

Multi-MTU subnets

draft-van-beijnum-multi-mtu-02

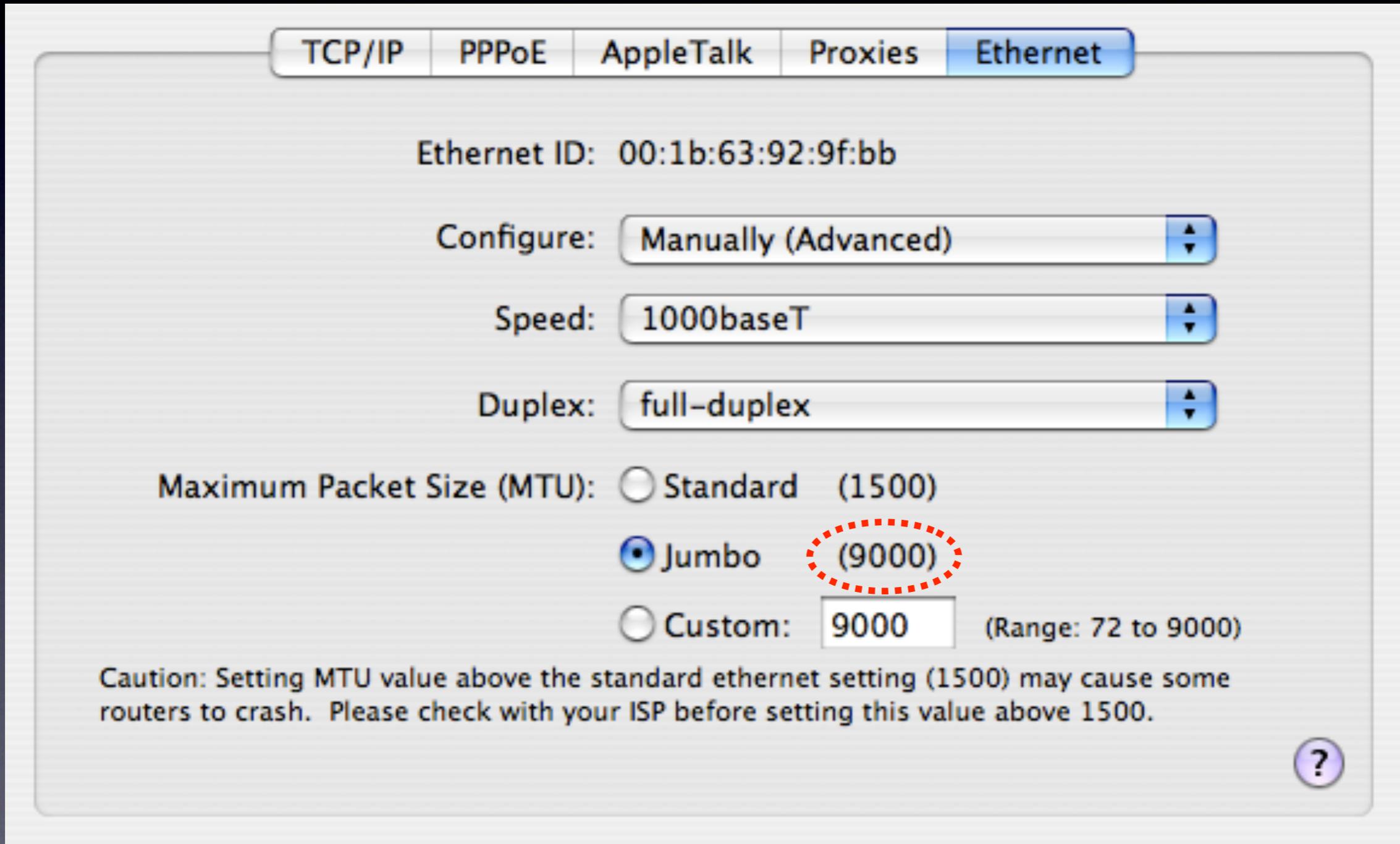
int-area @ IETF-71, March 11, 2008
Φιλαδέλφεια, PA, US

Ijtsch van Beijnum

Ethernet MTU = 1500

- IEEE 802 values interoperation
 - you can connect 10 Mbps and 10000 Mbps ethernets and it just works
- But: packets per second getting out of hand
 - 10 Mbps: 813 pkts/sec
 - 10 Gbps: 812744 pkts/sec

But...



Big Packet Advantages

- More room for additional headers without path MTU discovery breakage
- Lower overhead, especially with large headers
- Less per packet work in hosts = faster
- Less per packet work in routers = possible power/heat savings
- Better TCP performance

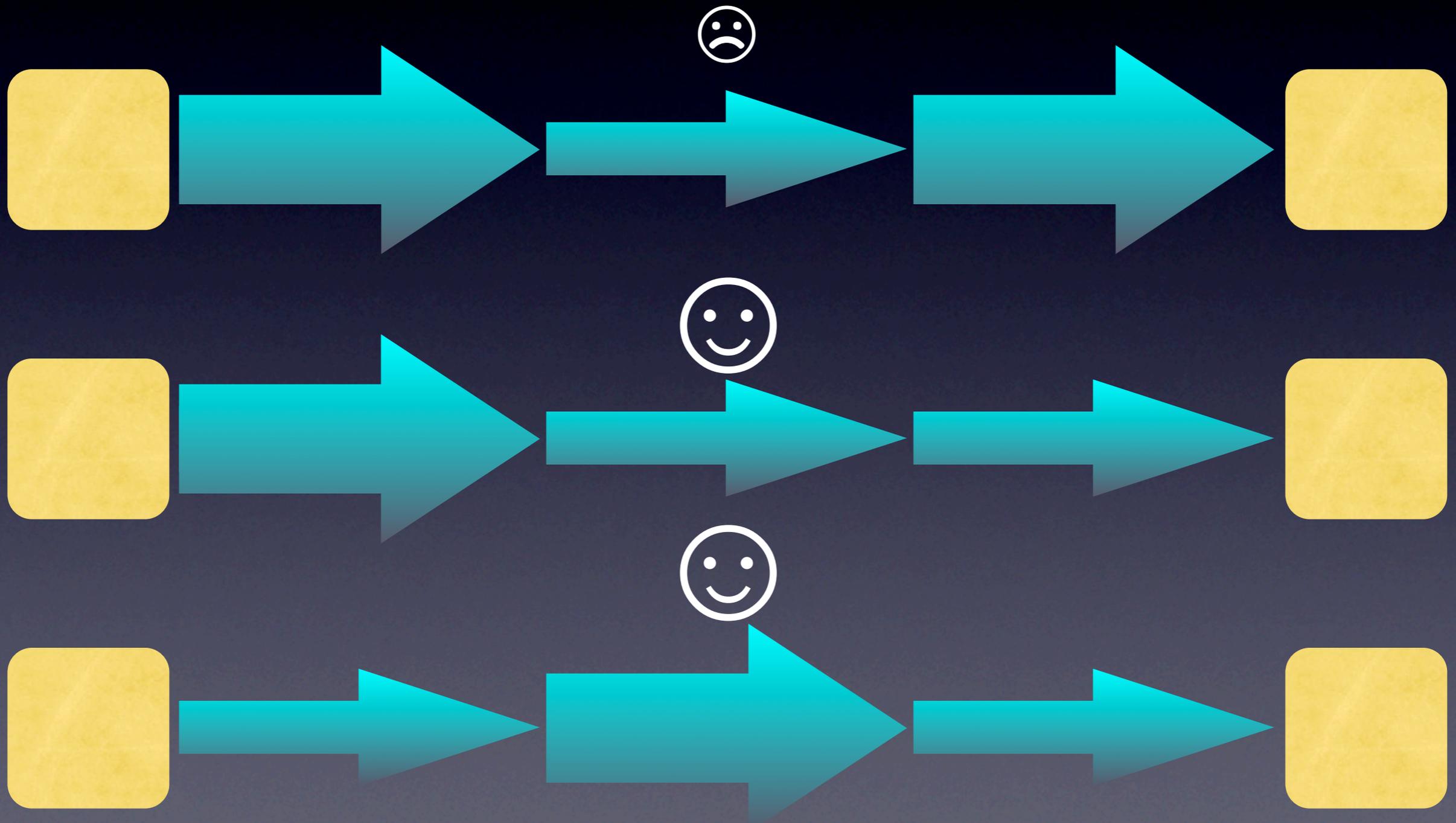
Jumboframes

- Lots of gigabit ethernet equipment supports larger packets: "jumboframes"
- Common value: ± 9000 bytes
 - but no standard non-standard size
- "mini jumbos" of upto ± 2000 bytes common in lower-speed switches

Disadvantages (I)

- More delay and jitter
 - so only do 1500+ at 1000 Mbps or faster
- Depend more on path MTU discovery.
However:
 - see the problem if **you** break PMTUD
 - can always reduce MTU (not increase...)
 - few problems with large MTU in middle

PMTUD problems



Disadvantages (2)

- More packet loss from bit errors
 - ideal pkt size = $\sqrt{(\text{overhead bytes} / \text{BER})}$
- More undetected bit errors (?)
 - naive: more errors/packet, but fewer packets = no difference
 - complex: hamming distance makes CRC32 much stronger than expected
 - use stronger FCS for jumboframes?

What we need

- Ability to turn on jumbos without touching all hosts on a subnet
- Take advantage of hardware improvements without protocol work
 - no more hardcoding of MTU sizes
- Be backward compatible!
 - also with current jumbo deployments

Three mechanisms

1. Just make administratively setting a jumboframe size on a subnet easier
2. Use bigger packets and depend on RFC 4821 path MTU discovery
3. Neighbor discovery options and jumbo ARP for communicating and testing per-neighbor MTUs

Admin settings (req)

- Configure router to send RA option with:
 - MAXMTU: no packets bigger than this
 - SAFEMTU: upto this size without probing
 - (compatible with current jumbo frame deployments)
 - SLOWMTU: upto this size if < 600 Mbps

RFC 4821 (opt)

- "Packetization Layer Path MTU Discovery"
- If a transport protocol does RFC 4821
 - simply send large packets
 - figure out what works and what doesn't
- But no larger than MAXMTU for QoS reasons etc

Explicit probing (opt)

- ND MTU option
 - tell neighbor what you can receive
- ND padding option
 - pad NS packet to jumbo size, see if anything comes back
- Jumbo ARP
 - simply pad ARP packet, watch for reply

To do

- Publish as experimental?
- But first:
 - consider DHCP, esp. for IPv4
 - talk about use cases / applicability
- Questions?