

Onze pakketten zijn te klein!

Amsterdam, 9 jan 2014

Iljitsch van Beijnum

Onze pakketten zijn te klein!

Amsterdam, 9 jan 2014

Iljitsch van Beijnum

Our packets are too small!

Onze pakketten zijn te klein!

Amsterdam, 9 jan 2014

Iljitsch van Beijnum

With subtitles.

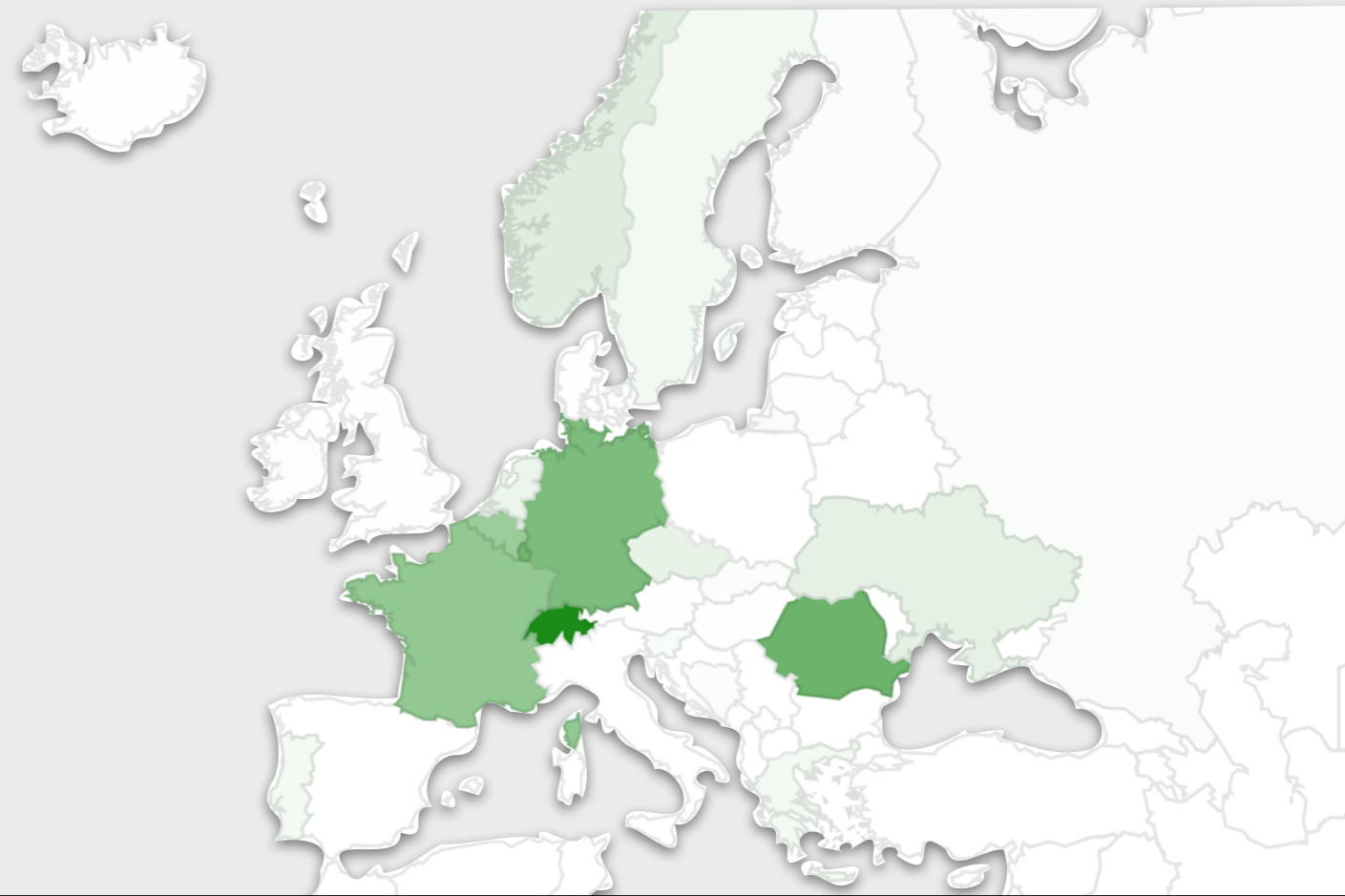
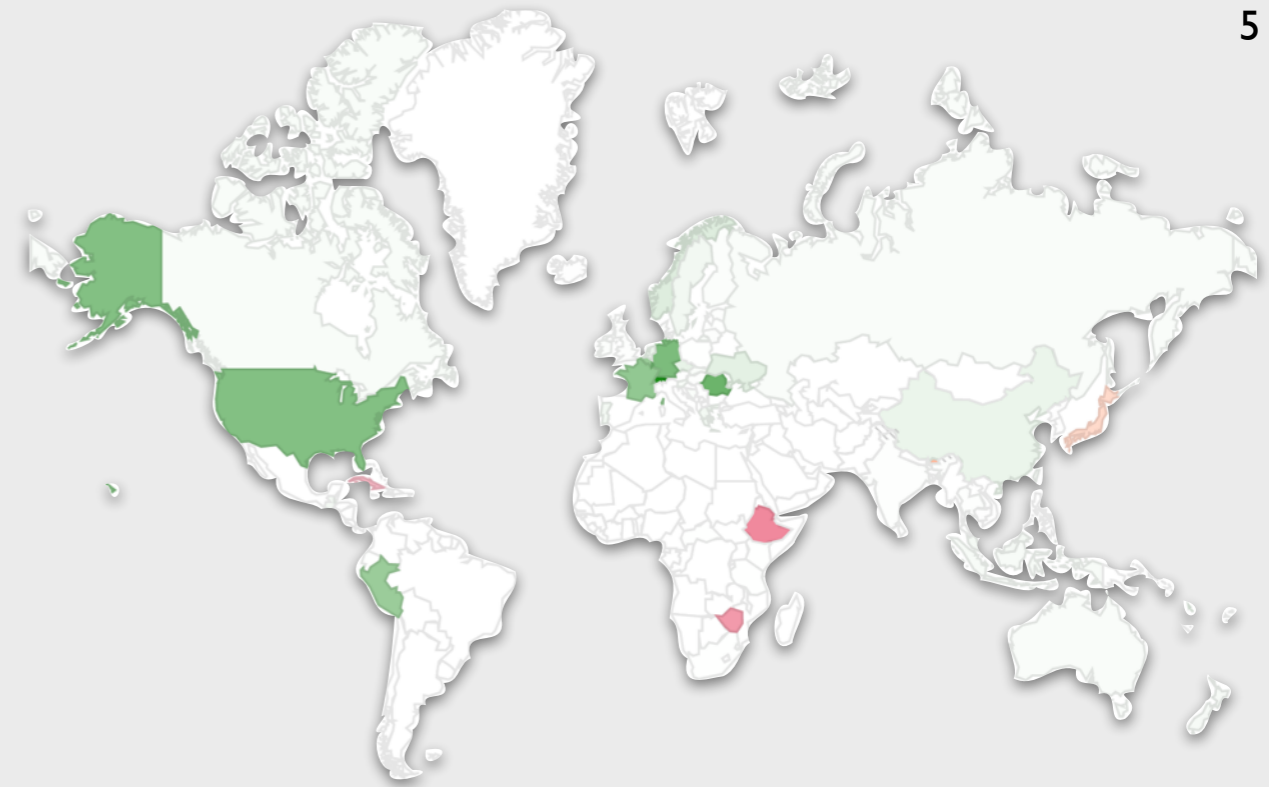
Maar eerst even kort IPv6

- In 2013:
 - Alexa top 500: 23% → 23%
 - AMS-IX: ± 4 Gbps → $\pm 8,5$ Gbps (= 0,5%)
 - Google: 1% → 2,6% (2,6 x)
 - Akamai: jun '13: 60k/sec → jan '14 300k/sec

But first a quick IPv6 update.

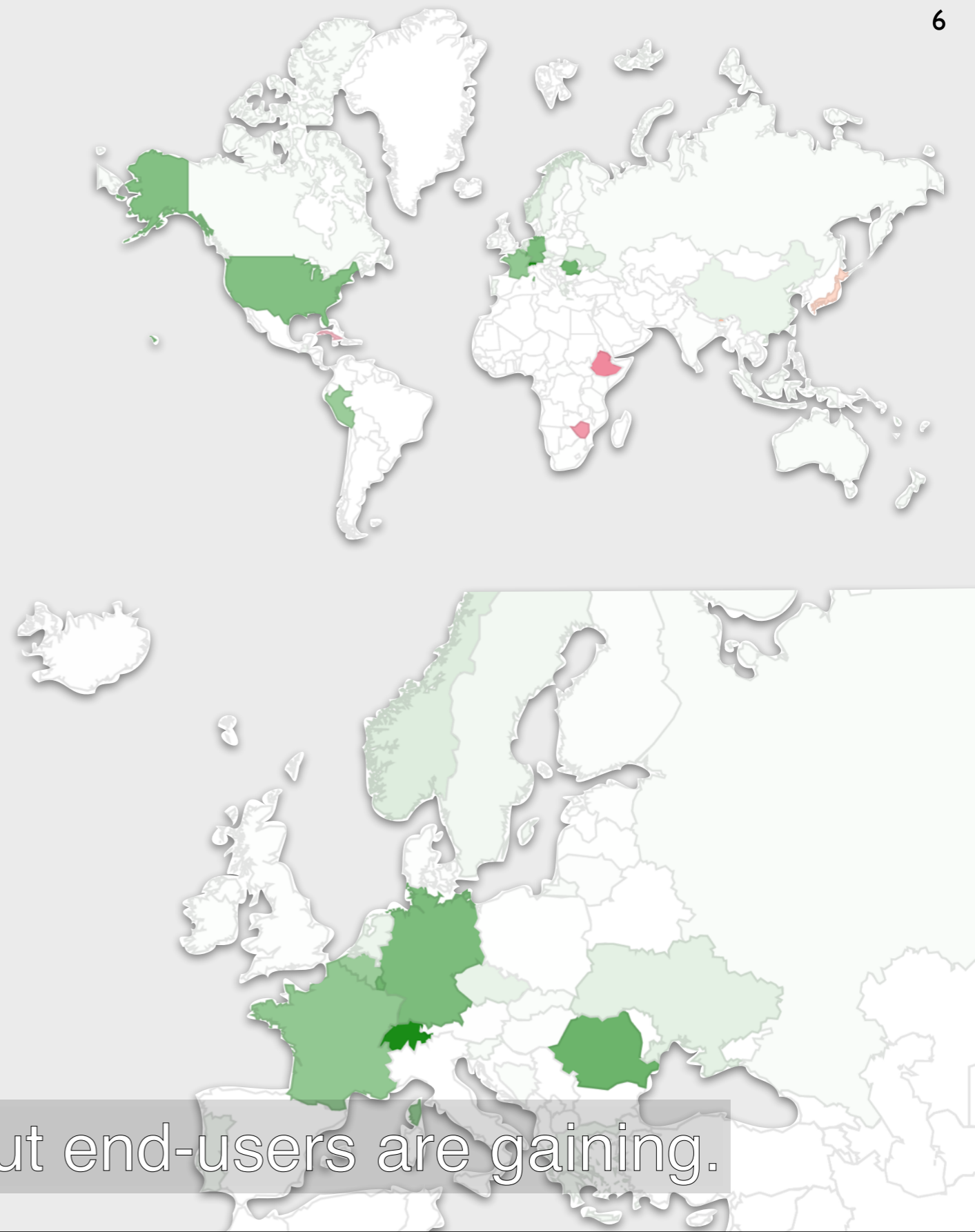
Conclusie

- Web: IPv6 stagneert
- Eindgebruikers: IPv6 in opkomst
- (één klein landje houdt moedig stand...)



Conclusie

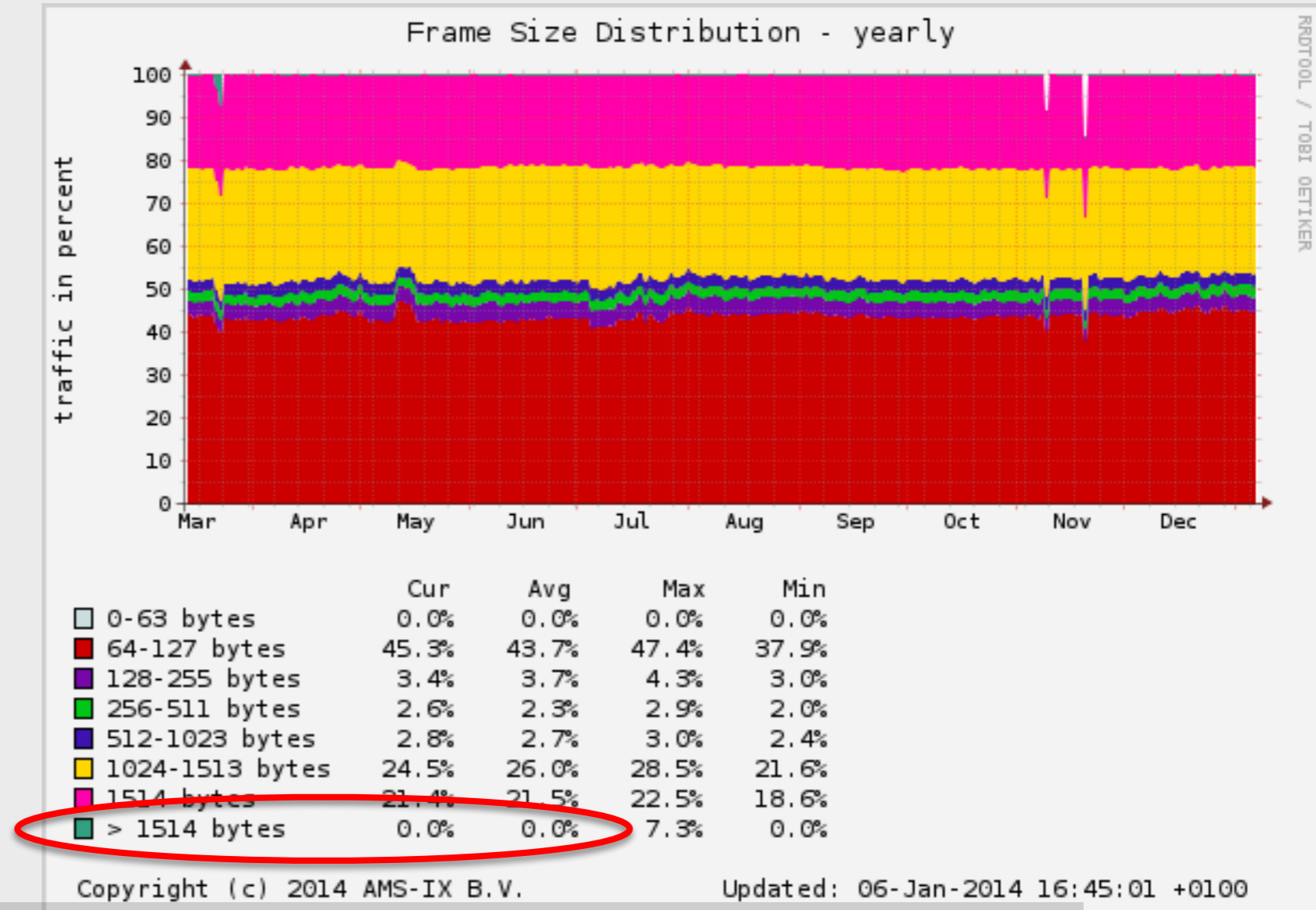
- Web: IPv6 stagneert
- Eindgebruikers: IPv6 in opkomst
- (één klein landje houdt moedig stand...)



Web is stagnating but end-users are gaining.

Maar...

- IPv6 of IPv4:
- het blijven veel te kleine pakketten!



But... IPv6 or IPv4: the packets are too small!

Waarom maar 1500 bytes?

- De originele Ethernet-standaard specificeert een MTU van 1500 bytes
- MTU = Maximum Transfer Unit
 - maximale grootte van een IP-pakket
 - (het Ethernetpakket is dan 1514 / 1518 bytes)
- Ofwel: ± 800 pakketten per seconde (PPS)

Original Ethernet standard specifies 1500 bytes.

Maar dat was 30 jaar geleden!

~ 1980	10 Mbps	Ethernet	800 PPS
~ 1995	100 Mbps	Fast Ethernet	8000 PPS
~ 1998	1000 Mbps	Gigabit Ethernet	80000 PPS
~ 2002	10000 Mbps	10 Gigabit Ethernet	800000 PPS
~ 2010	100000 Mbps	100 Gigabit Ethernet	8 MPPS

But that was 30 years ago!

Compatibiliteit

- Fast Ethernet moest interoperabel zijn met Ethernet = 1500 bytes
- Gigabit Ethernet moest interoperabel zijn met Fast Ethernet = 1500 bytes
 - (hoewel bijna alle GE hardware "jumboframes" aankan)
- Zelfde voor 10 en 100 Gigabit Ethernet

Newer Ethernet devices need to talk to older ones.

Compatibiliteit

- Fast Ethernet moest interoperabel zijn met Ethernet = 1500 bytes
- Gigabit Ethernet moest interoperabel zijn met Fast Ethernet = 1500 bytes
 - (hoewel bijna alle GE hardware "jumboframes" aankan)
- Zelfde voor 10 en 100 Gigabit Ethernet

Hence it was never possible to increase the packet size.

Compatibiliteit

- Fast Ethernet moest interoperabel zijn met Ethernet = 1500 bytes
- Gigabit Ethernet moest interoperabel zijn met Fast Ethernet = 1500 bytes
 - (hoewel bijna alle GE hardware "jumboframes" aankan)
- Zelfde voor 10 en 100 Gigabit Ethernet

Even though most GE hardware can do "jumboframes".

Het probleem

- Hoeveelheid werk is bijna hetzelfde ongeacht MTU
- Dus kleinere pakketten = meer CPU-gebruik
 - (of, bij switches en routers: snellere ASIC)
- Dus: lagere performance en/of hoger energieverbruik!



Small packets means more work means more energy.

Wat doen we eraan?

- Nieuwe pakketgrootte afspreken?
 - die is over 10 jaar weer te klein...
- In plaats daarvan: **flexibiliteit!**
 - iedereen heeft z'n eigen MTU
 - geef jouw MTU door aan je burens
 - zij sturen pakketten van de juiste grootte

New packet size is only a short term solution.

Wat doen we eraan?

- Nieuwe pakketgrootte afspreken?
 - die is over 10 jaar weer te klein...
- In plaats daarvan: **flexibiliteit!**
 - iedereen heeft z'n eigen MTU
 - geef jouw MTU door aan je burens
 - zij sturen pakketten van de juiste grootte

Instead: negotiate MTU size with neighboring systems.

Maar... IEEE kan dit niet

- Bij Ethernet staat ieder pakket op zich
 - je weet dus niks van de ontvanger
- Maar IP kan dit wel:
 - eerst ARP of Neighbor Discovery voordat er data uitgewisseld wordt
 - dus: stop MTU in ARP of ND optie

With Ethernet receiver capabilities are unknown.

Maar... IEEE kan dit niet

- Bij Ethernet staat ieder pakket op zich
 - je weet dus niks van de ontvanger
- Maar IP kan dit wel:
 - eerst ARP of Neighbor Discovery voordat er data uitgewisseld wordt
 - dus: stop MTU in ARP of ND optie

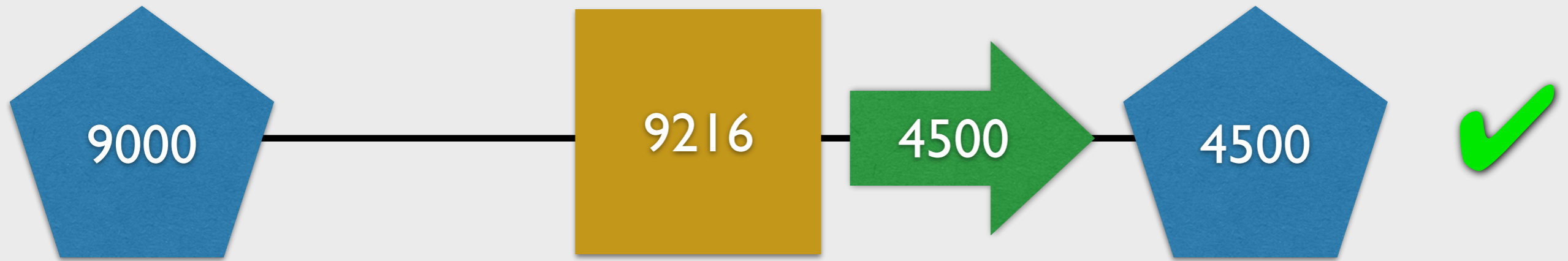
But not with IP: ARP or Neighbor Discovery first, then data.

Maar... IEEE kan dit niet

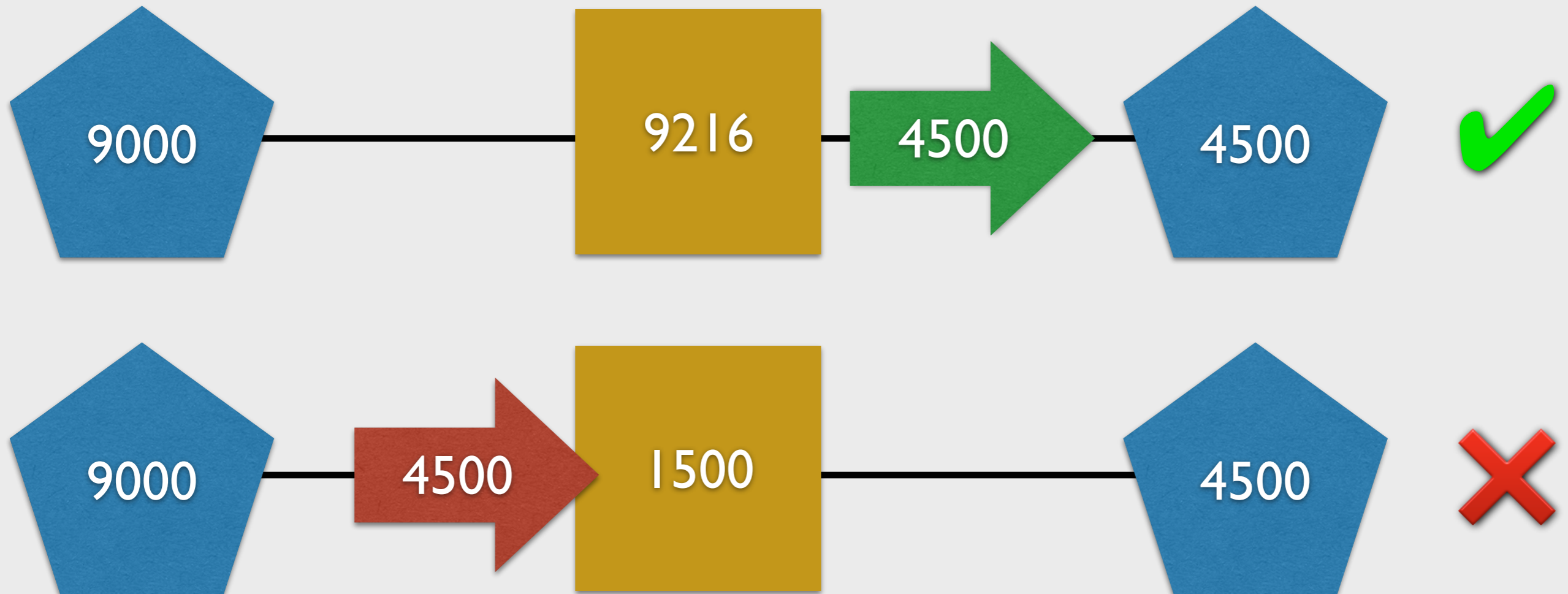
- Bij Ethernet staat ieder pakket op zich
 - je weet dus niks van de ontvanger
- Maar IP kan dit wel:
 - eerst ARP of Neighbor Discovery voordat er data uitgewisseld wordt
 - dus: stop MTU in ARP of ND optie

So put MTU in ARP or ND option.

Complicaties...



Complicaties...



Test packets to see if switches can handle big packets.

Vragen?

- Dat was mijn verhaal!
- Vragen?
- Zie ook:
- <http://tools.ietf.org/html/draft-van-beijnum-multi-mtu>
- Of mail me: iljitsch@muada.com

That's it! Any questions?