CS7NS5/CSU44032

Security & privacy

Stephen Farrell

stephen.farrell@cs.tcd.ie

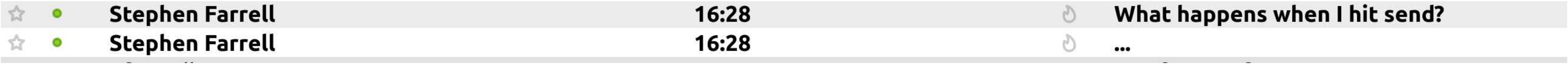
What happens when you hit “send”?
(Things about how email works.)

https://down.dsg.cs.tcd.ie/cs7053/
https://github.com/sftcd/cs7053

There are **many** slides in this deck

I will be skipping over many of them very quickly but happy to go back as needed (just yell at me or send email later)

Let’s see what happened so these mails could traverse the Internet



☆ ● Stephen Farrell 16:28 What happens when I hit send?
☆ ● Stephen Farrell 16:28 ...

- Both emails are from stephen.farrell@cs.tcd.ie to stephen@jell.ie
- The “from” address is my main TCD email
- The “to” address is (also me:-) at another mail server I operate myself
- The full content of the 1st mail is at: https://down.dsg.cs.tcd.ie/cs7053/lectures/mail1.html

How did that mail get to its destination?

- Sent from my `cs.tcd.ie` account on my laptop on the **TCD** SCSS WiFi n/w using Thunderbird
- ... to `outlook.office365.com`
  - That could be in **Paris**: https://www.lookip.net/ip/52.97.233.18 but probably not (26ms vs. 37ms average ping time for inria.fr)
- ... then it bounced around some internal Microsoft servers:
  - `DB6PR07CA0182.eurprd07.prod.outlook.com`
  - `DB7PR02MB5113.eurprd02.prod.outlook.com`
  - `HE1PR0202MB2746.eurprd02.prod.outlook.com`
- ... ‘till it reached a Microsoft egress server: `EUR02-VE1-obe.outbound.protection.outlook.com`
  - That could be in **Vienna** https://www.lookip.net/ip/213.199.154.47 but probably not
- ... which sent it to `jell.ie`’s mail exchange server `vps.jell.ie`
  - That is in **CityWest** https://www.lookip.net/ip/185.24.234.243
- ... where I read it from my laptop also via Thunderbird back in **TCD**

My mail got to routed to it’s destination via the Internet, so let’s talk about that a bit...

First though...

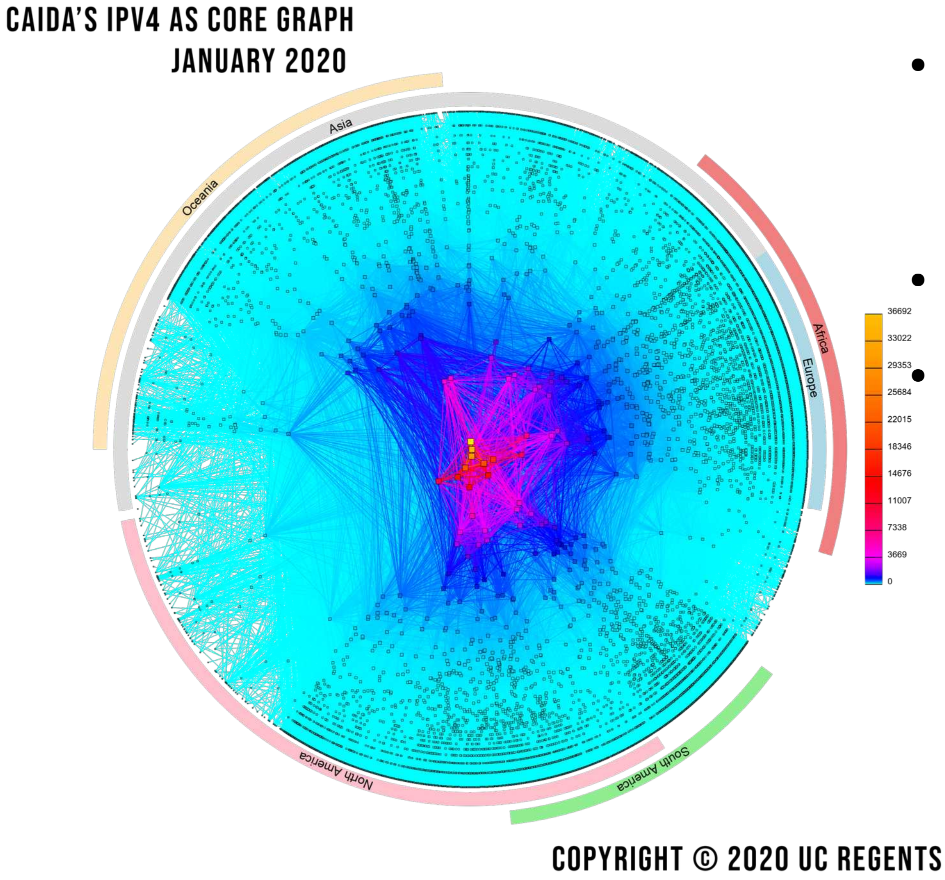
Is the Internet a network?

Is the Internet a network?
(hint: the answer is “no”:-)

A network of networks

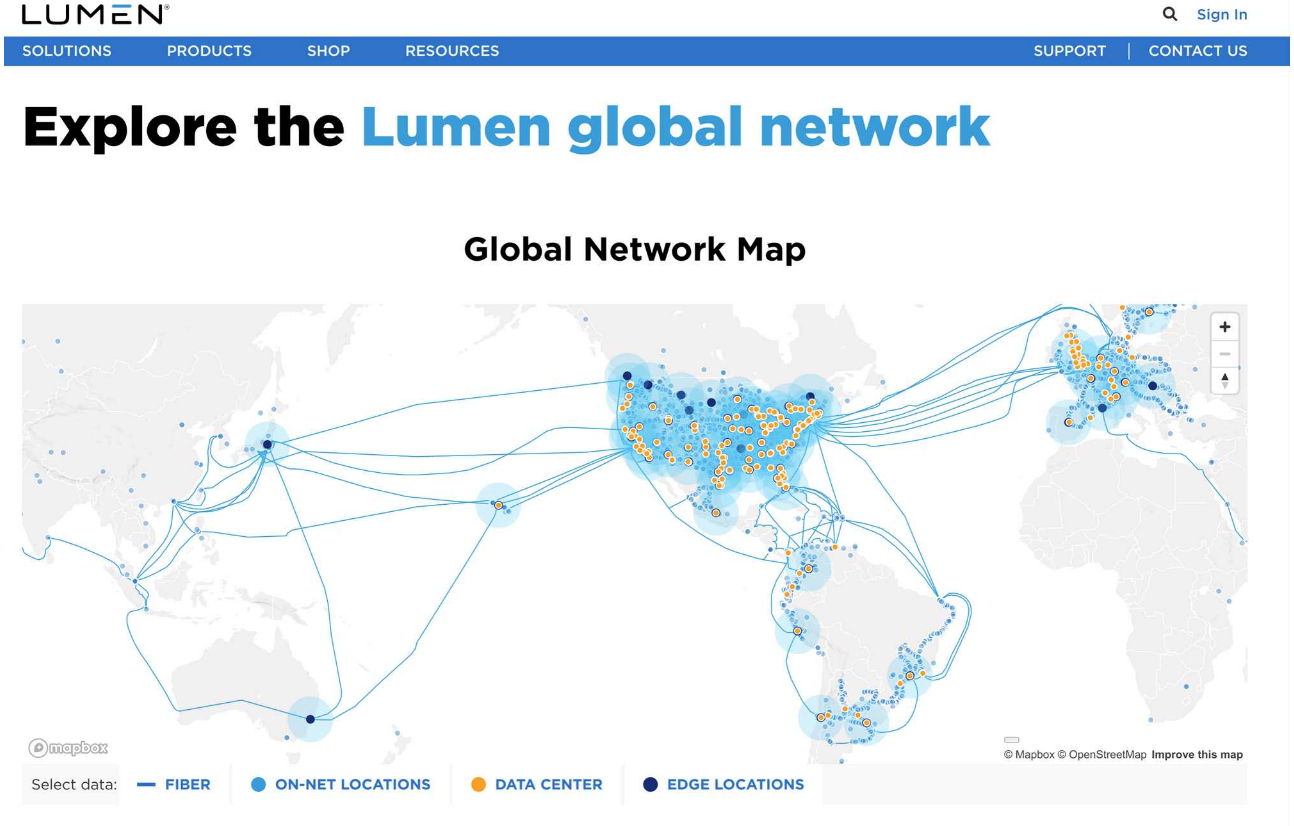
- The Internet is made up of tens of thousands of Autonomous Systems (ASes)
  - https://en.wikipedia.org/wiki/Autonomous_system_%28Internet%29
  - 72951 ASes as of 2022-02-02, (https://www.cidr-report.org/as2.0/)
    - Was: 65,428 in Aug 2019
- Think of these as the set of Internet Service Providers (ISPs, like Eircom, Vodafone, Virgin), other networks (e.g. HEANET which is TCD’s “ISP”), big companies (e.g. Microsoft, Google, IBM) and oddities like Internet eXchange Points (IXPs, like INEX)
- Each is (in principle and often in practice) an independent network (or set of networks) and their operators can do whatever they want
  - They’re essentially defined by sets of numbers: Static: AS number (ASN); Dynamic: sets of IP address prefixes
- They interact using Internet protocols (like IP, TCP, BGP) and peering arrangements
  - IP: Internet Protocol; TCP: Transmission Control Protocol; BGP: Border Gateway Protocol
  - Peering: agreeing to transit/handle traffic (and for how many €€€)

CAIDA Map of ASes



- CAIDA (Center for Applied Internet Data Analysis) is a UC San Diego Internet measurement organisation
  - You can measure **a lot** of what happens on the Internet as it happens!
- This is a 2020 map of the ASes as they were then
  https://www.caida.org/projects/cartography/as-core/2020/
- More central => more connected, serving more people
  - In the middle, are the highly connected ASes such as level3

Lumen (was Level3 etc.) is one of those (a BIG one)



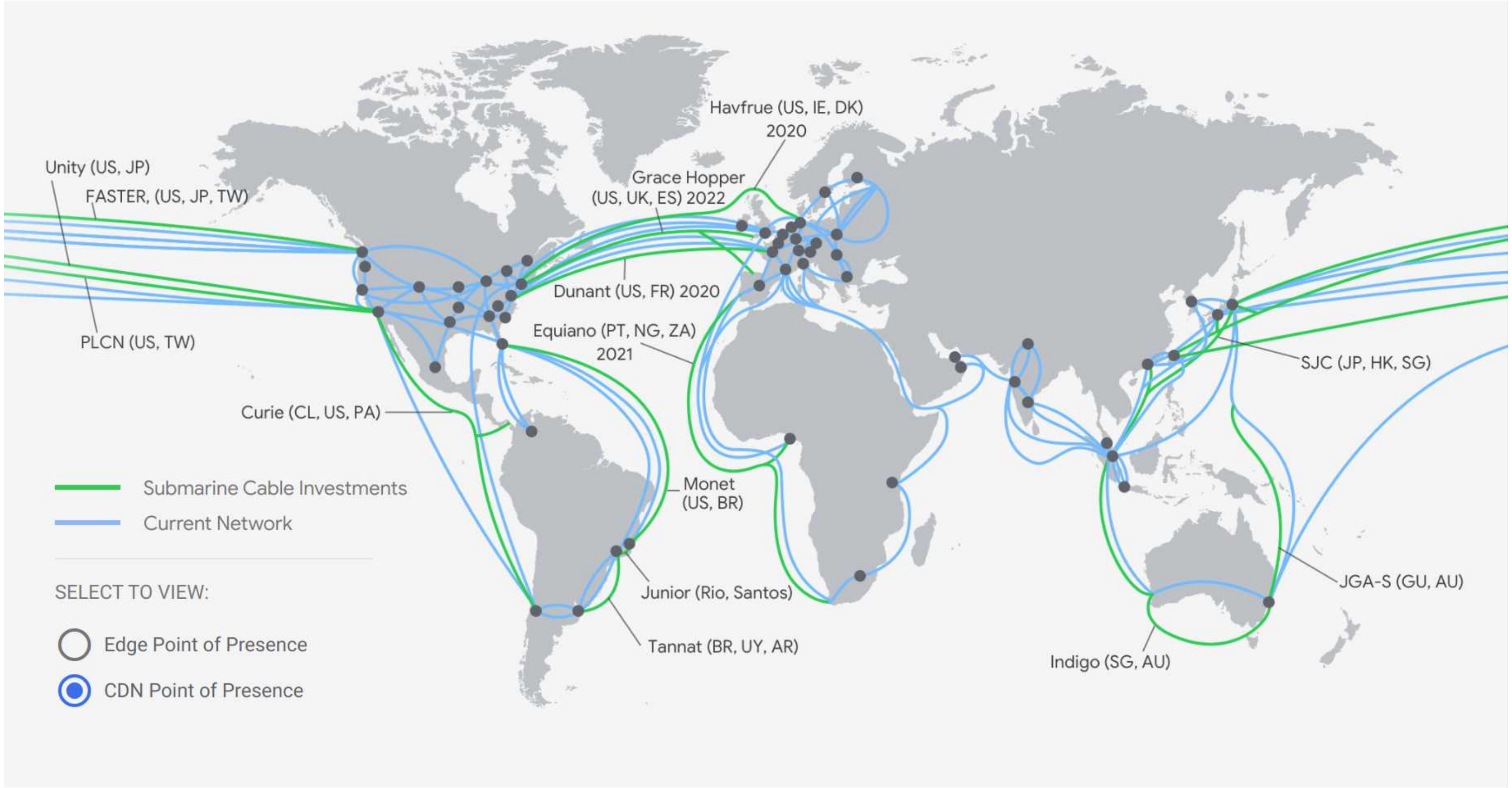
Explore the Lumen global network

Global Network Map

Select data: — FIBER ● ON-NET LOCATIONS ● DATA CENTER ● EDGE LOCATIONS

https://www.lumen.com/en-us/resources/network-maps.html

Google cloud



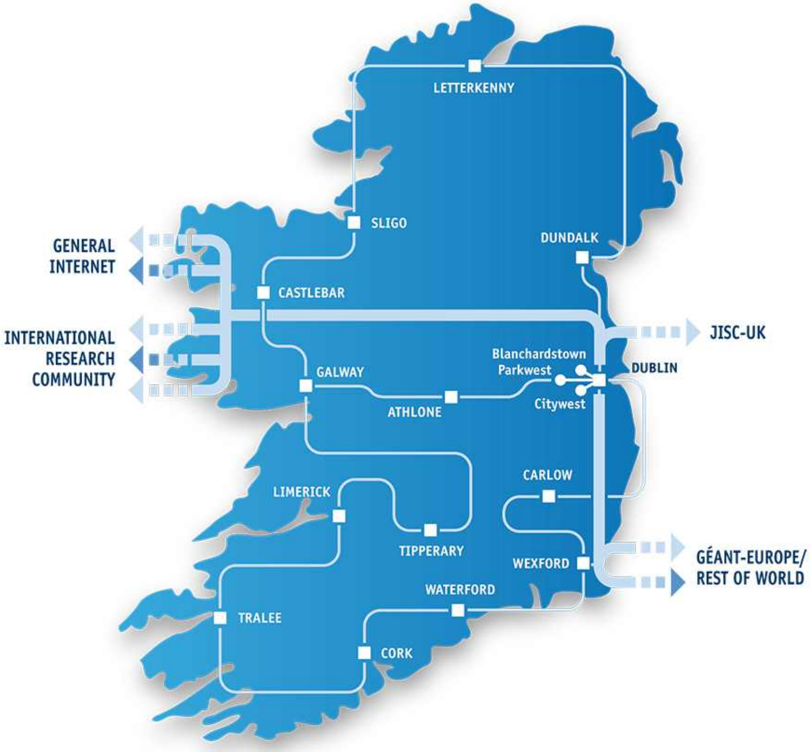
SELECT TO VIEW:

○ Edge Point of Presence

● CDN Point of Presence

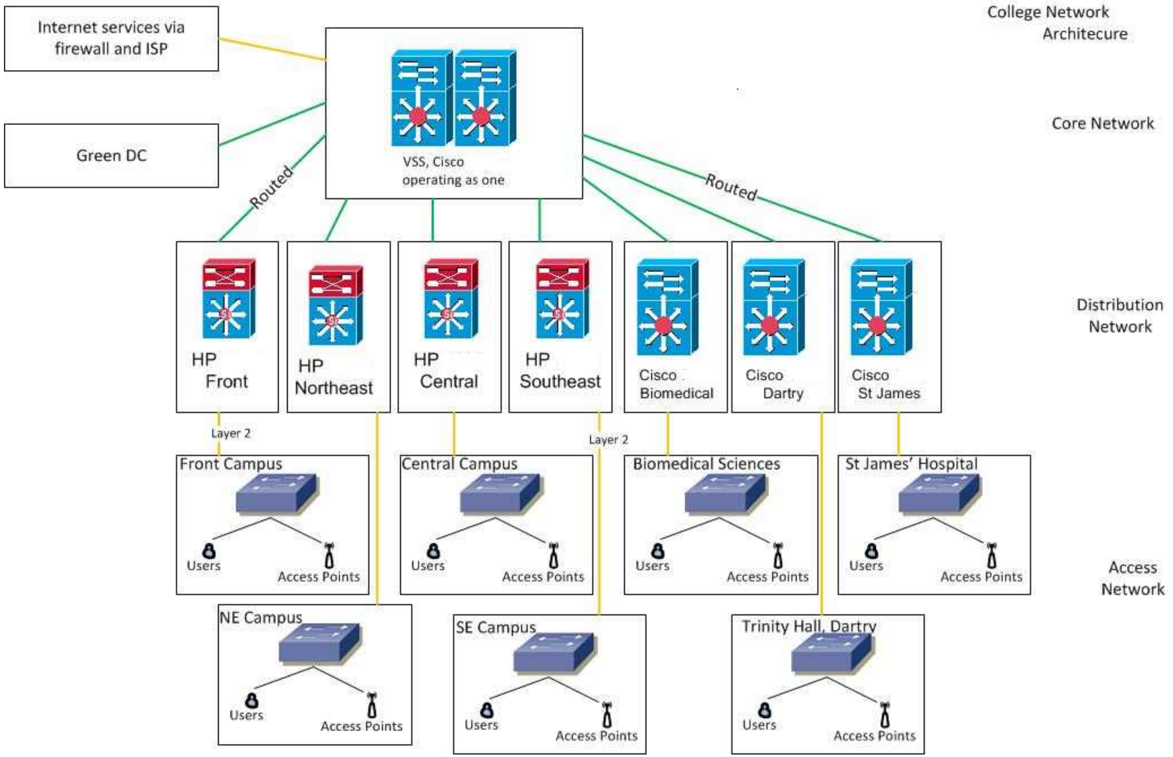
https://cloud.google.com/about/locations/#network note this is just google cloud, not all their stuff

Heanet national n/w



https://www.heanet.ie/network-maps

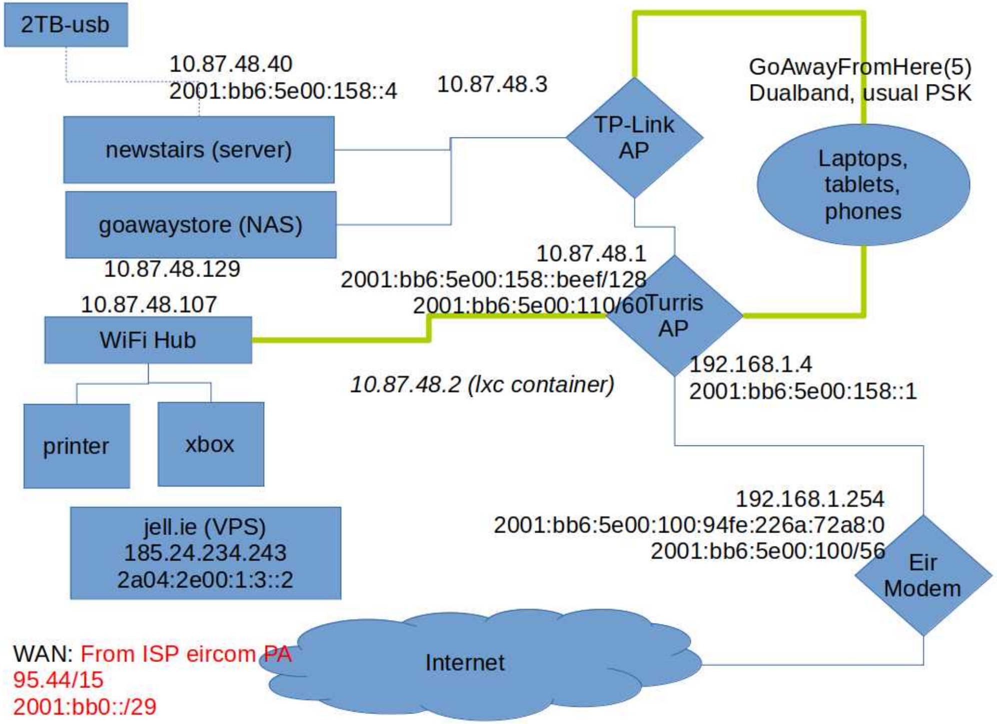
The TCD network (circa 2019)



- Network Core - Routers (High-availability pair)
- Distribution Layer - routed Layer 3 switches serving x7 campus zones
- Access Layer - Layer 2 Ethernet switches in building comms rooms and wireless Access Points
- External internet connectivity via L3 WAN block to ISP - Border Routers and Firewalls, DMZ hosting web services
- Data Centre network connectivity - central server and application hosting

Thanks to TCD IT for providing this anonymised diagram – doing that’s not a no-brainer btw!

My home network

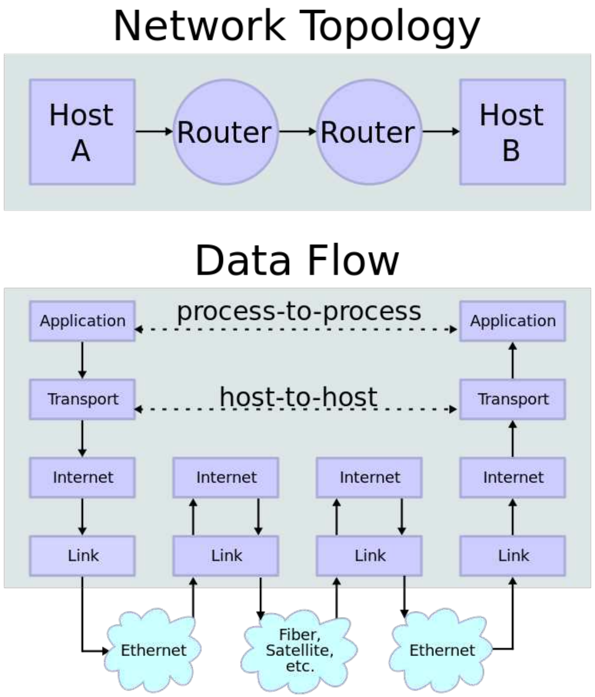


That’s from a few years ago – so is somewhat out of date:-)
Biggest change since: there’s now >12TB storage capacity!

Interoperability

- To make the Internet work, with all those networks at different scales, we need to agree on how to **interoperate** for some basic/minimal set of things
  - That means defining/agreeing on Internet Protocols
  - Where we need to agree on how to interoperate, a lot of that is done by the Internet Engineering Task Force (IETF) and other Internet standards bodies (IEEE SA, W3C)
- But we do not aim to agree about everything in everyone’s network
  - An awful lot happens at the “application layer” in code written by people and organisations, e.g. FB, Google, banks, Netflix, ...
  - Those services are **not the Internet** – they depend on the Internet!
- And yet more happens when people configure services that use generic code

What’s a network protocol?



https://en.wikipedia.org/wiki/File:IP_stack_connections.svg

“Permissionless innovation”

- One important point is: in principle each network operator can do whatever it wants so long as it interoperates “nicely” with others (and even when it doesn’t act particularly nicely;-)
  - That also applies to your home network (if you want and are able)
  - There are no protocol police (yet!)
- This is one of the main reasons why the Internet has been so successful

“Tussles”

- Repeating: we do not aim to agree about everything in everyone’s network...
- When the “policies” reflected in those collide then “fun” follows;-)
  - If protocols or application code constrains what operators can do then people complain
  - If what n/w operators are doing breaks (esp changes to) applications then people complain
  - In both cases people often complain at the wrong place;-)

The Internet is not the web

- The web is (roughly) the set of computers that speak the HTTP protocol
  - HTTP == HyperText Transfer Protocol (http://example.com)
  - HTTPS == HTTP/Transport Layer Security (https://example.com)
- Email doesn’t use HTTP, but rather (mostly) the Simple Mail Transfer Protocol (SMTP) which is a couple of decades older than HTTP
- Mobile network internals (3G, 4G, 5G...) mostly run over IP using a bunch of protocols you’d prefer to never have to know about

Back to my mail...

- From my laptop the packets making up the mail (submission) probably went via `134.226.254.93` (a TCD router) to the destination address of `52.97.214.146` (Outlook ingress server)
- TCD (who “are” `134.226/16`) know how to route to the outside world via Heanet
- Heanet know how to route to `52.96/12` (Msft), via INEX, our local Internet Exchange Point (IXP)
- There are about 12 routers on that path
- A (possibly different) reverse path also needs to work, for acknowledgements, to handle errors and to prevent denial-of-service and reduce spam (so packets from `52.97.214.146` need to find their way back to `134.226/16` as needed).
- But how did anything know that `outlook.office365.com` is at `52.97.214.146`?

Domain Name System (DNS)

How do we find the destination?

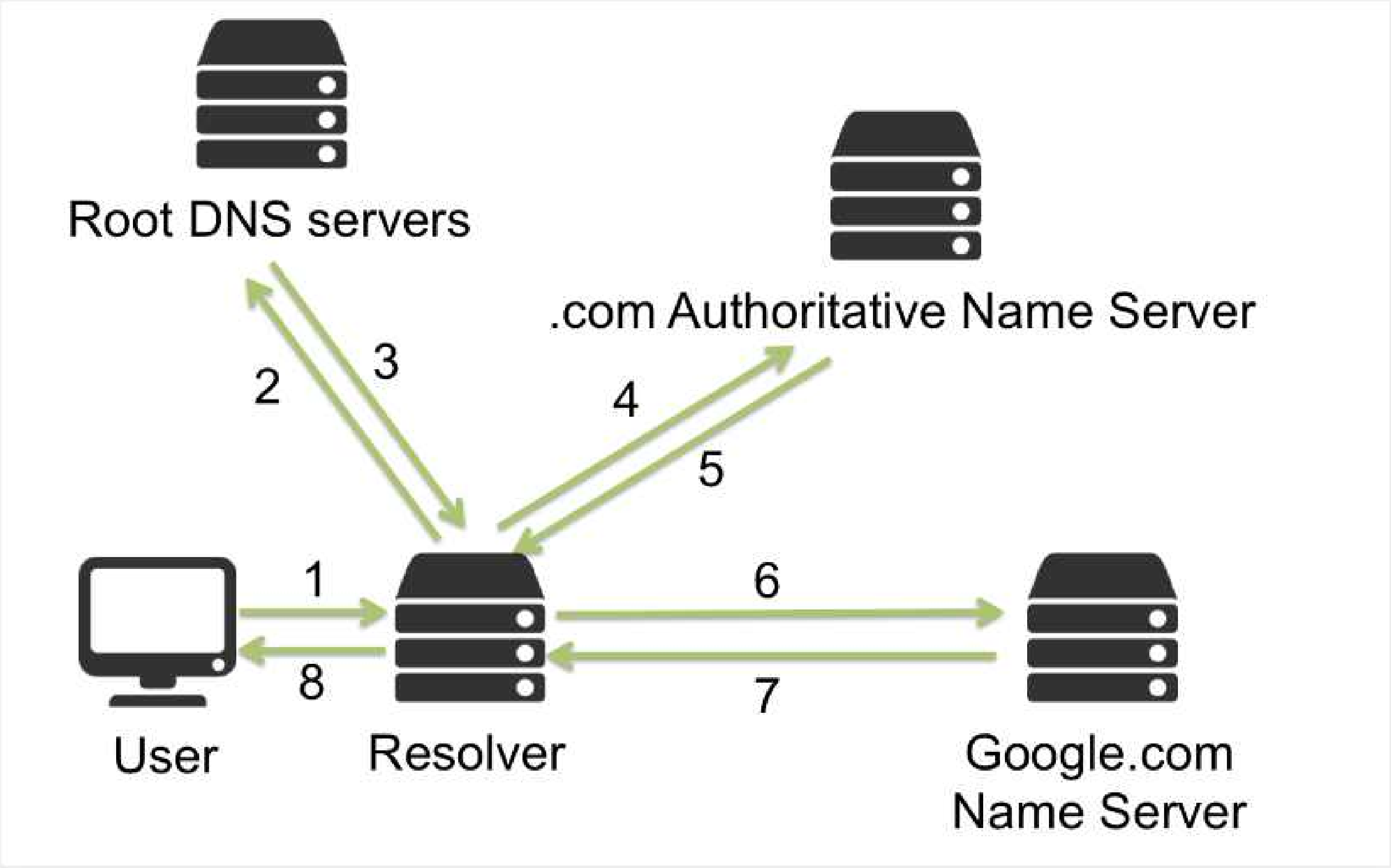
- All those names we saw earlier are DNS names:
  - `cs.tcd.ie`
  - `outlook.office365.com`
  - `DB6PR07CA0182.eurprd07.prod.outlook.com`
  - `DB7PR02MB5113.eurprd02.prod.outlook.com`
  - `HE1PR0202MB2746.eurprd02.prod.outlook.com`
  - `EUR02-VE1-obe.outbound.protection.outlook.com`
  - `jell.ie`
  - `vps.jell.ie`

What’s the DNS?

- **The** single world-wide distributed naming system (that’s worked so far)
- Main purpose: map names to IP addresses so applications and humans can deal with names rather than addresses
  - tcd.ie is easier to type/remember than something within `134.226/16` or `2001:770:10::/48`
  - IPv4 addresses use A resource record (RR) type; IPv6 addresses use AAAA resource record (RR) type
- The DNS is also used for many other purposes
  - Mail address right-hand-side to mail server name(s) via MX RR
  - DNS block lists of spam sources (and other block lists)
  - “Passive DNS” monitoring for various security purposes
  - Telling lies (RPZ) to help people avoid “bad” domains
  - State- or enterprise-level censorship
  - ...

Traditional DNS recursive diagram

(resolving google.com)



DNS Ecosystem

- Top Level Domains (TLDs)
  - Country-code TLDs (ccTLDs): .ie, .uk, .is,...
    - Each more or less do what they want
    - IEDR manage .ie zone, CZ.nic manage .cz, ...
  - Generic TLDs (gTLDs): .com, .org, .net,...
    - Run under ICANN’s oversight (https://icann.org)
    - There are ~1000 of those now (because $$$)
- Second level domains (2LDs), or effective Top Level Domains (eTLD)
  - Comply with parental controls (to some extent)
  - Examples: example.com, tcd.ie, amazon.co.uk
  - .com TLD has ~158M names, .ie has ~300k, .org has ~10M, Total: ~365M (as of Q3 2021) https://www.verisign.com/en_US/domain-names/dnib/index.xhtml
- Third level and below: controlled by parent/2LD/eTLD, e.g. down.dsg.cs.tcd.ie

The root zone

- The root zone “.” is special – it’s content is (carefully) managed by IANA and handed over to the root server operators…
    - https://www.iana.org/domains/root/servers
- The root server operators serve the root zone – about 1000 instances worldwide in about 130 countries, with subsets of those managed each of by 12 organisations via13 named root servers (there was one merge in the last couple of decades)
- Most root zone instances are accessed using anycast IP routing
    - Other public authoritative and even recursive servers (e.g. QuadN’s such as 8.8.8.8 or 1.1.1.1 or 9.9.9.9) also use anycast for better performance
- The root zone is pretty stable – and the Internet really needs that to be the case
- Every Recursive needs at least one root server IP to start

Registry/Registrar/Registrant

- Top Level Domains (TLDs) are operated by registries,
  - IEDR manage .ie; Affilias operate a bunch of ccTLDs and gTLDs https://afilias.info/global-registry-services
  - Public Interest Registry (PIR) operate .org (and feed $$$ to Internet Society, which feeds $$ to IETF and RFC editor)
- Registrars accredited by registries and deal with registration of names (and transfer and de-registration); registry-specific rules may apply, e.g. “connection to Ireland” for .ie
- Registrant is the entity that wants/has a name registered
- Registries handle name conflicts, e.g. when trademark issues arise via some dispute resolution process (can involve $$$)
- Registration costs to registrants from registrars vary from “free” to ~$1000, but mostly ~$10 per year
  - Some money flows up from registrar to registry (ccTLD or gTLD) and to ICANN (for gTLDs)
- ICANN auction new gTLDs now and then :-(
  - Costs ~$1M+ to play that game, ICANN have ~$150M resting in an account as a result

DNS Servers/Protocol

- A (logical) zone file for a domain is served by an Authoritative DNS server
- Recursive DNS servers query the Authoritatives to resolve names, e.g. starting at "." ask "where is .ie"; get answer; at ".ie" ask "where is tcd.ie"; at "tcd.ie" ask "what is the IP for www.tcd.ie"
- Clients (your laptop/phone) ask Recursive servers to resolve names e.g. "what is the IP for tcd.ie" and if it doesn't already know the answer the Recursive will do as much of the dance above as needs (re-)doing
- The protocol spoken between clients, Recursives and Authoritatives is the DNS protocol

Changing DNS...

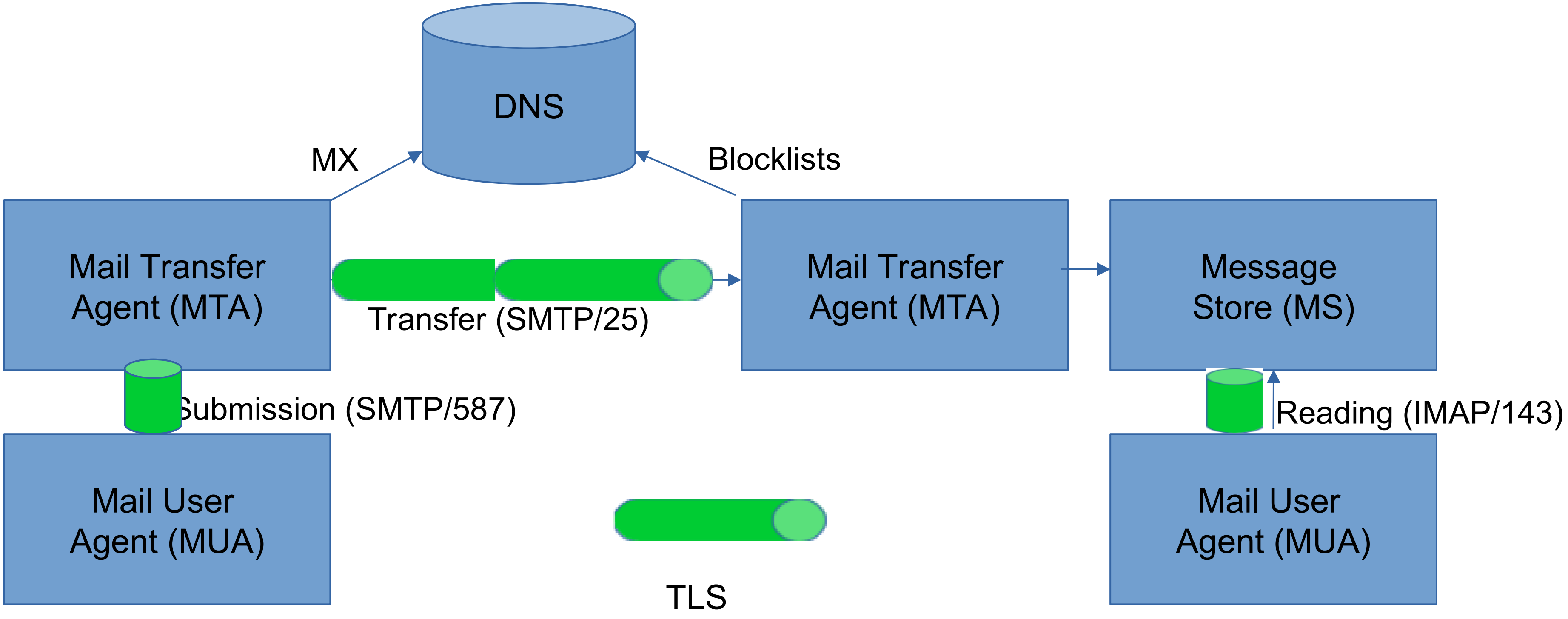
- “Public” recursives such as 8.8.8.8 (google), 9.9.9.9 (quad9) or 1.1.1.1 (cloudflare) are moving a control point (n/w location of recursive)
- Stub <-> recursie DNS encryption mechanisms (DoT/DoH) are moving another control point (visibility of queries/answers *and* the placement of the stub on the host)
- We’re starting to figure out how to encrypt recursive <-> authoritative
- Affected people get very cranky when such changes happen:-)
- These are real changes – recent EU proposal for funding a “Euro” public recursive (quad9 is in .ch btw)

Back to my mail...

- I don’t have logs for these but the following is likely/possible...
- My laptop maybe used `89.234.186.112` as a name server to resolve `outlook.office365.com`
  - That may be in **Brittany, France** https://ipinfo.io/89.234.186.112
- Microsoft maybe accessed `a.iedr.ie/77.72.78.91` to find the name server for `jell.ie`
  - That is in **Dublin, Ireland** https://ipinfo.io/77.72.78.91 but has “mirrors” in Holland, US etc.
- Microsoft accessed one of the `jell.ie` name servers (maybe `ns1.frobbit.se/45.56.92.19`) to resolve `jell.ie`
  - That may be in **Freemont, CA, USA** https://ipinfo.io/45.56.92.19
- The authoritative zone file for `jell.ie` is “homed” **in my home:-)**
  - DNSSEC re-signed daily on a server in my home and uploaded to various places including `ns1.frobbit.se`
- `vps.jell.ie` used it’s own (local) recursive DNS server to get Microsoft’s DKIM public key so may have done the full “DNS dance” shown earlier, that may have accessed `ns1-208.azure-dns.com/40.90.4.208`
  - That may be in **Redmond, WA, USA** https://ipinfo.io/40.90.4.208

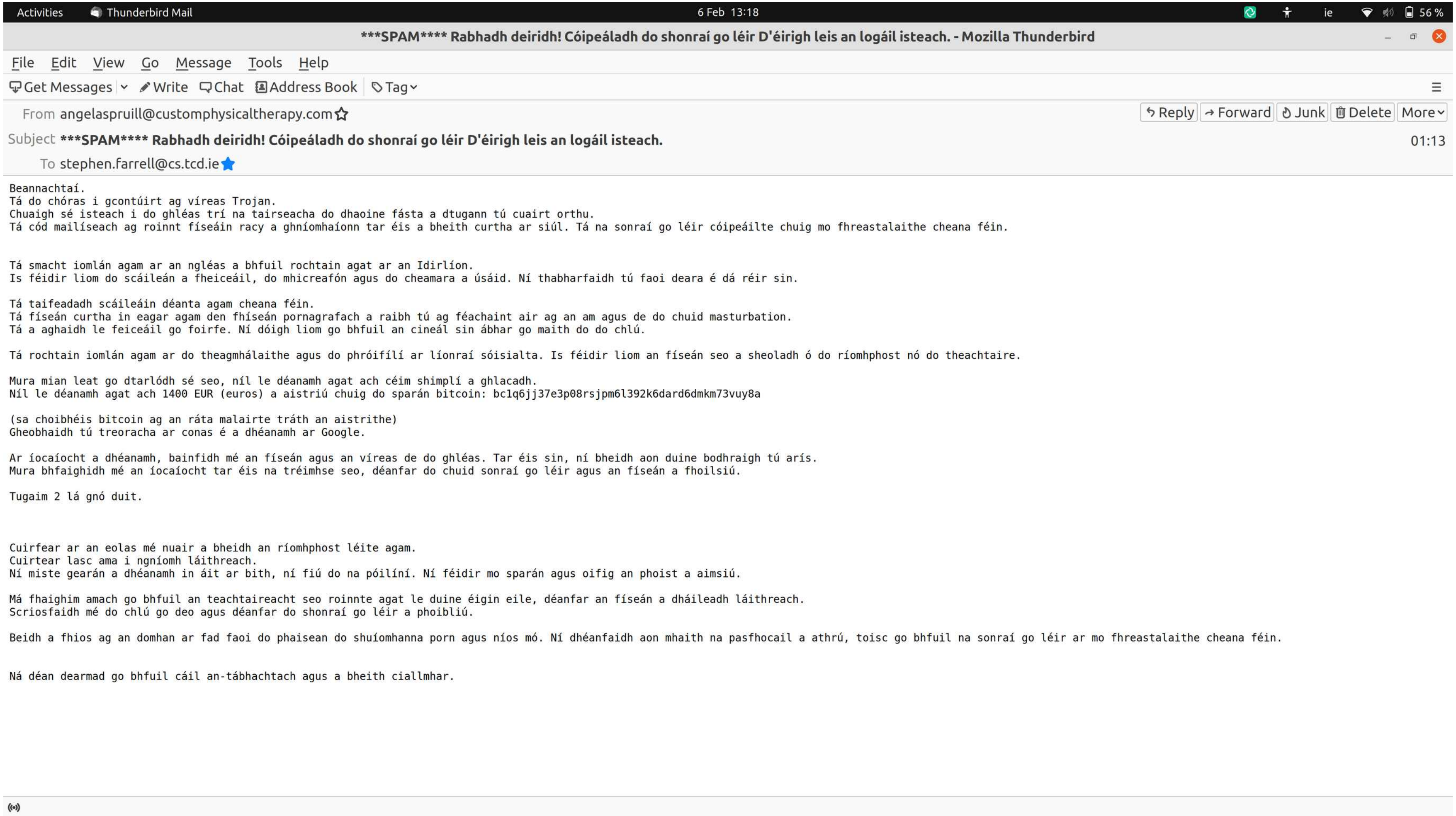
Email Protocols

Mail Transport Security



Mail Transport Security sometimes described as “hop-by-hop security”

Skipping over Spam...



From angelaspruill@customphysicaltherapy.com

Subject ***SPAM**** Rabhadh deiridh! Cóipeáladh do shonraí go léir D'éirigh leis an logáil isteach.

To stephen.farrell@cs.tcd.ie

Beannachtaí.
Tá do chóras i gcontúirt ag víreas Trojan.
Chuaigh sé isteach i do ghléas trí na tairseacha do dhaoine fásta a dtugann tú cuairt orthu.
Tá cód mailíseach ag roinnt físeáin racy a ghníomhaíonn tar éis a bheith curtha ar siúl. Tá na sonraí go léir cóipeáilte chuig mo fhreastalaithe cheana féin.

Tá smacht iomlán agam ar an ngléas a bhfuil rochtain agat ar an Idirlíon.
Is féidir liom do scáileán a fheiceáil, do mhicreafón agus do cheamara a úsáid. Ní thabharfaidh tú faoi deara é dá réir sin.

Tá taifeadadh scáileáin déanta agam cheana féin.
Tá físeán curtha in eagar agam den fhíseán pornagrafach a raibh tú ag féachaint air ag an am agus de do chuid masturbation.
Tá a aghaidh le feiceáil go foirfe. Ní dóigh liom go bhfuil an cineál sin ábhar go maith do do chlú.

Tá rochtain iomlán agam ar do theagmhálaithe agus do phróifílí ar líonraí sóisialta. Is féidir liom an físeán seo a sheoladh ó do ríomhphost nó do theachtaire.

Mura mian leat go dtarlódh sé seo, níl le déanamh agat ach céim shimplí a ghlacadh.
Níl le déanamh agat ach 1400 EUR (euros) a aistriú chuig do sparán bitcoin: bc1q6jj37e3p08rsjpm6l392k6dard6dmkm73vuy8a

(sa choibhéis bitcoin ag an ráta malairte tráth an aistrithe)
Gheobhaidh tú treoracha ar conas é a dhéanamh ar Google.

Ar íocaíocht a dhéanamh, bainfidh mé an físeán agus an víreas de do ghléas. Tar éis sin, ní bheidh aon duine bodhraigh tú arís.
Mura bhfaighidh mé an íocaíocht tar éis na tréimhse seo, déanfar do chuid sonraí go léir agus an físeán a fhoilsiú.

Tugaim 2 lá gnó duit.

Cuirfear ar an eolas mé nuair a bheidh an ríomhphost léite agam.
Cuirtear lasc ama i ngníomh láithreach.
Ní miste gearán a dhéanamh in áit ar bith, ní fiú do na póilíní. Ní féidir mo sparán agus oifig an phoist a aimsiú.

Má fhaighim amach go bhfuil an teachtaireacht seo roinnte agat le duine éigin eile, déanfar an físeán a dháileadh láithreach.
Scriosfaidh mé do chlú go deo agus déanfar do shonraí go léir a phoibliú.

Beidh a fhios ag an domhan ar fad faoi do phaisean do shuíomhanna porn agus níos mó. Ní dhéanfaidh aon mhaith na pasfhocail a athrú, toisc go bhfuil na sonraí go léir ar mo fhreastalaithe cheana féin.

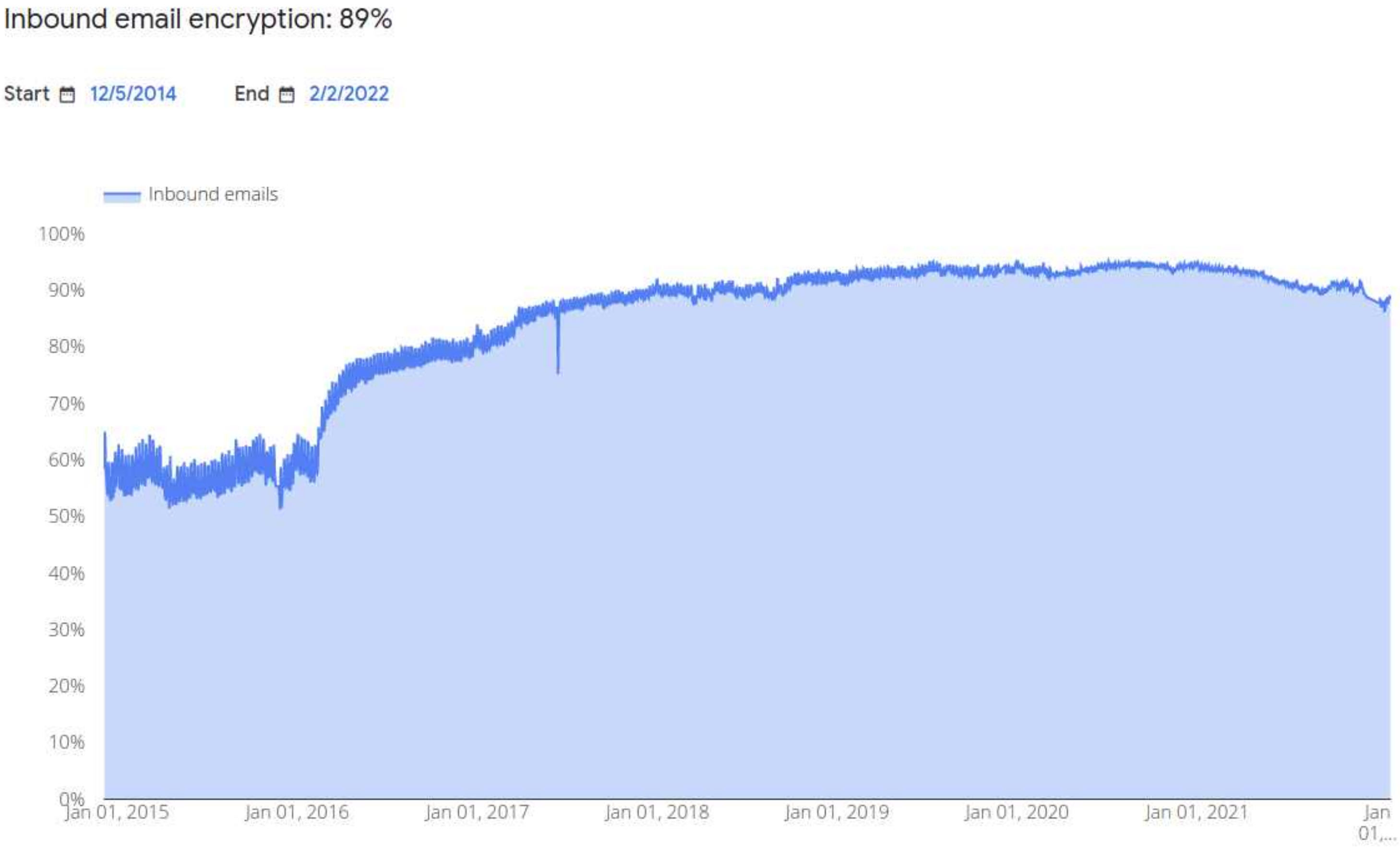
Ná déan dearmad go bhfuil cáil an-tábhachtach agus a bheith ciallmhar.

Mail Transport Security Stats

- Gmail publish “transparency report” stats

  https://transparencyreport.google.com/safer-email/overview
- Other service providers see commensurate numbers
- Recently: 89% of outbound and inbound are protected via some form of transport layer security
  - TLS: same security protocol that makes https from http!
  - Note: “Some form” == possibly less well configured
- Regional variations remain

HBH over time (still @gmail)



Inbound email encryption: 89%

Start 12/5/2014 End 2/2/2022

End-to-End email security

- We’d like to encrypt mail content from sender to recipient and not just hop-by-hop
- There are two ways to do that:
  - S/MIME and PGP
- S/MIME is much more “commercial”
- PGP is much more “roll-your-own”
- Neither is in widespread use (sadly)

Wanna Try S/MIME?

- It’s been a few years since I had a working s/mime setup (I use PGP mostly), but for your benefit...
- For this year, I tried this “free” service https://www.actalis.com/s-mime-certificates.aspx
- Verification email was Italian-first:-)
- Then they sent me a pkcs#12 file!
- That means they know the private key!!
- I didn’t install that in my mail user agent
- Doesn’t look good for such ad-hoc use: https://kb.mozillazine.org/Thunderbird_:_FAQs:_Get_an_SMIME_certificate

Pretty Good Privacy (PGP)

- PGP can do all that S/MIME does
- PGPMime is RFC 3156
- PGP’s basic formats in RFC 4880
  - Currently being updated (I help co-chair that)
    - https://datatracker.ietf.org/doc/draft-ietf-openpgp-crypto-refresh/
- Web-of-trust model != X.509 PKI
  - But you don’t have to
- Most important use-case? Maybe package signing
- Now natively supported in Thunderbird

S/MIME and PGP Deployment

- Most mail clients suport S/MIME or PGP either built-in or as an option
  - There are also “plug-in” products
- And mostly then *can* work together
  - I’ve used both, PGP more usable (with Thunderbird)
- But e2e secure mail is not ubiquitous

E2E email security barriers

- Designs pre-date web user agent which changes trust model (where's the private key kept? Needs new infrastructure)
- Needs major mailbox providers (yahoo, hotmail, gmail) to deploy the same thing which also needs to be implemented by all major MUA developers (microsoft, mozilla, apple, google)
- Public key retrieval needed (doable if the above done, but a killer if not), but who's gonna pay?)
- We need to unify S/MIME and PGP or pick one or we'll lose interop (it's ok if the other soldiers on for some niches)
- Users don't care, so it has to be entirely transparent for them (needs significant UI work, co-ordinated across MUAs and significant web-UAs)
- New anti-spam deployments needed
- Note: this list used have one more entry, but header-encryption is on the way to being sorted!

Back to my mail...

- Just for that mail packets flew about to places that likely include:
  - TCD
  - Paris
  - Vienna
  - CityWest
  - Brittany, France
  - Dublin, Ireland and/or Holland, US (mirrors)
  - Freemont, CA, USA and/or CityWest/Sweden
  - My home:-)
  - Redmond, WA, USA

Back to my mail...

Organisations involved include...

- TCD
- Microsoft
- Heanet
- INEX
- Servebyte
- Google
- Frobbit
- IEDR

...and less obviously...

- IETF
- ICANN
- LetsEncrypt
- Mozilla
- Canonical
- Linux Kernel maintainers
- Postfix maintainers

...and many more...

Conclusion

- It’s not that complex but it is complicated
  - Billions of emails are sent every day, not all of which are spam:-)
- Almost all of the issues above apply also to the web
  - But the Internet is (still:-) not the web
- There are many “players” in the Internet game
  - Tussles abound and that will continue
- There may be some legal consequences...
  - Hands over gratefully to Maria Grazia :-)