Measuring the usable maximum packet size across Internet paths: *How can we make PMTUD work?*

Ana Custura
Iain Learmonth
Gorry Fairhurst

University of Aberdeen, Scotland, UK, EU

MONROE

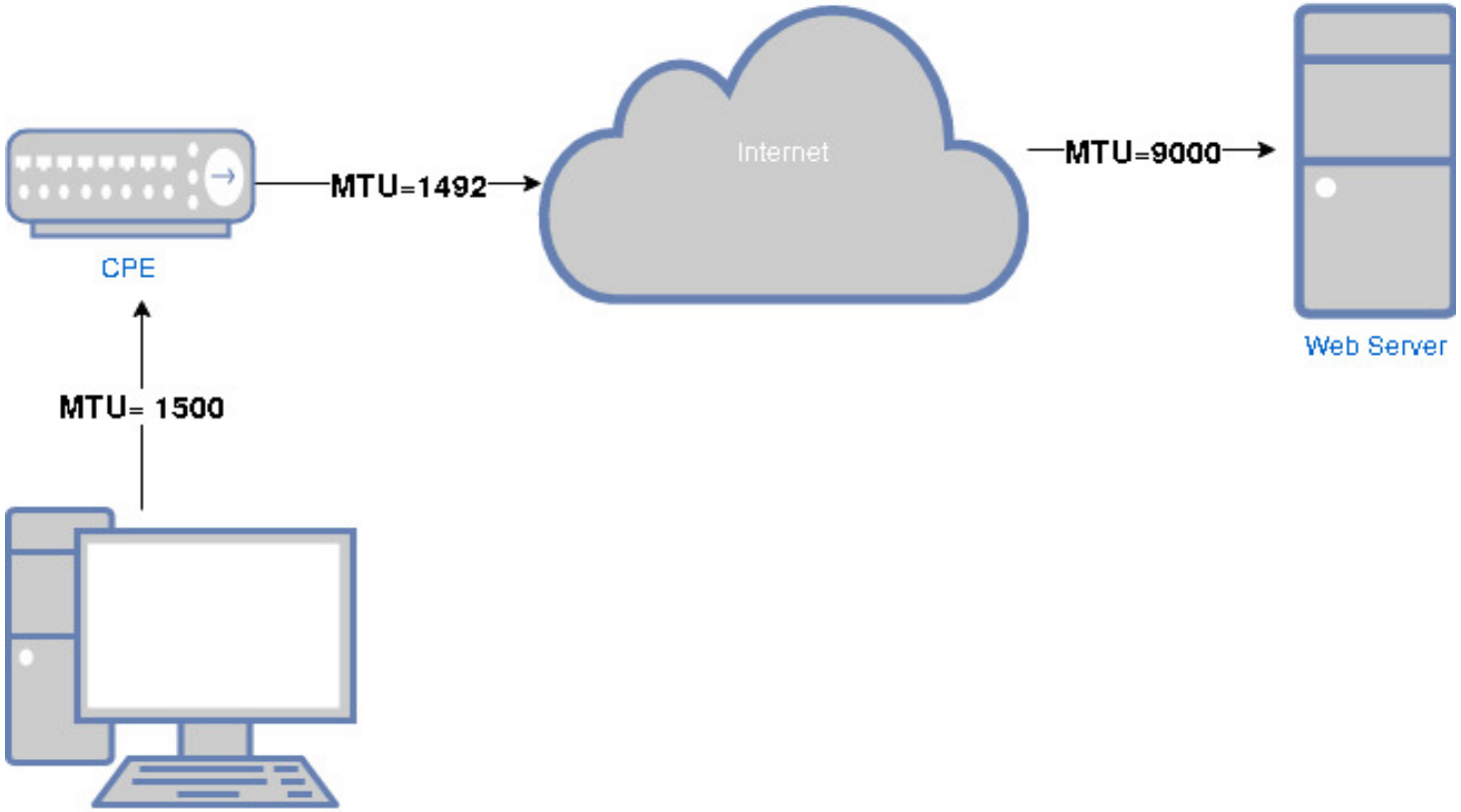
UNIVERSITY OF ABERDEEN

mami

It's "good" to send big packets

PMTU Discovery (PMTUD)

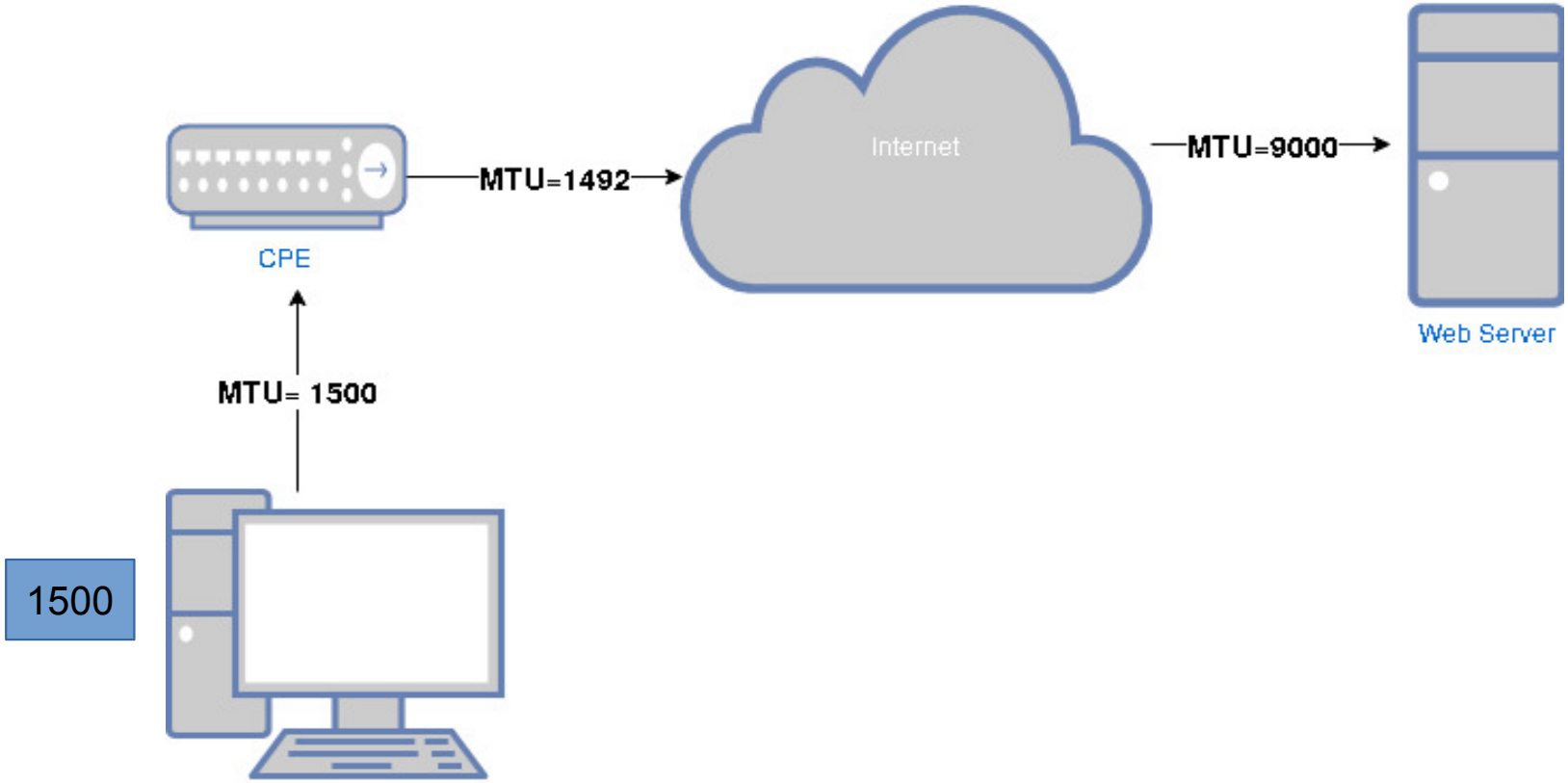
Network layer mechanism to determine the PMTU using PTB



It's "good" to send big packets

PMTU Discovery (PMTUD)

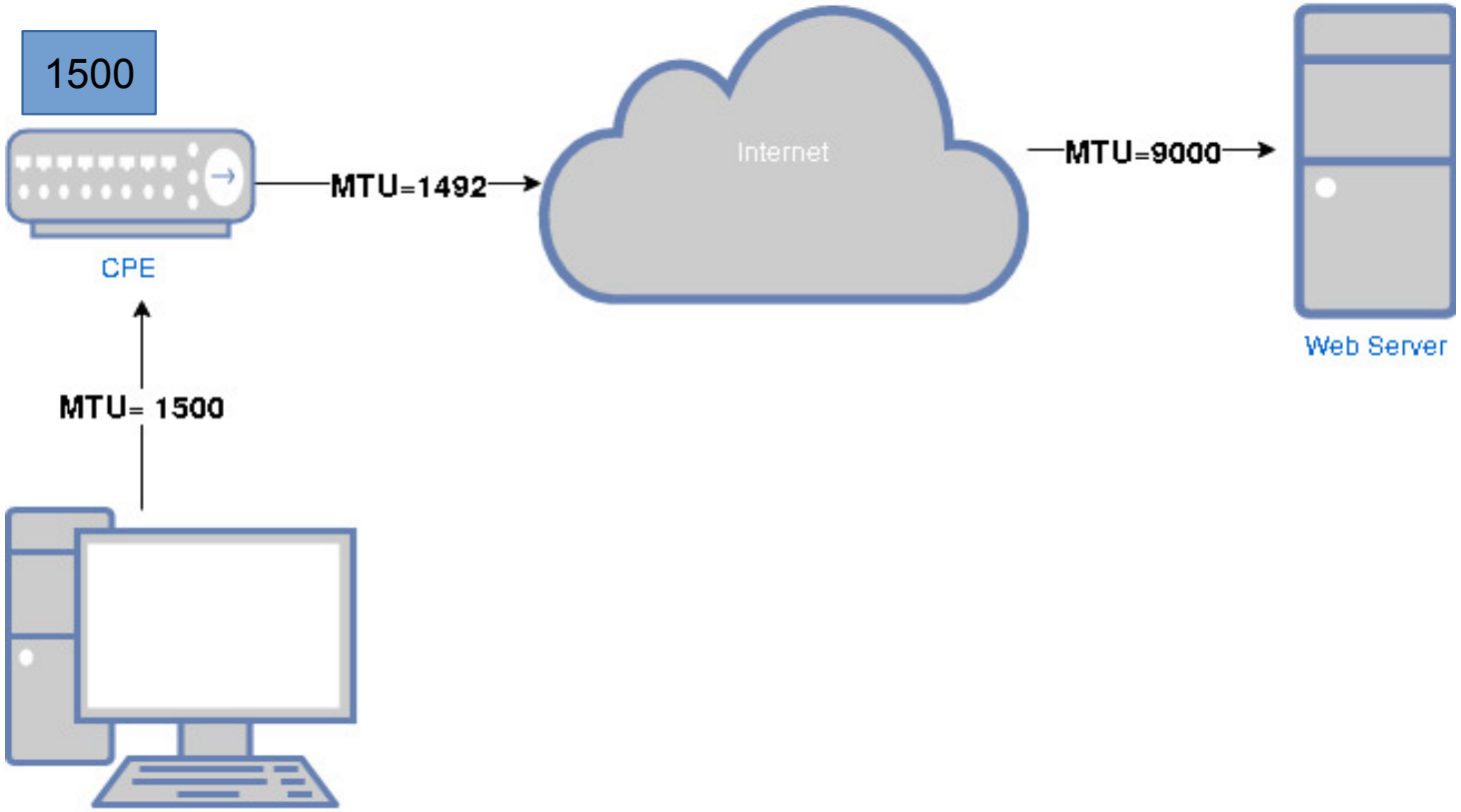
Network layer mechanism to determine the PMTU using PTB



It's "good" to send big packets

PMTU Discovery (PMTUD)

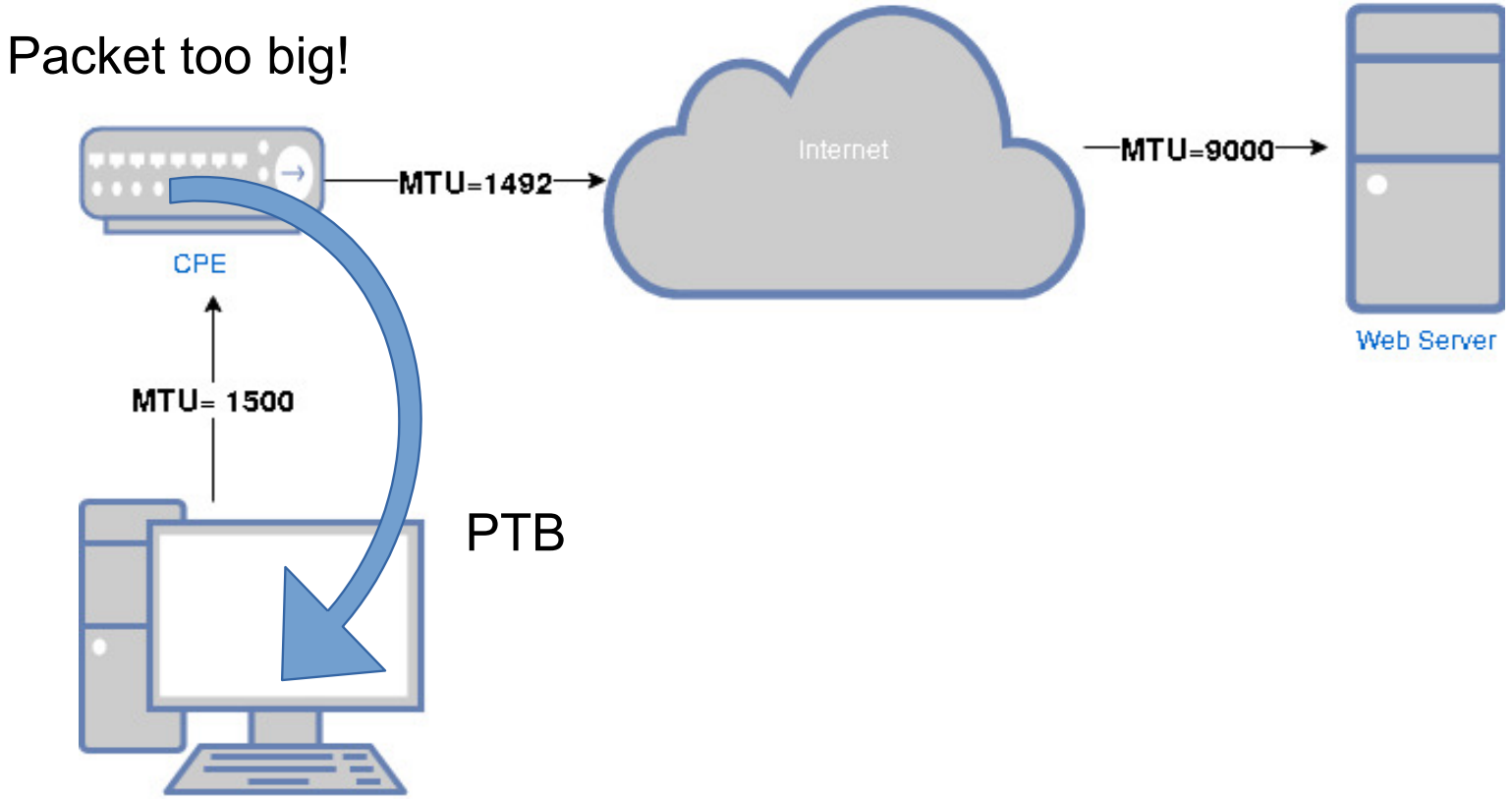
Network layer mechanism to determine the PMTU using PTB



It's "good" to send big packets

PMTU Discovery (PMTUD)

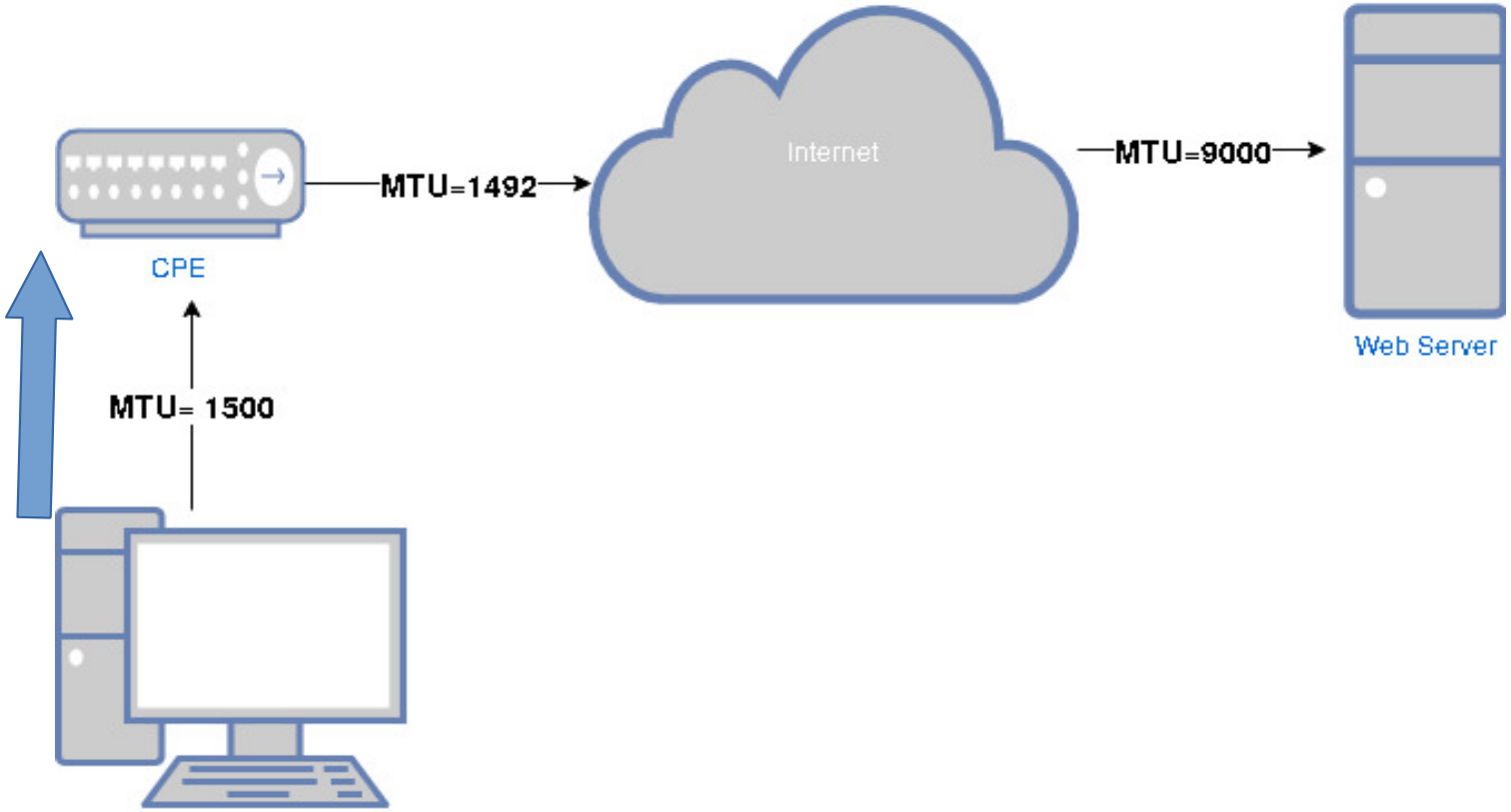
Network layer mechanism to determine the PMTU using PTB



It's "good" to send big packets

PMTU Discovery (PMTUD)

Network layer mechanism to determine the PMTU using PTB



It's even better to avoid black holes

PTB unreliable -> hence PMTUD actually doesn't work :-(

ICMP firewalls, CPE

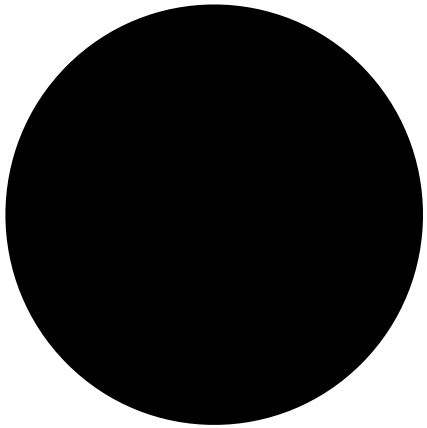
ECMP (+Firewalls and others) make ICMP unreliable

It's even better to avoid black holes

PTB unreliable -> hence PMTUD actually doesn't work :-(

ICMP firewalls, CPE

ECMP (+Firewalls and others) make ICMP unreliable



MONROE

UNIVERSITY OF ABERDEEN

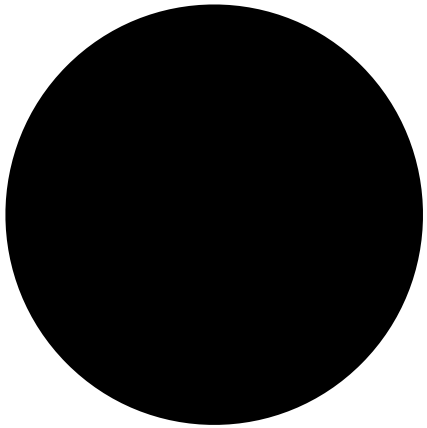
mami

It's even better to avoid black holes

PTB unreliable -> hence PMTUD actually doesn't work :-(

ICMP firewalls, CPE

ECMP (+Firewalls and others) make ICMP unreliable



The transport KNOWS when this happens…

(Packets > PMTU) do not arrive at the destination

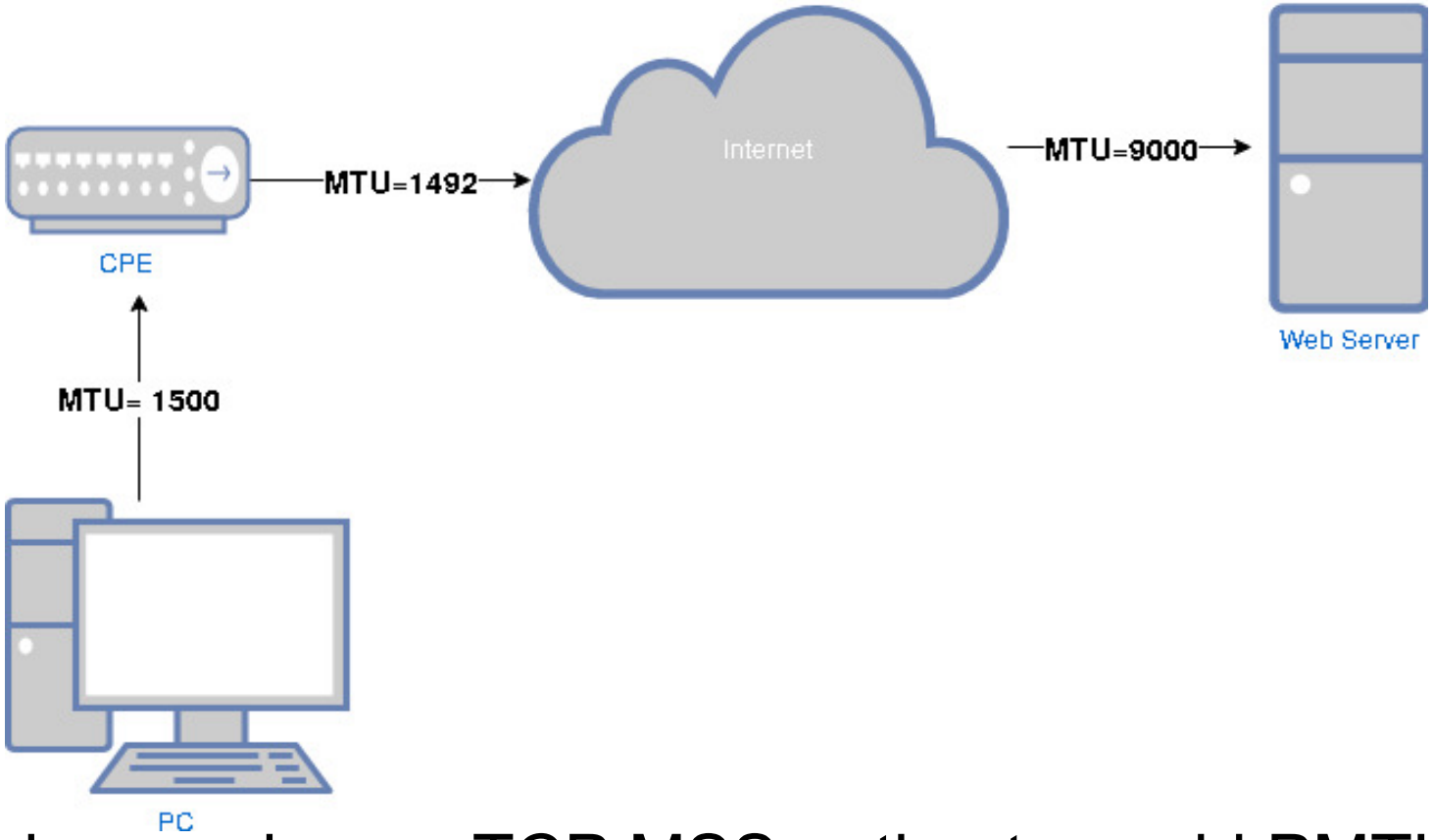
MONROE

UNIVERSITY OF ABERDEEN

mami

TCP Maximum Segment Size

TCP MSS option advertises remote link MTU



Middleboxes change TCP MSS option to avoid PMTUD failures: TCP MSS Clamping

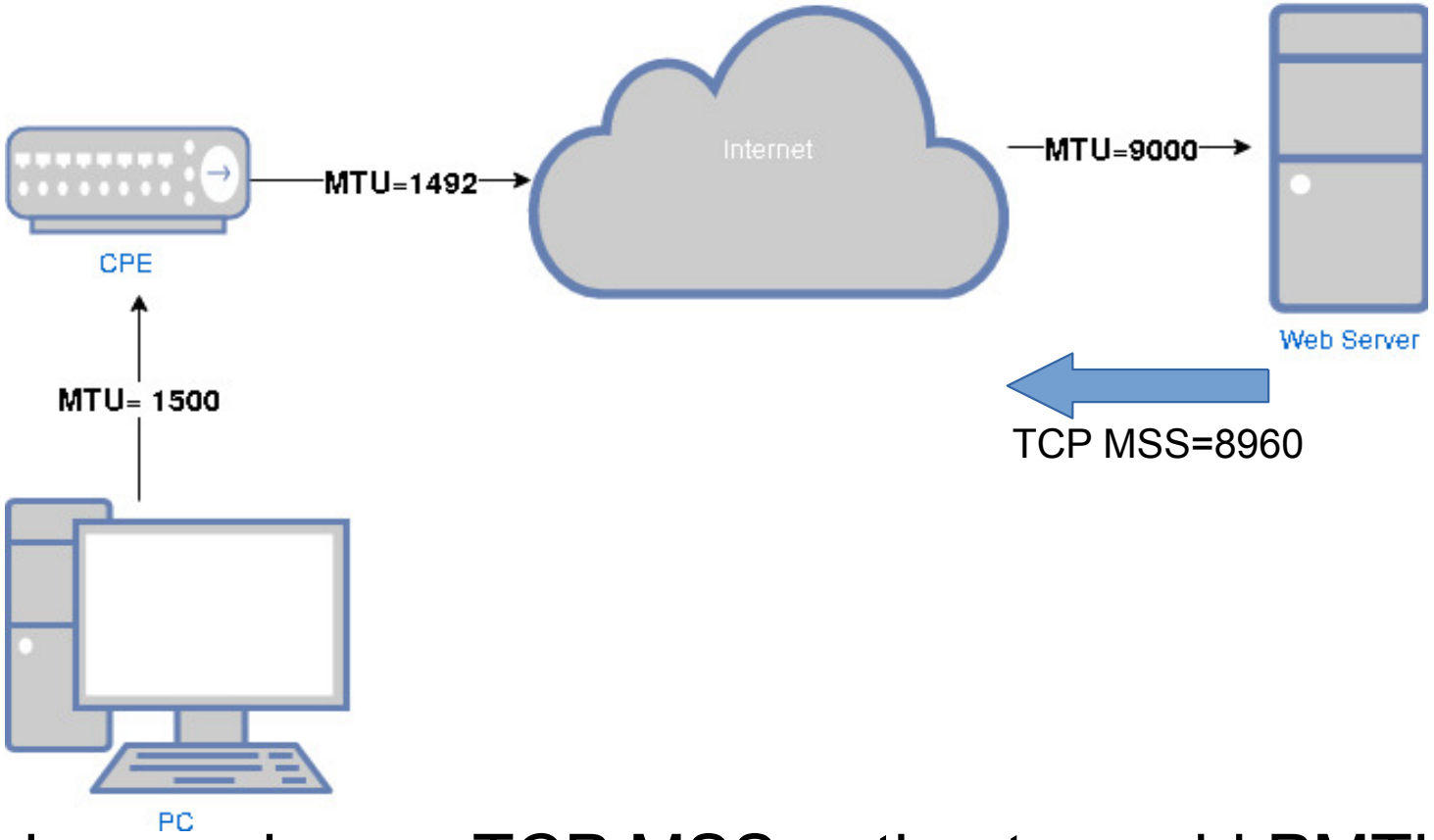
MONROE

UNIVERSITY OF ABERDEEN

mami

TCP Maximum Segment Size

TCP MSS option advertises remote link MTU



Middleboxes change TCP MSS option to avoid PMTUD failures: TCP MSS Clamping

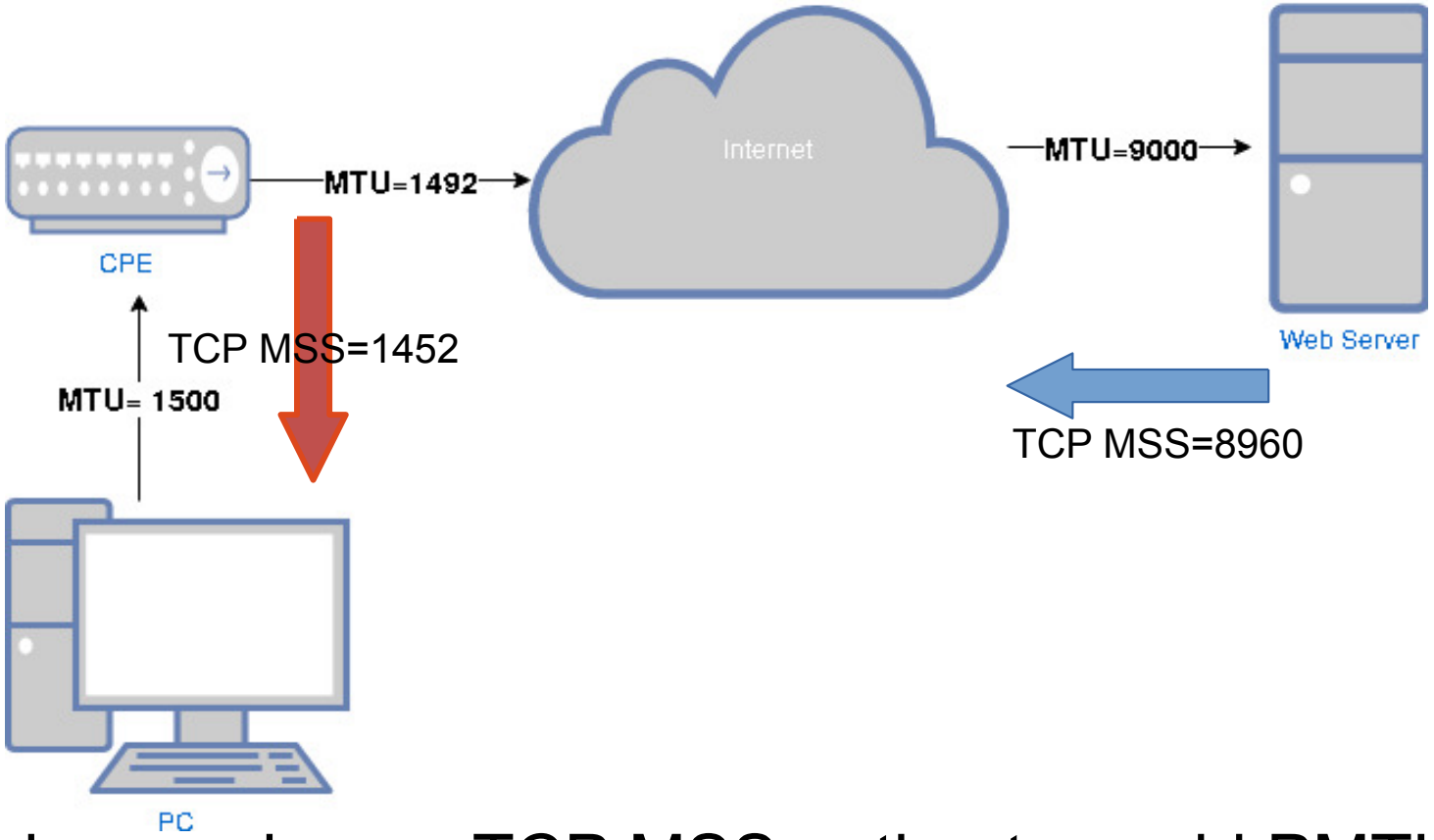
MONROE

UNIVERSITY OF ABERDEEN

mami

TCP Maximum Segment Size

TCP MSS option advertises remote link MTU



Middleboxes change TCP MSS option to avoid PMTUD failures: TCP MSS Clamping

MONROE

UNIVERSITY OF ABERDEEN

mami

Tools and datasets

Purpose	Tool used	Dataset name
Collect server advertised MSS	PATHspider	A.1 “PATHspider”
Validate server advertised MSS	Ping	A.2 “Ping”

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv4

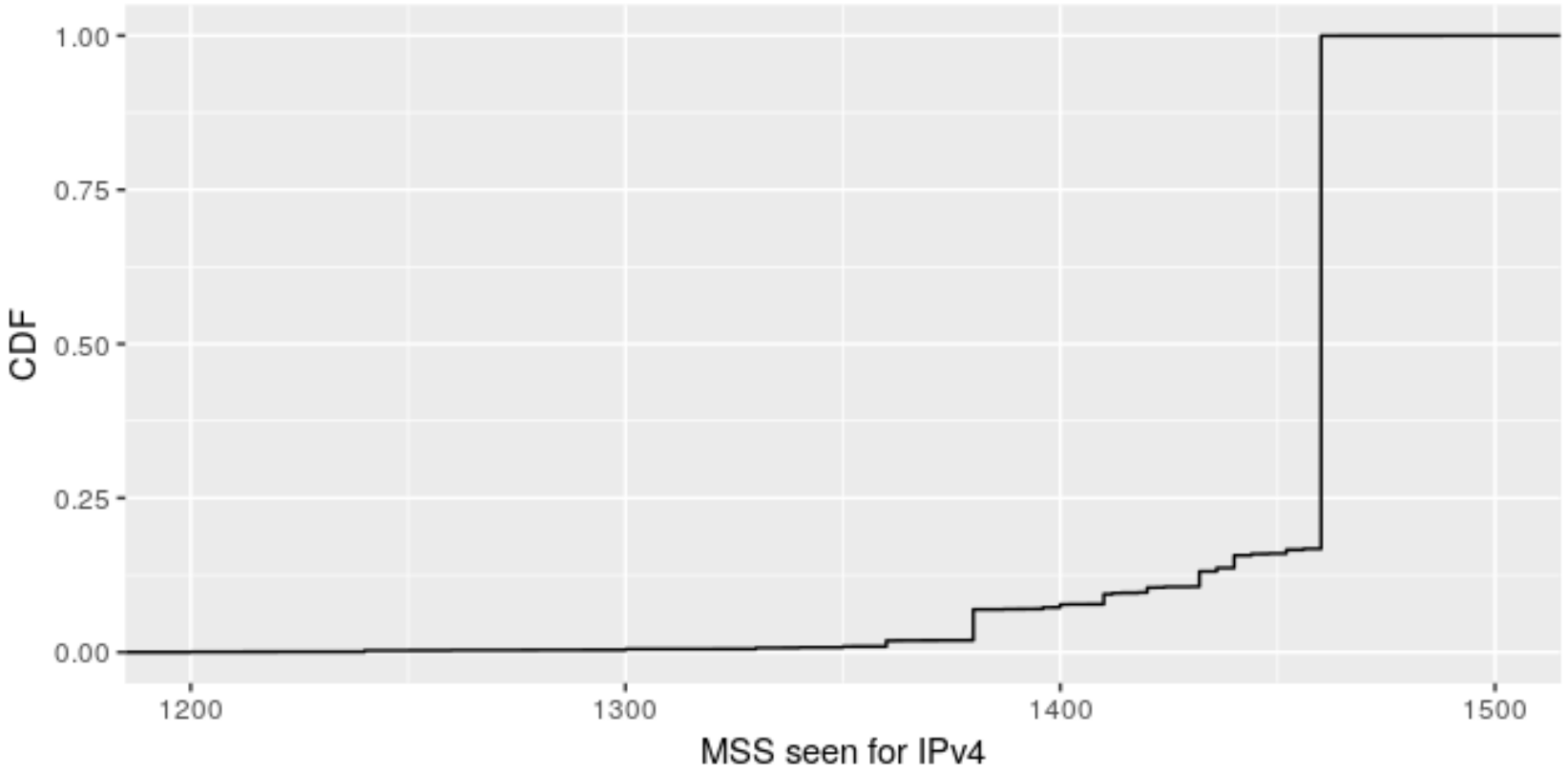


Figure: Avertised MSS (in bytes) on TCP SYN/ACK server response seen at Janet academic network

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv4

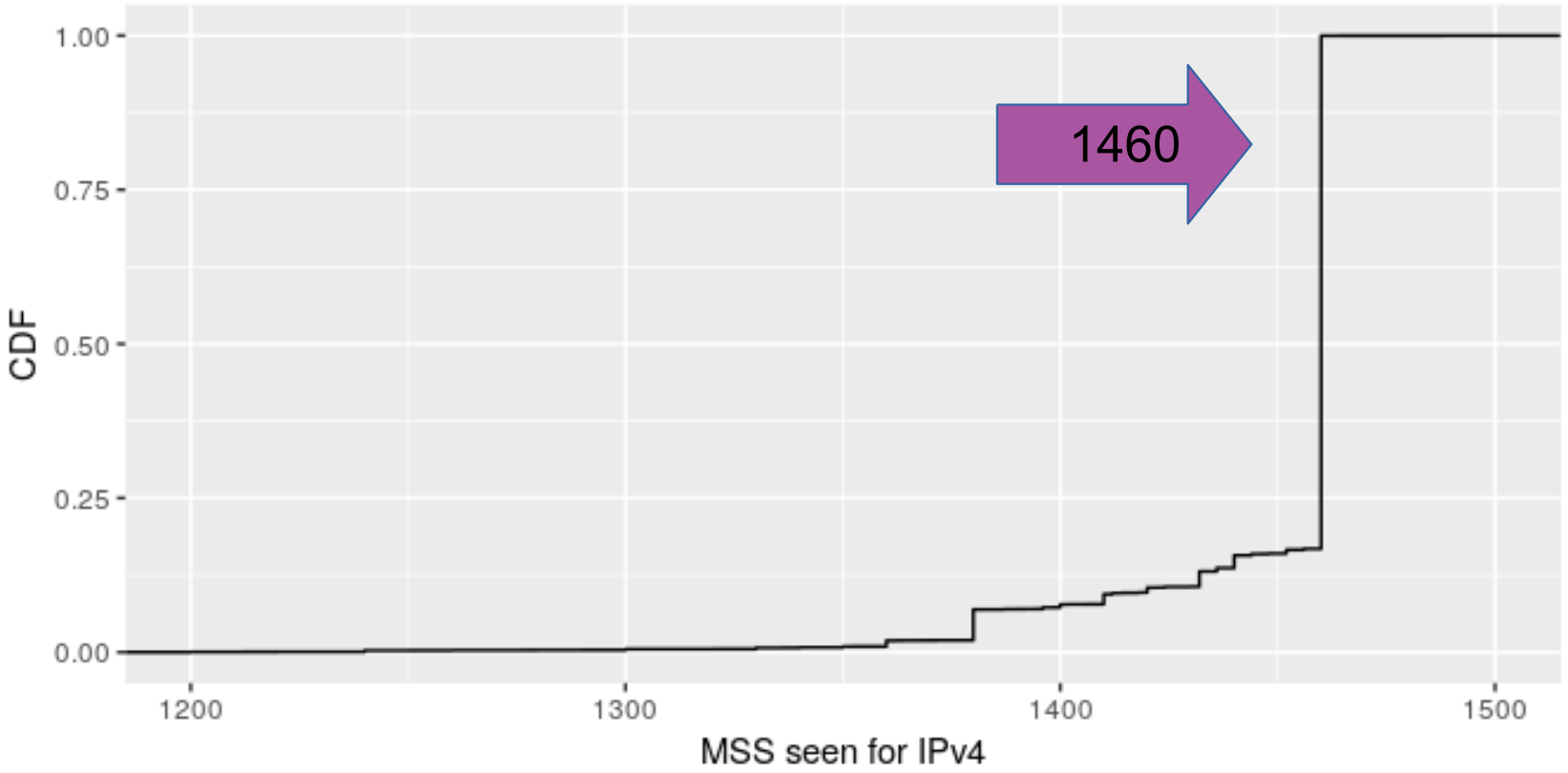


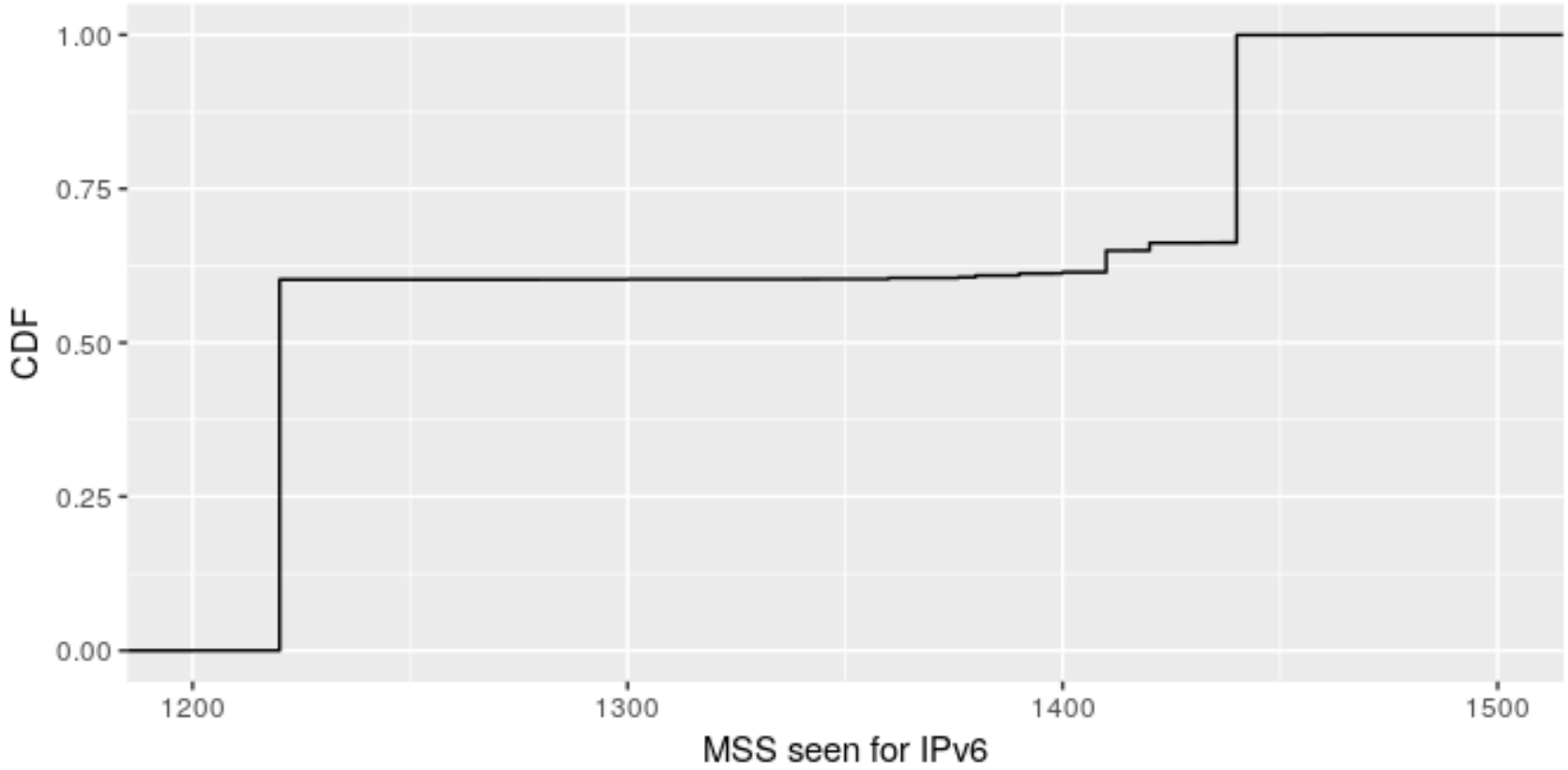
Figure: Avertised MSS (in bytes) on TCP SYN/ACK server response seen at Janet academic network

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv6



Advertised TCP MSS (B) on TCP SYN/ACK server response seen at Janet network

draft-v6ops-pmtud-ecmp-problem-00

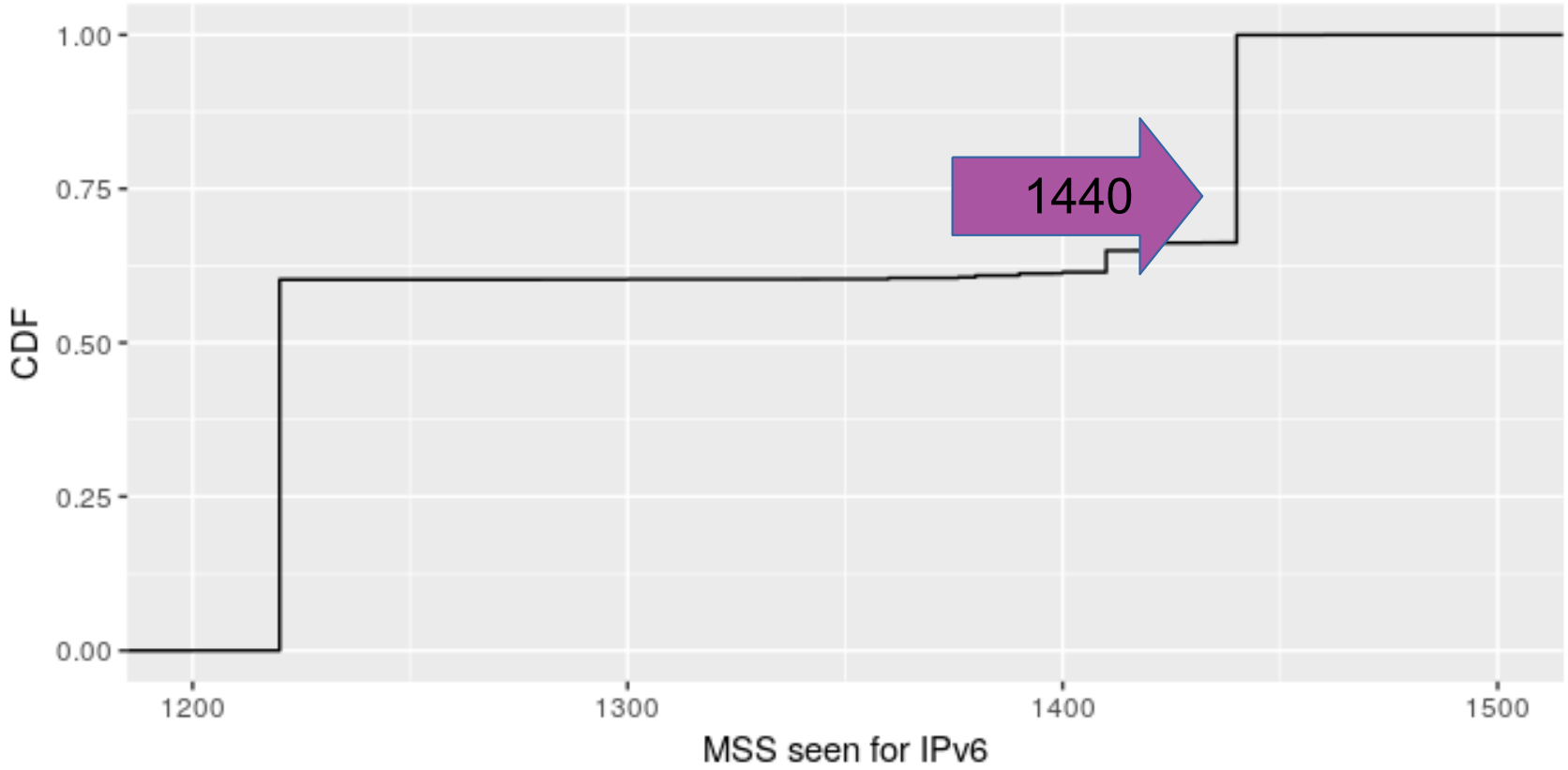
(1) See https://blog.cloudflare.com/path-mtu-discovery-in-practice/

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv6



Advertised TCP MSS (B) on TCP SYN/ACK server response seen at Janet network

draft-v6ops-pmtud-ecmp-problem-00

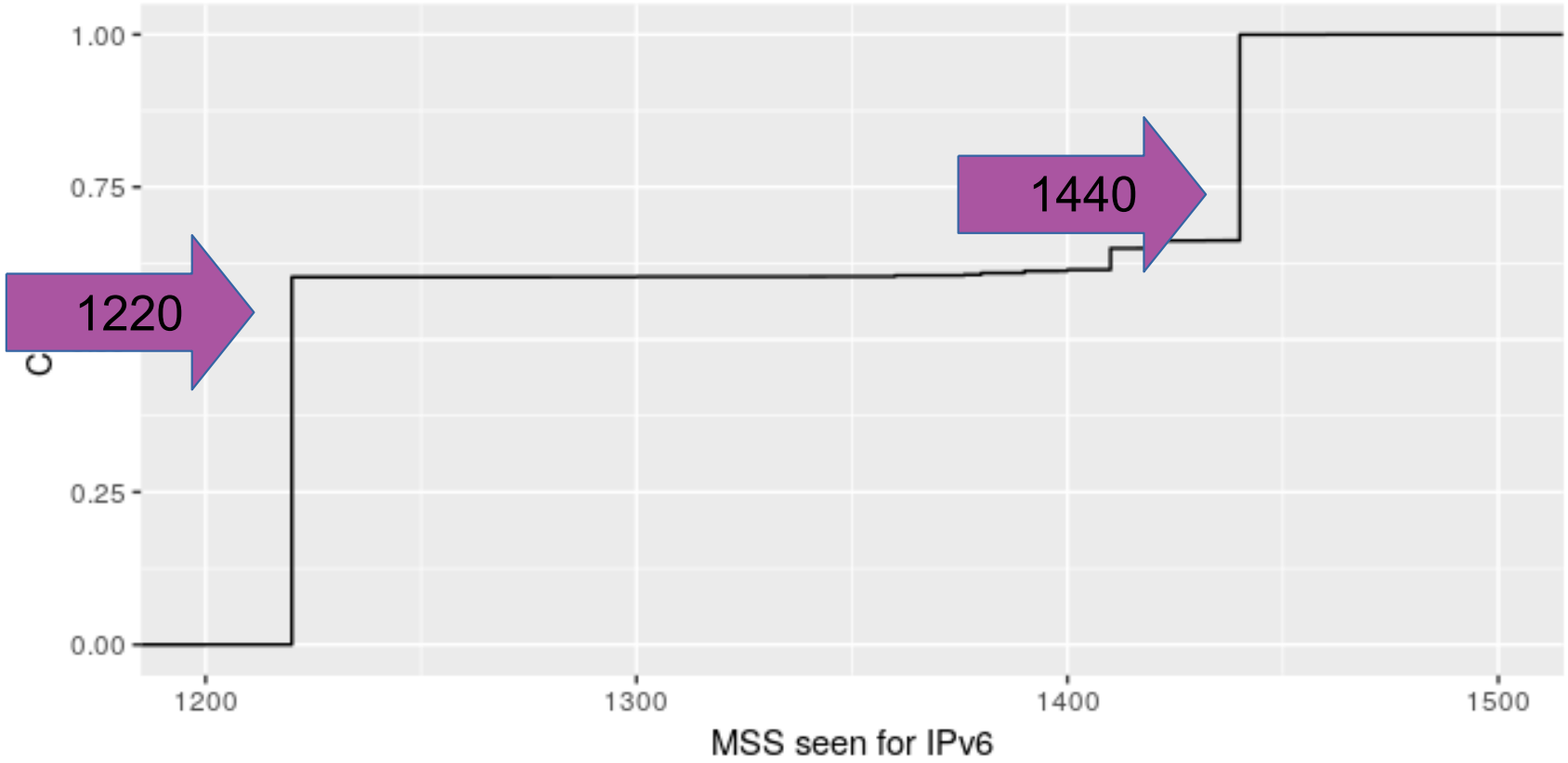
(1) See https://blog.cloudflare.com/path-mtu-discovery-in-practice/

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv6



Advertised TCP MSS (B) on TCP SYN/ACK server response seen at Janet network

draft-v6ops-pmtud-ecmp-problem-00

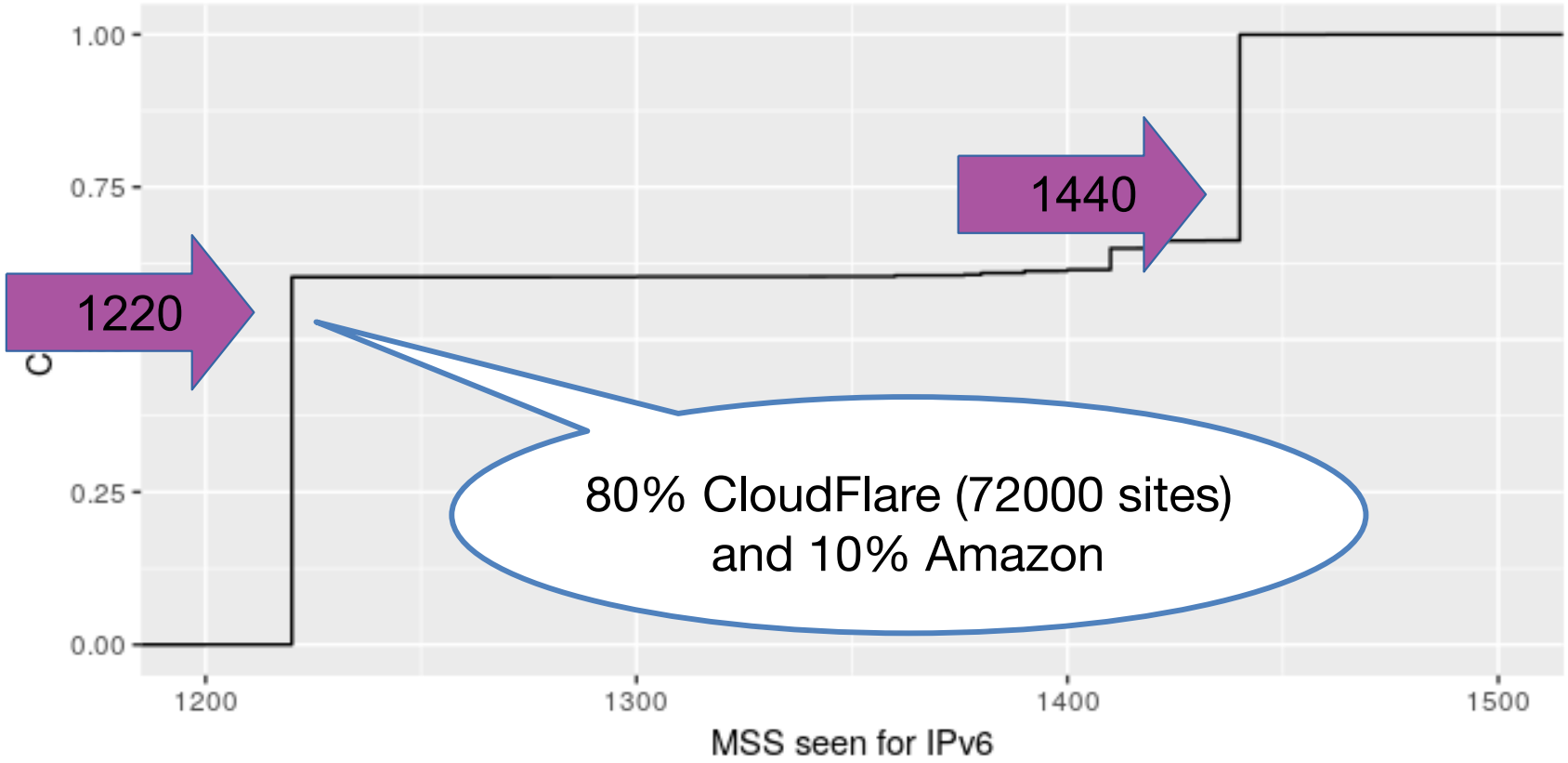
(1) See https://blog.cloudflare.com/path-mtu-discovery-in-practice/

MONROE

UNIVERSITY OF ABERDEEN

mami

Server advertised MSS- “PATHspider”- IPv6



Advertised TCP MSS (B) on TCP SYN/ACK server response seen at Janet network

draft-v6ops-pmtud-ecmp-problem-00

(1) See https://blog.cloudflare.com/path-mtu-discovery-in-practice/

MONROE

UNIVERSITY OF ABERDEEN

mami

Server-advertised MSS and "Ping" results

For 295,000 PATHspider targets:

Sent a probe the size of the advertised TCP MSS

Also sent a 1500B ICMP probe (A.1 "Ping")

Of the subset that advertised MSS < 1460B (34,920),

93% were reached with a 1500B probe.

MONROE

UNIVERSITY OF ABERDEEN

mami

Tools and datasets

Purpose	Tool used	Dataset name
Collect server advertised MSS	PATHspider	A.1 “PATHspider”
Validate server advertised MSS	Ping	A.2 “Ping”
Collect wireless/mobile client advertised MSS	Pathtrace	B.1 “MONROE”
Collect wired edge client MSS	RIPE Atlas Traceroute	B.2 “RIPE”

MONROE

UNIVERSITY OF ABERDEEN

mami

Client advertised MSS – Mobile edge

Dataset B.1, “MONROE”, consists of traceroute-style measurements from the MONROE platform no TCP MSS option

Network	Inserted MSS option
Telenor Norway	1410 bytes
Telia Sweden	1400 bytes
Vodafone Italy	1400 bytes
Wind Italy	1420 bytes

Inserted MSS options by mobile network, n = 10 paths

A total of 888 hops (21%) had an MSS Option

TCP MSS Clamping

MONROE

UNIVERSITY OF ABERDEEN

mami

Client advertised MSS - Wired edge

TCP traceroute from 3000 RIPE Atlas probes towards our server (Dataset B.2, “RIPE”)

4.8% of probes arrive carrying an MSS option, some ***larger*** than allowed by standard Ethernet

764 of the MSS values (23%) in received probes differed from the sent value of 1460 (MSS Clamping)

… Some box in the network is “trying” to help!

MONROE

UNIVERSITY OF ABERDEEN

mami

Tools and datasets

Purpose	Tool used	Dataset name
Collect server advertised MSS	PATHspider	A.1 “PATHspider”
Validate server advertised MSS	Ping	A.2 “Ping”
Collect wireless/mobile client advertised MSS	Pathtrace	B.1 “MONROE”
Collect wired edge client MSS	RIPE Atlas Traceroute	B.2 “RIPE”
Explore server PMTUD	Scamper	C.1 “Scamper”
Explore client PMTUD	Netalyzr Traceroute	C.2 “Netalyzr”
Inspect ICMP quotations	Pathtrace	D “ICMP”

MONROE

UNIVERSITY OF ABERDEEN

mami

Client PMTU - Mobile edge results

Sent a 1500 byte UDP probe to our server with the DF flag set on 10 paths

16 mobile operators tested from over 40 vantage points using the MONROE platform (Dataset C.2 - “Netalyzr”)

Both experiments consistently reported a PMTU of 1500 B

MONROE

UNIVERSITY OF ABERDEEN

mami

PMTU in the Internet - IPv4

60k Cisco Umbrella domains - Dataset C.1, “Scamper”

	1420 MTU	576 MTU	576 Black-hole
PMTU too small	7.45%	3.7%	0.95%
PMTUD success	68.2%	63.9%	8.2%
PMTUD failure	16.4%	19.5%	67.4%
No DF set*	12.5%	12.3%	15.2%
Clear DF	2.7%	4.1%	NIL

68% for IPv4 servers succeed in performing PMTUD

Up to 20% failed for IPv4, twice amount reported in 2010

* ~ 10% (to ~1 %) did not attempt PMTUD (no DF)

MONROE

UNIVERSITY OF ABERDEEN

mami

PMTU in the Internet - IPv6

60k Cisco Umbrella domains - Dataset C.1, “Scamper”

	1280 MTU	1280 Black-hole
PMTUD too small	59.6%	53.1%
PMTUD success	95.5%	32%
PMTUD failure	4.5%	67.9%

95% tested IPv6 succeeded in performing PMTUD

..but 60% of tested web servers did not attempt PMTUD

68% IPv6 and 76% IPv4 webservers failed PMTUD when local messages were blackholed

MONROE

UNIVERSITY OF ABERDEEN

mami

Does PMTUD work?

PMTUD doesn't work reliably

There are real obstacles:

- Unreliable delivery of PTB messages (ECMP, tunnel, filter…)
- PTB info needs to be validated
- A smaller MSS prevents using PMTUD for TCP
  - Many servers now lower their TCP MSS
  - MSS clamping in the network common
- ….

MONROE

UNIVERSITY OF ABERDEEN

mami

How can we make PMTUD work?

We continue to expand our measurement set

Reliable PMTU has to be found at the transport level

- TCP PLPMTUD *could* help, but not enabled/tested
  - (RFC4821, 2007)
- DPLPMTUD being specified for UDP
  - (see TSVWG: draft-ietf-tsvwg-datagram-plpmtud)
- After, perhaps re-think PLPMTUD for TCP ?

MONROE

UNIVERSITY OF ABERDEEN

mami

Find out more at:

Exploring usable Path MTU in the Internet
Ana Custura, Gorry Fairhurst and Iain Learmonth
TMA, Vienna, 2018
http://tma.ifip.org/2018/wp-content/uploads/sites/3/2018/06/tma2018_paper57.pdf

This work is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644399 (MONROE). It used tools supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 688421 (MAMI).