

MTU is Hard

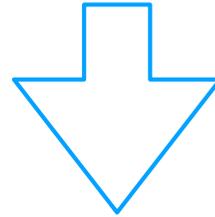
- packets bigger simply don't get there.

Gorry Fairhurst
Bob Hindem
Ana Custura



One thing, so many layers

Application Data



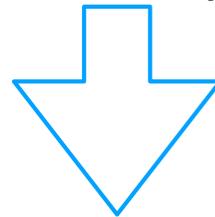
User of Transport: MSS / Max datagram / etc

Flow: PL PTMTU

Transport: PMTU

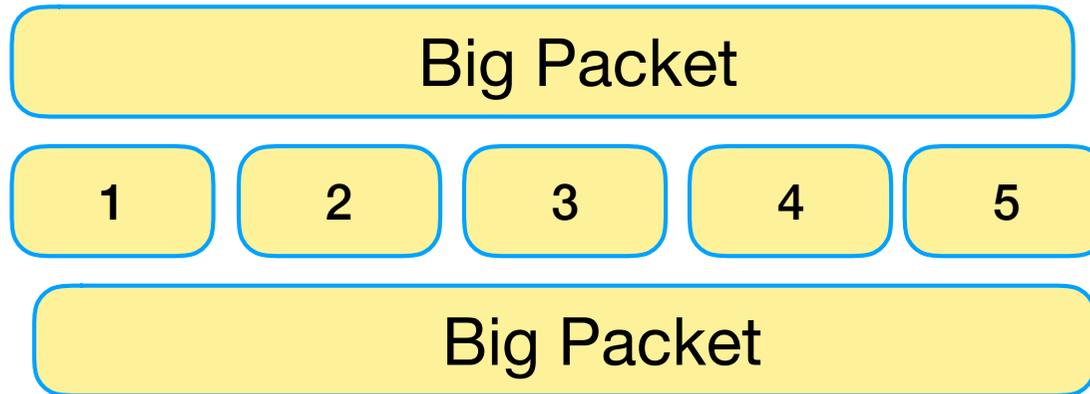
IP-Network: Link-MTU (biggest IPv6 packet in RA)

Network: IP-MTU (e.g., 1500 B)



Interface: Ethernet MTU

Fragments save the day



Many used to think:

Packets start out with one size

Fragmented as needed by routers

Reassembled at endpoint

[RFC791], 1981

Fragmentation was harmful



Why not make the endpoint smarter:

lots of frags

Operators hate fragments

Endpoints hate fragments

Routers (often) drop!

Network Layer: PMTUD



Packets start out with one size

Routers try to tell sender, if packet was too big (ICMP)

Endpoint changes size it allows

Transport adapts size or endpoint fragments

ICMP was not helpful



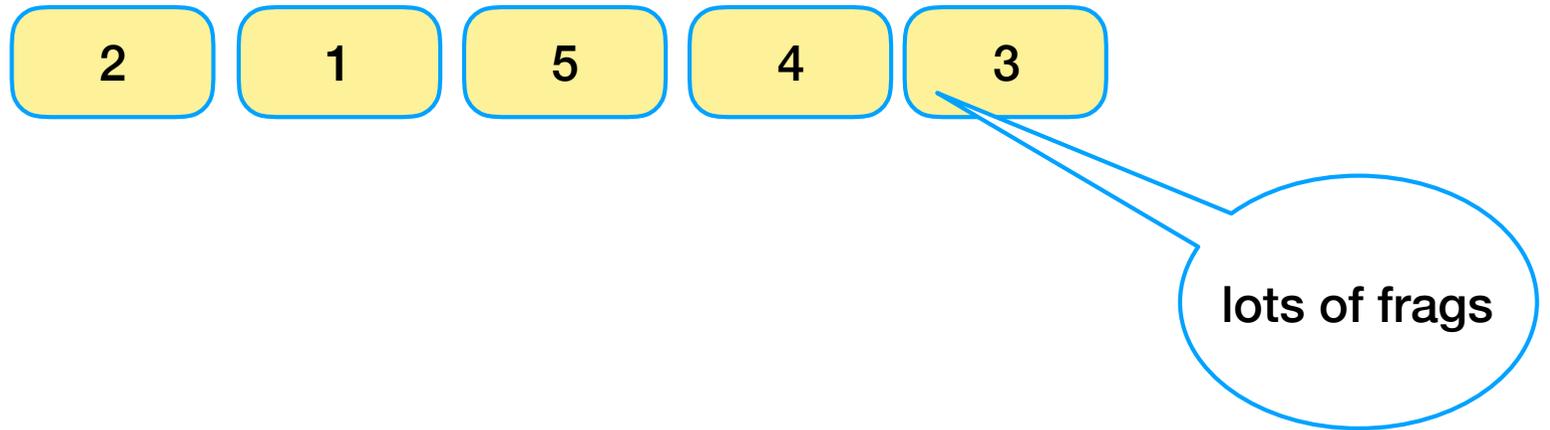
Send **transport** probes with data packets

Verify probe packets were received, PL PMTUD

Sadly tricky to implement in TCP

MSS-Clamping (from path) was deployed instead

Fragmentation was still harmful



Getting this right is hard

ICMP was harmful

PMTUD accepts any ICMP PTB message

Need a trust model to believe ICMP!

Verify sender on path!!!

Signals should be treated as “hints”

e.g. please try a probe of size “n” to see if it works

path change panic... check that “m” still works

... check before you (mis)use!

...will need to ignore spurious hints

... maybe some natural size that works best for the app

PL PMTUD saves the MTU



(D) PL PMTUD has arrived:
SCTP, QUIC, STUN, UDP-Opt, etc

New signals come:

HBH Options

Other interesting things

Measure... Design ... Measure

How much of the Internet supports 1500B (or so)

How much supports more?

Much more?

Can we use HBH packets to find a PMTU hint?

Are there paths that forward Hop-by-Hop packets?

Are there other hints?

PATHspider and Pathtrace

	PATHSpider	Pathtrace
Type	A to B test of IPv6 HBH and Dest Options	Traceroute experiments using IPv6 HBH and Dest Opts
Protocol	DNS/UDP/IPv6 (with extension header)	ICMP/IPv6 (with extension header)
Test	Send a DNS query and tests for valid responses	Records ICMP replies from routers along the path
Implementation	Raw Sockets	Packet forging with Scapy
Results format	JSON	PCAP

Early data - prior to running the real experiments - PATHSpider

	Aberdeen	Slough	California
Forwards HBH Opts	18.9%	18.9%	0%
Forwards Destination Opts	24.7%	25.8%	24.1%
None Forwarded	65.2%	63.9%	75.9%
Both forwarded	8.8%	8.6%	0%
Total servers	3266	3281	3280

???

Results - PATHSpider IPv6 HBH

A quarter of the tested servers replied with a valid DNS response for Destination options

1/8 for HBH options

Loss could be at destination or along the path - need to investigate failures with pathtrace

Deploy *more* probes in *more* locations :-)



MTU is Hard

- It's taken a long time to get here.

It's not a network problem

- unless you're talking about tunnels!

It's a transport problem:

The only thing you can TRUST is your own probes

Lots of helpful hints you can also use

Now we can innovate again!