TEU00311
What is the Internet doing to me?
(witidtm 2021/2022)

Stephen Farrell
stephen.farrell@cs.tcd.ie

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

URLs accessed 20210907 (not all content from URLs updated)

What’re we here for?

- We all use the Internet all the time
- You may or may not know what’s happening under the hood, and shouldn’t need to know all the nitty-gritty detail
- But, to make better decisions as to what you do, it’s good to know something about some of those details
- This module aims to help you learn enough to make better decisions about what you want, and how to get it, as you interact with the Internet
- I hope: you’ll apply those lessons, tell others about it all and maybe agitate for a better Internet for a better society (but you won’t fail the module if you don’t agitate:-)

Administrivia

TCD Personnel/Contacts

- Lectures:
  - Dr. Stephen Farrell, stephen.farrell@cs.tcd.ie
  - Dr. Dave Lewis, dave.lewis@cs.tcd.ie
    - You'll see Dave after Reading week
  - Dr. Eoin O'Dell, eoin.odell@tcd.ie
    - You'll see Eoin next week
- For generic stuff, I guess you email me – try include "witidtm" in the subject line

about:me

- SCSS research fellow
- Research topics: Internet security & privacy and delay-tolerant networking
- Other courses taught, pubs, CVish stuff etc:
  - https://www.scss.tcd.ie/Stephen.Farrell/

Who are we? (SITS as of Sep 7th)

- 1 Business
- 1 Economics Single Honors
- 1 Film Studies and English Literature Joint Honors
- 1 French and Italian Joint Honors
- 1 Irish and French Joint Honors
- 1 Law and Business Joint Honors
- 1 Medicinal Chemistry
- 1 Music Single Pathway
- 1 Zoology
- 2 Electronic and Computer Engineering
- 2 Law Single Pathway
- 2 Management Science and Information Systems Studies Single Pathway
- 2 Microbiology
- 2 Theoretical Physics Single Pathway
- 3 Business and Economics
- 3 Global Business
- 5 Computer Engineering
- 6 Computer Science Single Pathway
- 6 Human Health and Disease Single Pathway
- 1 Me
- 1 Eoin
- 1 Dave

Schedule

- All online for week 1
  - But… see survey so we can find out who prefers what...
- Lecture slots, all available via BB collaborate “ultra”, so “live-with-recording”:
  - Tuesday 1000-1050
  - Thursday 1600-1750
- We’ll figure a way to handle “hybrid” slots with some people present in college and others online
  - That may be messy, but let’s see how we go (your ideas/comments welcome too btw!)
- “Office hours”:
  - Monday 1100-1130+ also via BB collaborate “ultra”
  - Lecturer(s) will hang out there/then, any of you welcome to join & chat
  - Might swap that time some weeks depending on conflicts with other schedules
- Reading week: October 25th

Assessments

- 3 Assignments, submit via BB
- GDPR request and anonymised report – 30 %
- “Ethics canvas” – 30 %
- Individual report on device & app tracking – 40 %
- Re-assessment (if needed) will be a couple of other (more difficult and boring) assignments

Module Materials

- There is no book – feel free to recommend some if you like
- Materials will be linked to from, or on, the module web page:
  - https://down.dsg.cs.tcd.ie/witidtm
- Content of module web page is also in Github at:
  - https://github.com/sftcd/witidtm
- Clone that repo and/or visit that page often, as it will change during the runtime of the module!
  - Who knows what “clone that repo” means?
  - I’ll be happy to take PRs, if offered – if **very** good I might even give some marks

Style

- Unless otherwise stated, all lecture slots will be recorded
  - If you don’t consent to that, let us know!
- Would love to see active participation via in-person or online chat and audio
  - Happy to take suggestions to improve things (probably won’t accept all suggestions though;-)
- For some lecture slots, we’ll minimise the presentation to encourage interaction
- Please don’t sit there and say nothing!
- It is entirely ok to ask what might appear to be less-than-clever questions, e.g. “Who makes money from YouTube?” - supposedly naive questions can be good and the answers might be quite subtle
- It is entirely ok to comment on what we tell you, e.g. “That’s nonsense – I use <foo> all the time and it’s fine afaics” - this is about you after all, so (dis)agreeing with us and one another is desirable (but don’t be an ass, and do listen)
- If you don’t comment or ask questions, we’ll all be more bored and I’ll get cranky!

Butting in...(a good thing!)

- If I’m waffling on and you’re in class just yell at me (nicely:-)
- If you’re online have a question or comment then we can try a virtual “mic line”:
  - Type “+q” in the chat room to join the line (easier to spot and order than “raised hands”)
  - It’s ok to e.g. “+q to ask if tcd.ie is a DNS domain?” or “+q to follow up on that point about FB” or “+q to say that I didn’t get that last bit”
- I’ll go to the “mic line” either then or in a bit, we can do the umute dance and chat about the questions one at a time
- If there’s a bunch of people in the “mic line” and your point has been covered already then it’s good to type “-q ...” to get out of the mic line
- It is entirely fine to chat in the chat room all the time – just keep it relevant and considerate
- If (subsets of) the class have their own external chat room(s)... I don’t care:-)

So, with administrivia out of the way,
let’s start...

Here’re some questions we’ll come back to at the end of the slide-deck (today or next day, whenever) but please start to ponder them...

I think the Internet is great
(and have for 30+ years)
but
are the 3-4 billion people connected
all your friend?

How do you interact with the Internet?

Do you care about your, my, or all of our, security on the Internet?

If so, what do you care about most?

Do you care about your, my, or all of our privacy?

That’s all for now, we’ll be returning to those questions as we go.

But… what other questions should we be considering?

What else?

- Your topic here... what’d you like to cover?
-
-
-

(we can revisit this multiple times)

Meanwhile… let’s start with…

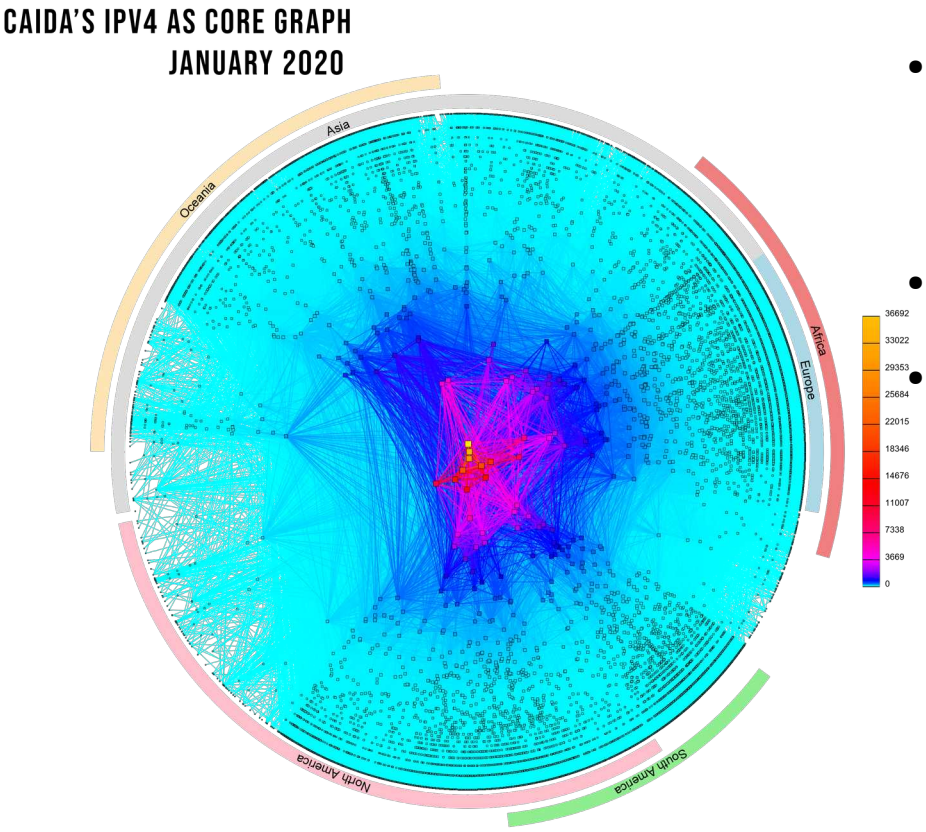
Is the Internet a network?

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

A network of networks

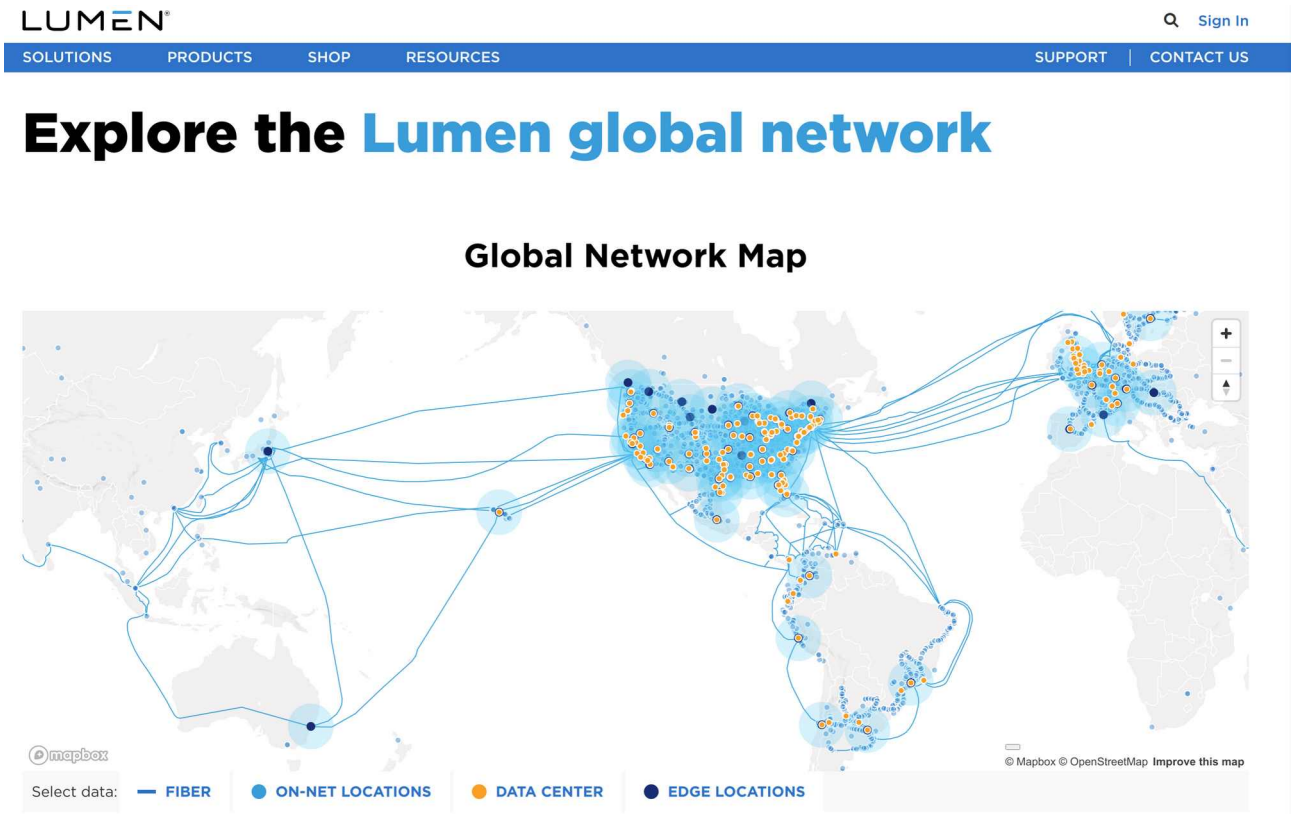
- The Internet is made up of tens of thousands of Autonomous Systems (ASes)
  - https://en.wikipedia.org/wiki/Autonomous_system_%28Internet%29
  - 72173 ASes as of 20210907, (https://www.cidr-report.org/as2.0/)
    - Was: 65,428 in Aug 2019
- Think of these as the set of Internet Service Providers (ISPs, like Eircom, Vodafone, Virgin), other networks (e.g. HEANET which is TCD's "ISP"), big companies (e.g. Google, FB) 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)
  - IP: Internet Protocol; TCP: Transmission Control Protocol; BGP: Border Gateway Protocol
- We'll delve more into all that later, but first… some pretty pictures

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 and cogent

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



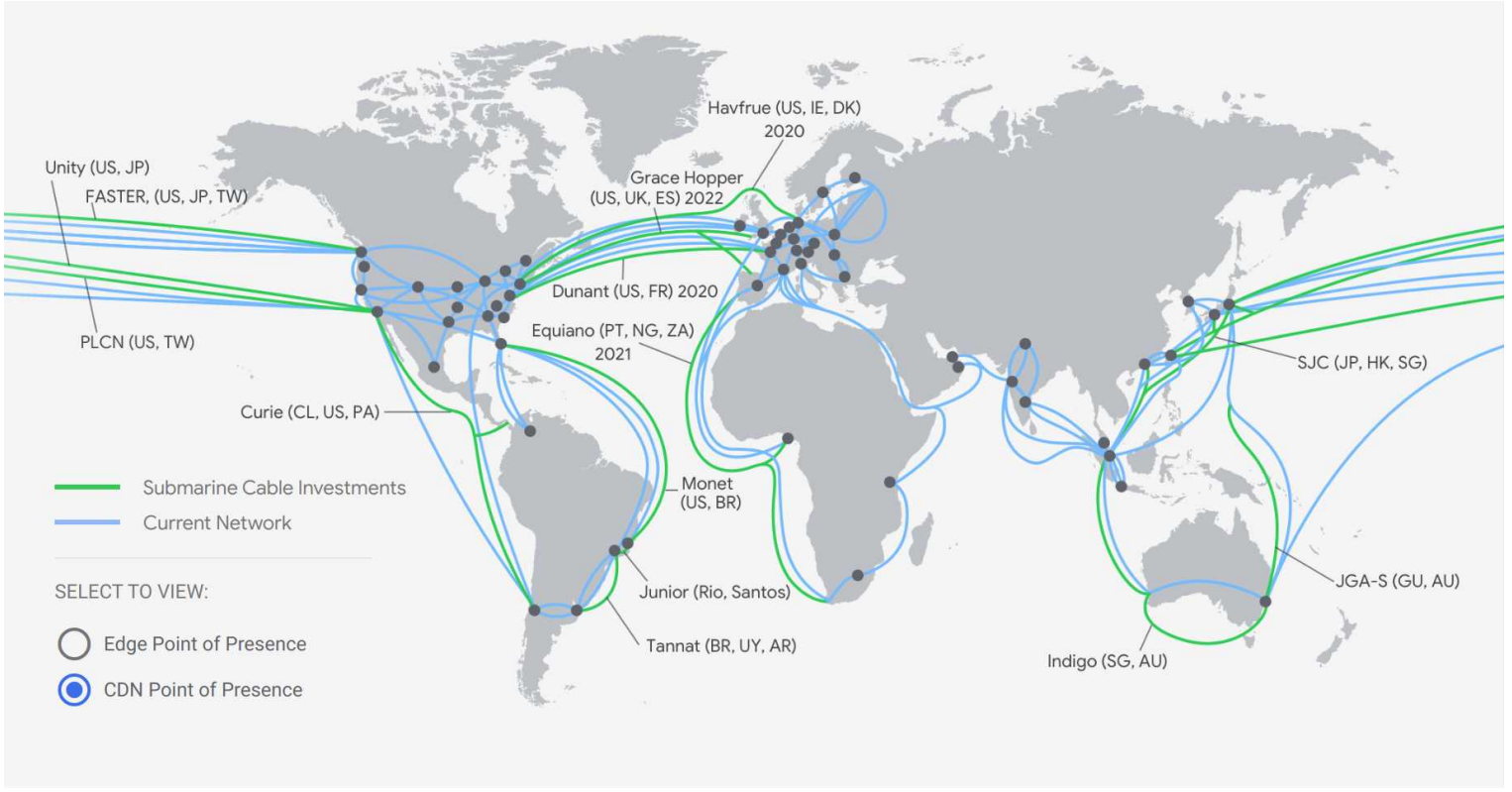
https://www.lumen.com/en-us/resources/network-maps.html

Cogent similarly

Our primary service offering consists of Internet access and data transport offered over our award-winning fiber optic, IP data-only network, along with colocation in any of our **54 Internet data centers**. We service two customer segments: "**corporate**" (small businesses to Fortune 100 companies) and "netcentric" (**carriers** / service providers and application / content providers, whose businesses rely primarily on Internet access). Our innovative, facilities-based network spans six continents: North America, South America, Europe, Asia, Australia and Africa. With over **59,700 route miles of intercity fiber** and more than **38,300 metro fiber miles**, we provide service to over **211** major markets and interconnect with over **7,530 other networks**.

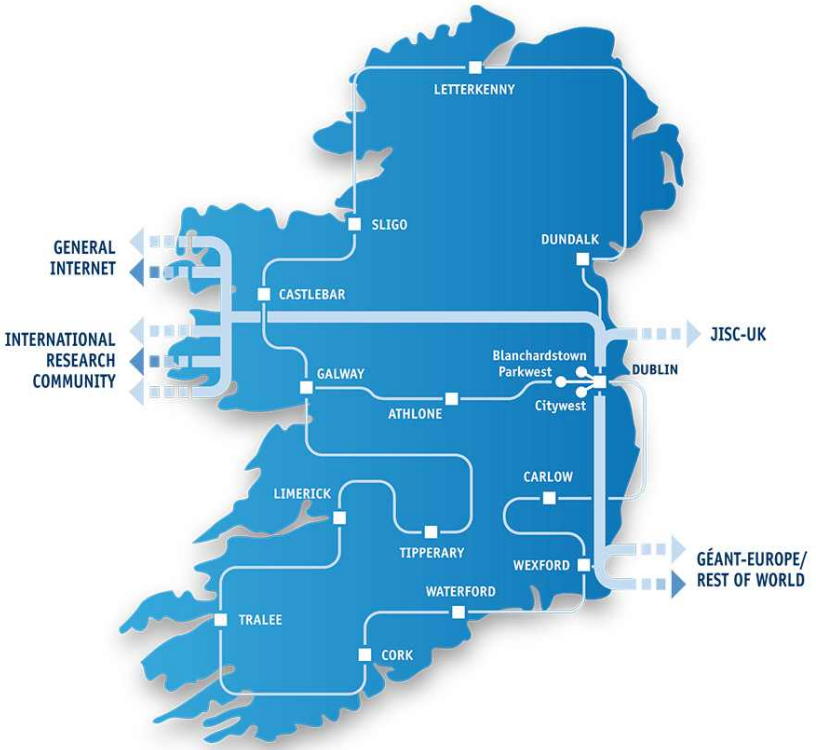
https://www.cogentco.com/en/network/network-map

Google cloud



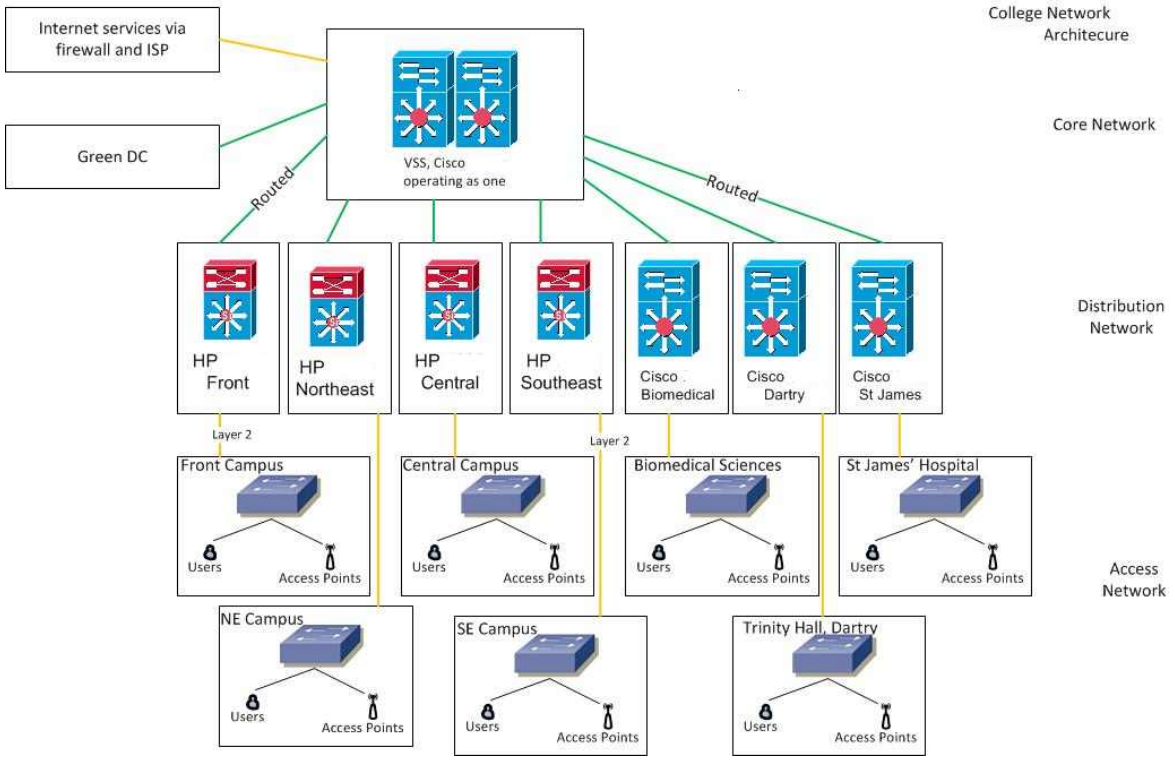
https://cloud.google.com/about/locations/#network note this is just google cloud, not all their stuff

Heanet national n/w



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

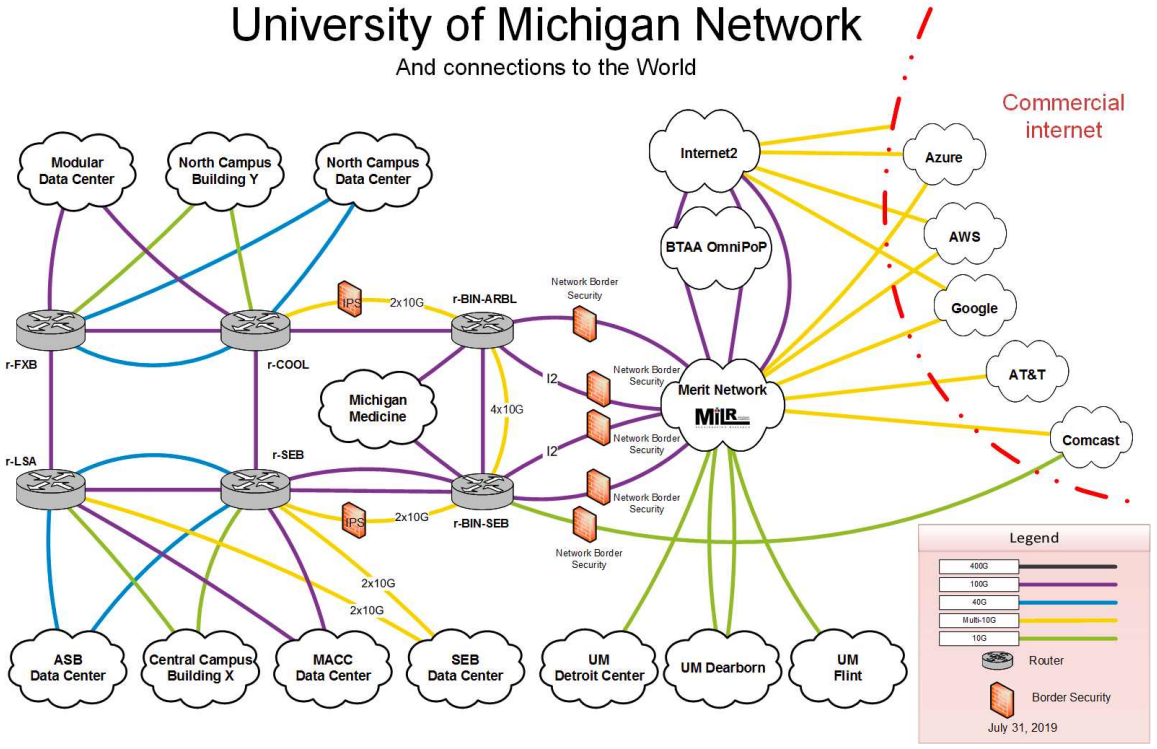
The TCD network (circa 2019)



- 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

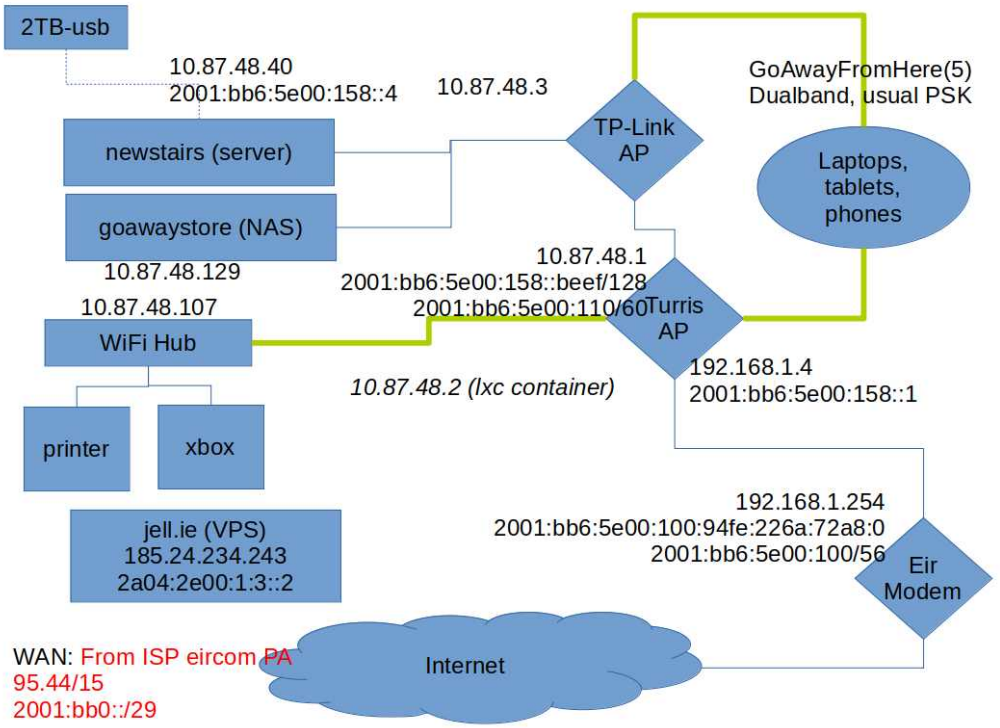
Thanks to TCD IT for providing this anonymised diagram – doing that's not a no-brainer btw!

A 2nd campus example



https://its.umich.edu/enterprise/wifi-networks/campus-network-diagram-description

My home network

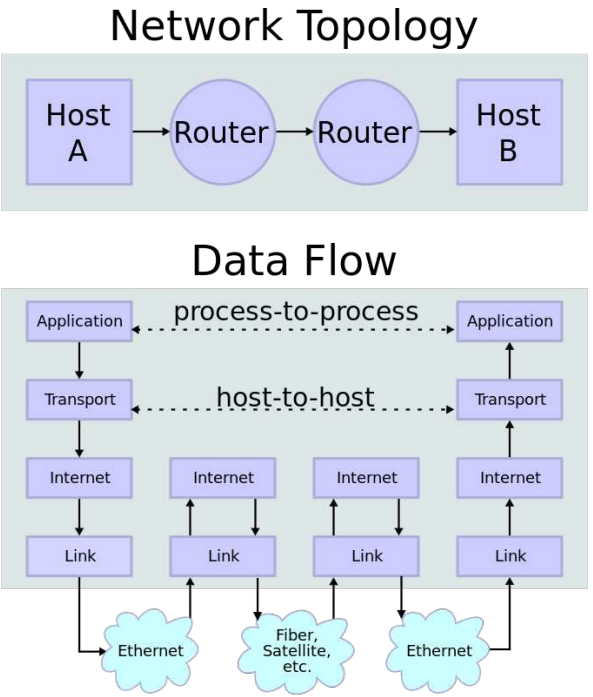


That's from a few years ago – so is somewhat out of date:-)

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)
    - I'm quite involved with IETF stuff, so consider me biased there:-)
- But we do not aim to agree about everything in everyone's network
  - So 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
- Related: the classic “End-to-end argument” paper
  - Salzer, Reed. Clark,“End-to-end arguments in system design” ACM ToCS (1984).
    - https://web.mit.edu/Saltzer/www/publications/endtoend/endtoend.pdf
  - I recommend a read of that!
  - Don’t consider it as gospel though – it’s the end-to-end **argument** and not really the end-to-end **principle** even though it gets called the latter a lot

“Tussles”

- Repeating: we do not aim to agree about everything in everyone’s network...
  - So an awful lot happens at the “application layer” in code written by people and organisations
  - And yet more happens when people configure services that use that code
- 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;-)
- Another paper:
  - Clark, David D., et al. "Tussle in cyberspace: defining tomorrow's Internet." ACM SIGCOMM Computer Communication Review. Vol. 32. No. 4. ACM, 2002.
    - https://www2.cs.duke.edu/courses/compsci514/cps214/spring09/papers/p347-clark.pdf
  - Same “Clark”, but older:-) Interestingly, the 2002 paper is IMO far more dated (and wrong!) than the 1984 paper!
- We’ll consider a “topical” example later (Search for “DNS over HTTPS” if you want to check it out); don’t worry if this issue seems a bit too hard to grok at this point

Aside: Cyber<blah>

- Be wary of anyone who uses a term like “Cyber<blah>”
- ~90% of the time, that’s a strong indicator that they don’t really know what they’re talking about (if they did, they’d use a more precise and well-defined term)
- Sadly, about 10% of the time (and increasing) such terms are used because “industry” keeps on doing it and people just repeat stuff thoughtlessly
- Don’t be afraid to ask someone to define “Cyber<blah>” if they use such a term, and don’t be surprised if they find that hard!
  - E.g. “Does cyberspace include a person driving a car that’s had it’s license plate automatically scanned?” or “What’s not included in cybersecurity?”

The Internet is not the web

- Another important point!
- 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
- But lots of our interactions with the Internet are via the web

Some of the things we'll do later...

- Understand what happens under the hood when your browser loads a web site
- Get an overview of advertising networks
- Talk a bit about spam and phishing (email badness)
- Learn how to watch network traffic
- We'll look more at the web soon, meanwhile back to you...

How do you interact with the Internet?

Do you care about your, my, or all of our, security on the Internet?

If so, what do you care about most?

Do you care about your, my, or all of our privacy?

That’s all for now, we’ll be returning to those questions as we go.

But… what other questions should we be considering?

What else?

- List from 20190912 lecture:
  - Use of personal data in politics
  - Banking
  - HOWTO do coding
  - Encryption
  - U18 and their rights
  - Cryptocurrency BTC
  - Spyware
-

  (we can revisit this multiple times)

What else?

- List from 20201001 lecture:
  - From earlier:
    - COVID-19 tracing apps
    - Alexa etc (smart speakers)
  - Who has power
  - Censorship
  - Schrems
  - Voting
  - Cloud storage
-

  (we can revisit this multiple times)