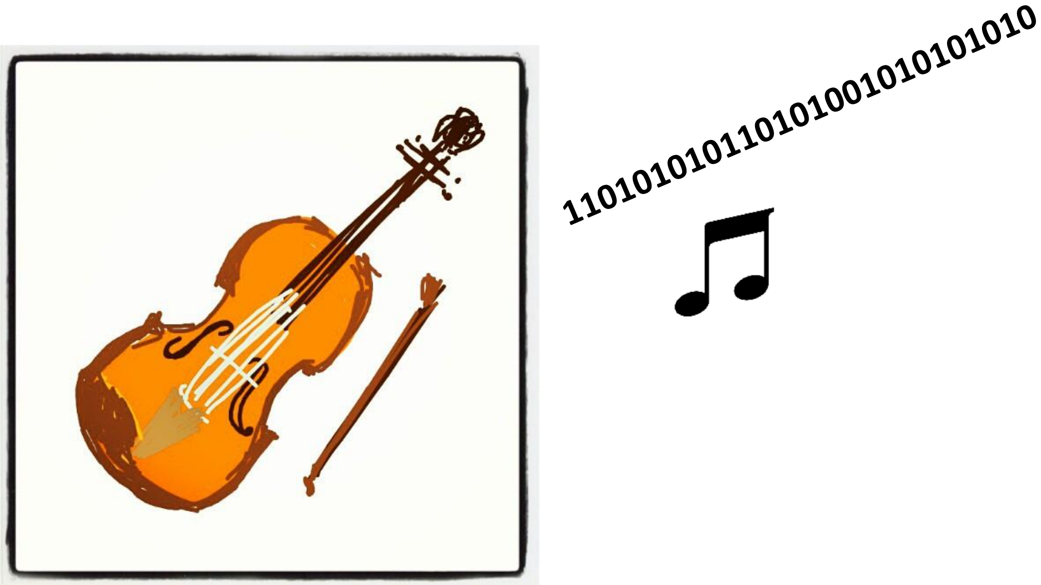
Who’s fiddling with my bits?

Gareth Tyson

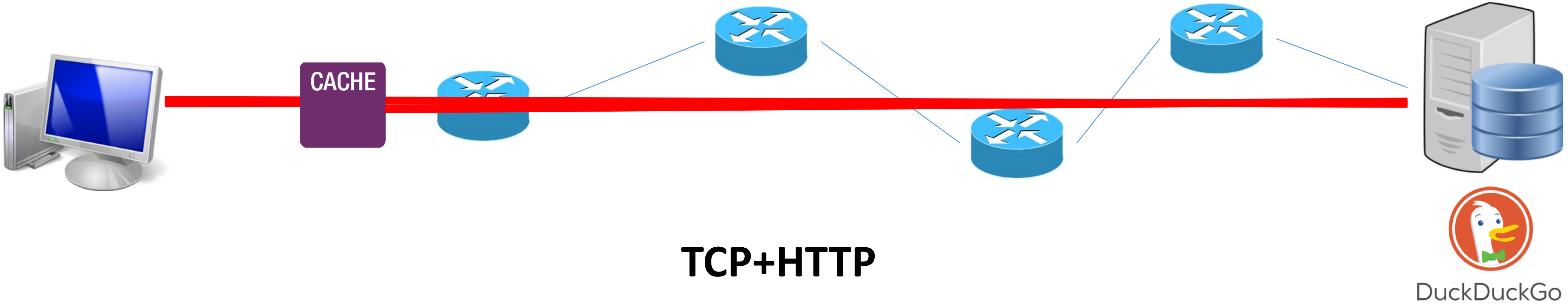


Exploring HTTP Header Manipulation in the Wild.

G. Tyson, S Huang, F. Cuadrado, I. Castro, V. Perta, A. Sathiaseelan and S. Uhlig.

In Proc. 26th World Wide Web Conference (WWW), Perth, Australia (2017)

Let’s remind ourselves how HTTP web requests work…



TCP+HTTP

Do middleboxes tell us anything Juicy ?

Do middleboxes tell us anything Juicy ?

What can middleboxes in a network tell us about the region where it is based?

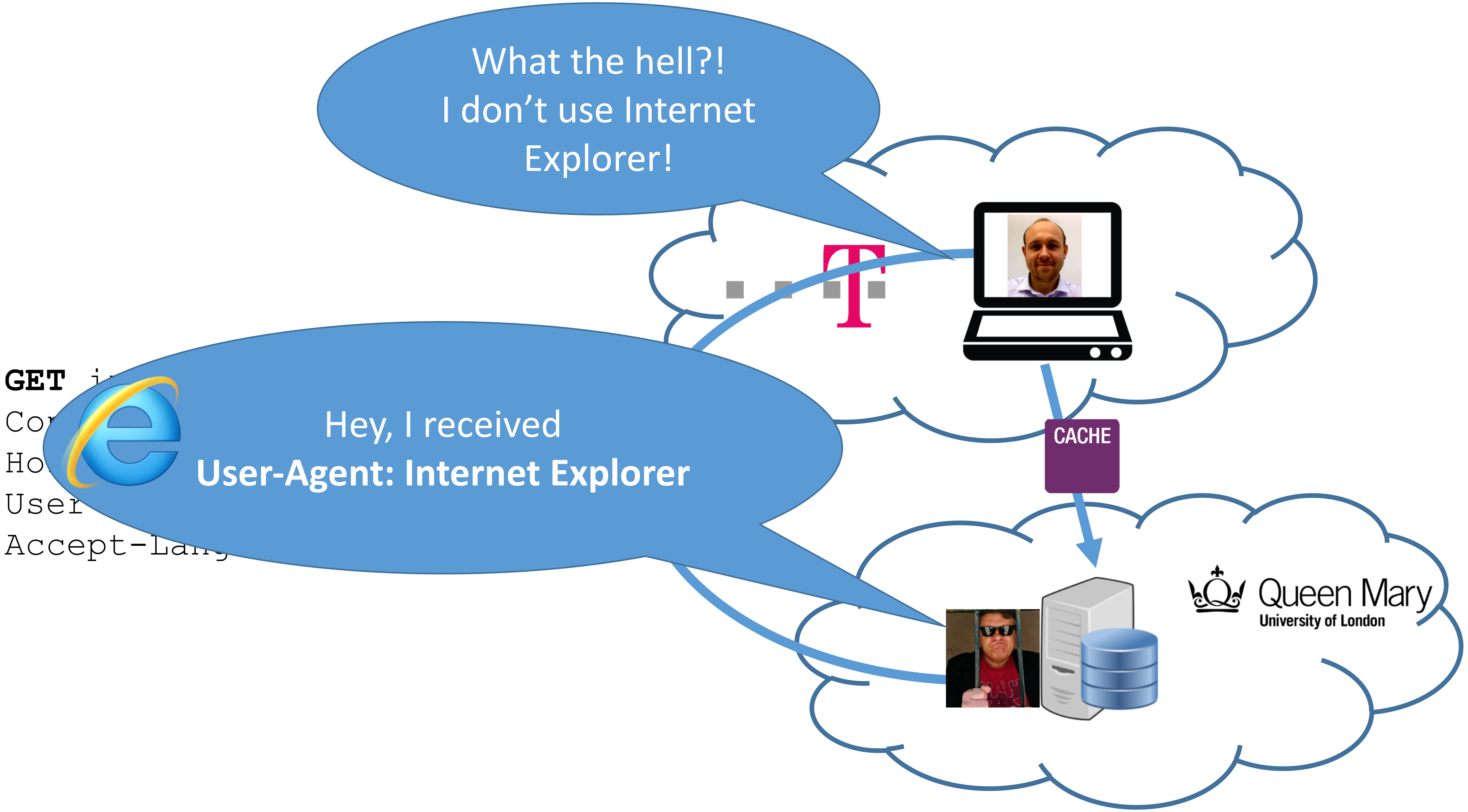
What could we learn?

- Perhaps we see a web cache?
  - The network has expensive transit? Slow downlink?
- Perhaps we see a web firewall?
  - Has security concerns? Has censorship?
- Perhaps we see a WiFi authenticator?
  - Certainly has a WiFi network!

Exploratory data collection methodology

1) Launch HTTP requests from around the world
2) Monitor/detect when HTTP messages are intercepted by middleboxes
3) Split samples into different countries
4) Explore the differences

So, we need to recruit users from around the world…

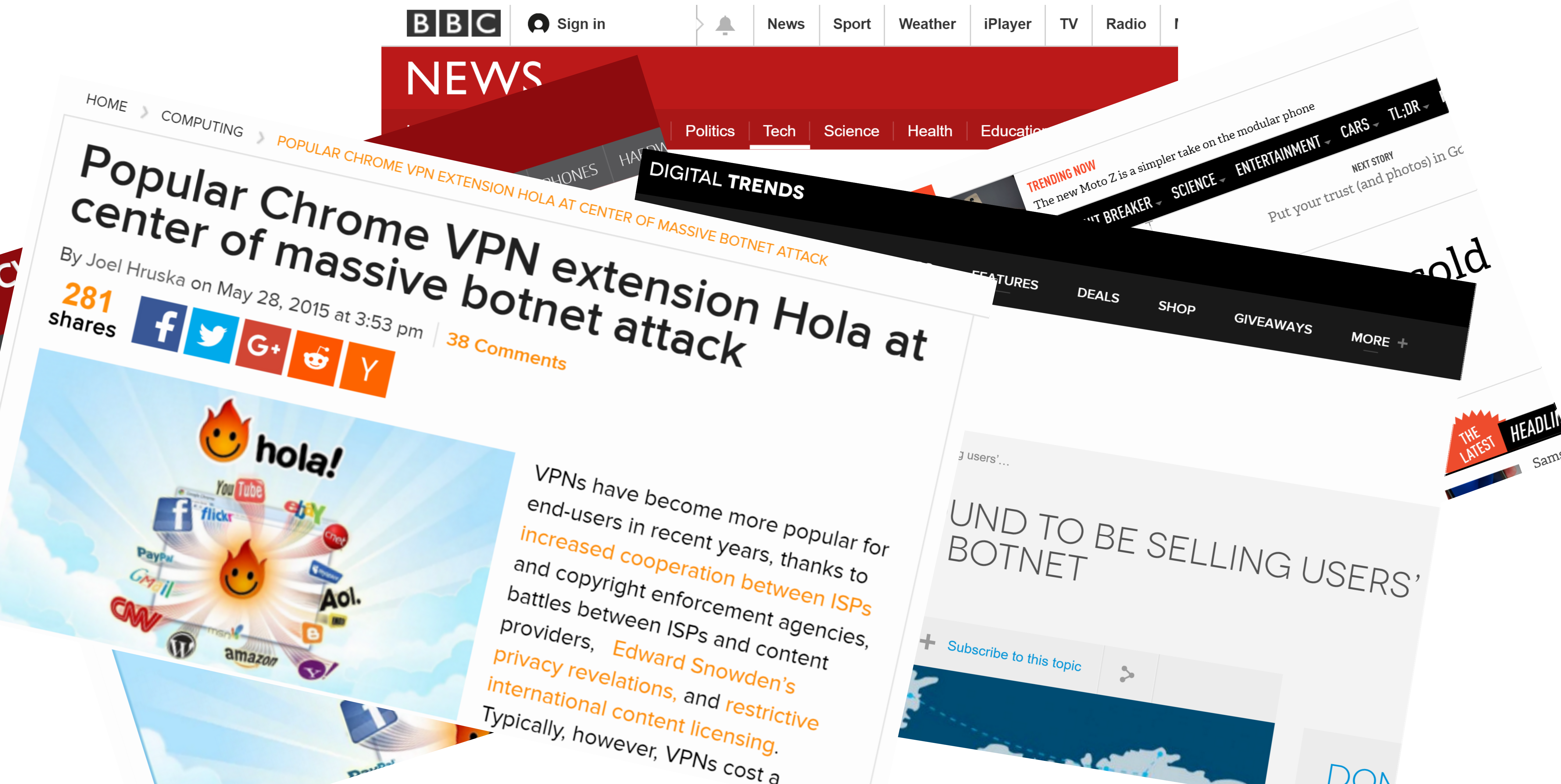


What the hell?!
I don't use Internet Explorer!

Hey, I received
User-Agent: Internet Explorer

```
GET i
Co
Ho
User
Accept-Lan
```

But how do we scale this up?



BBC Sign in News Sport Weather iPlayer TV Radio

NEWS

Politics Tech Science Health Educatio

HOME › COMPUTING › POPULAR CHROME VPN EXTENSION HOLA AT CENTER OF MASSIVE BOTNET ATTACK

Popular Chrome VPN extension Hola at center of massive botnet attack

By Joel Hruska on May 28, 2015 at 3:53 pm | 38 Comments

281 shares

VPNs have become more popular for end-users in recent years, thanks to increased cooperation between ISPs and copyright enforcement agencies, battles between ISPs and content providers, Edward Snowden's privacy revelations, and restrictive international content licensing. Typically, however, VPNs cost a

DIGITAL TRENDS

FEATURES DEALS SHOP GIVEAWAYS MORE +

TRENDING NOW The new Moto Z is a simpler take on the modular phone

IT BREAKER SCIENCE ENTERTAINMENT CARS TL;DR

NEXT STORY Put your trust (and photos) in Go

old

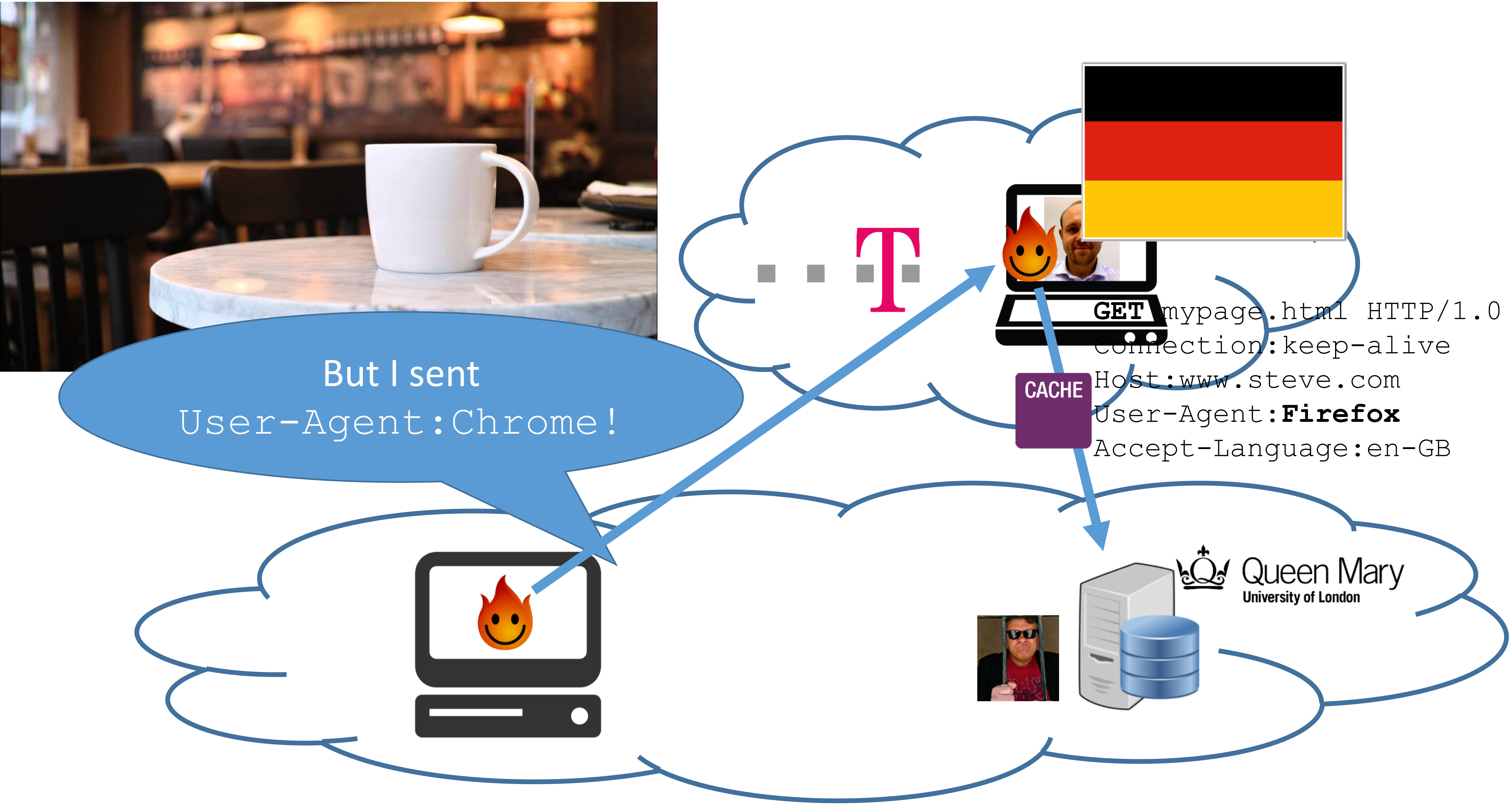
g users'...

UND TO BE SELLING USERS' BOTNET

+ Subscribe to this topic

THE LATEST HEADLIN

Sams

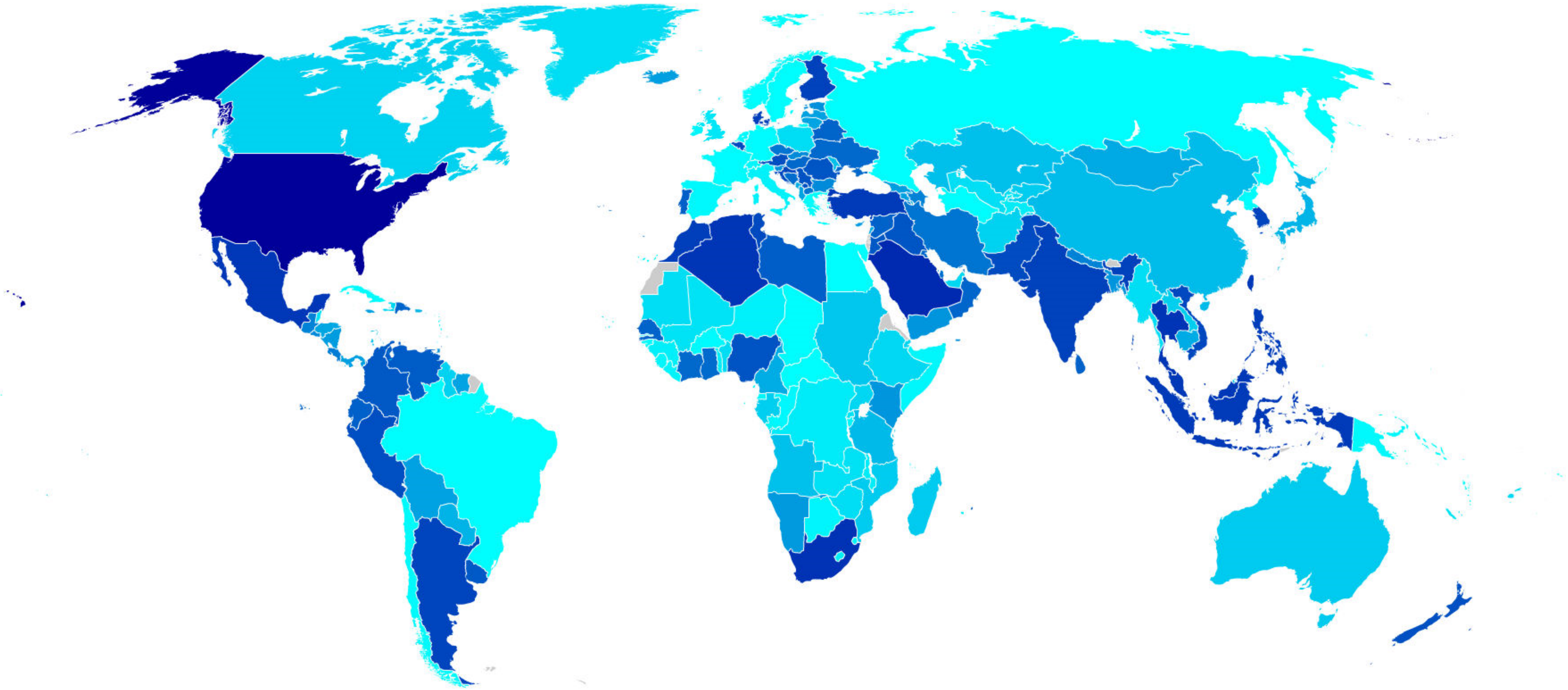


But I sent
`User-Agent:Chrome!`

```
GET mypage.html HTTP/1.0
Connection:keep-alive
Host:www.steve.com
User-Agent:Firefox
Accept-Language:en-GB
```

What does Hola give us?

Our geo coverage is pretty comprehensive



Back to HTTP headers…

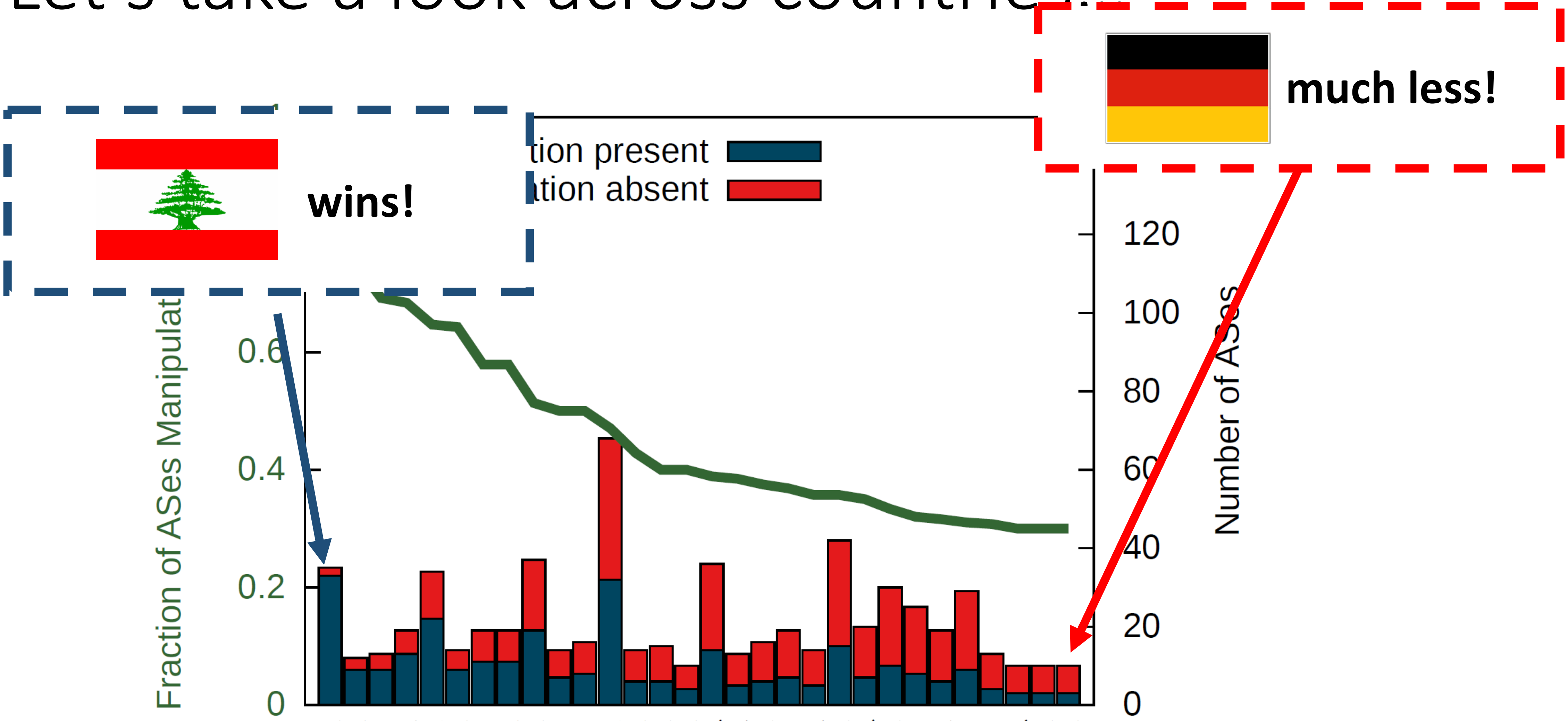
A few “highlights”

How many networks manipulate?

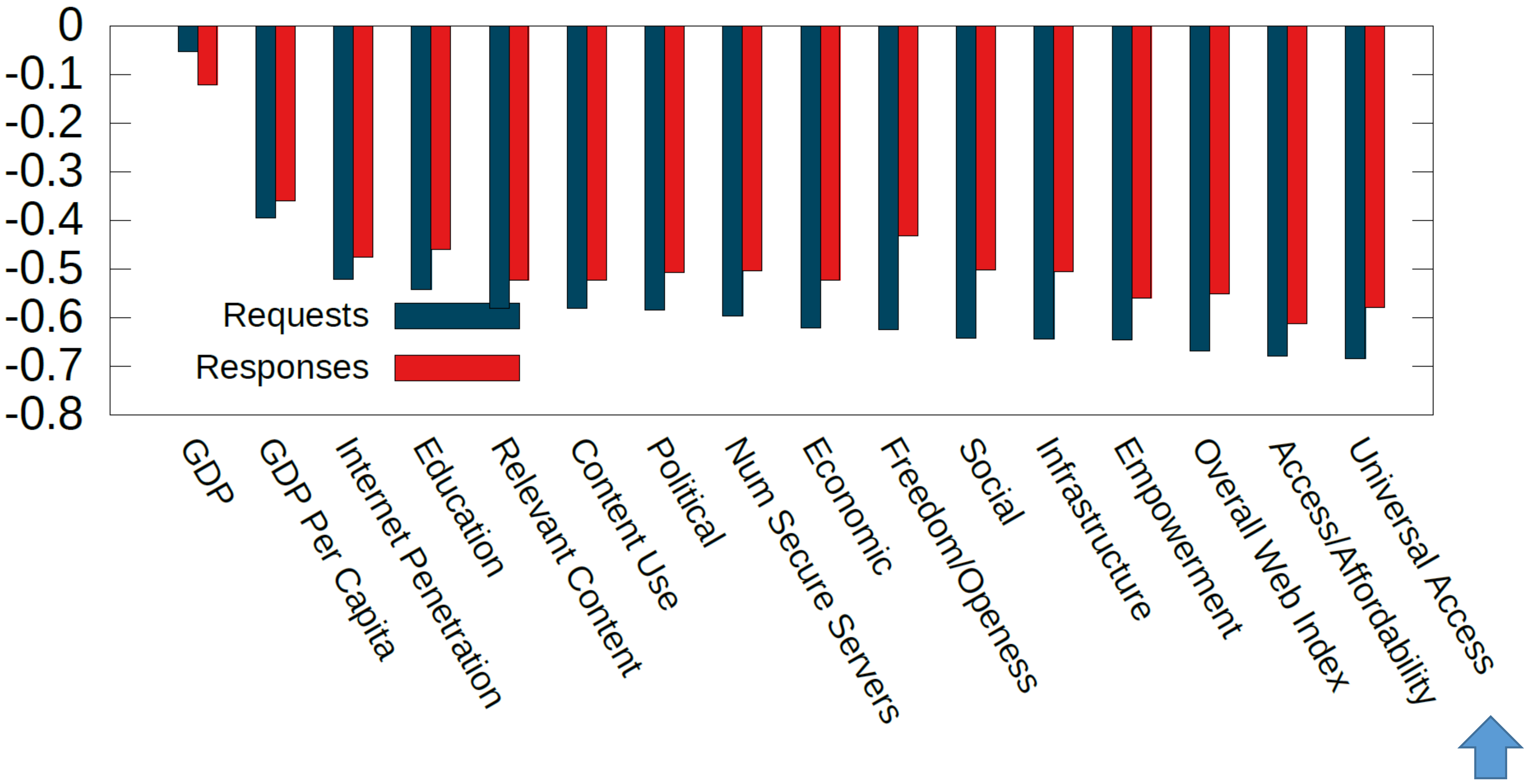
Quite a lot!

- 21% of the ASes have requests manipulated
- 19% for responses
- 25% of ASes contain sessions that manipulate headers at least once

Let’s take a look across countries…



Can we find patterns here?

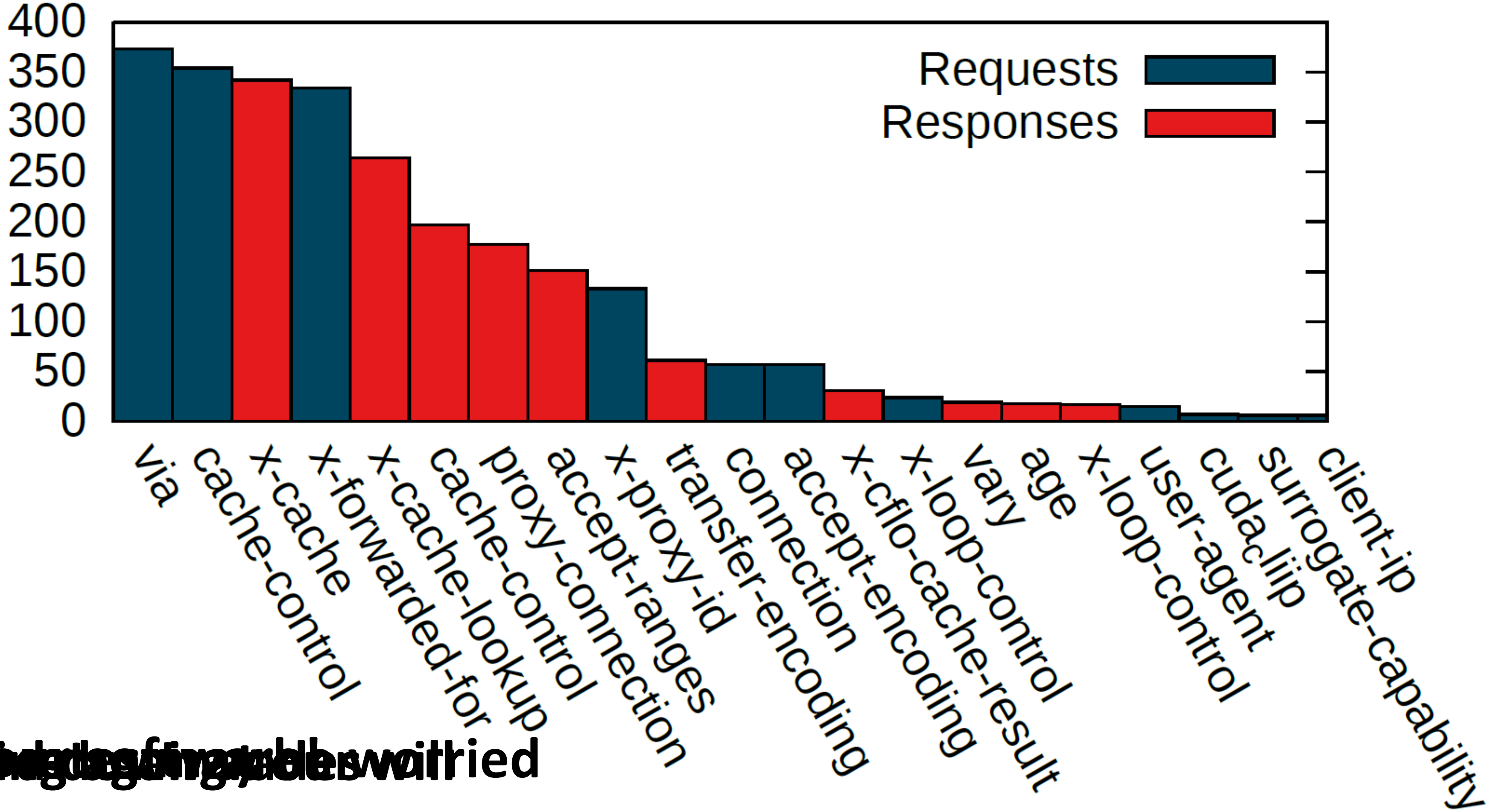


High Web Index == few middleboxes? Why?

- Internet transit cheaper in regions like Europe and US
  - IXPs are fewer in places like Africa
  - Therefore less necessary for caches in Europe/US
- HTTPS has been kicking caches hard
  - The hypergiants (Google, Netflix, Facebook) have been deploying their own dedicated cache boxes
- ….all translates to business common sense
  - EU: Cheaper to contact origin server than run caches
  - AF: Cheaper to run cache than contact the origin web server (via transit)

Lets look at the changes

So, what are the changes?



[illegible] ried

18 ASes injecting Age, 1 AS in SA reporting cache hits,

3 ASes removing cache-control settings!

15 ASes have their User-Agent downgraded, we even see HTTP/1.1 being downgraded

1 AS injecting cookies, 1 AS injecting phone numbers

26 ASes running Squid 2.5 or ealier!

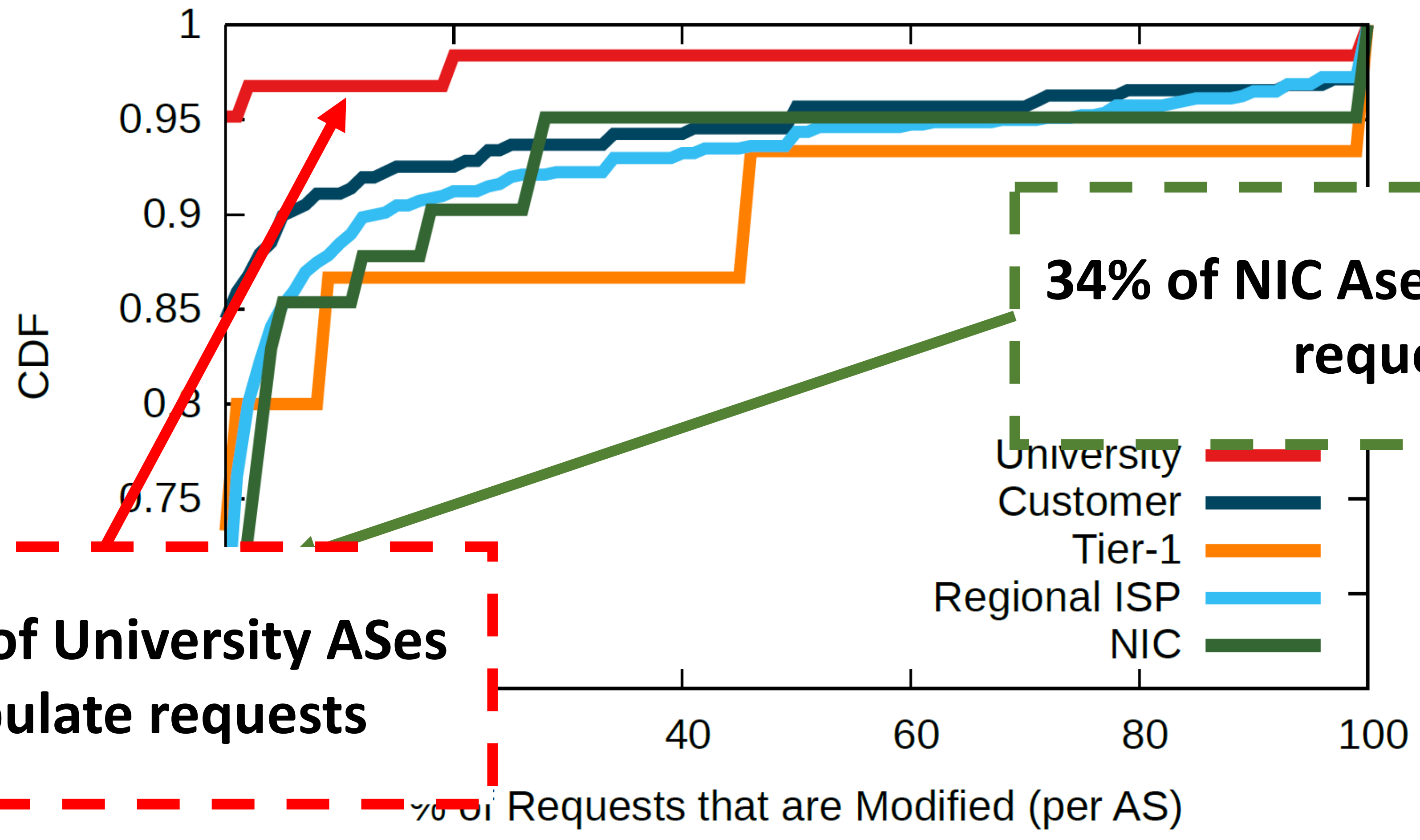
Conclusion

- Plenty of “meddleboxes” around
- Network operators are (usually) rational folks
    - So, observations can often tell us a lot about networks & regions

- Middleboxes are software too – they need T&C
- Take notice of your samples (networks and geo!)
- What happens with HTTP/2.0?

Any questions?

How prevalent is manipulation across network types?



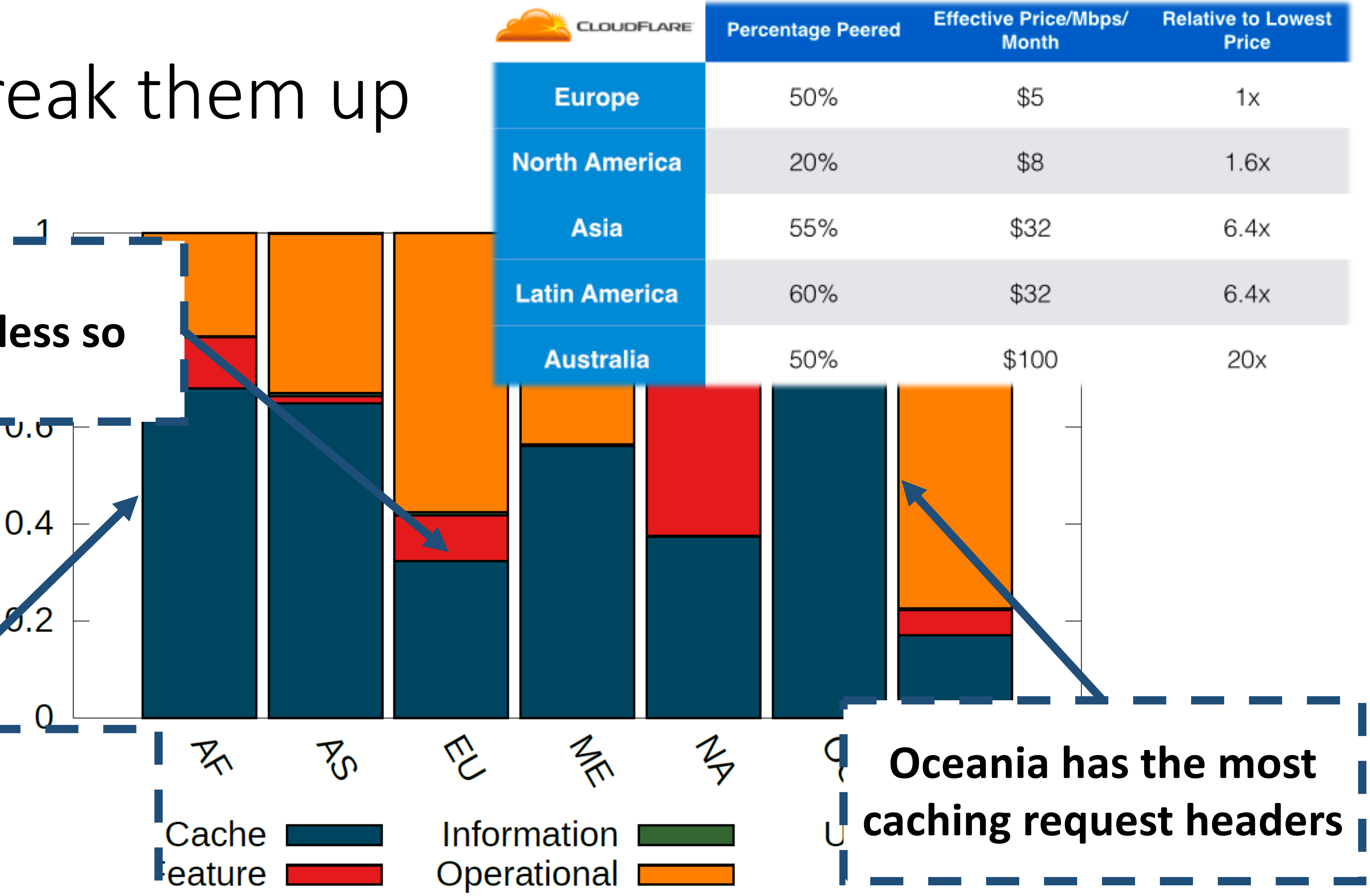
34% of NIC Ases manipulate requests

Only 5% of University ASes manipulate requests

Let’s break them up

- Cache
  - E.g. X-Cache, Cache-Control
- Operational
  - E.g. Via, X-Forwarded-For, CUDA_CLIP
- Feature Request/Advert
  - E.g. Connection, Accept-Encoding
- Informational
  - E.g. User-Agent, Set-Cookie
- Unknown
  - X-Client-TOS, SFID, X-TMV-Type, X-DG-TaggedAs, X-IMForwards
  - - - - - - - -

Let’s break them up



CLOUDFLARE	Percentage Peered	Effective Price/Mbps/ Month	Relative to Lowest Price
Europe	50%	$5	1x
North America	20%	$8	1.6x
Asia	55%	$32	6.4x
Latin America	60%	$32	6.4x
Australia	50%	$100	20x

Europe much less so

Africa too

Oceania has the most caching request headers