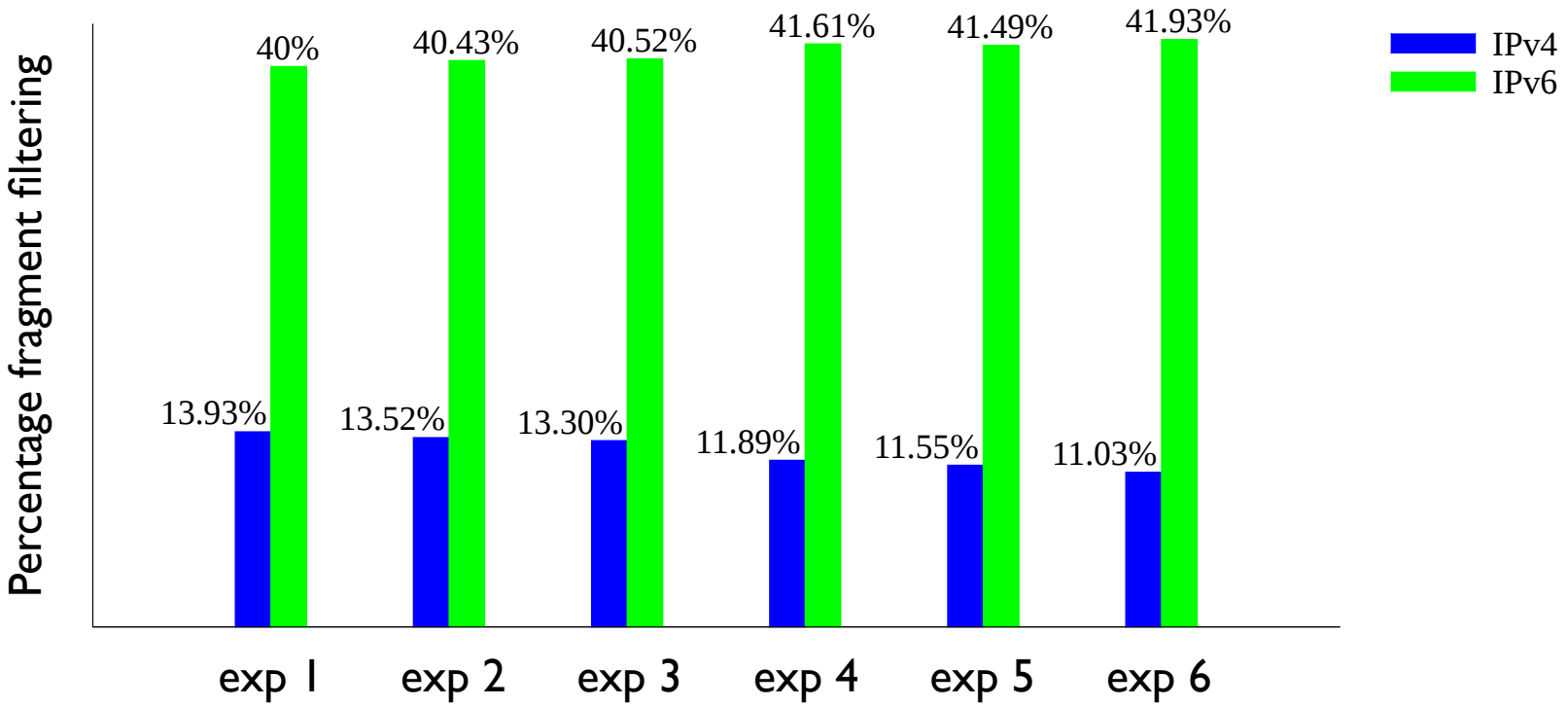
Results

Fragment filtering IPv6 – MTU 1500



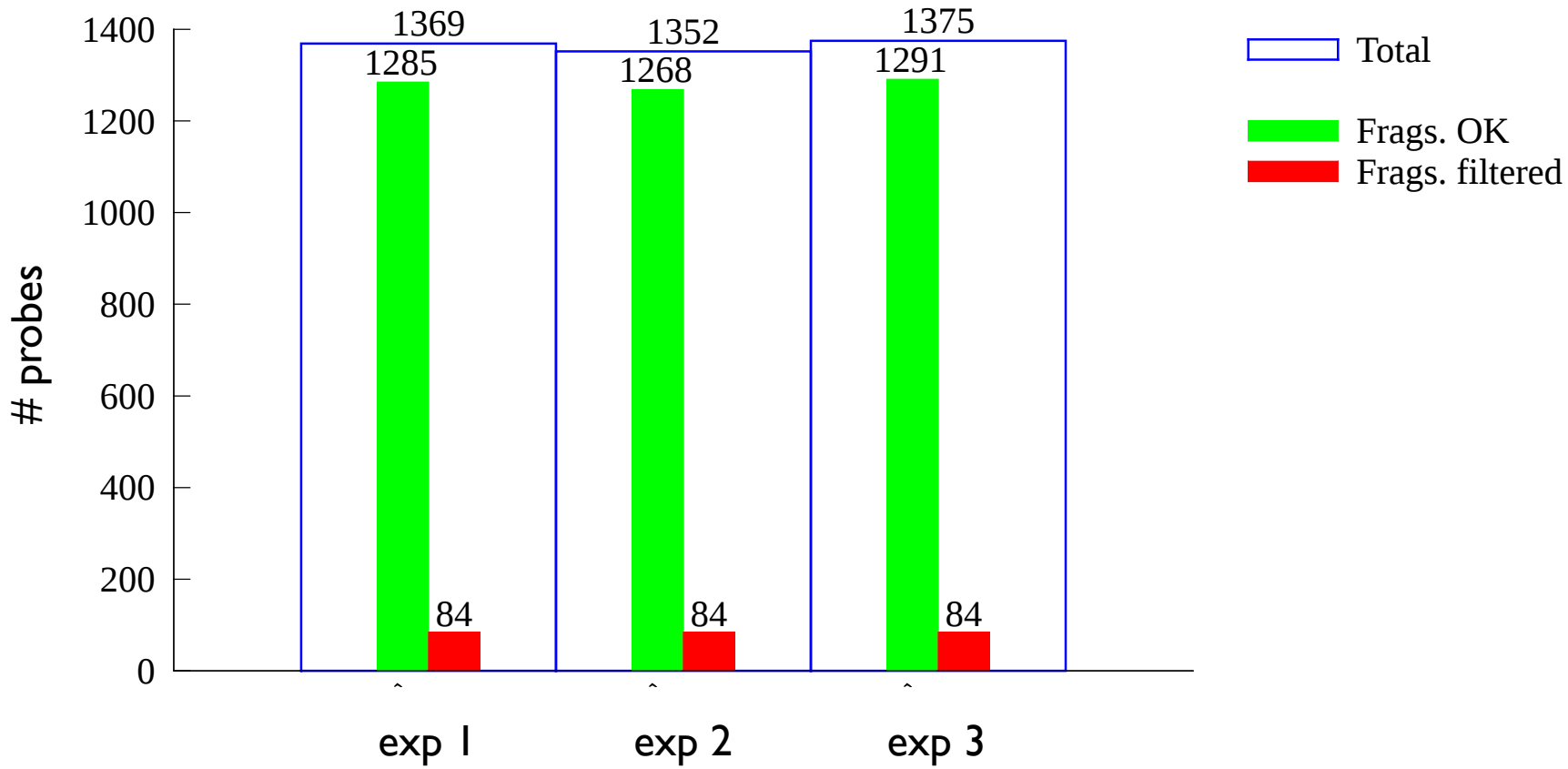
Results

Fragment filtering percentages – MTU 1500



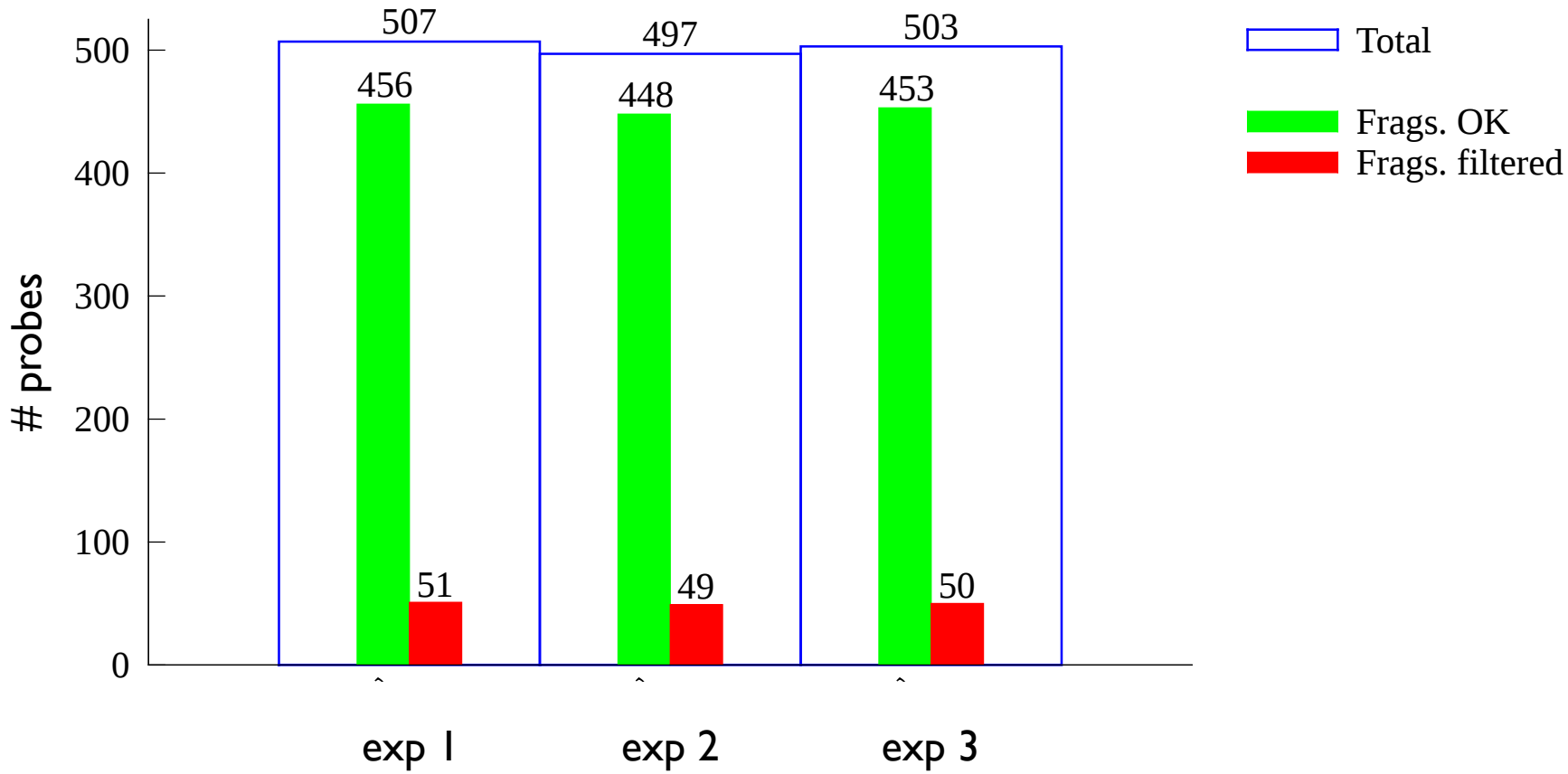
Results

Fragment filtering IPv4 – MTU 576



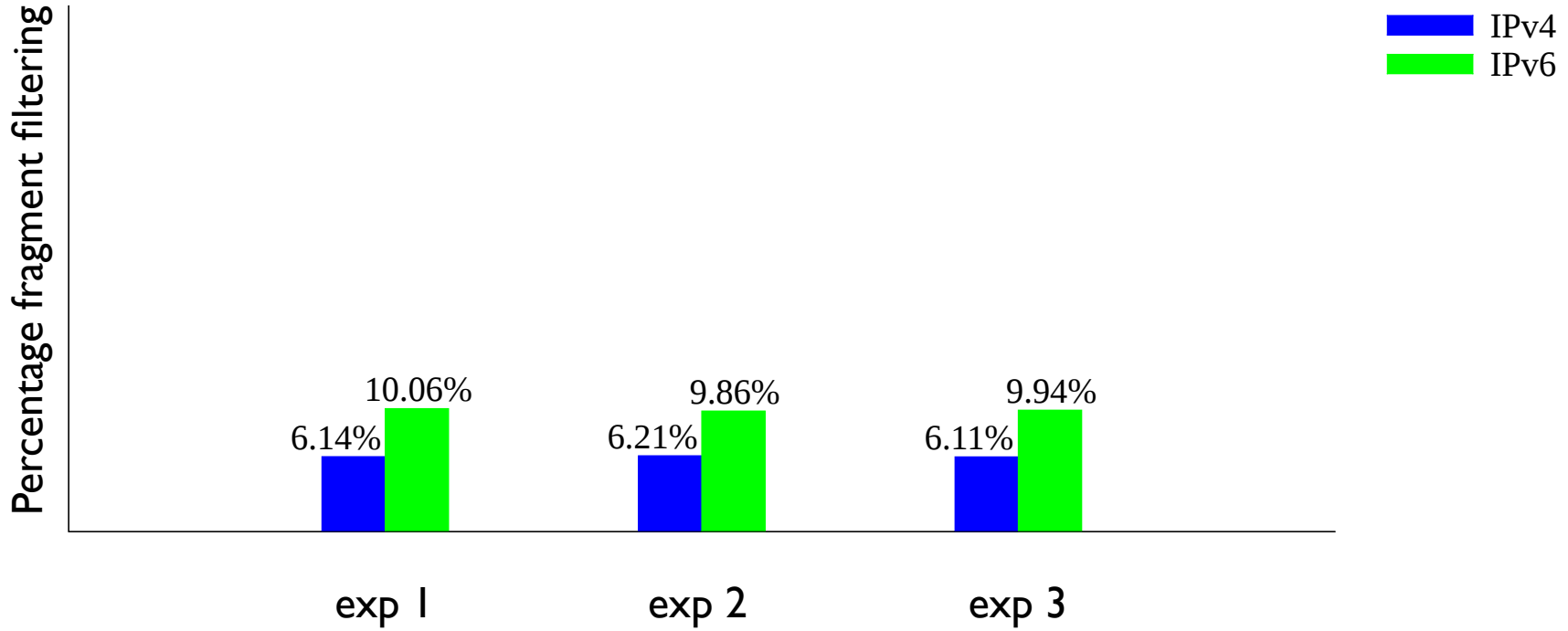
Results

Fragment filtering IPv6 – MTU 1280

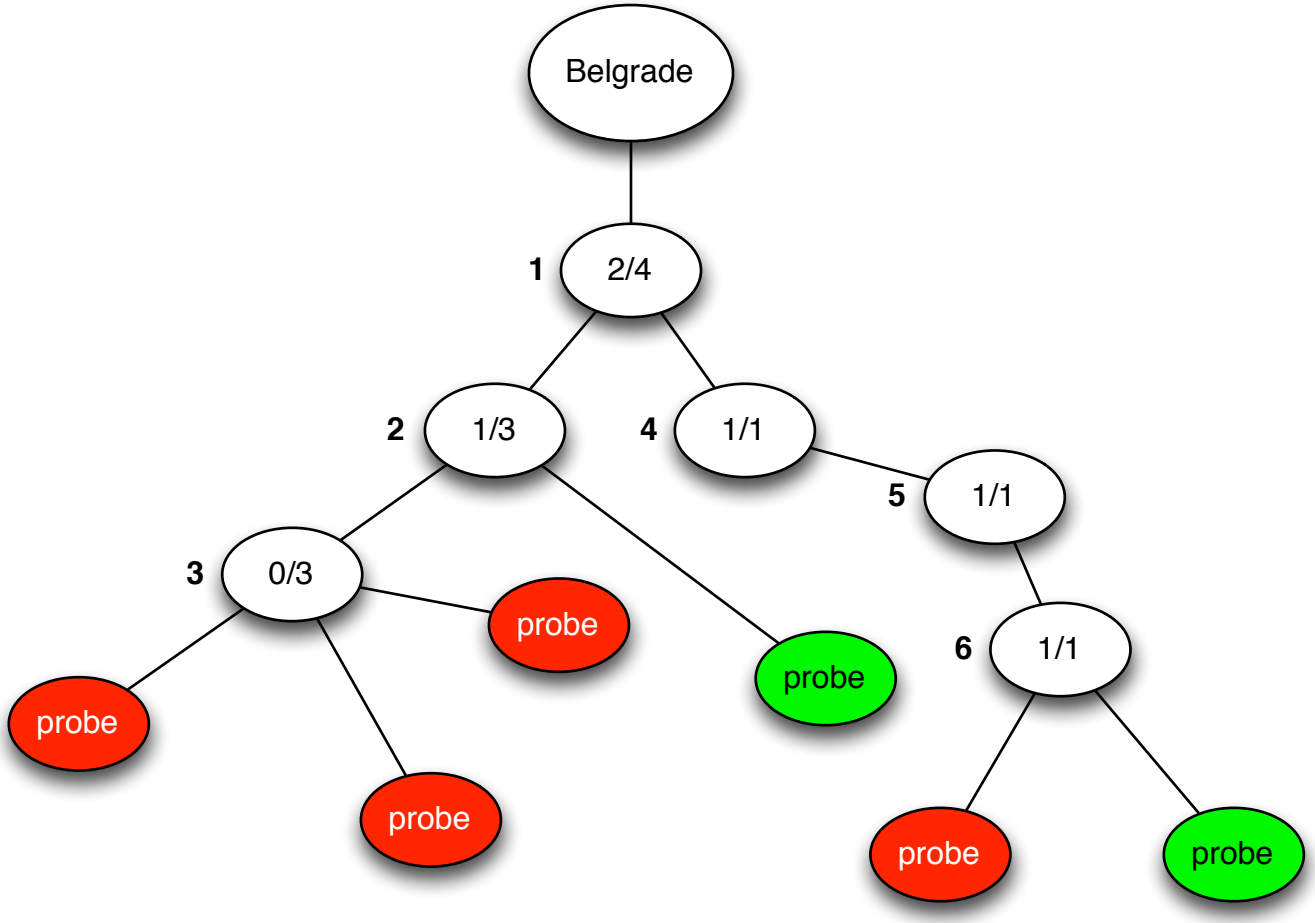


Results

Fragment filtering percentages – MTU 576/1280



Hop counting



Results

Where do IPv4 ICMP PTB messages get filtered?

Bad	Total	Error percentage	Ip
69	1126	6.1%	145.145.19.190
53	810	6.5%	145.145.80.65
16	311	5.1%	145.145.80.73
13	214	6.1%	77.67.72.109
7	199	3.5%	109.105.98.33
2	60	3.3%	62.40.124.157
...			
2	2	100.0%	203.50.6.78
2	2	100.0%	203.50.6.89
2	2	100.0%	61.10.0.118
2	2	100.0%	80.231.159.10
2	2	100.0%	84.116.238.49

Results

Where do IPv6 ICMP PTB messages get filtered?

Bad	Total	Error percentage	Ip
3	391	0.8%	2001:610:158:1916:145:100:99:17
2	292	0.7%	2001:610:e08:64::65
2	131	1.5%	2001:7f8:1::a500:6939:1
1	9	11.1%	2001:470:0:217::2
1	6	16.7%	2001:470:0:67::2
1	46	2.2%	2001:470:0:3f::1

...

No routers with 100% failure rate

Results

Where do IPv4 fragments get filtered?

Bad	Total	Error percentage	Ip
84	1369	6.1%	145.145.19.190
56	983	5.7%	145.145.80.65
28	381	7.3%	145.145.80.73
14	256	5.5%	109.105.98.33
21	247	8.5%	77.67.72.109
9	62	14.5%	62.40.124.157
...			
3	3	100.0%	212.188.22.158
2	2	100.0%	146.97.33.137
2	2	100.0%	158.64.16.189
2	2	100.0%	174.35.131.38
2	2	100.0%	188.230.128.10

Results

Where do IPv6 fragments get filtered?

Bad	Total	Error percentage	Ip
181	435	41.6%	2001:610:158:1916:145:100:99:17
138	322	42.9%	2001:610:e08:64::65
74	146	50.7%	2001:7f8:1::a500:6939:1
28	53	52.8%	2001:470:0:3f::1
27	91	29.7%	2001:610:e08:72::73
21	53	39.6%	2001:948:2:6::1
...			
6	6	100.0%	2001:610:f01:9012::14
4	4	100.0%	2001:16d8:aaaa:5::2
4	4	100.0%	2001:7f8:1::a503:9326:1
4	4	100.0%	2a01:348::10:0:1
4	4	100.0%	2a01:348::27:0:1

Conclusion

- ICMP PTB messages get dropped
  - More for IPv4 but nobody notices
  - Complete path is 1500 and DF bit helps
- Fragments get dropped
  - More in IPv6
  - DNS servers do not respond to ICMP PTB
- Path MTU black holes
  - Occur on the edges of the Internet, not in the core

Recommendations

- Recommendations for Filtering ICMPv6 Messages in Firewalls – RFC4890
- Don't filter IPv4 ICMP type 3 code 4
- Packetization Layer Path MTU Discovery – RFC4821
- Don't filter fragments (problems for DNSSEC)
- Don't reduce MTU on interface
- No MSS clamping

Acknowledgements

- RIPE NCC
  - Philip Homburg
  - Andreas Strikos
  - Vesna Manojlovic
  - Emile Aben

MSc. Thesis available

http://www.nlnetlabs.nl/publications/

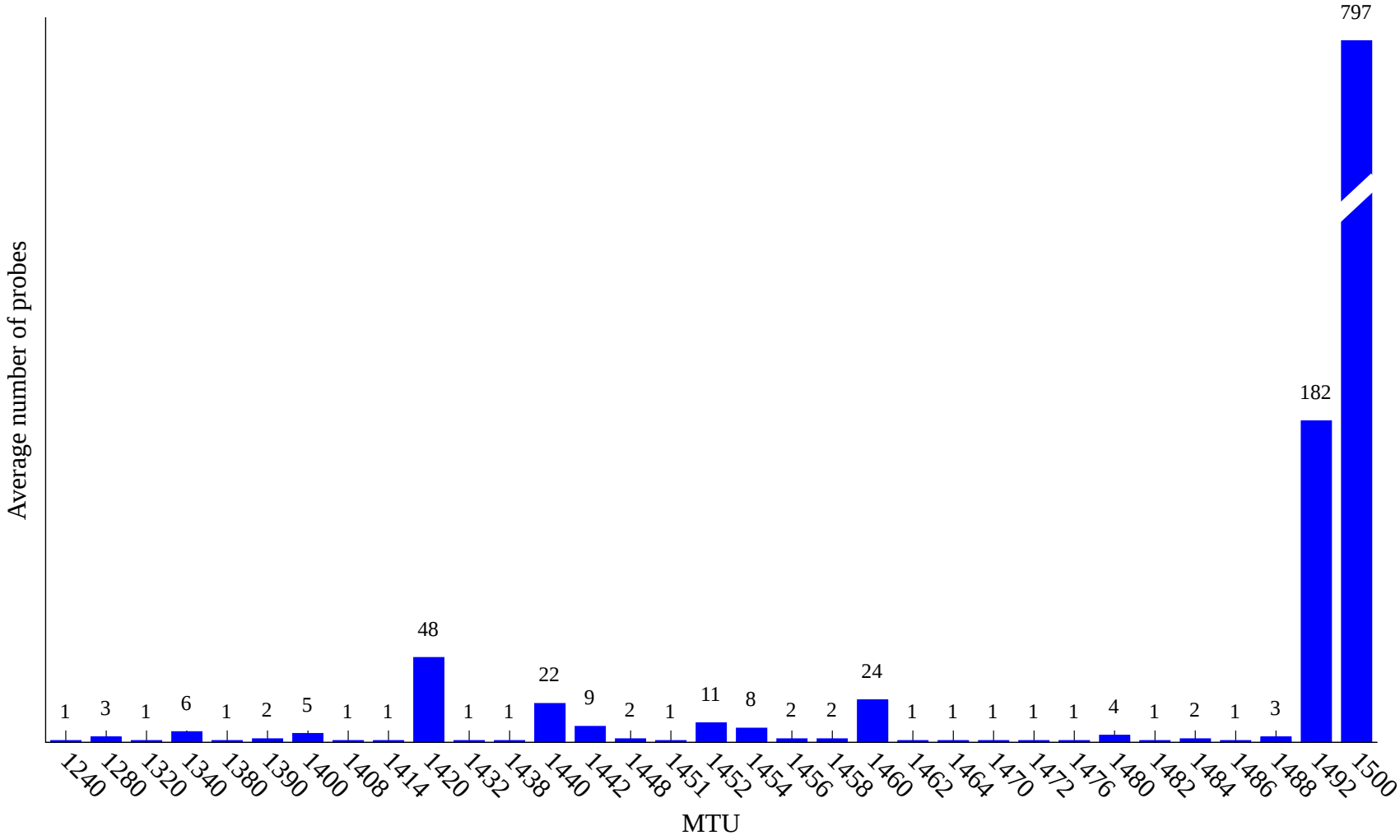
Questions?

Maikel de Boer – maikel.deboer@os3.nl
Jeffrey Bosma – jeffrey.bosma@os3.nl

PMTU DETERMINATION

Results

Path MTU determination IPv4



Results

Path MTU determination IPv6