

The internet is fine
I'm on Facebook
right now!



A discussion on how networking in support of data intensive research is not at all the same as networking for general use.



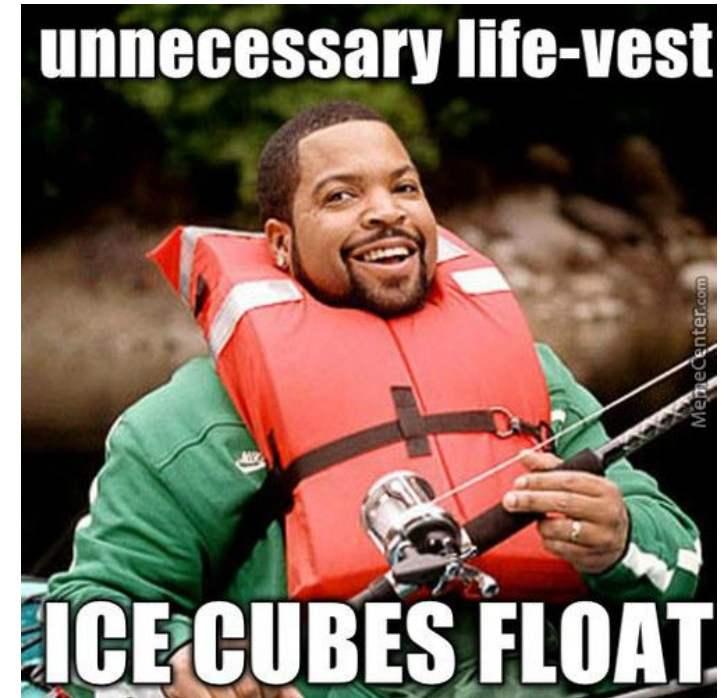
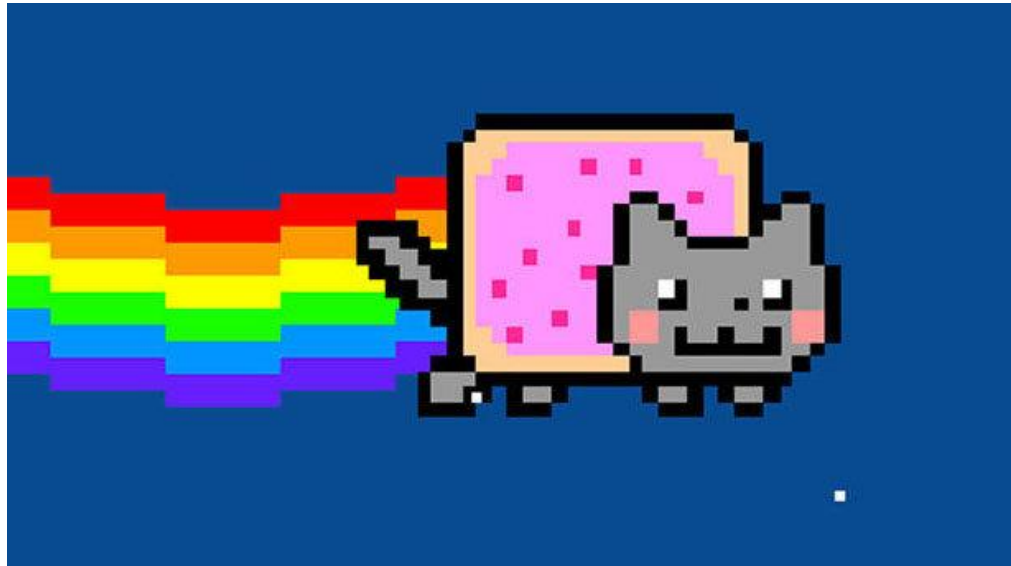
Try Not To Laugh Animals | Funniest Cat Videos In The World | Funny Animal Videos #128

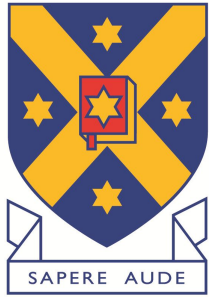
370K views • 3 days ago



Try Not To Laugh Animals | Funniest Cat Videos In The World | Funny Animal Videos #128, Funny Dogs, Cute Pets, Funniest ...

New





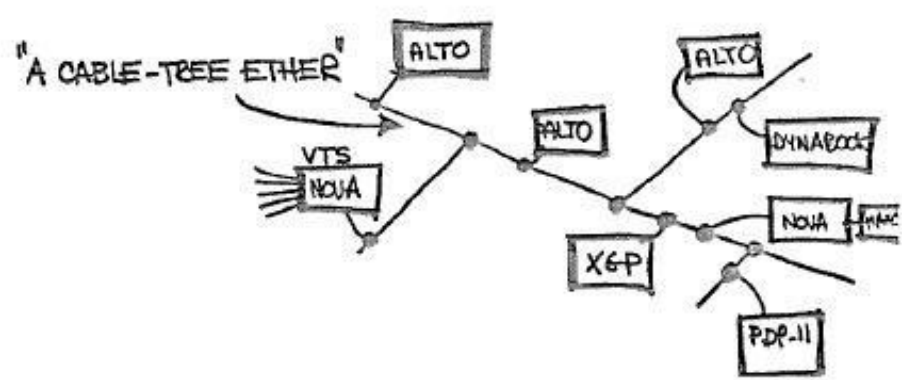
UNIVERSITY
of
OTAGO

Te Whare Wānanga o Otāgo
NEW ZEALAND

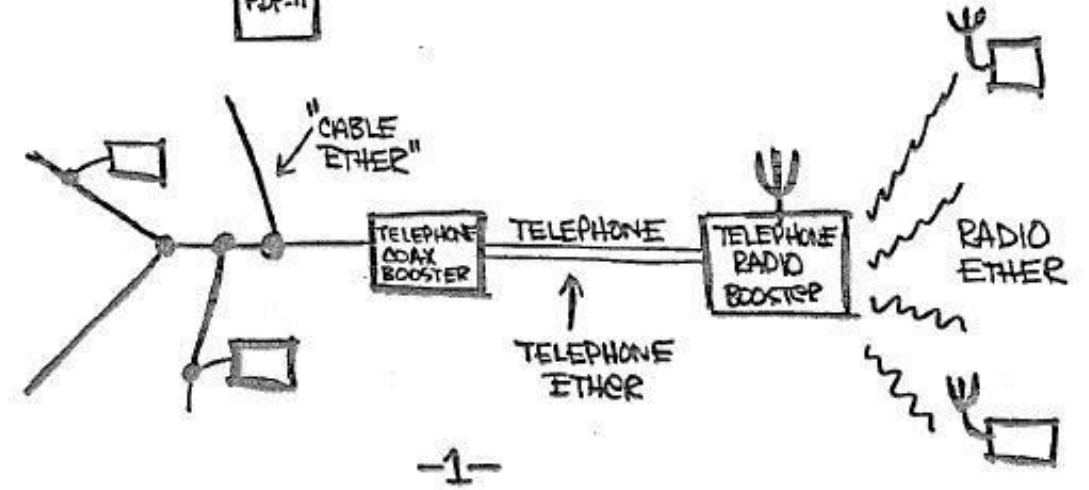
Wallace A. Chase

Head of Department, ITS
wallace.chase@otago.ac.nz

@bmtfr



ETHER!



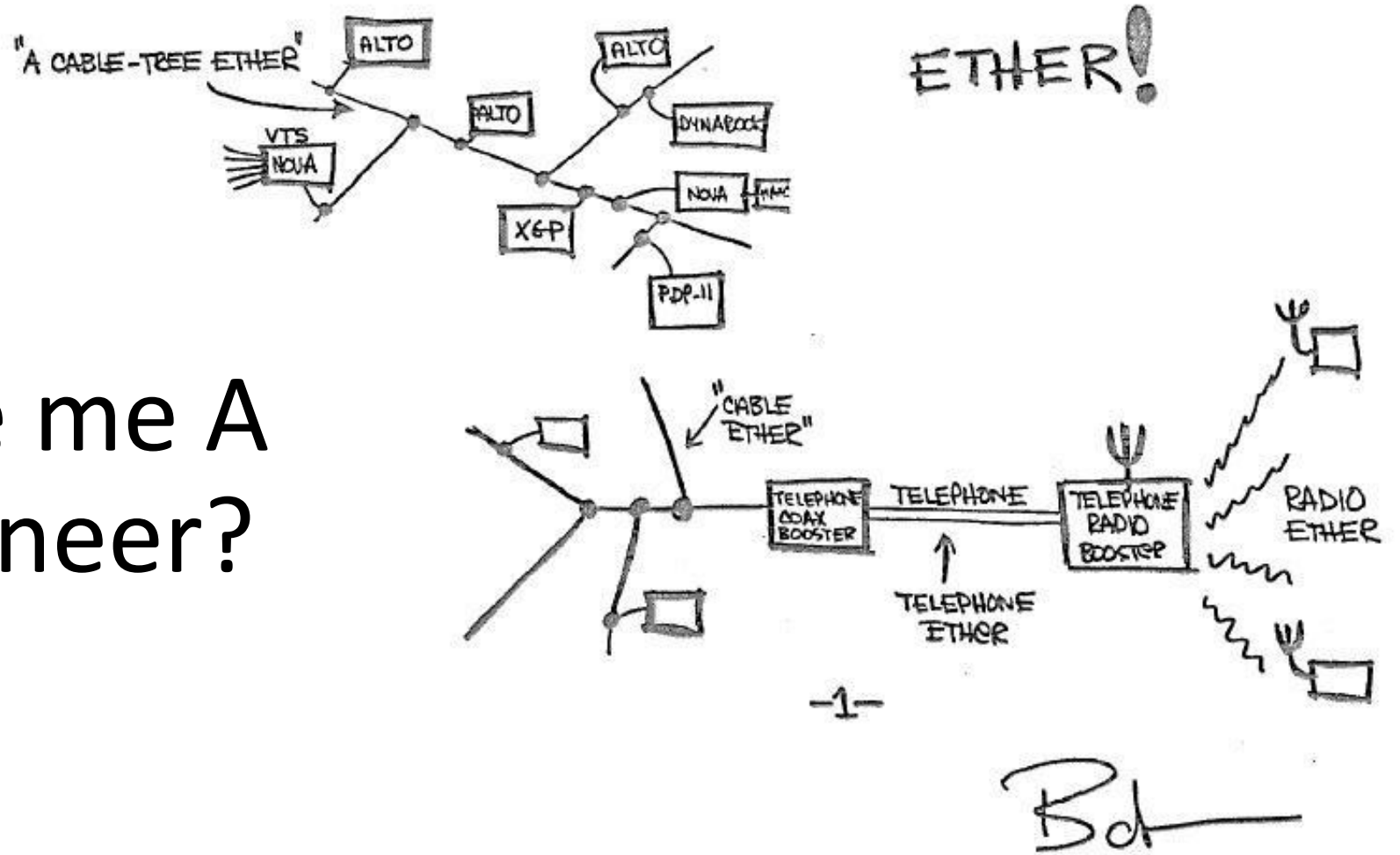
Bch



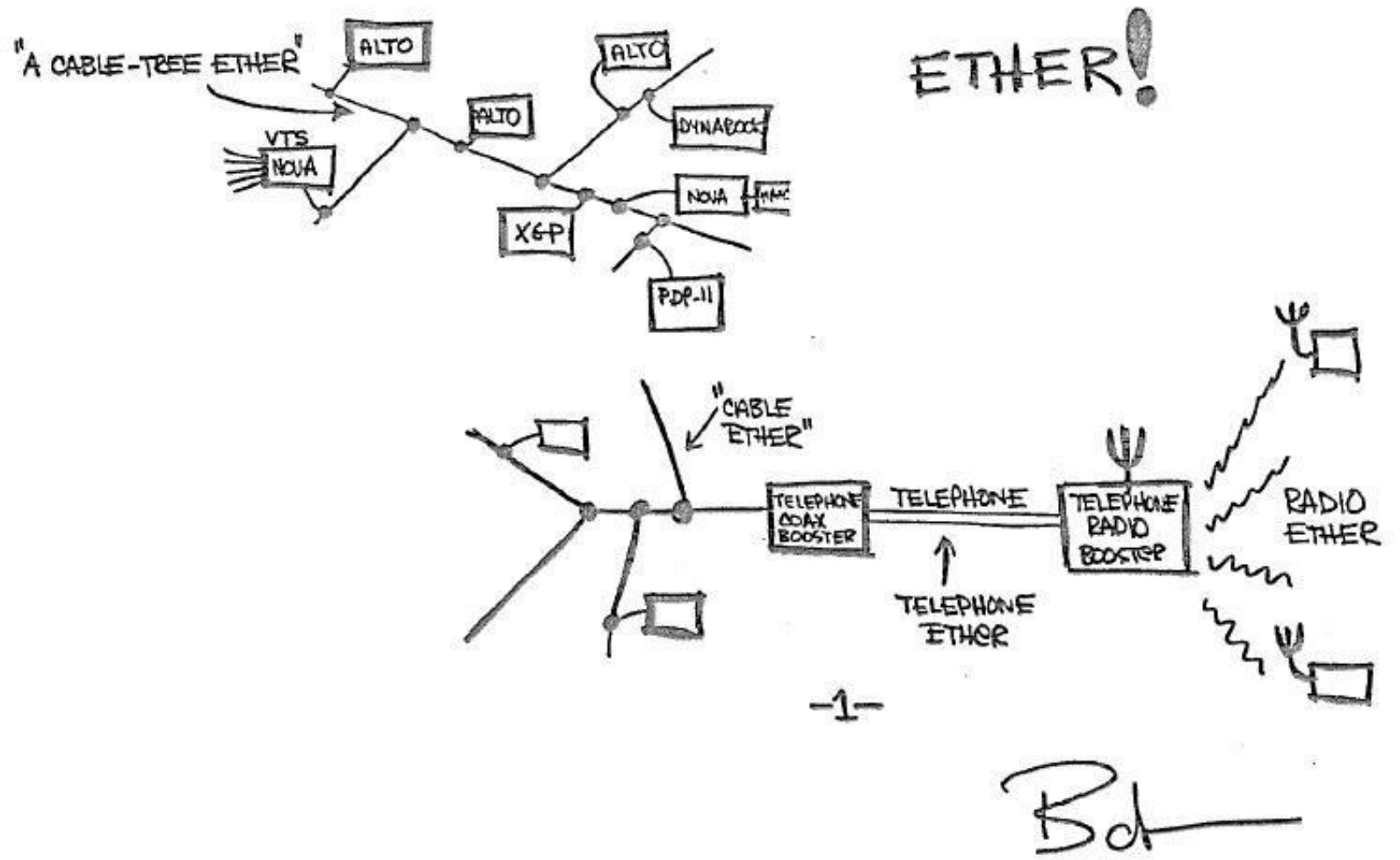
Lets learn some networking!



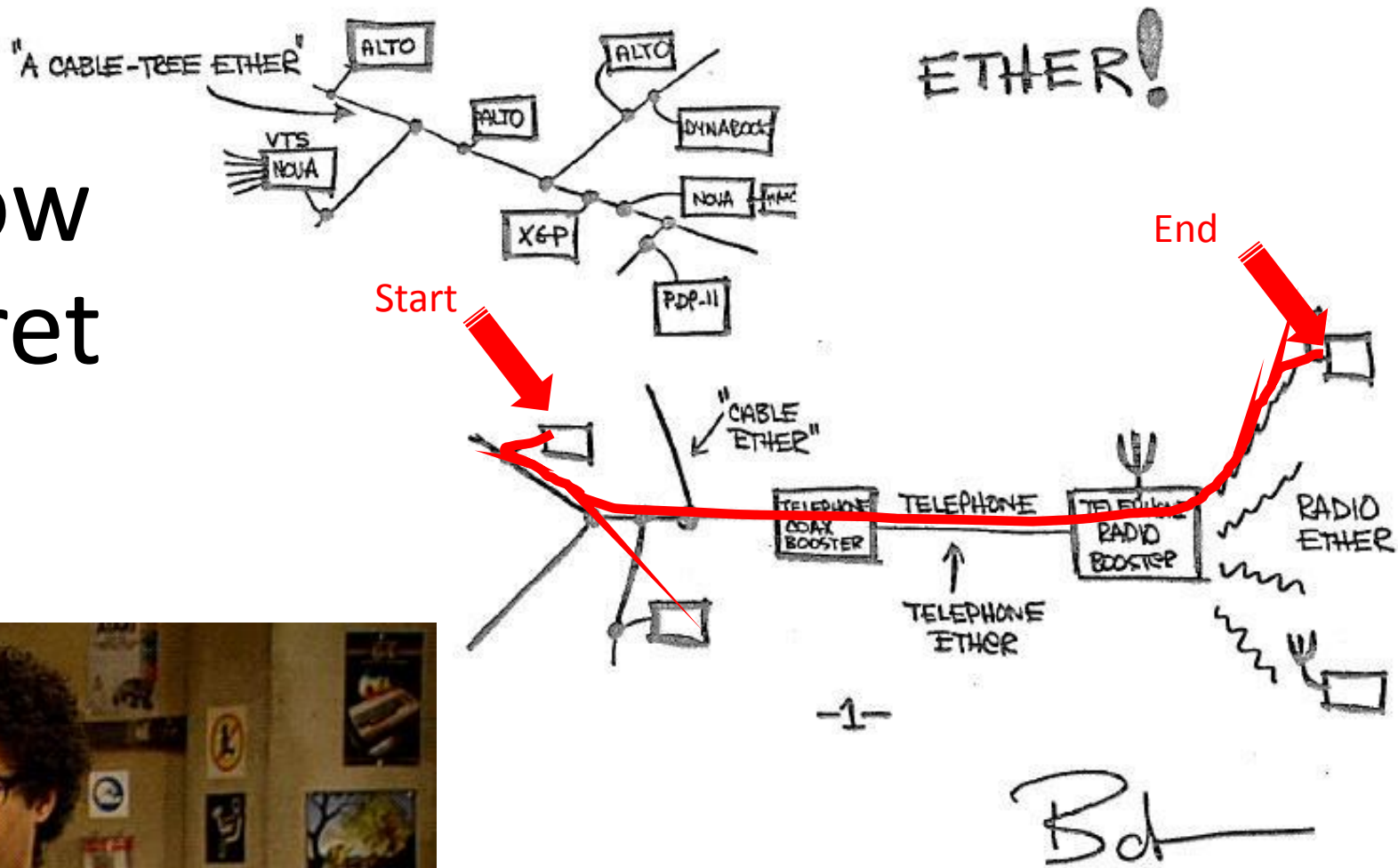
Will this make me A Network Engineer?



No.



But it will allow you to interpret their world.

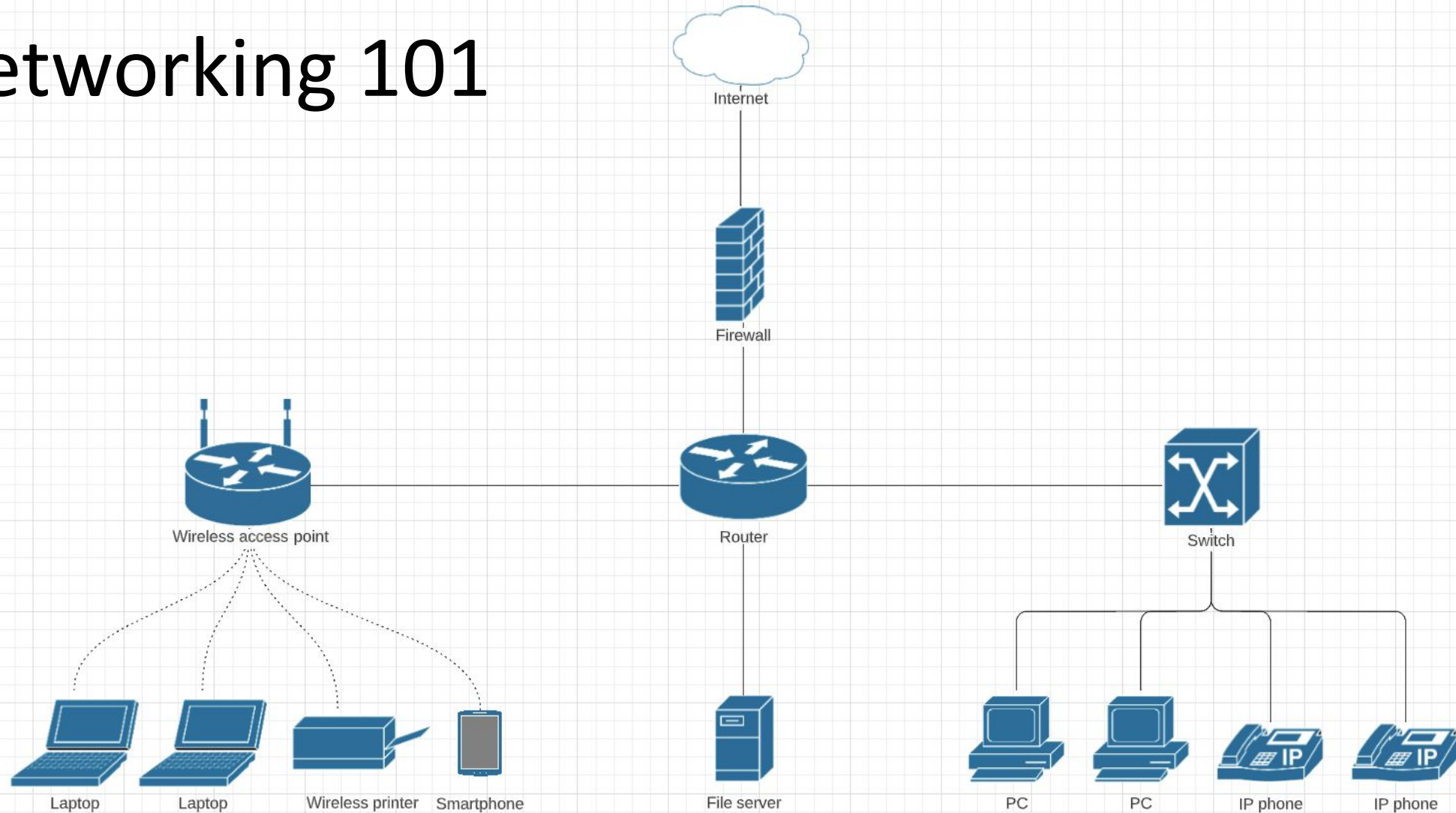


I'll just put this over here with the rest of the fire

Networking 101

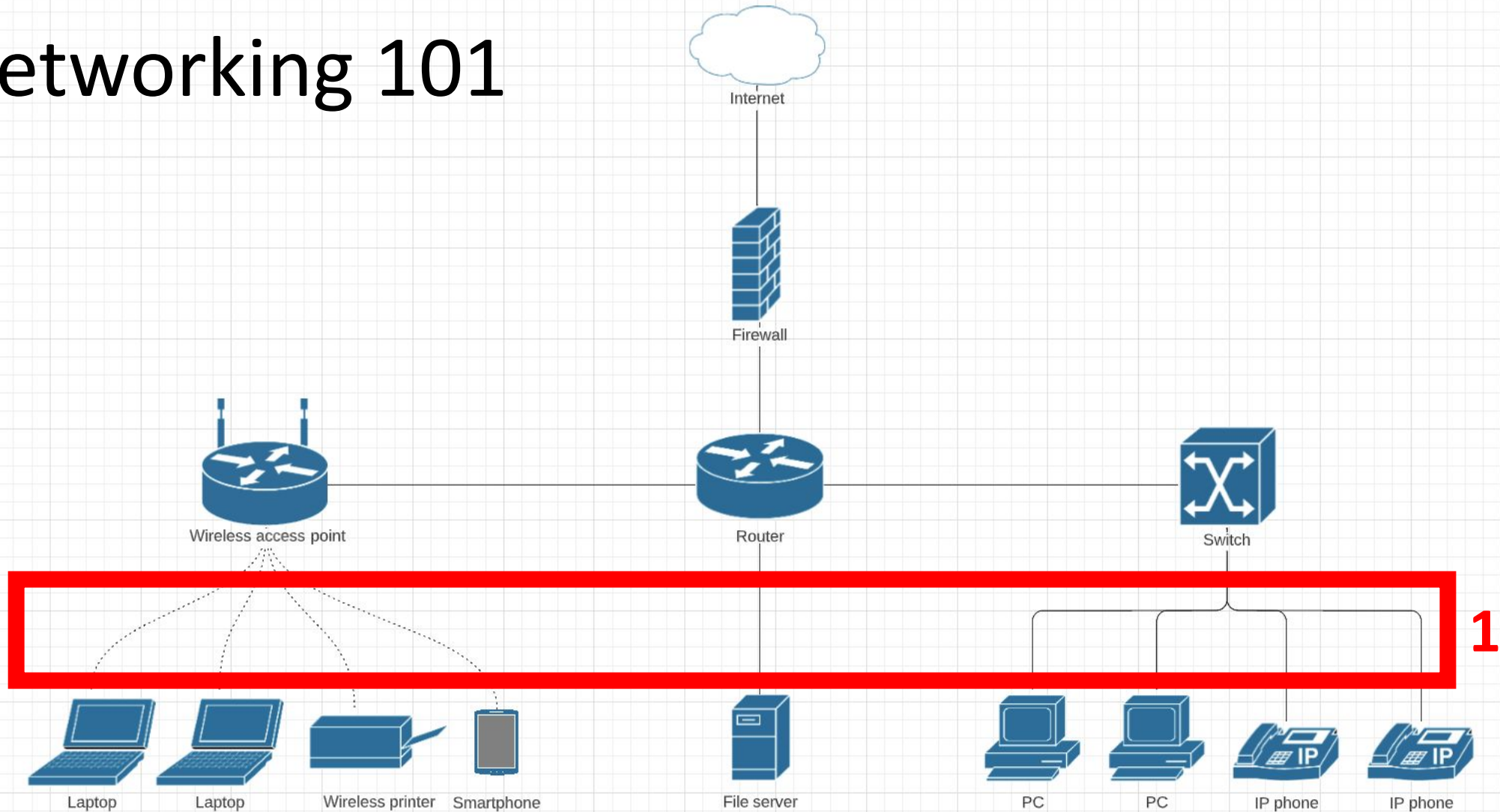
Layer		Protocol data unit (PDU)	Function ^[23]	
Host layers	7	Application	Data	High-level APIs, including resource sharing, remote file access
	6	Presentation		Translation of data between a networking service and an application; including character encoding, data compression and encryption/decryption
	5	Session		Managing communication sessions, i.e., continuous exchange of information in the form of multiple back-and-forth transmissions between two nodes
	4	Transport	Segment, Datagram	Reliable transmission of data segments between points on a network, including segmentation, acknowledgement and multiplexing
Media layers	3	Network	Packet	Structuring and managing a multi-node network, including addressing, routing and traffic control
	2	Data link	Frame	Transmission of data frames between two nodes connected by a physical layer
	1	Physical	Bit, Symbol	Transmission and reception of raw bit streams over a physical medium

Networking 101



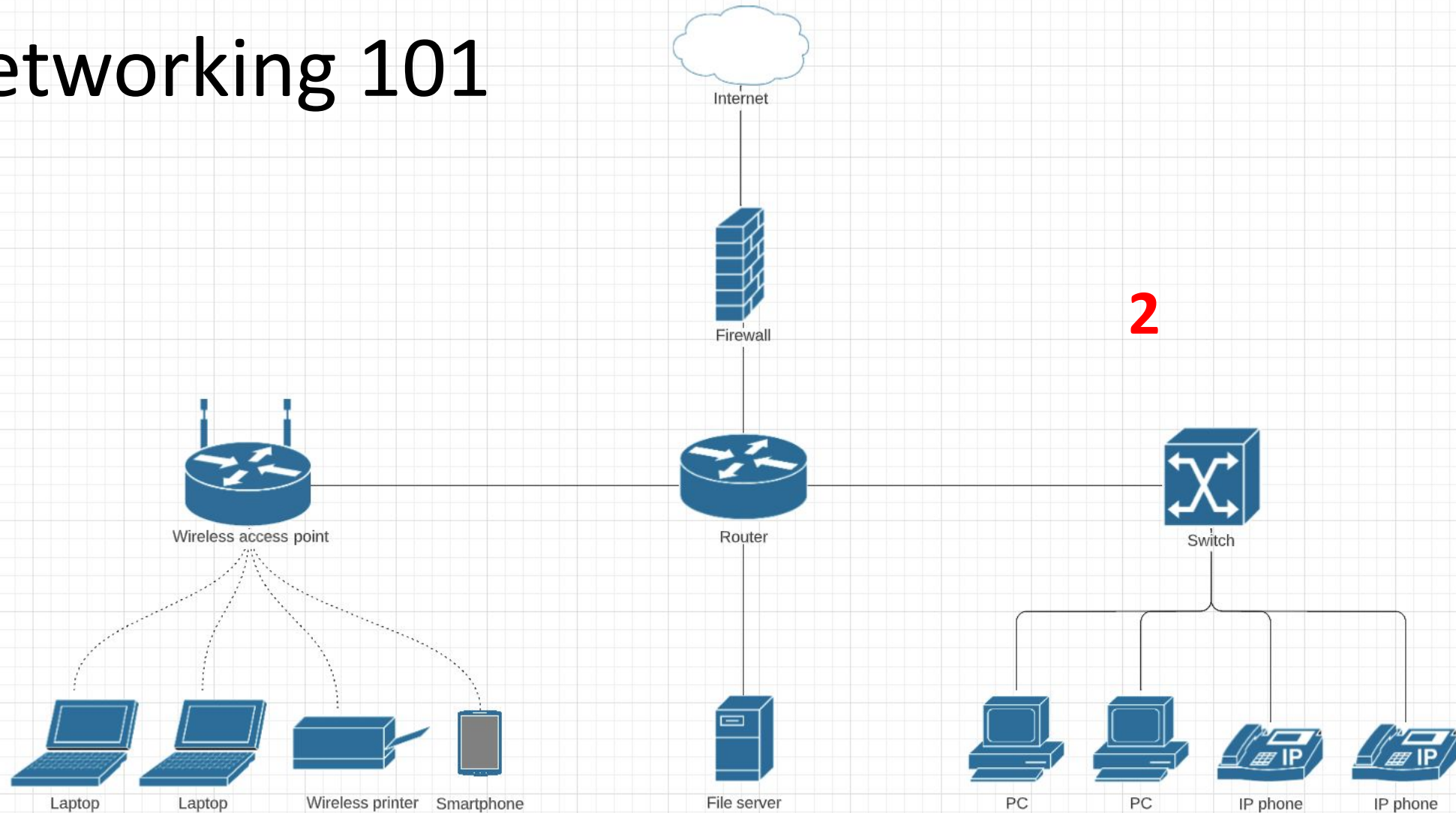
Media layers	3	Network	Packet	Structuring and managing a multi-node network, including addressing , routing and traffic control
	2	Data link	Frame	Transmission of data frames between two nodes connected by a physical layer
	1	Physical	Bit, Symbol	Transmission and reception of raw bit streams over a physical medium

Networking 101



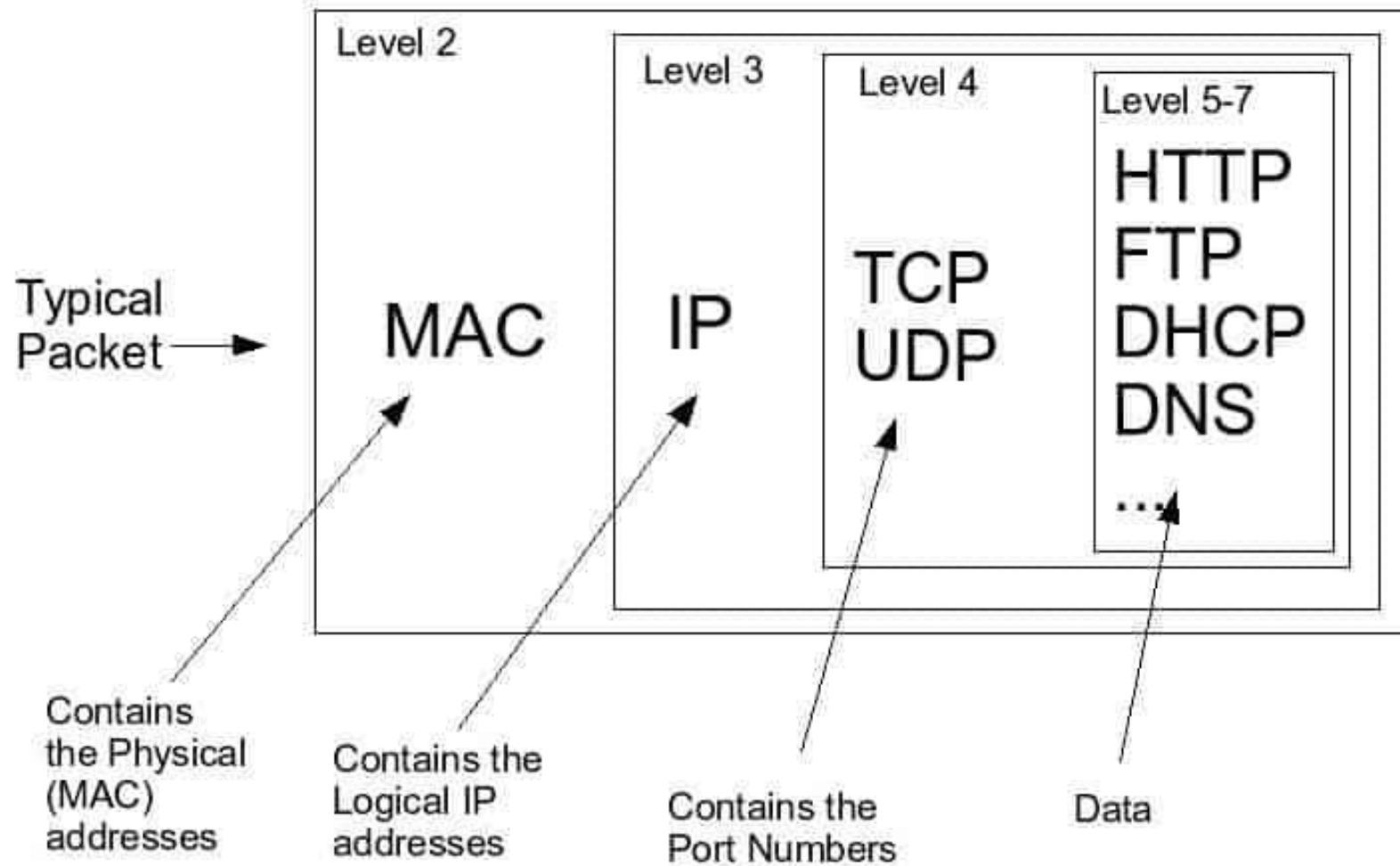
Media layers	3	2	1
	Network	Data link	Physical
	Packet	Frame	Bit, Symbol
	Structuring and managing a multi-node network, including addressing, routing and traffic control	Transmission of data frames between two nodes connected by a physical layer	Transmission and reception of raw bit streams over a physical medium

Networking 101



Media layers	3	Network	Packet	Structuring and managing a multi-node network, including addressing , routing and traffic control
	2	Data link	Frame	Transmission of data frames between two nodes connected by a physical layer
	1	Physical	Bit, Symbol	Transmission and reception of raw bit streams over a physical medium

Networking 101



IP: 192.168.1.8
MAC: 00-07-E9-A6-F2-53

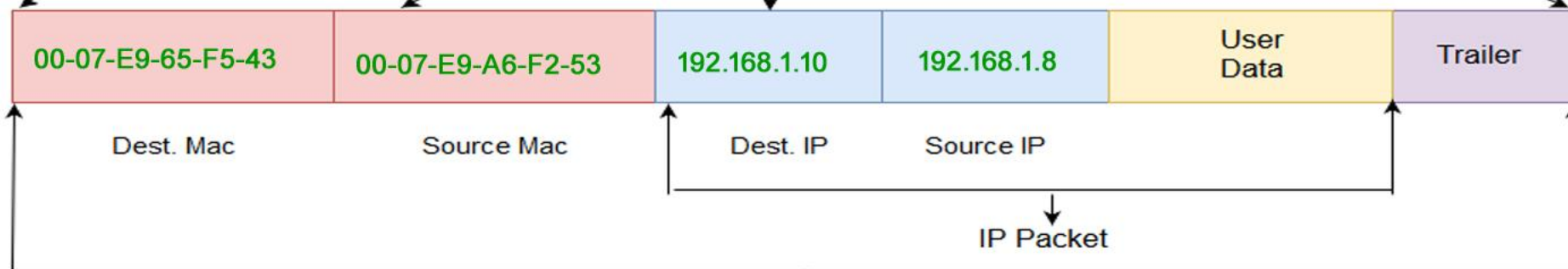


Host



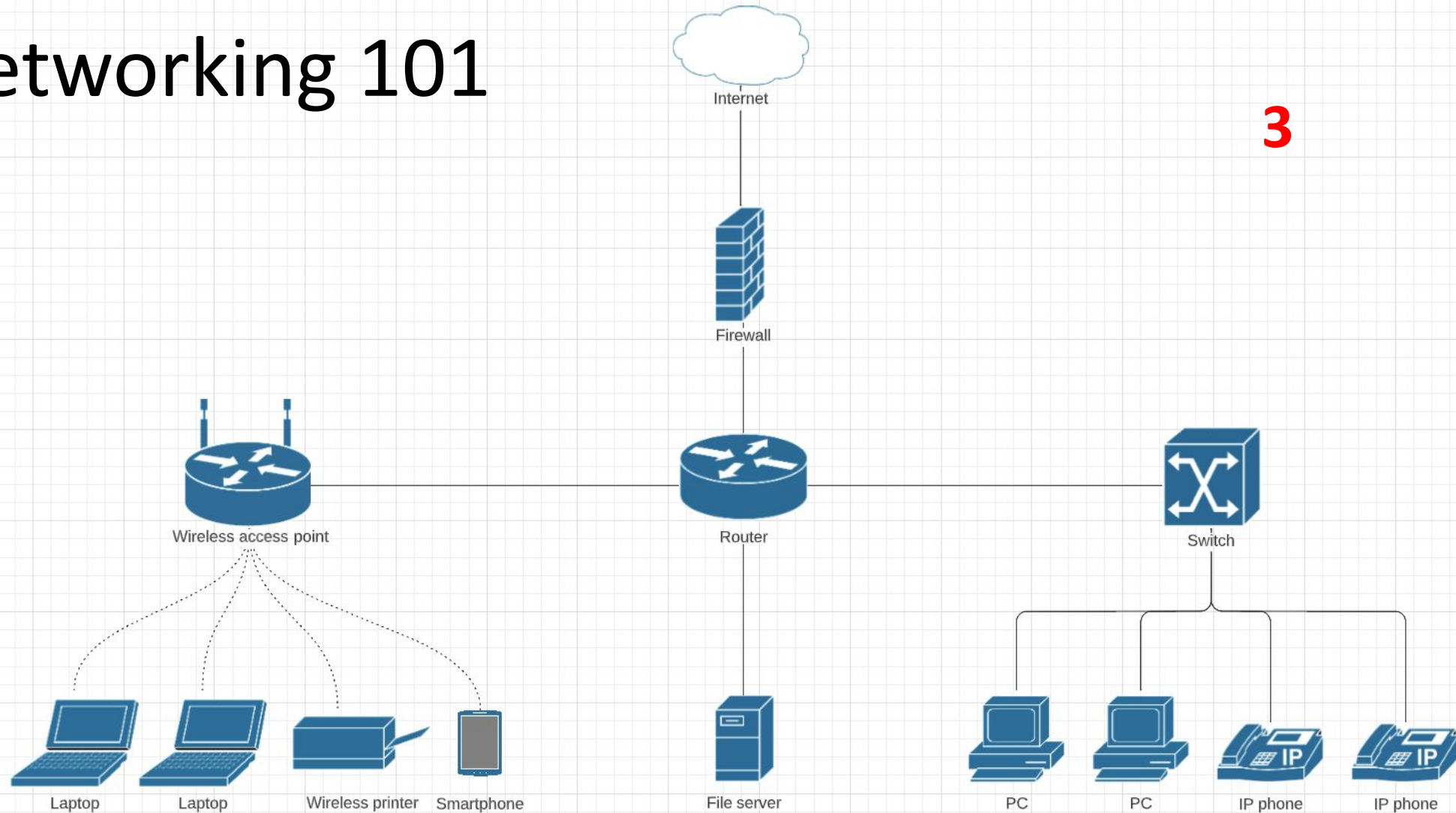
Server:
IP: 192.168.1.10
MAC: 00-07-E9-65-F5-43

Unicast IP and MAC destination addresses are used by the source to forward a packet.



Ethernet Frame

Networking 101

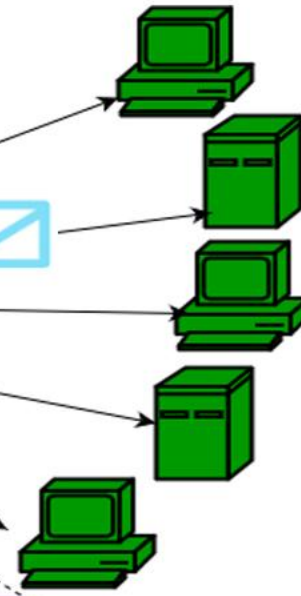
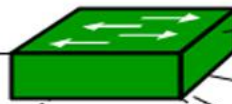


Media layers	3	Network	Packet	Structuring and managing a multi-node network, including addressing , routing and traffic control
	2	Data link	Frame	Transmission of data frames between two nodes connected by a physical layer
	1	Physical	Bit, Symbol	Transmission and reception of raw bit streams over a physical medium

IP: 192.168.1.8
MAC: 00-07-E9-A6-F2-53

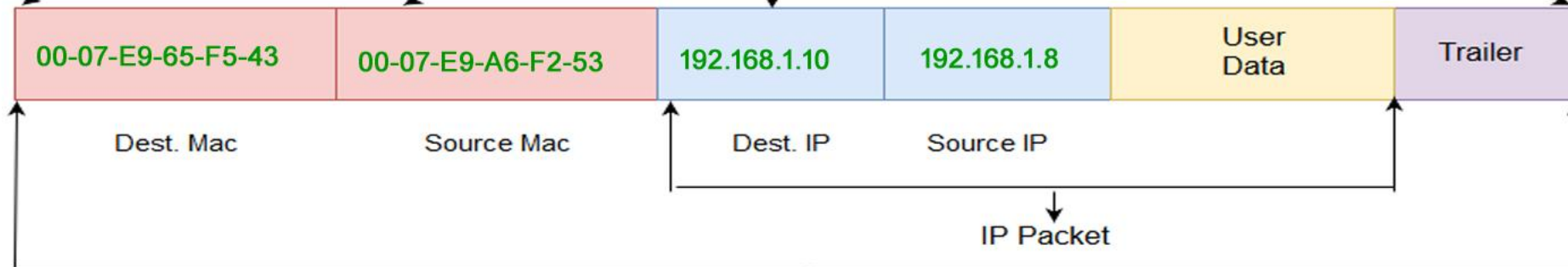


Host



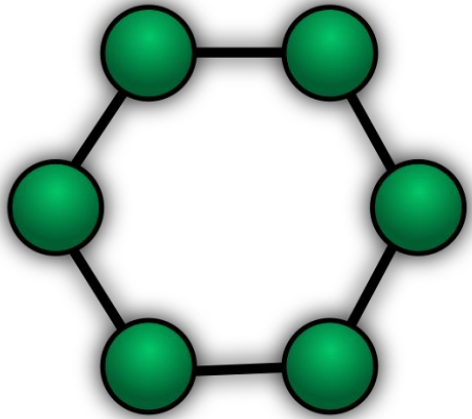
Server:
IP: 192.168.1.10
MAC: 00-07-E9-65-F5-43

Unicast IP and MAC destination addresses are used by the source to forward a packet.

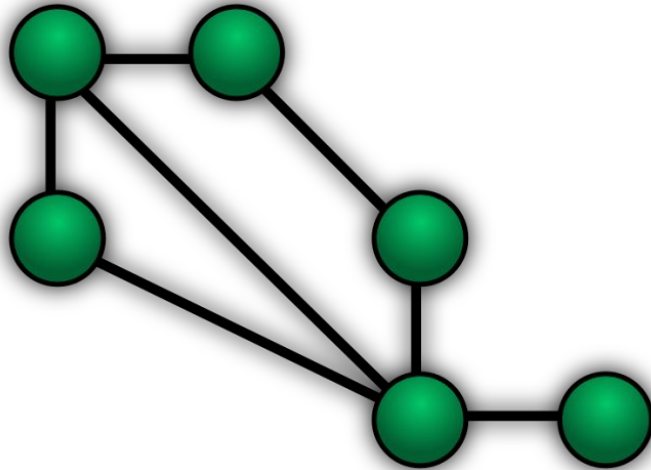


Ethernet Frame

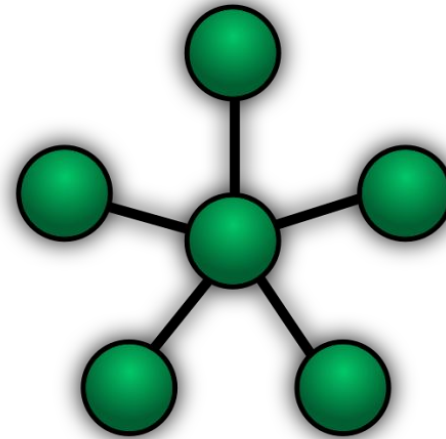
Networking 101



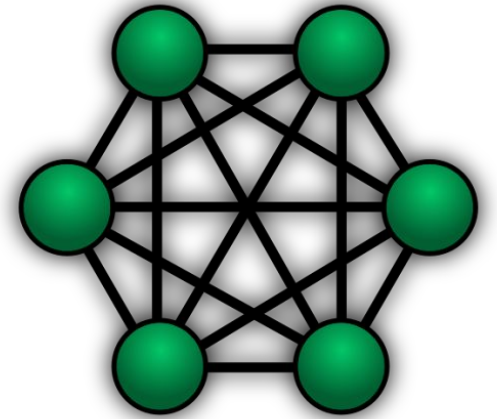
Ring



Mesh



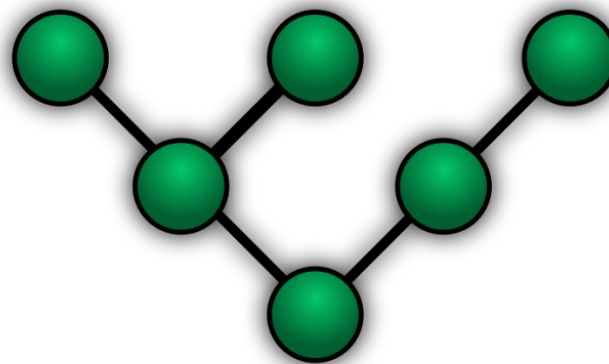
Star



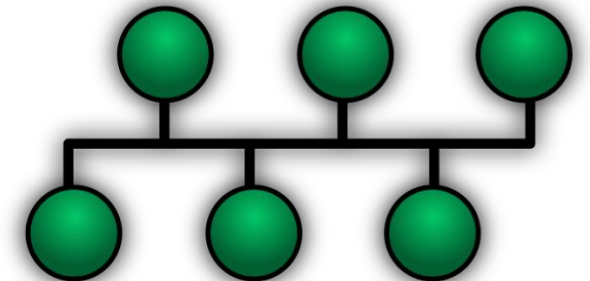
Fully Connected



Line



Tree



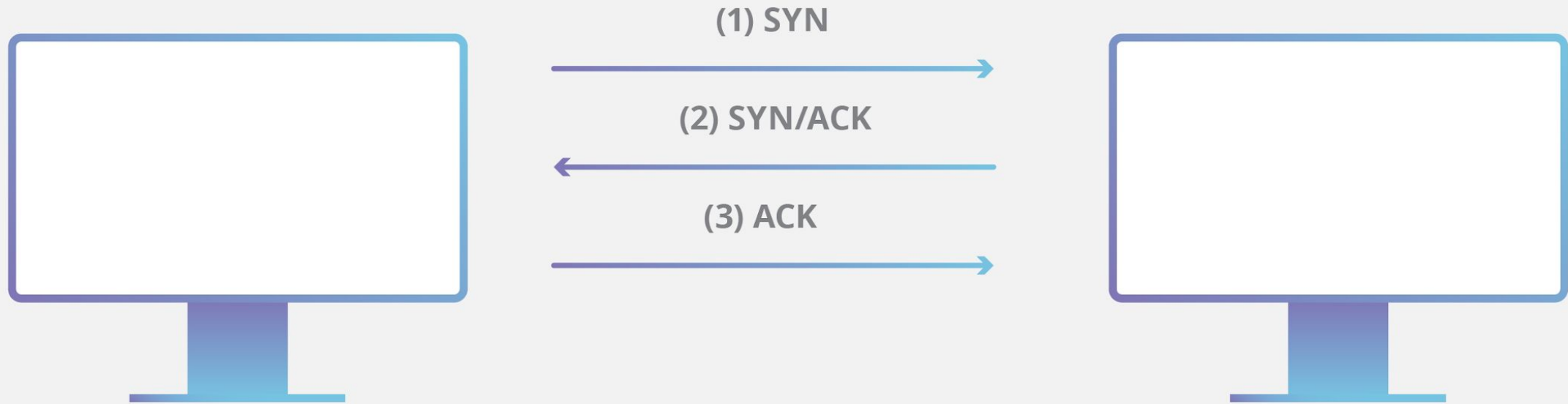
Bus

Networking 101



Networking 101

THREE - WAY HANDSHAKE (TCP)



SYN = SYNCHRONIZATION

ACK = ACKNOWLEDGEMENT

Networking 101

What are IP & TCP?

The Internet Protocol (IP) is the address system of the Internet and has the core function of delivering packets of information from a source device to a target device. IP is the primary way in which network connections are made, and it establishes the basis of the Internet.

IP does not handle packet ordering or error checking. Such functionality requires another protocol, typically TCP.

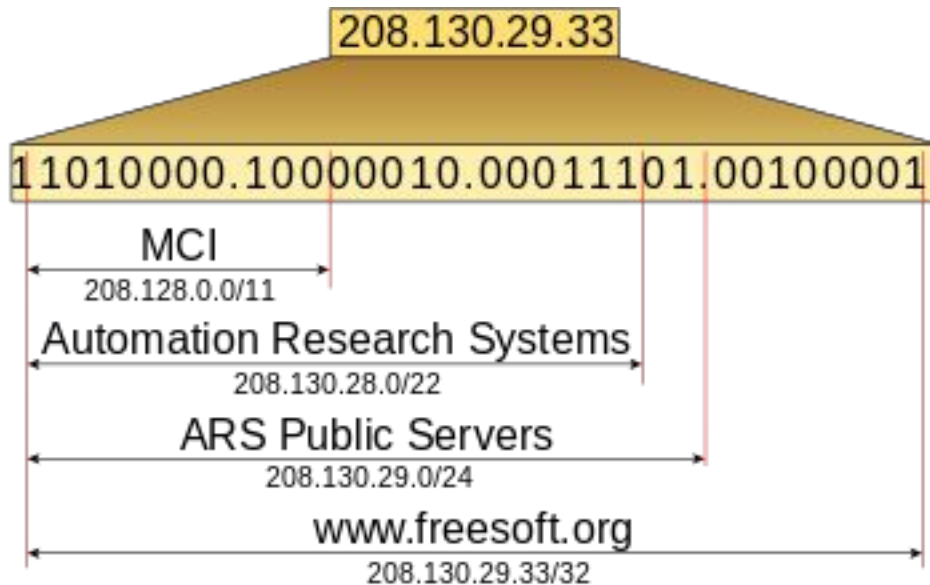
IPv4 address in dotted-decimal notation

172 . 16 . 254 . 1

↓ ↓ ↓ ↓
10101100.00010000.11111110.00000001

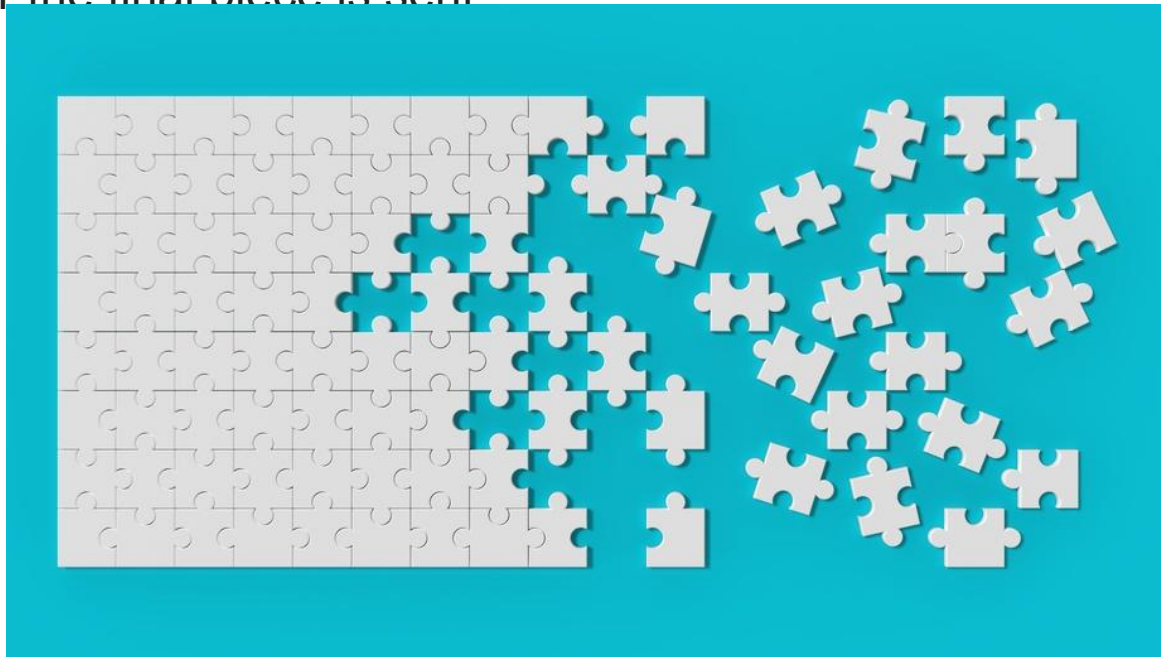
8 bits

32 bits (4 bytes)

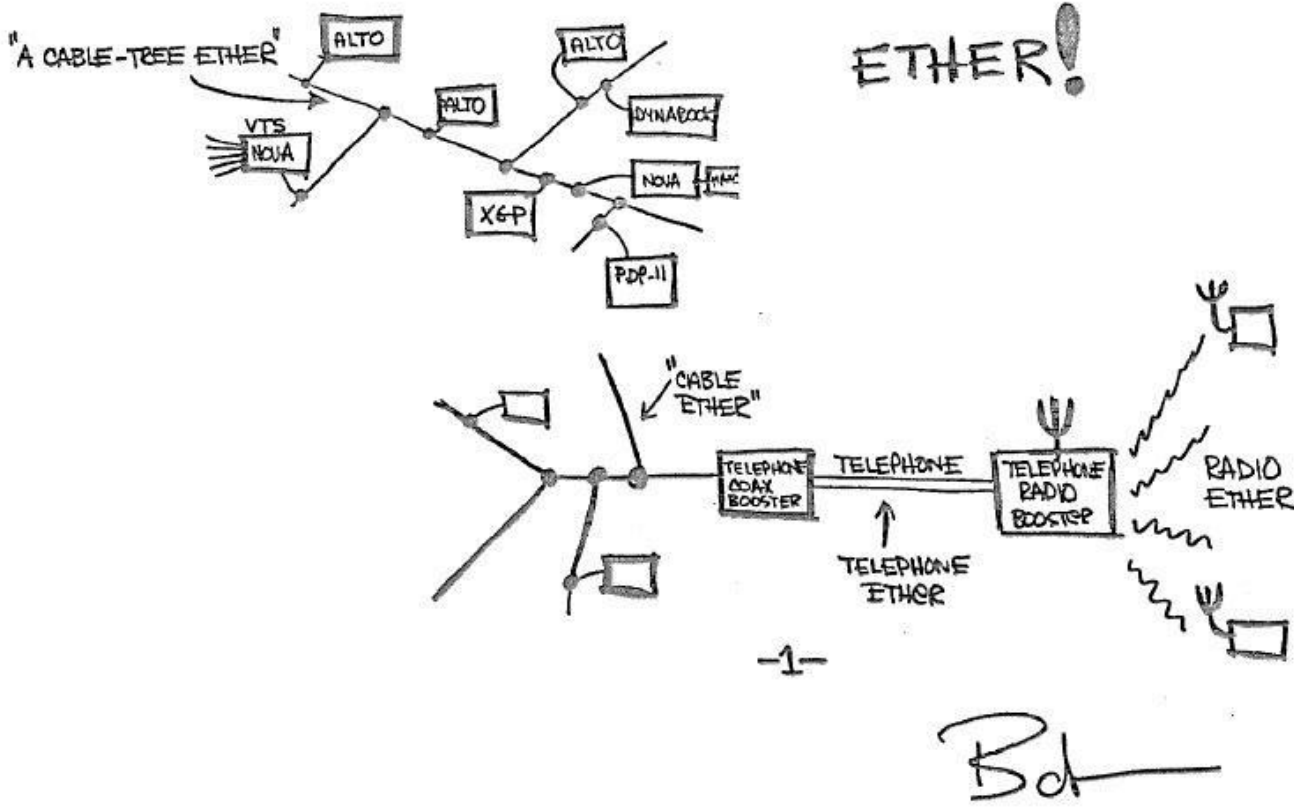


Networking 101

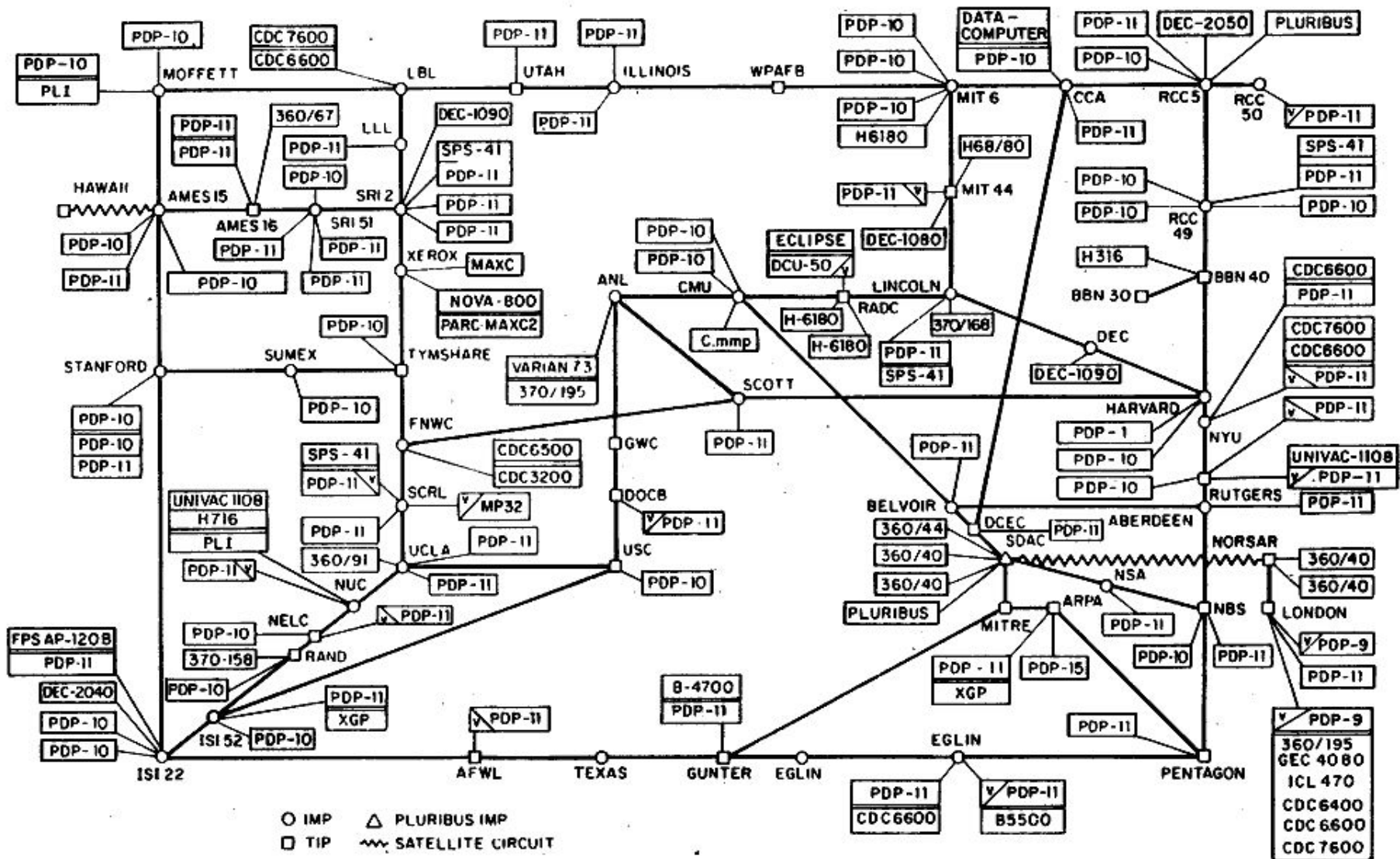
The TCP/IP relationship is similar to sending someone a message written on a puzzle through the mail. The message is written down and the puzzle is broken into pieces. Each piece then can travel through a different postal route, some of which take longer than others. When the puzzle pieces arrive after traversing their different paths, the pieces may be out of order. The Internet Protocol makes sure the pieces arrive at their destination address. The TCP protocol can be thought of as the puzzle assembler on the other side who puts the pieces together in the right order, asks for missing pieces to be resent, and lets the sender know the puzzle has been received. TCP maintains the connection with the sender from before the first puzzle piece is sent to after the final piece is sent



A history lesson – how did we get here?



ARPANET LOGICAL MAP, MARCH 1977



(PLEASE NOTE THAT WHILE THIS MAP SHOWS THE HOST POPULATION OF THE NETWORK ACCORDING TO THE BEST INFORMATION OBTAINABLE, NO CLAIM CAN BE MADE FOR ITS ACCURACY)

NAMES SHOWN ARE IMP NAMES, NOT (NECESSARILY) HOST NAMES

The goal was to exploit new computer technologies to meet the needs of military command and control against nuclear threats, achieve survivable control of US nuclear forces, and improve military tactical and management decision making.

Stephen J.
Lukasik



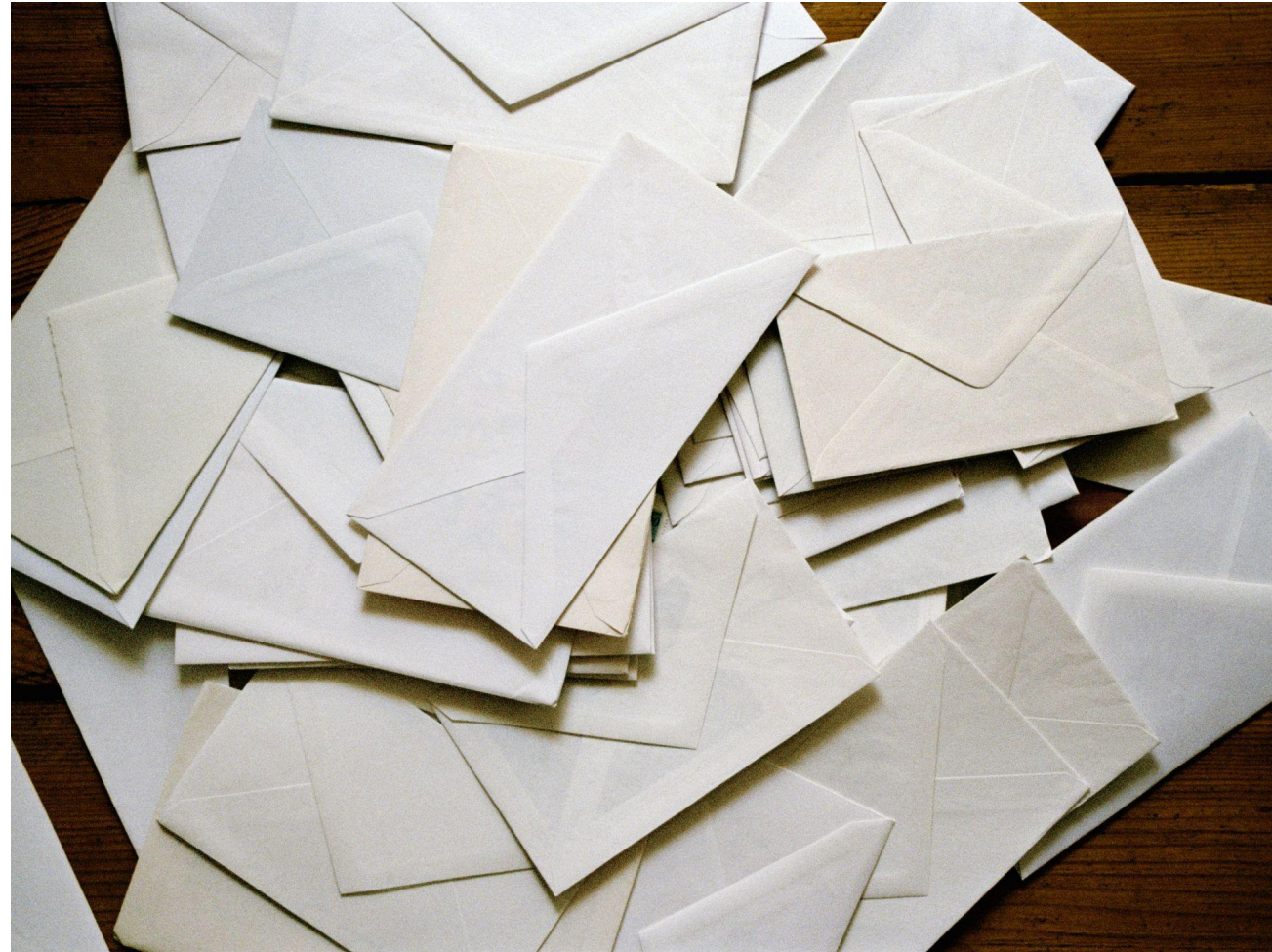
A history lesson – how did we get here?

- Survivable
- Resilient
- Decentralized
- Fault tolerant

Why this works well for the “internet”

- Decentralized
- Fault tolerant
- Delay not problem for most
- Precision not required
- Consume vs publish
- CDN (Content Delivery Network)
- Profit!

Think “lots of little”



Why this does not work well for research data

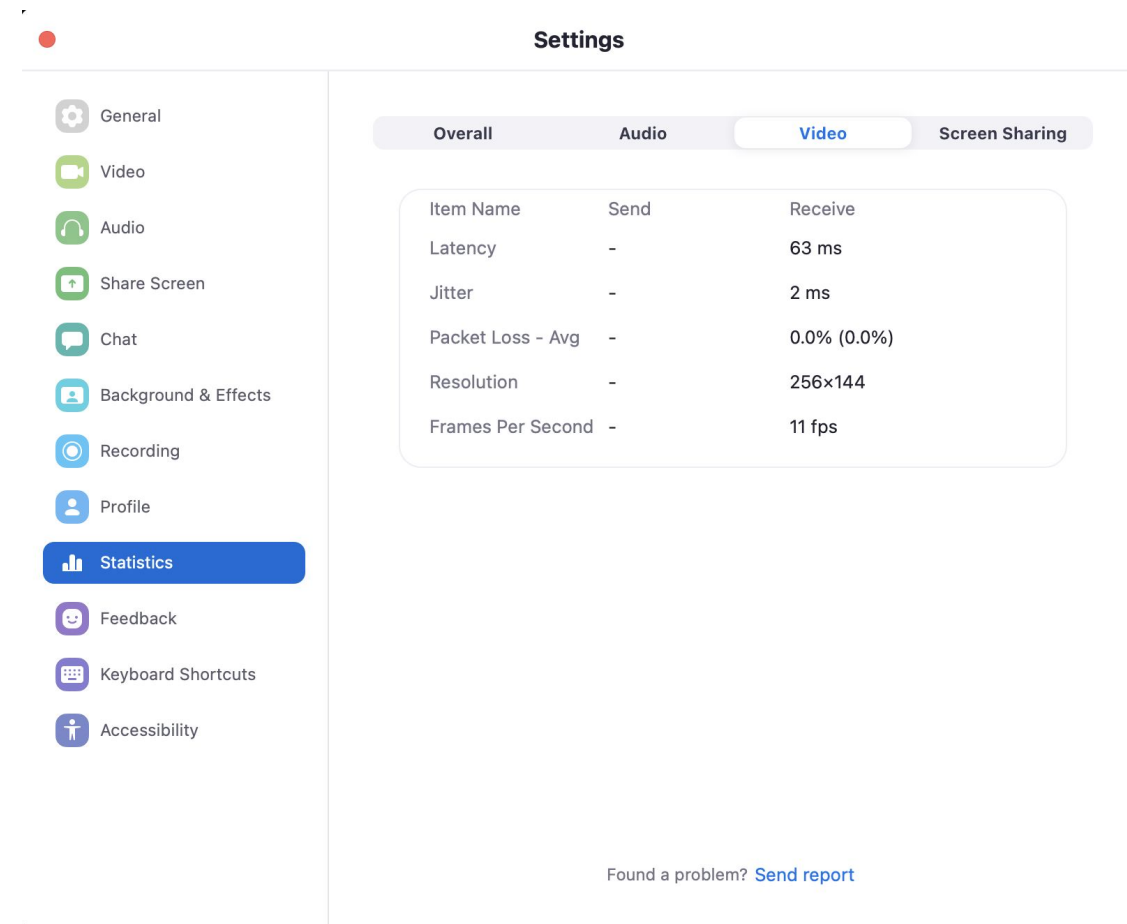
- Decentralized
- *Not* Fault tolerant
- Delay ~~not~~ *is* problem for most
- Precision ~~not~~ *is* required
- Consume vs publish
- ~~CDN (Content Delivery Network)~~
- ~~Profit!~~

Think “fewer large”



The big three...

- Packet loss
- Latency (or RTT -Round Trip Time)
- Buffer/window size



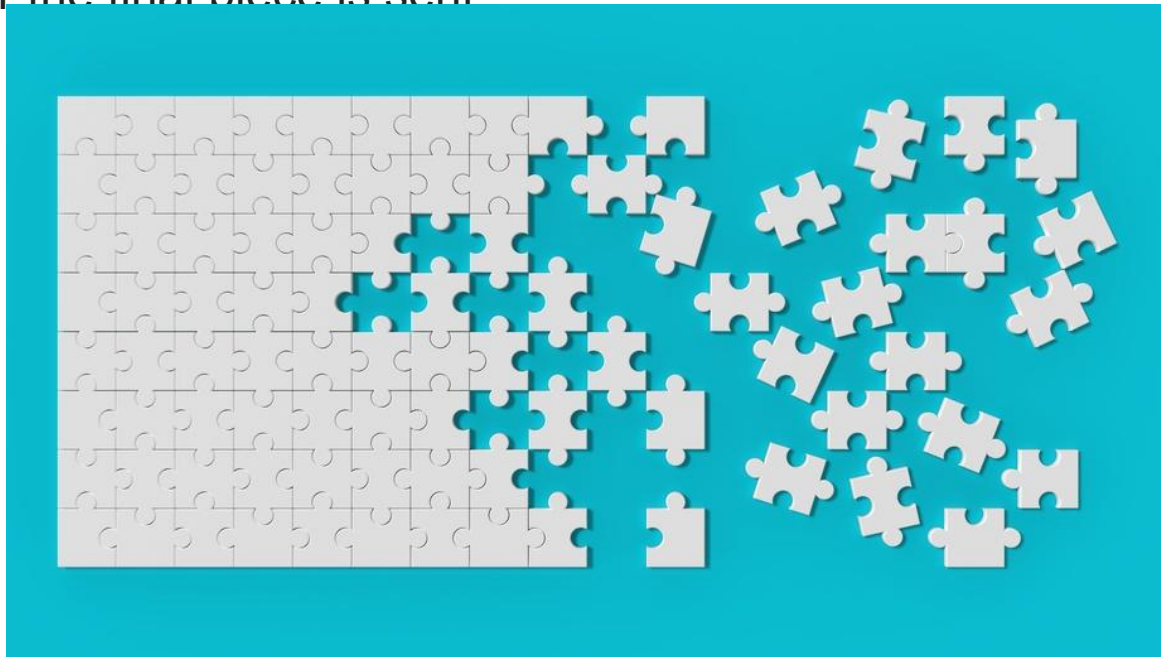
The screenshot shows the Zoom Settings application with the 'Video' tab selected. A 'Statistics' sidebar item is highlighted in blue. The main content area displays a table with video performance metrics.

Item Name	Send	Receive
Latency	-	63 ms
Jitter	-	2 ms
Packet Loss - Avg	-	0.0% (0.0%)
Resolution	-	256x144
Frames Per Second	-	11 fps

Found a problem? [Send report](#)

Networking 101

The TCP/IP relationship is similar to sending someone a message written on a puzzle through the mail. The message is written down and the puzzle is broken into pieces. Each piece then can travel through a different postal route, some of which take longer than others. When the puzzle pieces arrive after traversing their different paths, the pieces may be out of order. The Internet Protocol makes sure the pieces arrive at their destination address. The TCP protocol can be thought of as the puzzle assembler on the other side who puts the pieces together in the right order, asks for missing pieces to be resent, and lets the sender know the puzzle has been received. TCP maintains the connection with the sender from before the first puzzle piece is sent to after the final piece is sent



The big three...

- Packet loss

When TCP encounters packet loss, it backs off on its sending rate. The mechanism for doing this is to reduce the sender's notion of the window so that it attempts to send less data. TCP then ramps its sending rate back up again, in hopes that the loss was transitory.

When TCP encounters loss, it has to recover - but it starts with a small window and opens it back up again over time. The longer the latency, the longer the control loop is for doing this. So, all other things being equal, the time necessary for a TCP connection to recover from loss goes up as the round-trip time goes up.

This is a case where things are a lot better now than they were for many years....algorithms such as htcp and cubic are much more aggressive about ramping back up again than the old scheme (TCP Reno).

The big three...

- Buffer/window size

Buffer and window size determine both the amount of data that the kernel will keep in buffers for the connection, and the "window" that is advertised over the TCP connection (the window information sent via TCP reflects the size of the available buffers). The larger the window, the more data can be in flight between the two hosts. Note that if the window is smaller than the available bandwidth multiplied by the latency (the Bandwidth Delay Product) then the sender will send a full window of data and then sit and wait for the receiver to acknowledge the data.

The big three...

- Buffer/window size

Large windows are required for adequate throughput when latency is large (anything over about 10 milliseconds starts becoming an issue, and 20+ milliseconds is where it gets really tricky). When the window is large, TCP can send a lot of data all at once.

The network card typically doesn't know or care about TCP. The network card just knows that when there are packets to send, it throws packets onto the link until it doesn't have packets to send. This happens at whatever link speed the card is configured for, e.g. 10Gbps.

So, if the TCP window is 4MB, TCP will pass 4MB of data to the network card, and the network card will slam it onto the link at 10Gbps. This means that every device in the path between the sender and the receiver sees a high-speed burst of packets. If there are any buffering or queuing problems anywhere on the path, some of those packets will be dropped, causing TCP to back off.

Thus packet
loss...

The big three...

- Latency (or RTT -Round Trip Time)

Latency/RTT is the amount of time it takes for a packet to go from the sender to the receiver and back (a "round trip"). This is the minimum amount of time for one host to get information back from the other host about data that was sent.

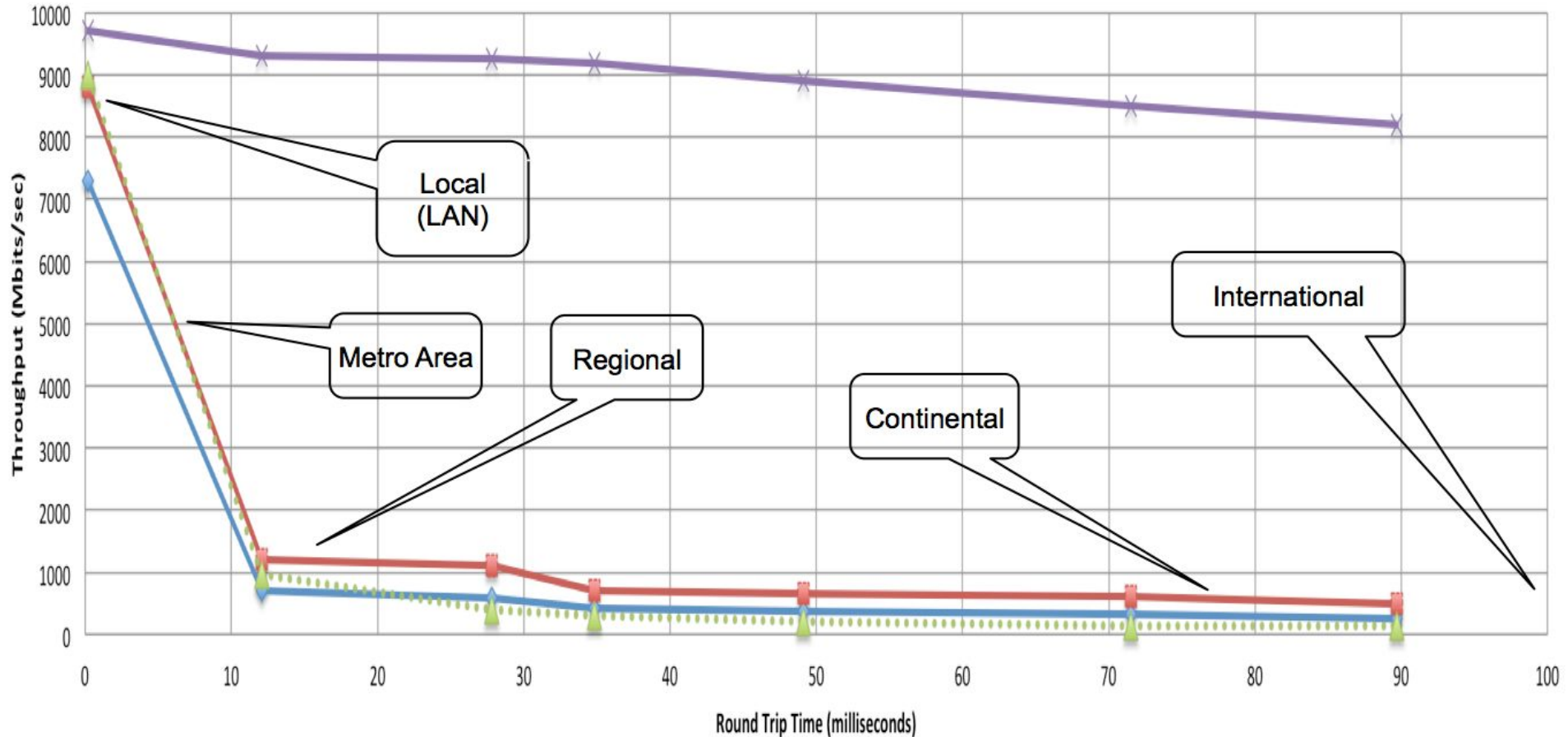
The big three...

All together
now...

Since the TCP windows are much smaller for low-latency connections, these issues often are not noticed at low latencies. If the packet loss is due to random error (e.g. dirty fiber, marginal optics, longer-than-spec cable length) then there will be a bit of loss here and there, and with low latency TCP will recover so quickly that people typically won't notice a performance hit. If there are small switch buffers in the path, they typically don't cause loss for low-latency connections, because TCP is not sending big enough bursts to cause loss. However, if you use the same infrastructure to send data to a host a long way away, you will see dramatic performance degradation. In the case of random error, TCP will always be recovery mode, and so will always have a small window. In the case of small buffers, TCP will ramp up, encounter loss, reduce its window, ramp up again, and so on - so TCP will effectively always have a small window and perform poorly.

The big three...

Throughput vs. increasing latency on a 10Gb/s link with **0.0046%** packet loss



Measured (TCP Reno)

Measured (HTCP)

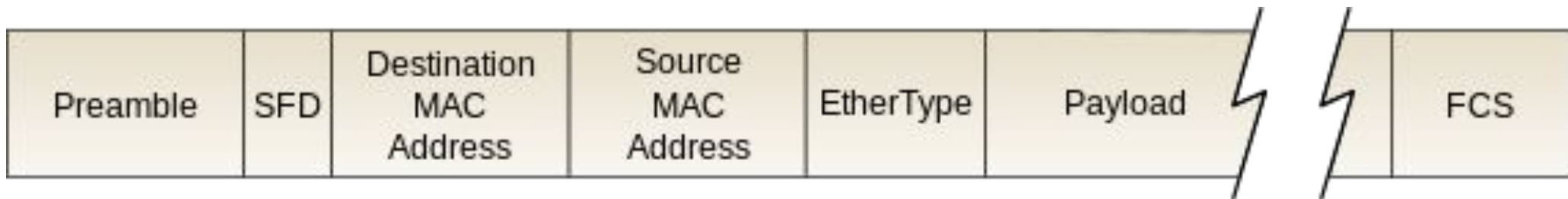
Theoretical (TCP Reno)

Measured (no loss)

MTU & Jumbo frames

Jumbo frames are Ethernet frames with more than 1500 bytes of payload, the limit set by the IEEE 802.3 standard.

Jumbo frames can carry up to 9000 bytes of payload



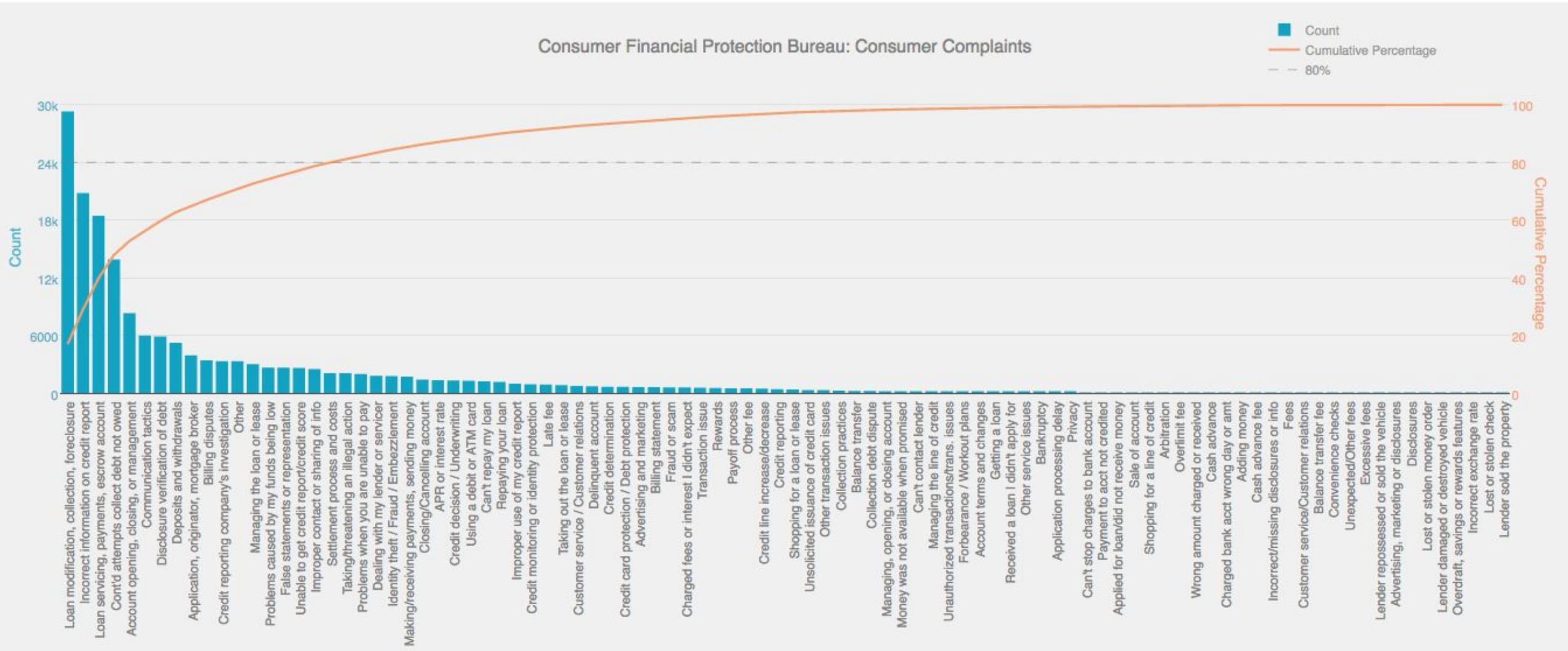
Lets talk about using protection...

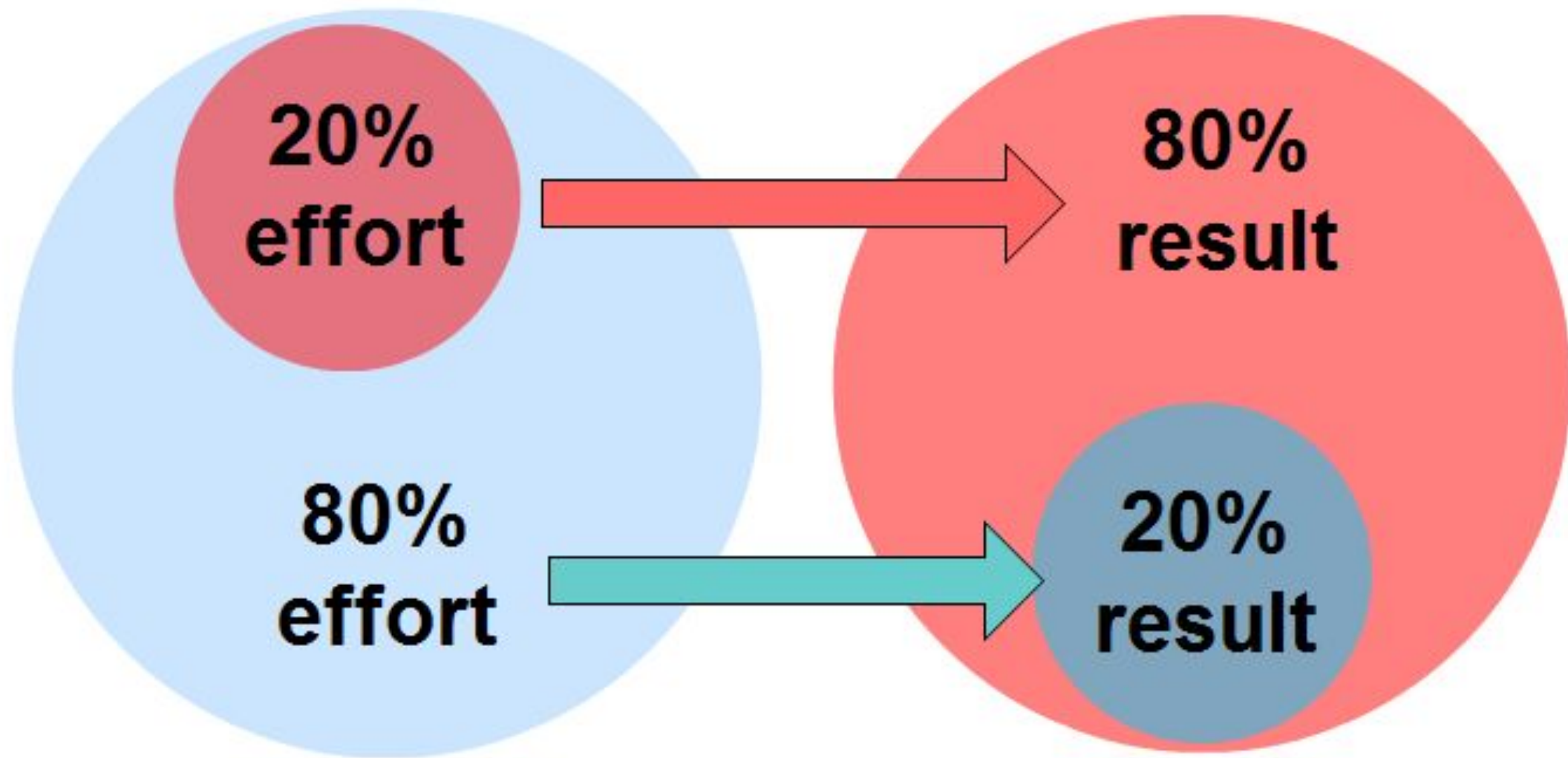
Lets talk about using protection...

- DDoS
- Intrusion Prevention
- Firewalls
- Access control
- Auditing



What works for the 95% does not necessarily work for all...





**20%
effort**

**80%
result**

**80%
effort**

**20%
result**



DEPTH
GAUGE

Commodity networks

The good

- Great for “normal” traffic
- Resilient by design
- Can move lots of “small” things moving around
- Great if what you are doing is accessing/and on a CDN (Content Delivery Network)
- Available almost everywhere

The not so good

- Not at all optimized for large flows
- Can be very expensive at scale
- **Often sub optimal routing and peering for point to point research traffic**
- Throttling , queuing, traffic shaping destroy throughput (and “they” don’t care)
- Commodity networks assume, and are designed for, “lots of small stuff”
- High speeds are not always available, or cost effective (10G, 40G, 100G)
- **If you have issues, good luck getting help**





Take the sloth and the giraffe...

Each are extremely efficient at consuming leaves...



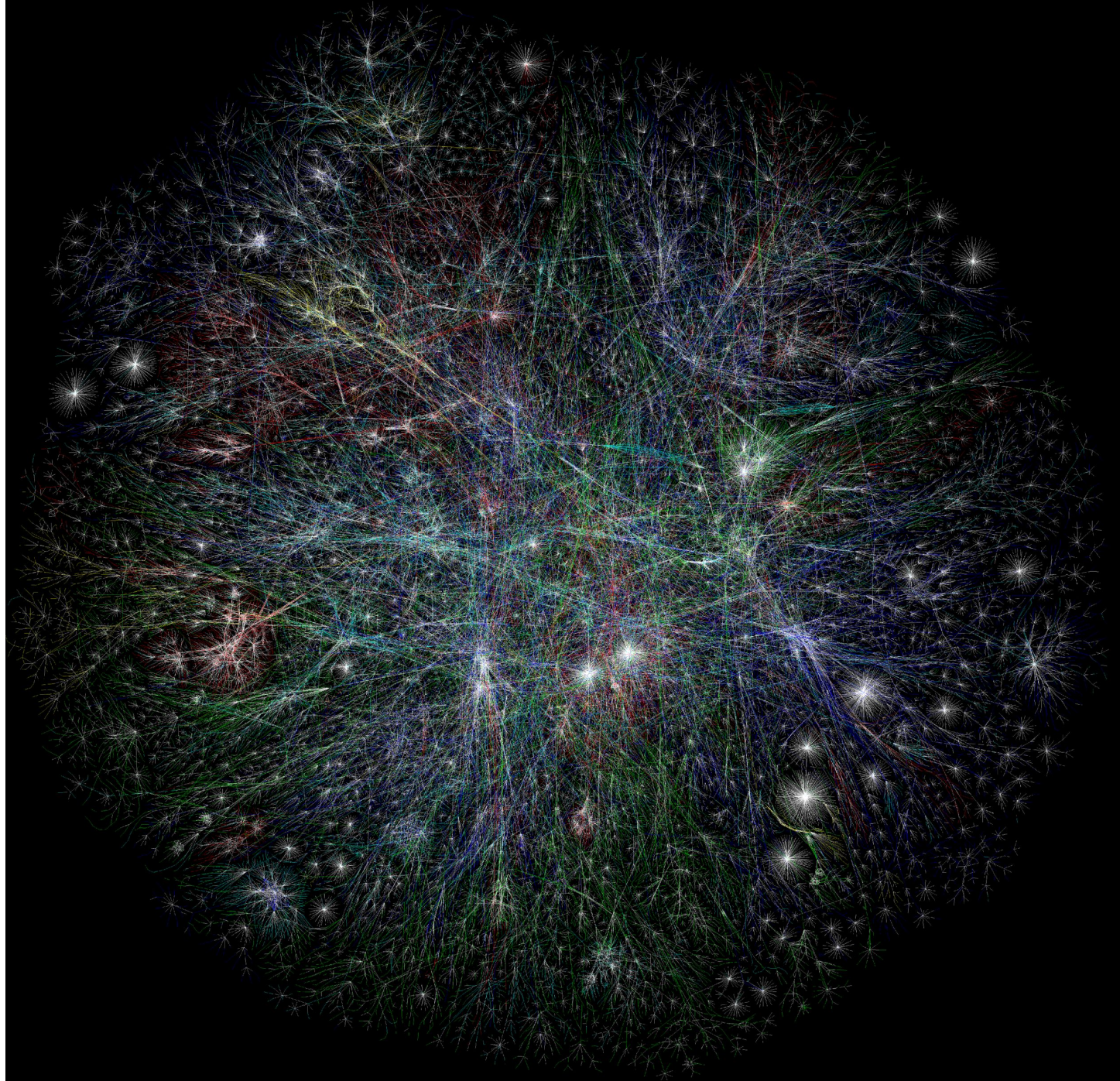
But that does not
make them
interchangeable...



The solution!

Networks designed, built and operated by and for the research community

Lets think
about
networks of
networks
connecting to
networks of
networks...



International Networks

REANNZ



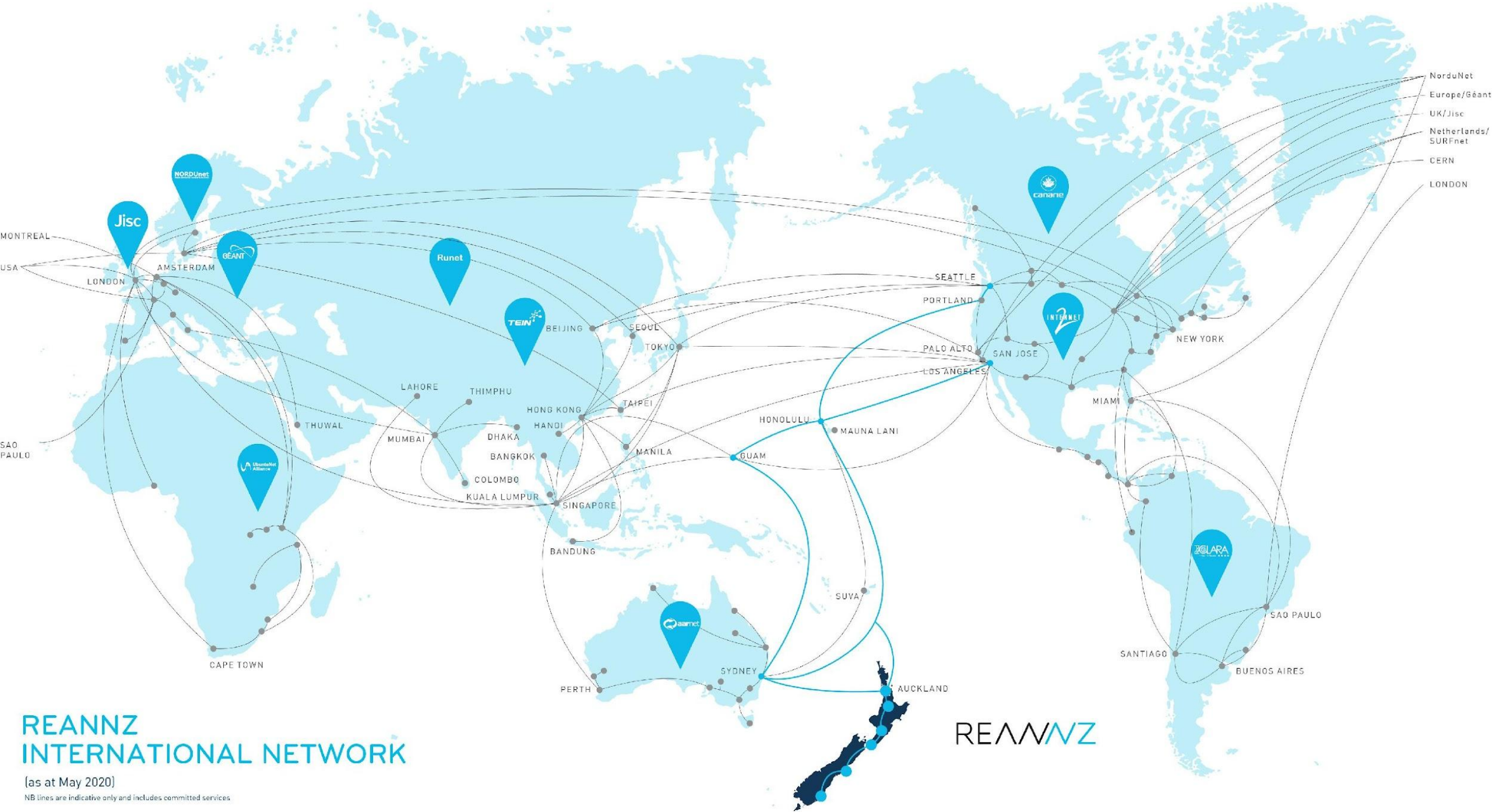
SCReN

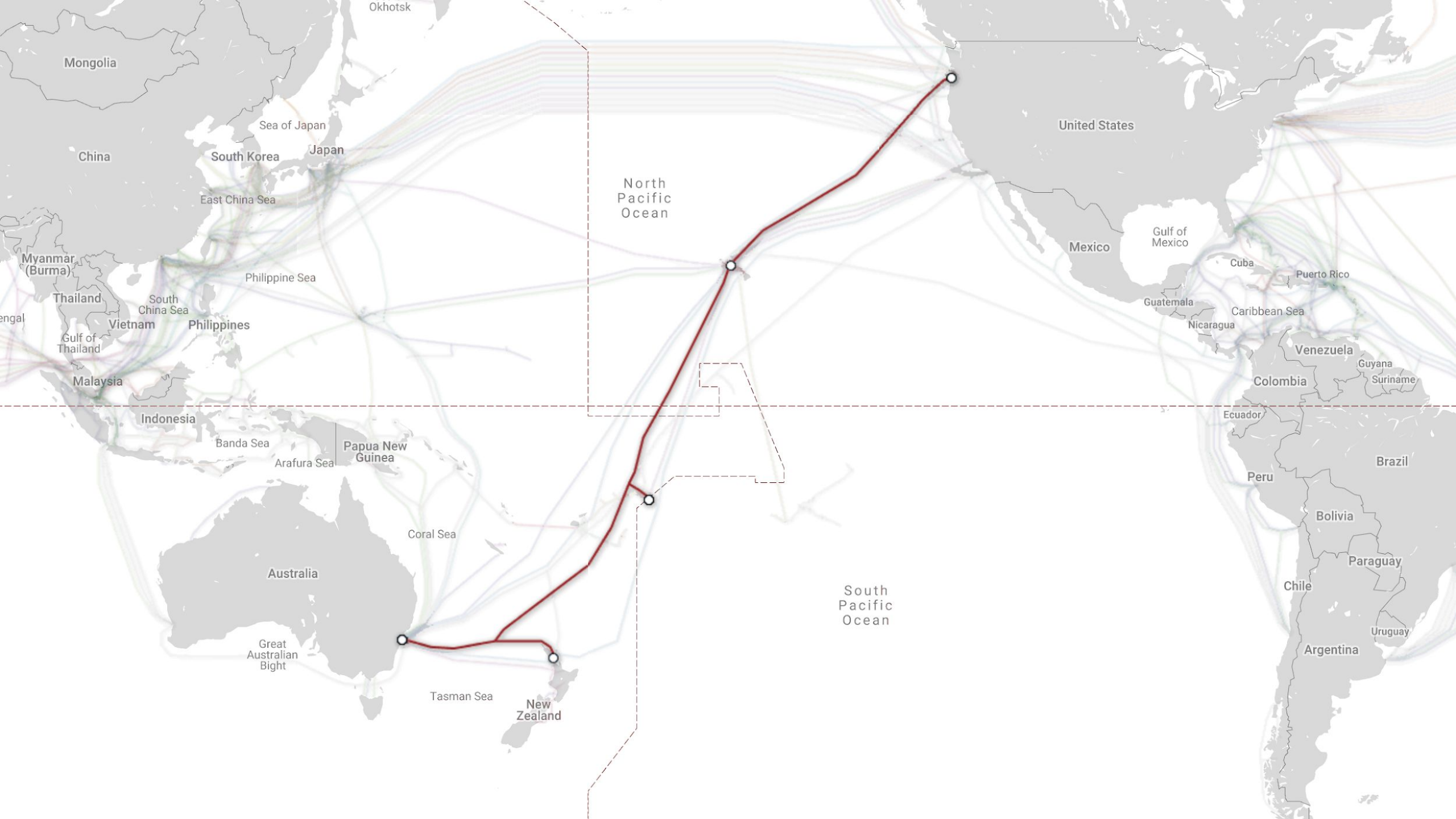
★
TEIN3



canarie

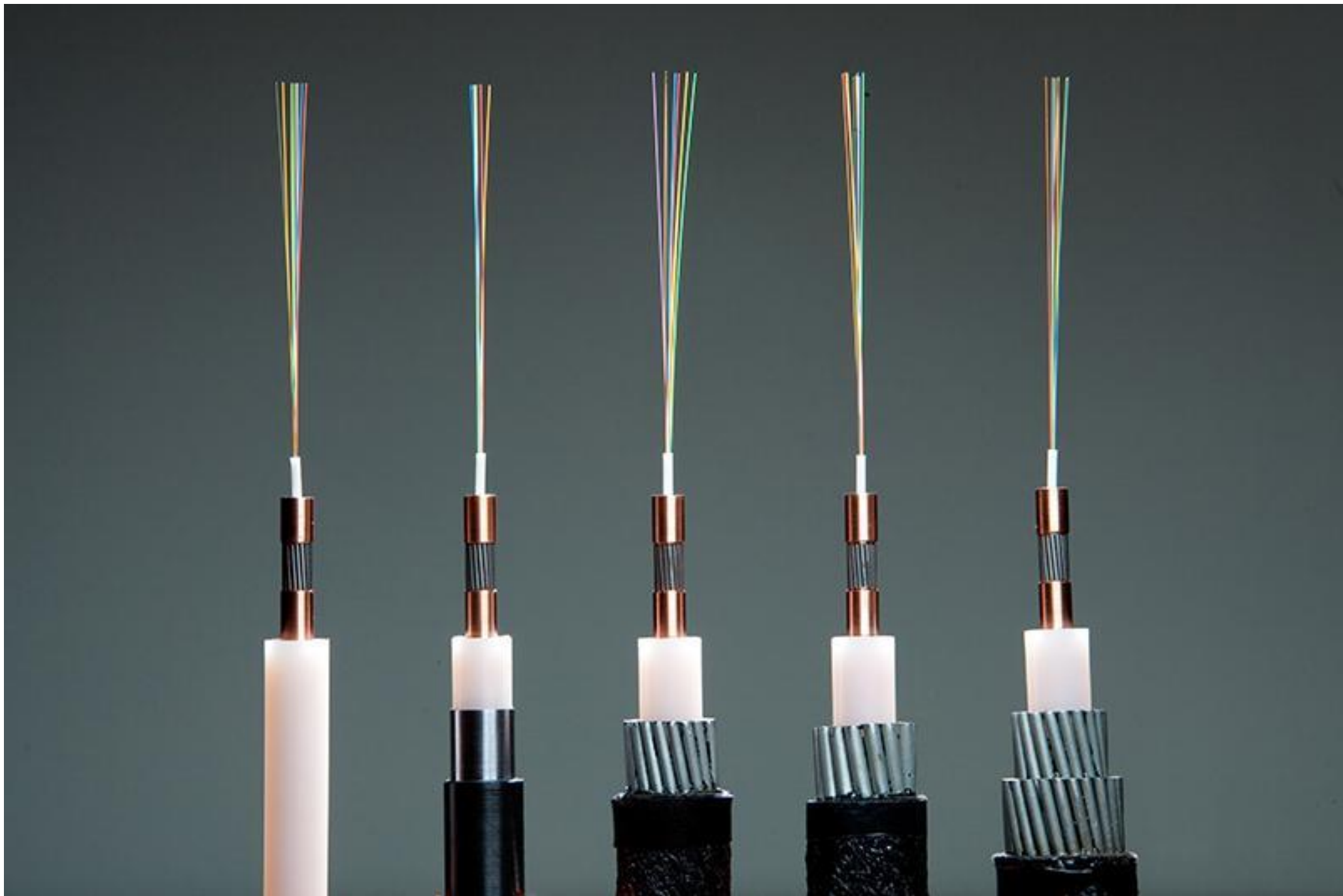






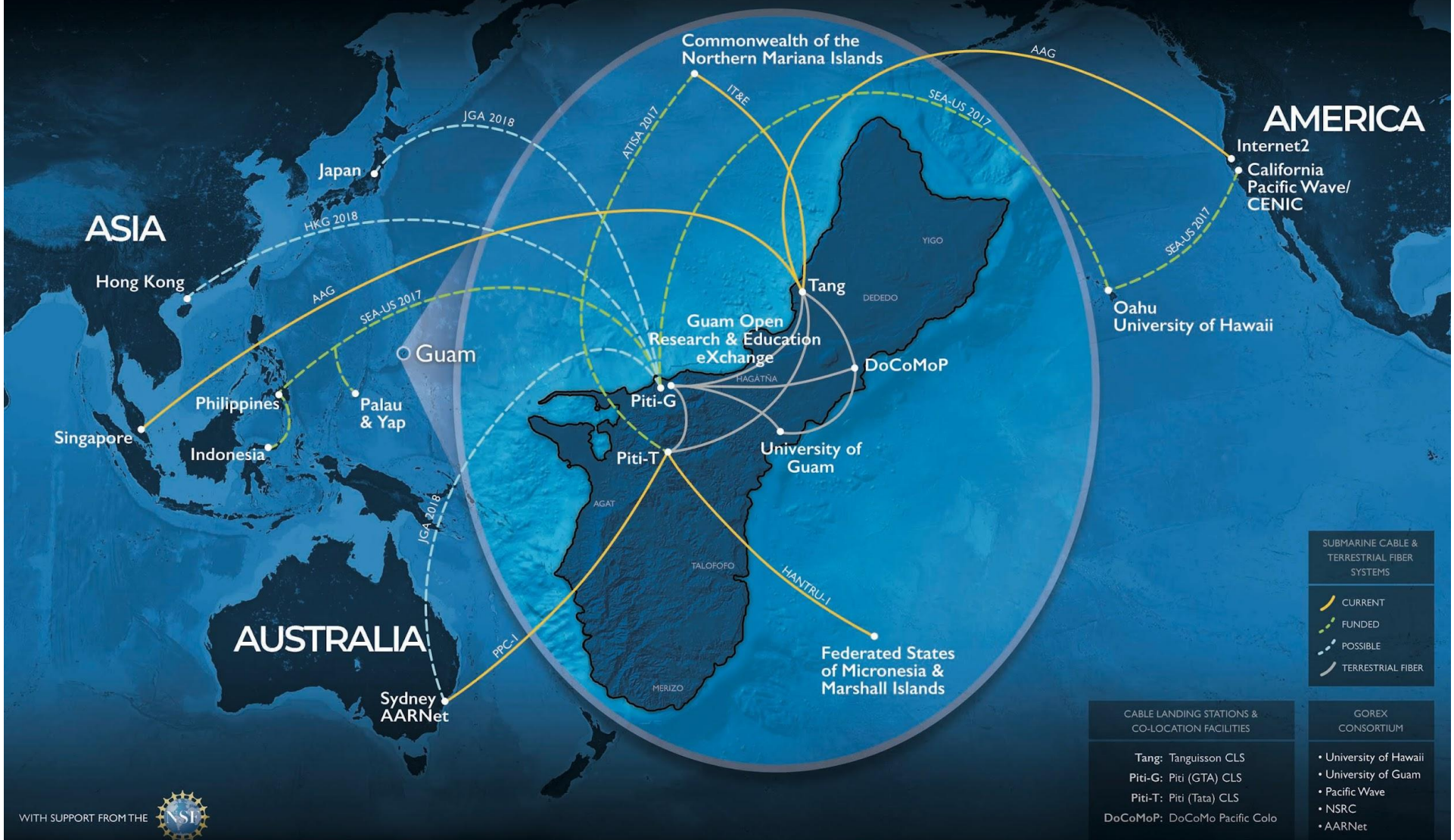
REANNZ INTERNATIONAL NETWORK

- Pacific Wave peering in Seattle
- Connection in Hawaii with UoH
- AARNet peering in Sydney
- GOREX peering in Guam
- Direct peering with providers and with other NRENs





GOREX: Guam Open Research & Education eXchange



WITH SUPPORT FROM THE 

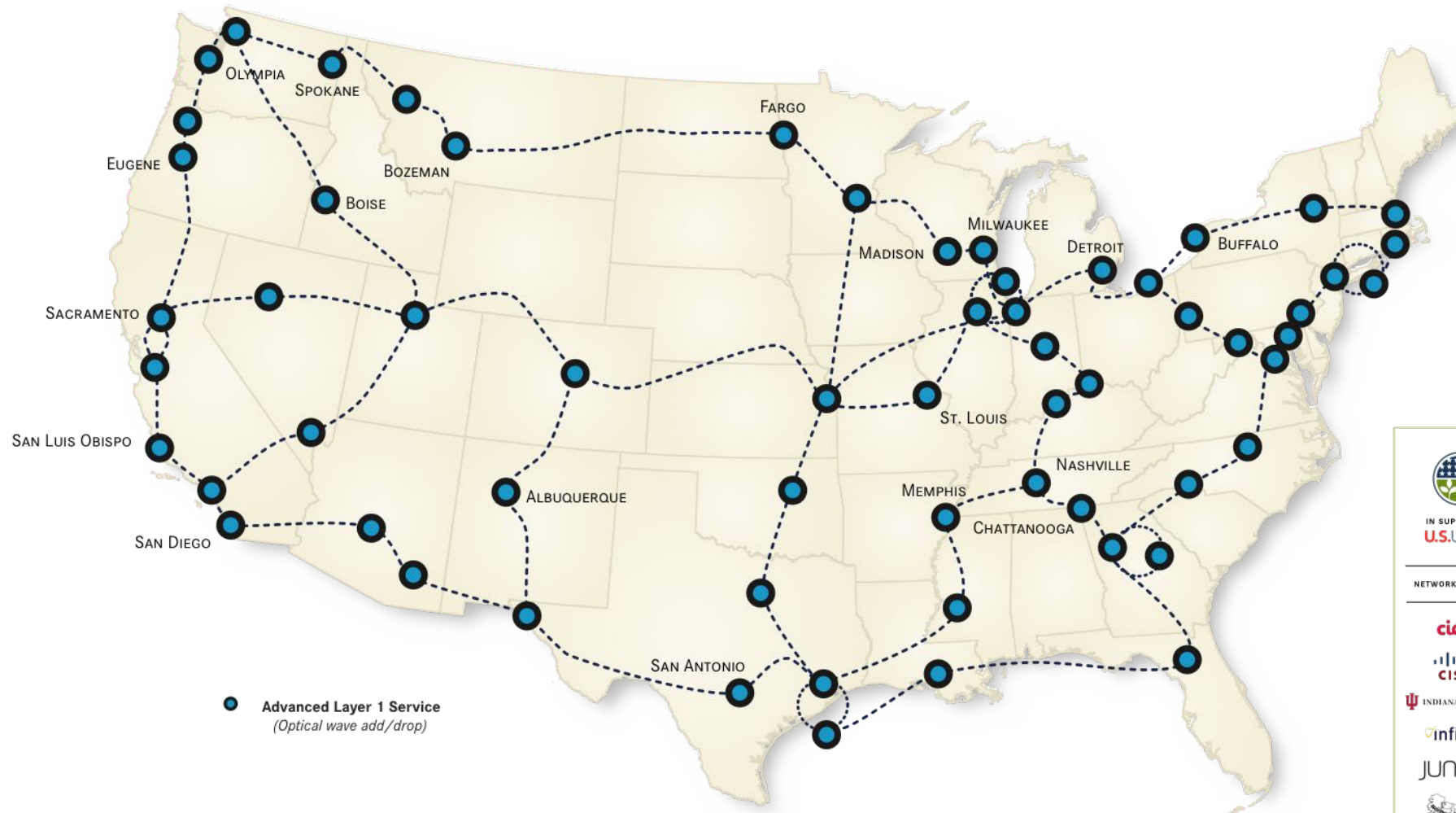
National Networks (US)





INTERNET2 NETWORK ADVANCED LAYER 1 SERVICE

MAY 2017



IN SUPPORT OF
U.S.UCAN

NETWORK PARTNERS

ciena



INDIANA UNIVERSITY

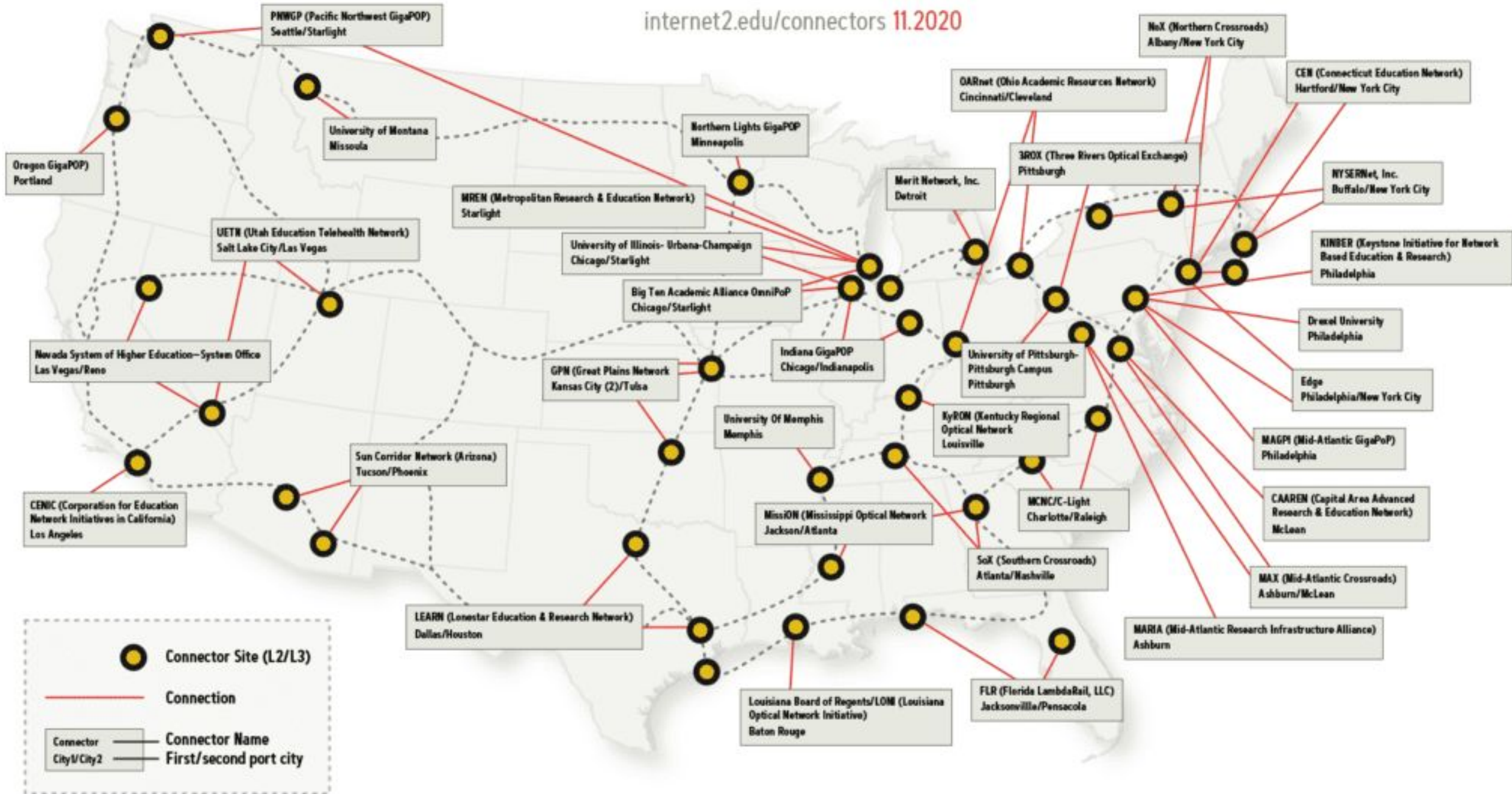
infinera

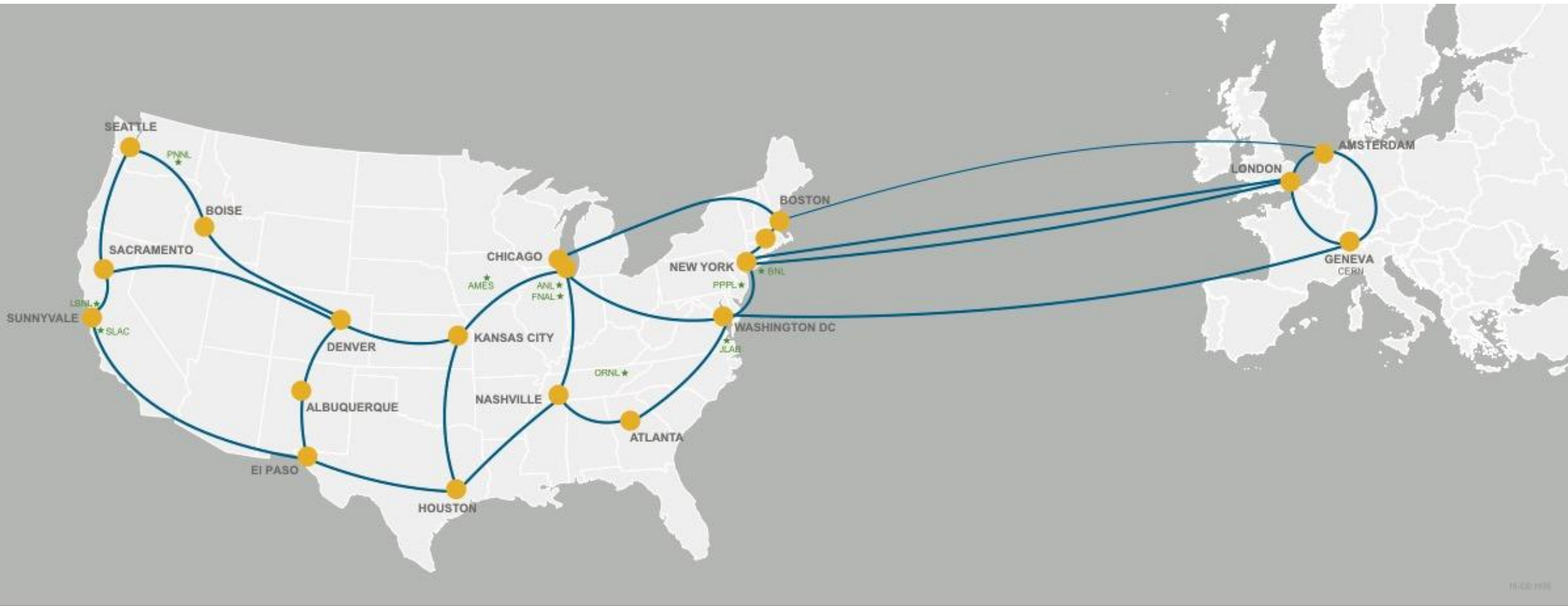
JUNIPER
NETWORKS



Internet2 Network Connections

internet2.edu/connectors 11.2020





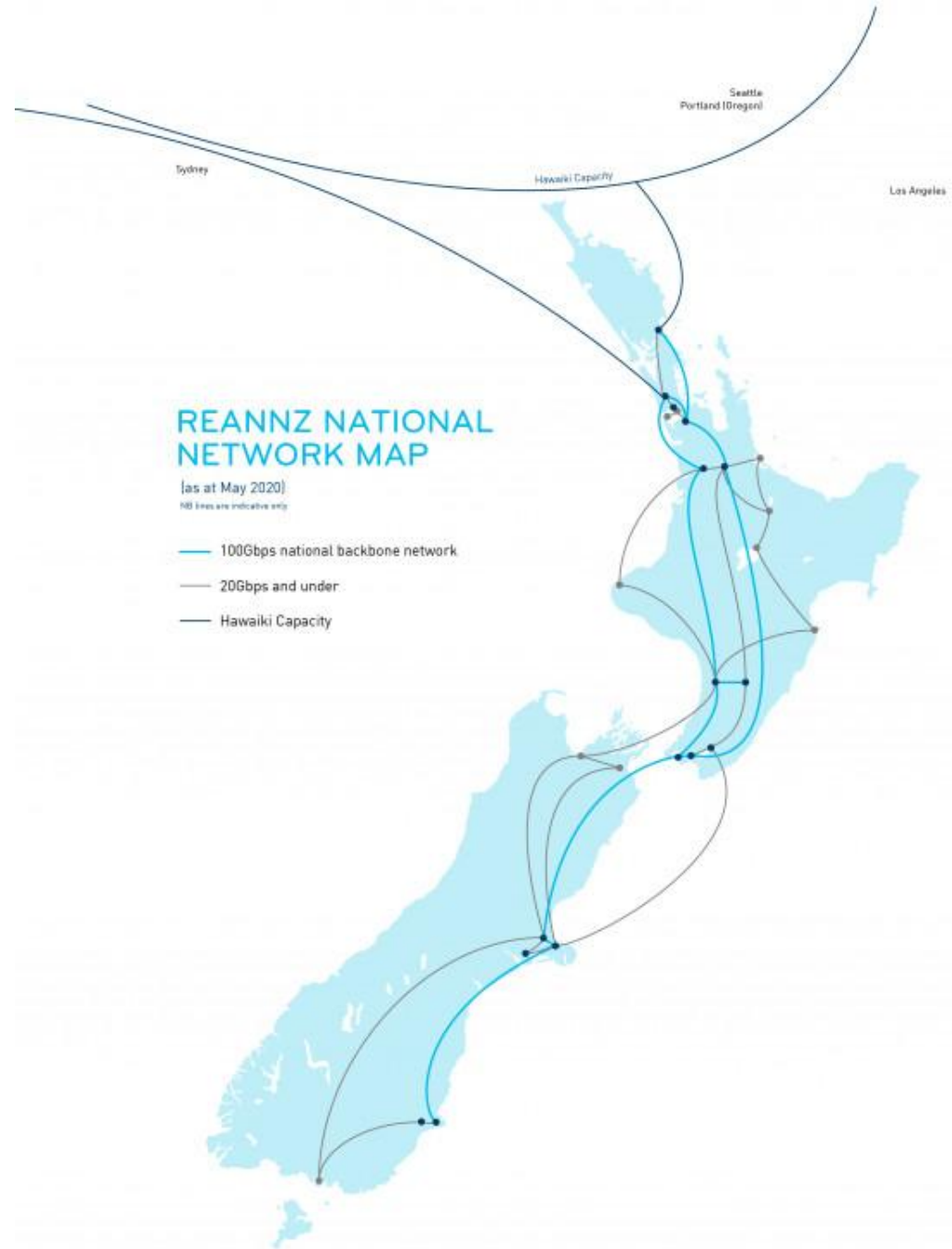
ESnet

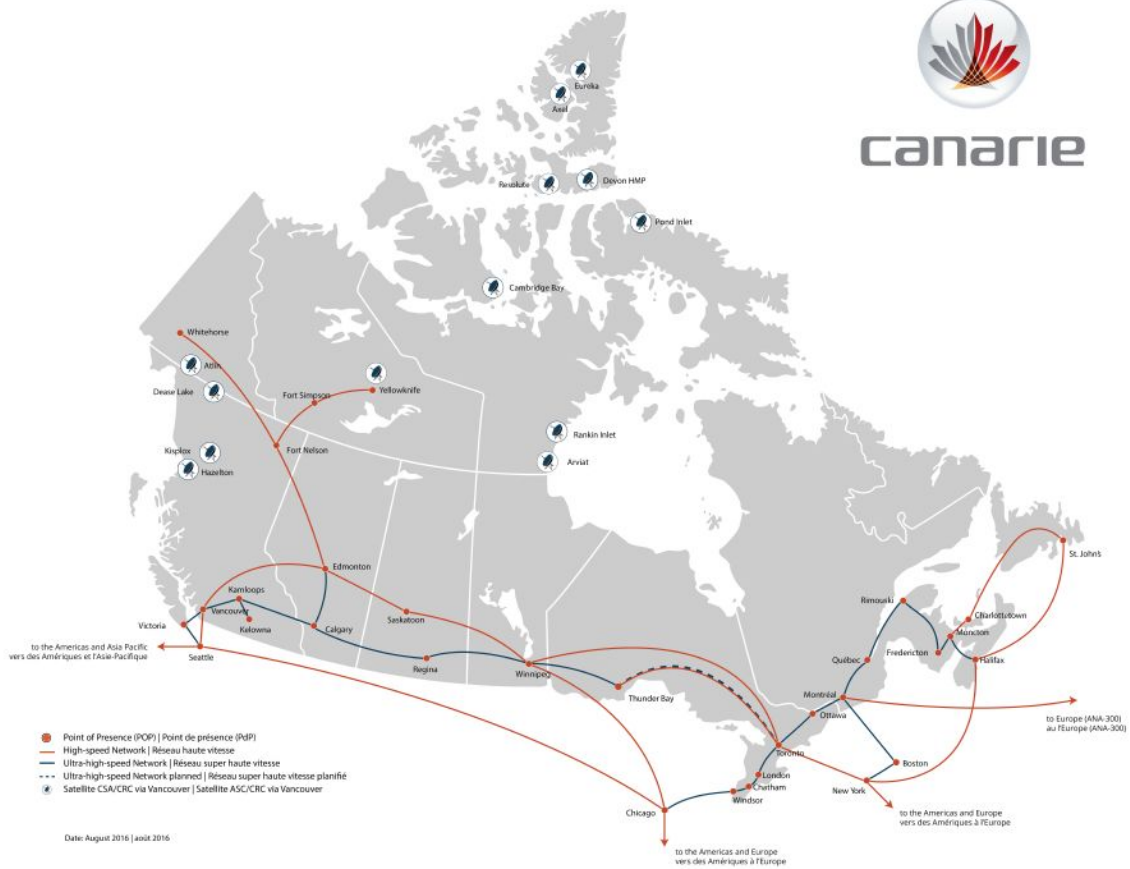
ENERGY SCIENCES NETWORK

★ Department of Energy Office of Science National Labs

- Ames** Ames Laboratory (Ames, IA)
- ANL** Argonne National Laboratory (Argonne, IL)
- BNL** Brookhaven National Laboratory (Upton, NY)
- FNAL** Fermi National Accelerator Laboratory (Batavia, IL)
- JLAB** Thomas Jefferson National Accelerator Facility (Newport News, VA)

- LBL** Lawrence Berkeley National Laboratory (Berkeley, CA)
- ORNL** Oak Ridge National Laboratory (Oak Ridge, TN)
- PNNL** Pacific Northwest National Laboratory (Richland, WA)
- PPPL** Princeton Plasma Physics Laboratory (Princeton, NJ)
- SLAC** SLAC National Accelerator Laboratory (Menlo Park, CA)





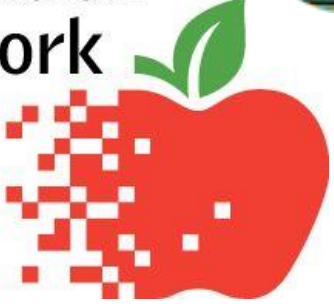
AARNET NATIONAL NETWORK

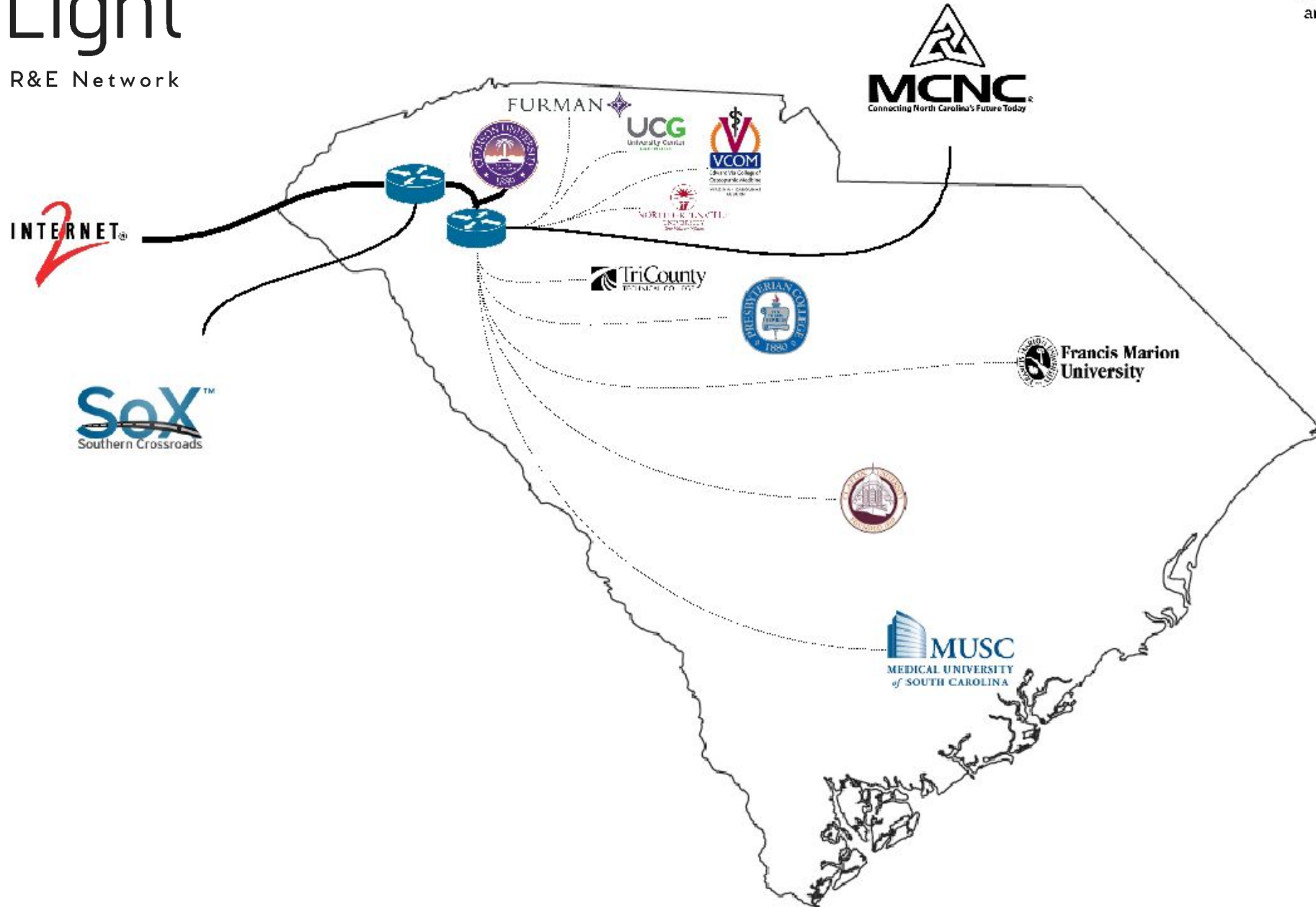


State/regional networks aka "your ISP"



K-20
Education
Network





Exchanges and PoPs



REANNZ

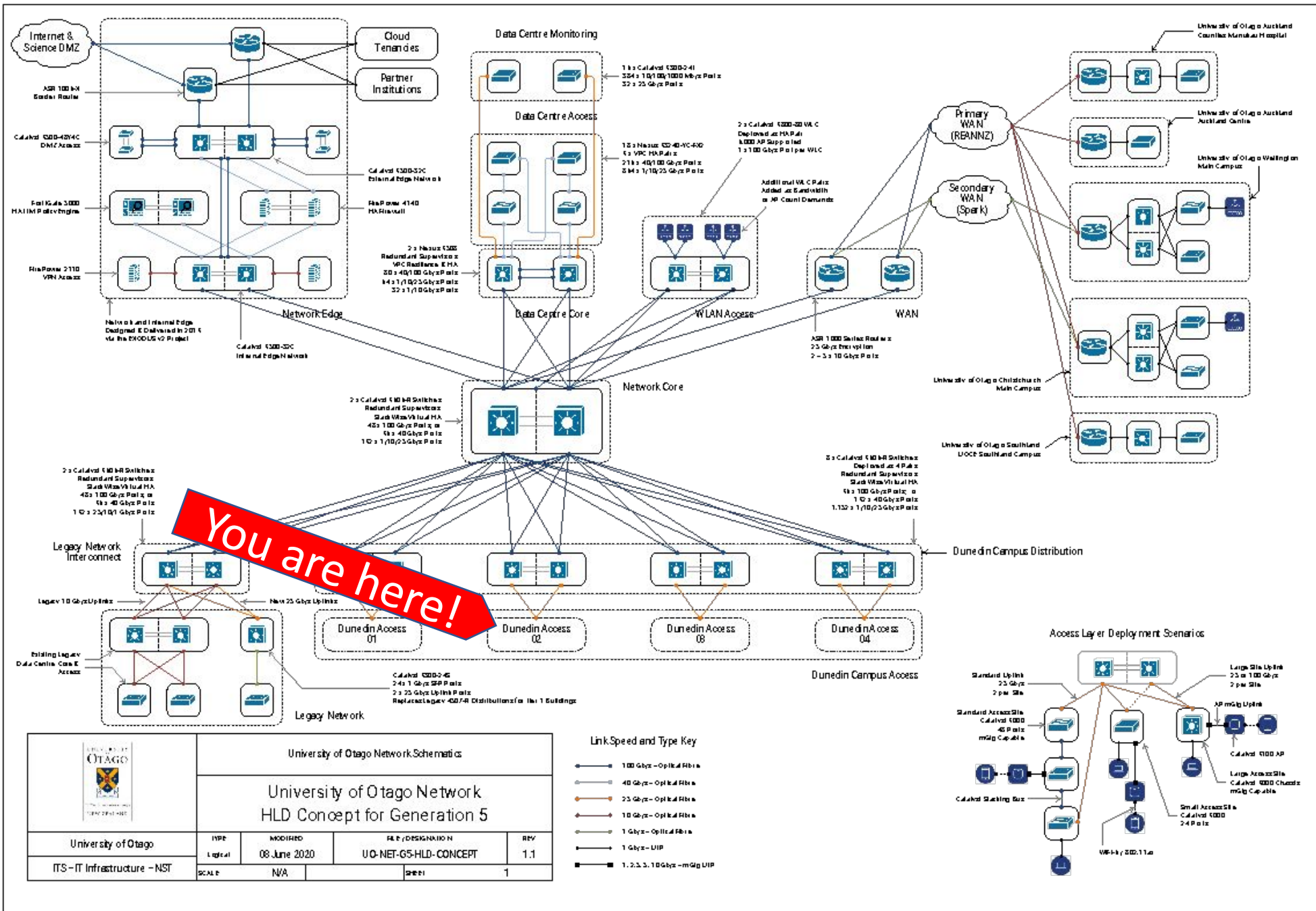


NNENIX
NORTHERN NEW ENGLAND
NEUTRAL INTERNET EXCHANGE

Local networks



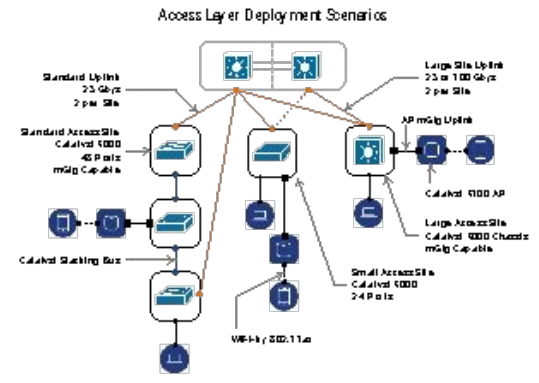
(Insert your logo here)



You are here!

University of Otago Network Schematics				
University of Otago Network HLD Concept for Generation 5				
UNIVERSITY OF OTAGO	TYPE	MODIFIED	FILE / DESIGNATION	REV
University of Otago	Logical	08 June 2020	UD-NET-GS-HLD-CONCEPT	1.1
ITS-IT Infrastructure - NST	SCALE	N/A	sheet	1

- Link Speed and Type Key**
- 100 Gbps - Optical Fibre
 - 40 Gbps - Optical Fibre
 - 25 Gbps - Optical Fibre
 - 10 Gbps - Optical Fibre
 - 1 Gbps - Optical Fibre
 - 1 Gbps - UTP
 - 1, 2, 3, 10 Gbps - mGig UTP



The local area network – spoiled for choice

- LAN networks are like Legos – no wrong way to build them (sort of)
- Minefield of equipment
- Random security systems



What's the deal with DMZs?



What makes up an university network?



Student

s

What makes up an university network?

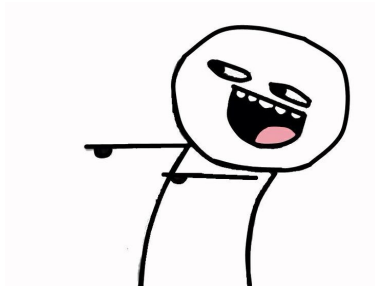


Student
s



Business
IOT

What makes up an university network?



**Student
s**



**Business
IOT**



Research and teaching

What makes up an university network?



**Student
s**



**Business
IOT**



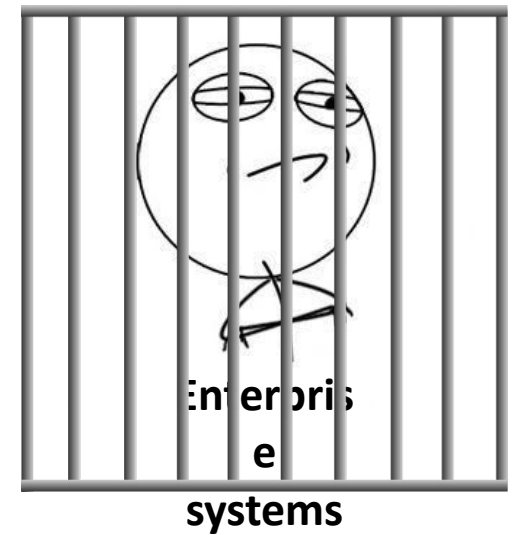
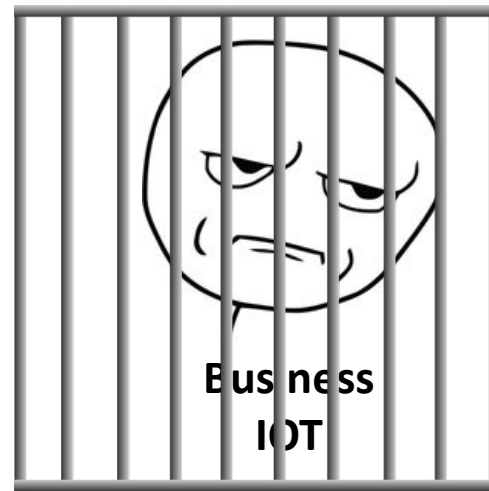
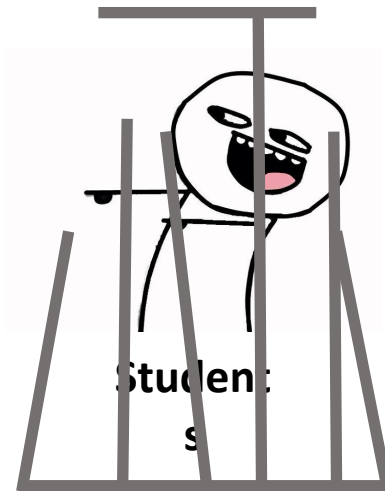
Research and teaching



**Enterpris
e
systems**



What makes up an university network?



The university enterprise network...



The university enterprise network...



"internet connection"

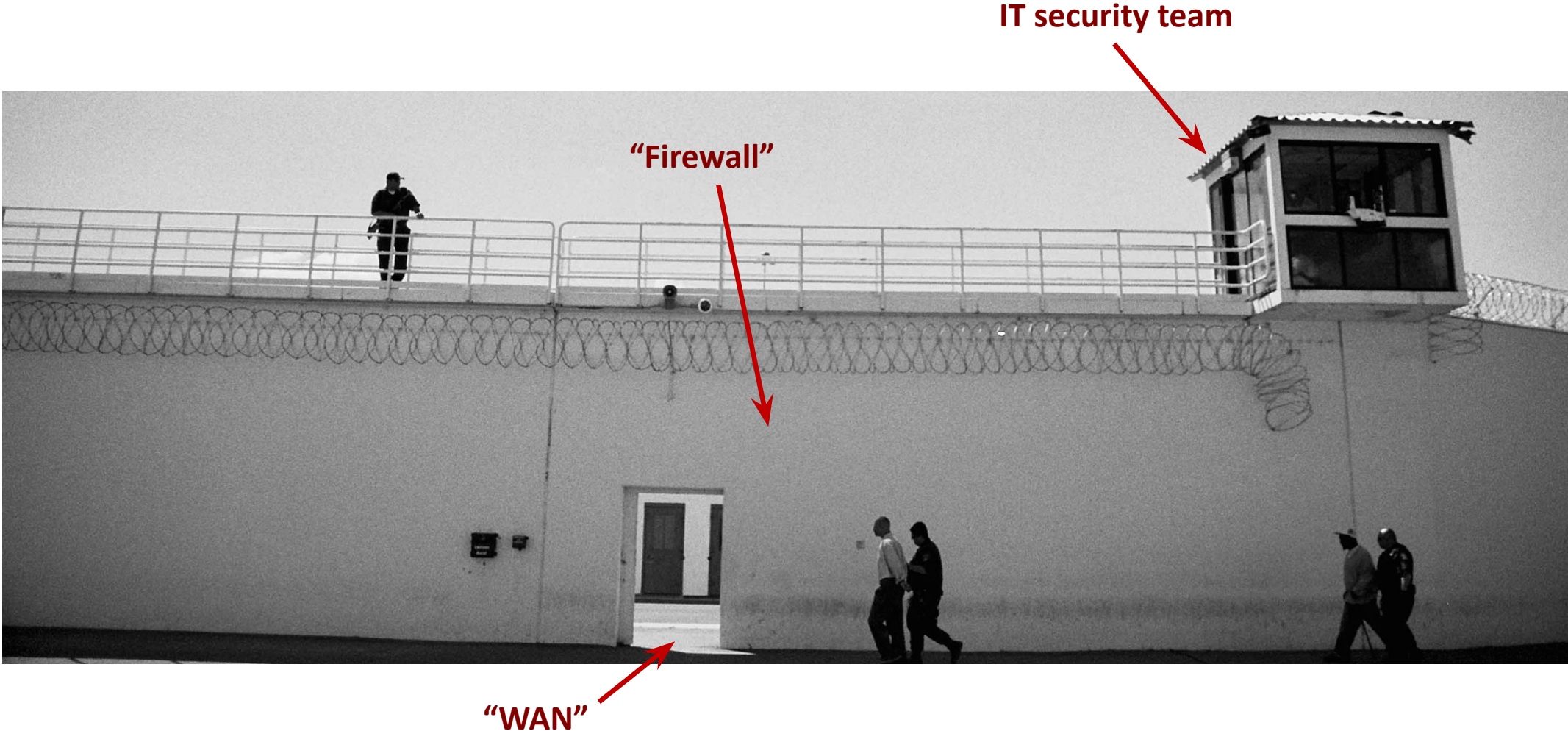
The university enterprise network...



"internet connection"

"Firewall"

The university enterprise network...



The academic network...

**Bureaucracy
(aka "Paperwork")**

IT security team



"Firewall"

"internet connection"

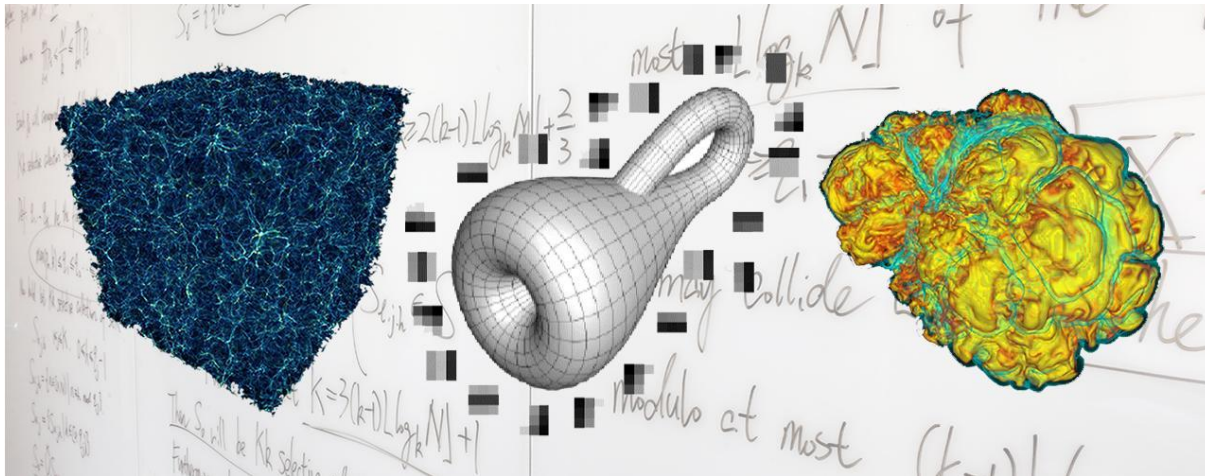
CAUTION

ANALOGIES AHEAD

PROTECTIVE HEADGEAR MUST BE WORN IN THIS AREA



Computational Research, an analogy...



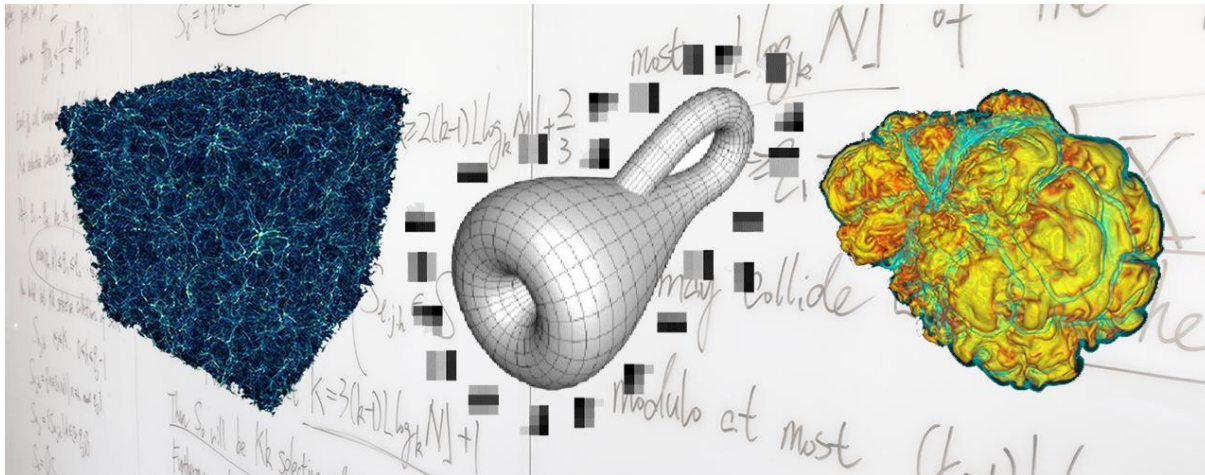
Computational Research

=



Velociraptor

Computational Research, an analogy...



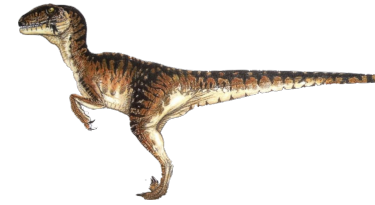
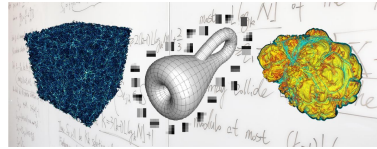
Computational Research

III



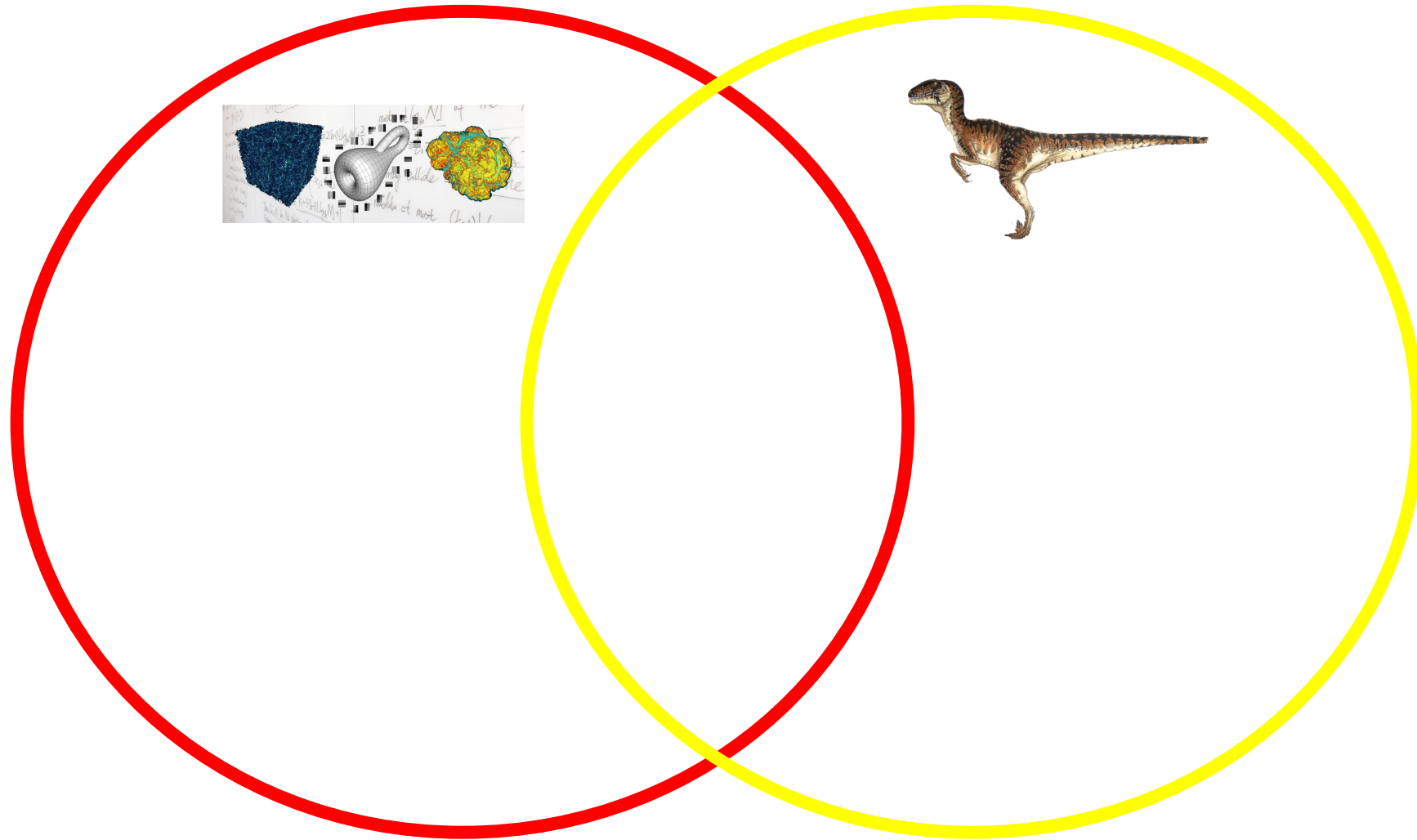
Velociraptor

Computational Research, an analogy...

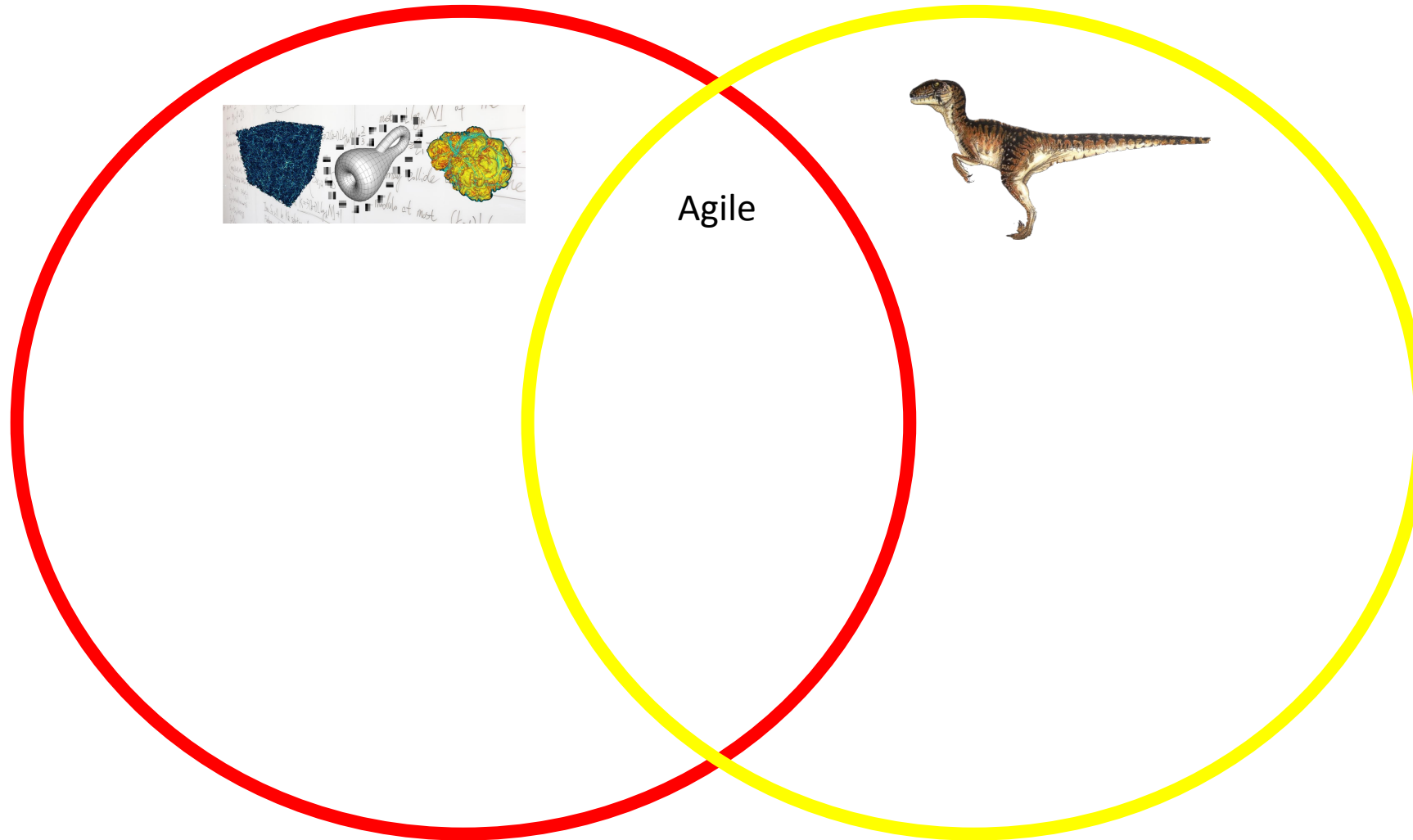


Irrefutable proof the analogy is valid...

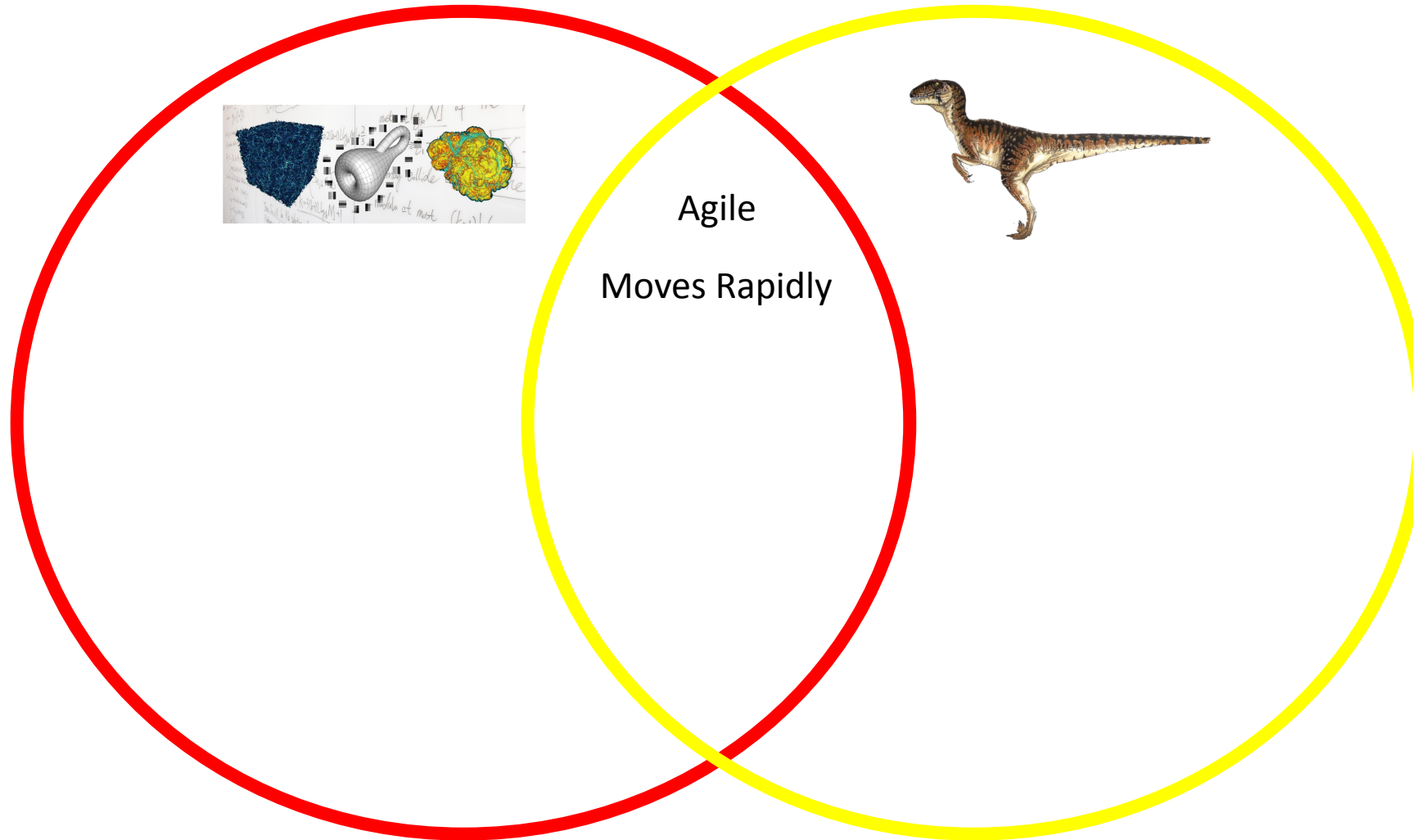
Computational Research, an analogy...



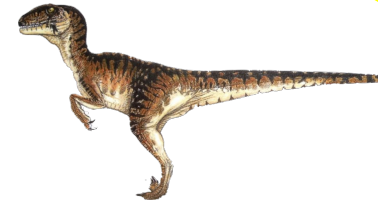
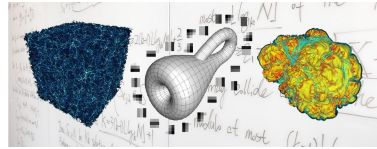
Computational Research, an analogy...



Computational Research, an analogy...



Computational Research, an analogy...

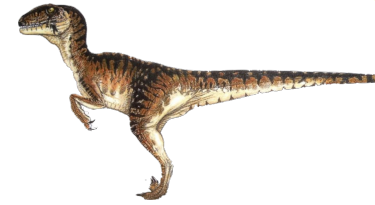
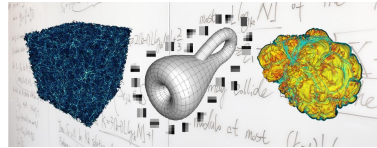


Agile

Moves Rapidly

More Effective
In Groups

Computational Research, an analogy...



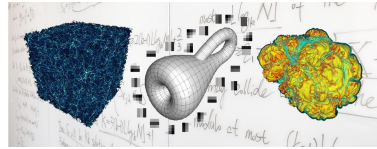
Agile

Moves Rapidly

More Effective
In Groups

Consumes All
Available Resources

Computational Research, an analogy...



Agile

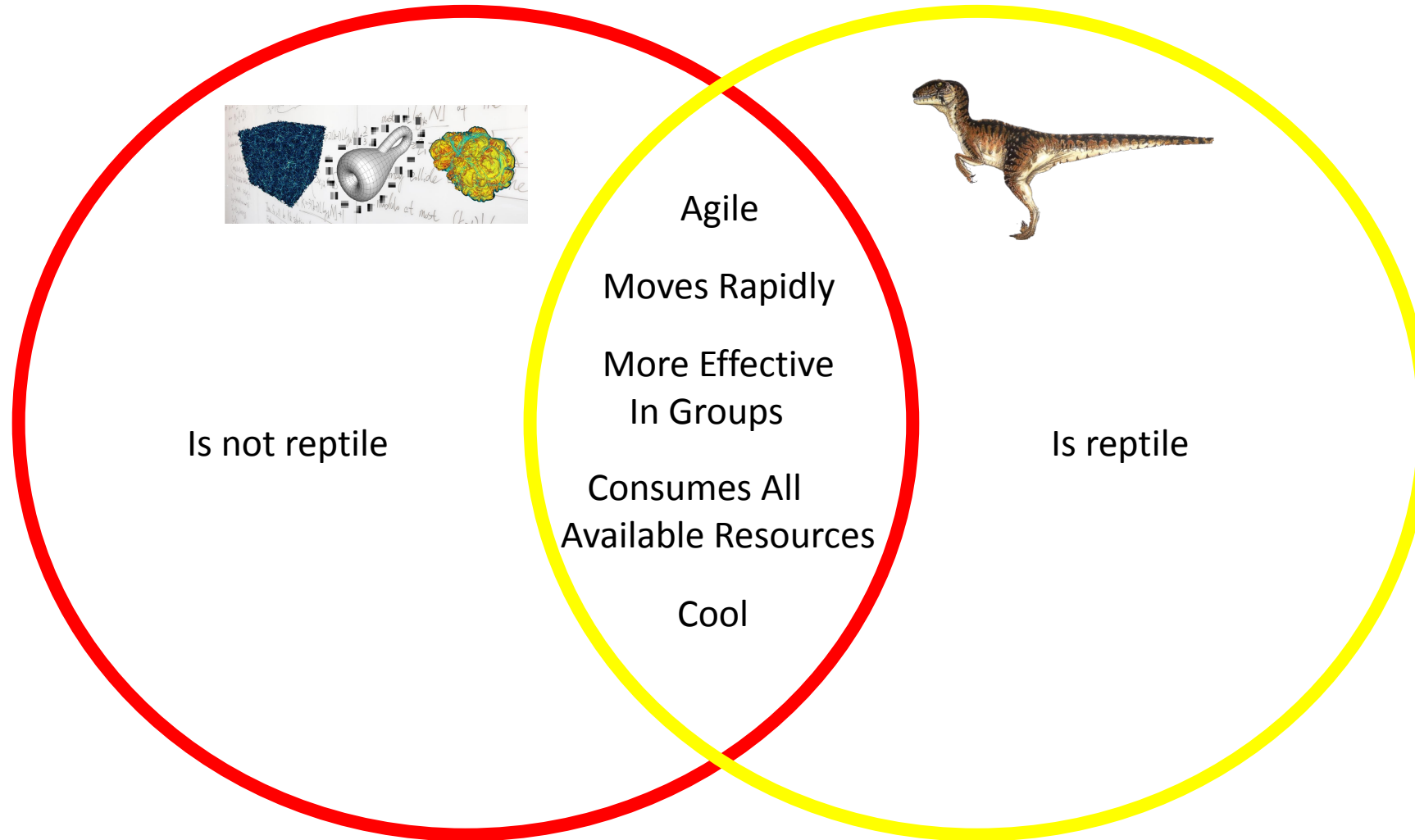
Moves Rapidly

More Effective
In Groups

Consumes All
Available Resources

Cool

Computational Research, an analogy...



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks

...that is highly important to myself, the educational community, and all of mankind as a whole. It is imperative that this data be *reasonably secured*; yet, *available* to my research peers. The *datasets are rather large*, and they may need to be shared across institutions.



When Computational Science Meets Traditional Networks

Would it be possible to place this in a *secure, reliable, flexible, accessible*, as well as *high performing* infrastructure?



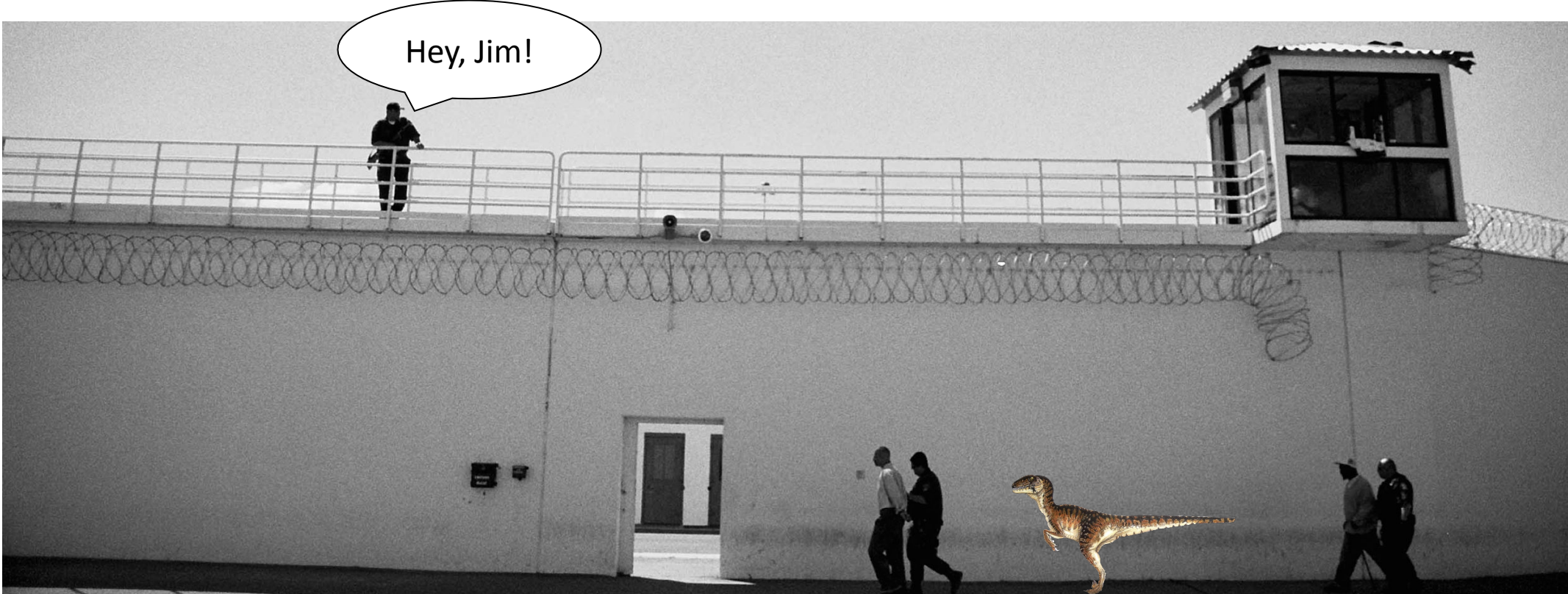
When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks

Gotta guy here.
Says he needs
stuff.



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks

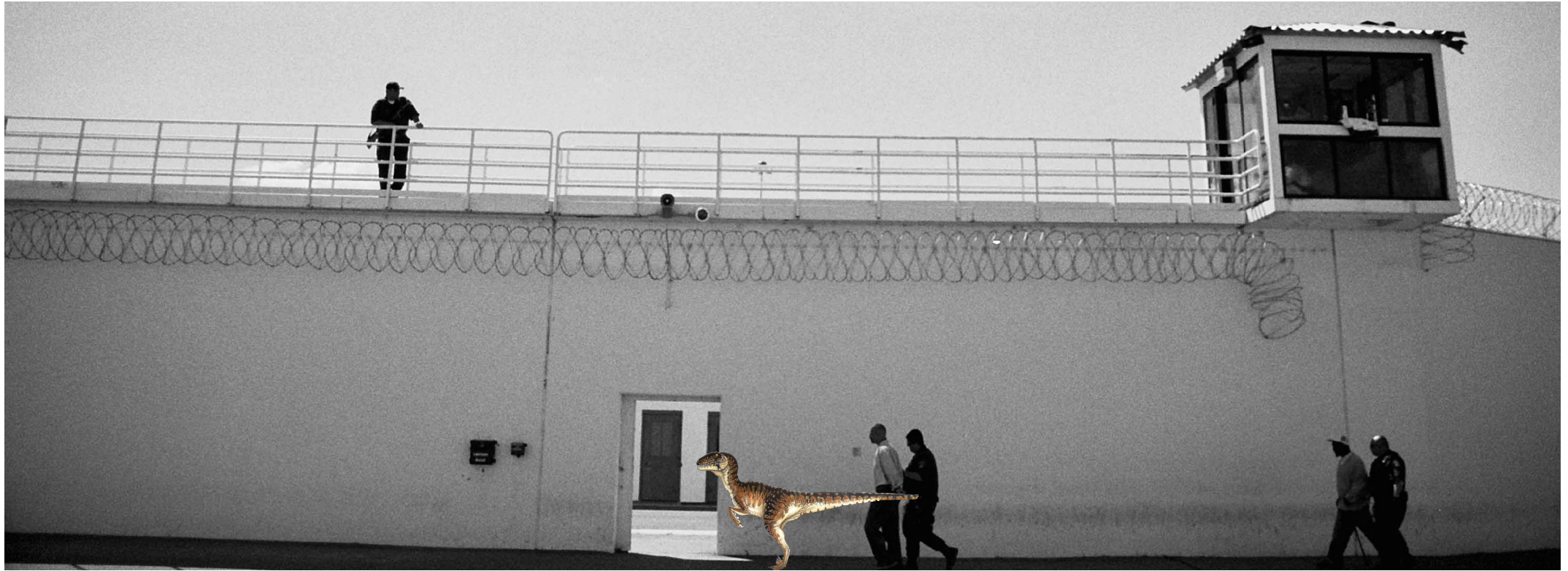
Something about data
and connectivity



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



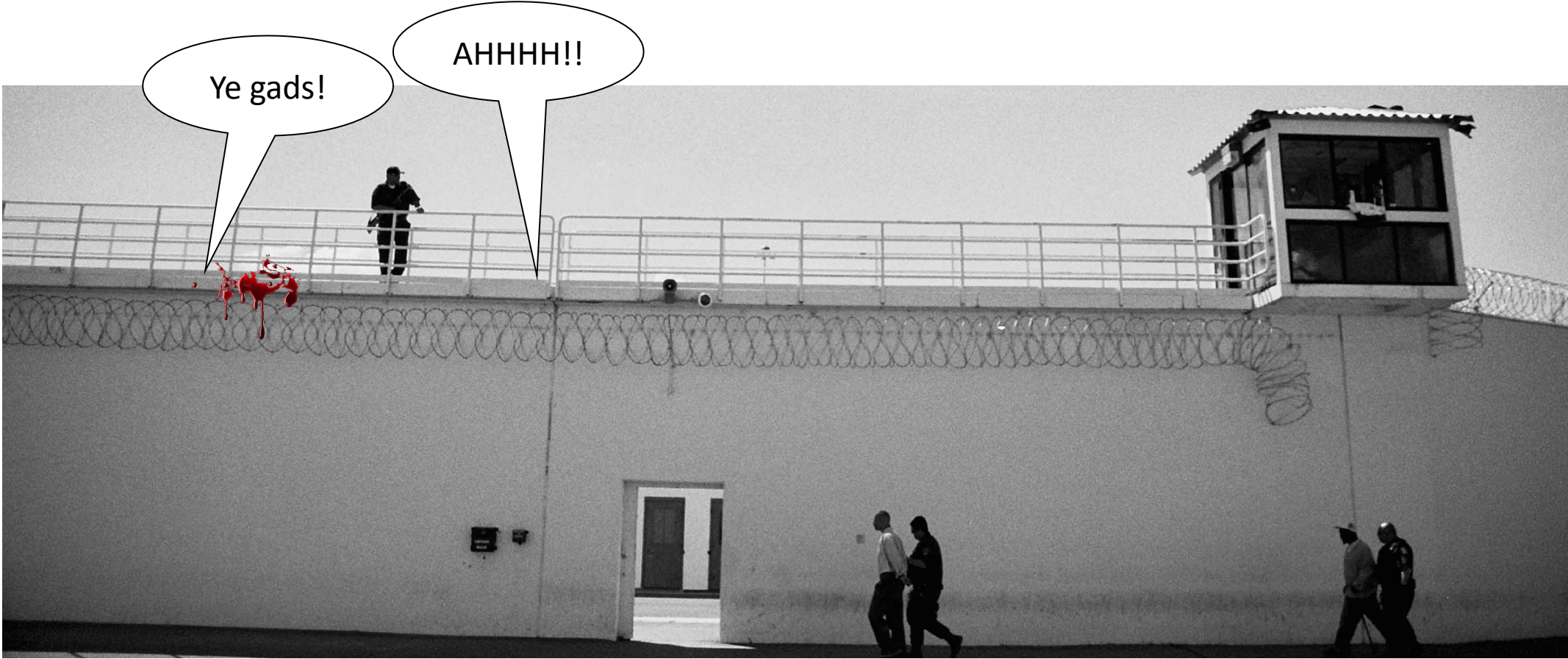
When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks



When Computational Science Meets Traditional Networks

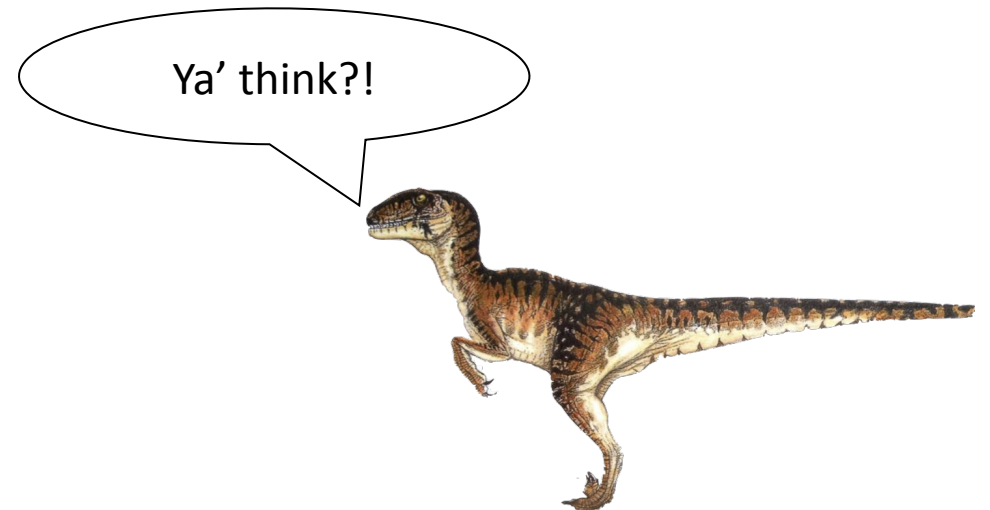


When Computational Science Meets Traditional Networks

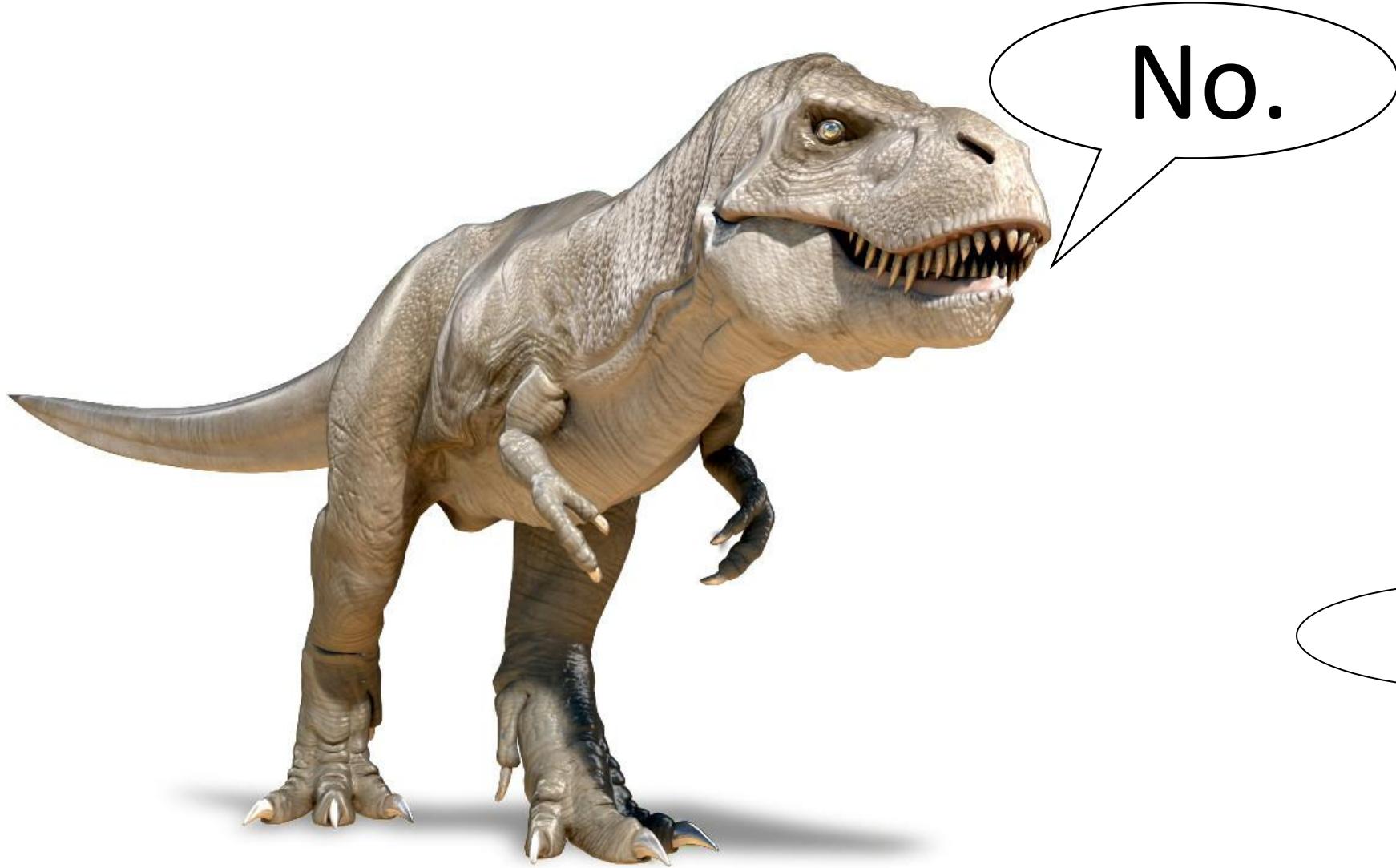


When Computational Science Meets Traditional Networks

OBSERVATION: The **requirements** of the computational researcher and the **service profile** of the traditional campus computer network (or other "commodity" networks) do not always align!



When Computational Science Meets Commercial Commodity Networks



When Computational Science Meets Commercial Commodity Networks



On second thought,
how much money yah
gots and how may friends
do you have I can eat, um,
I mean meet?

Eeek... Umm...



When Computational Science Meets Traditional Networks

This can result in adverse consequences:

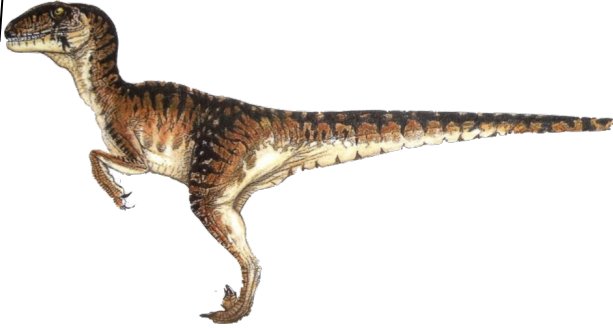
- Network performance issues for the researcher
- Network performance issues for everyone else
- Frustration for the researcher
- Frustration for IT staff



When Computational Science Meets Traditional Networks



Sigh. I guess cancer cures can wait.



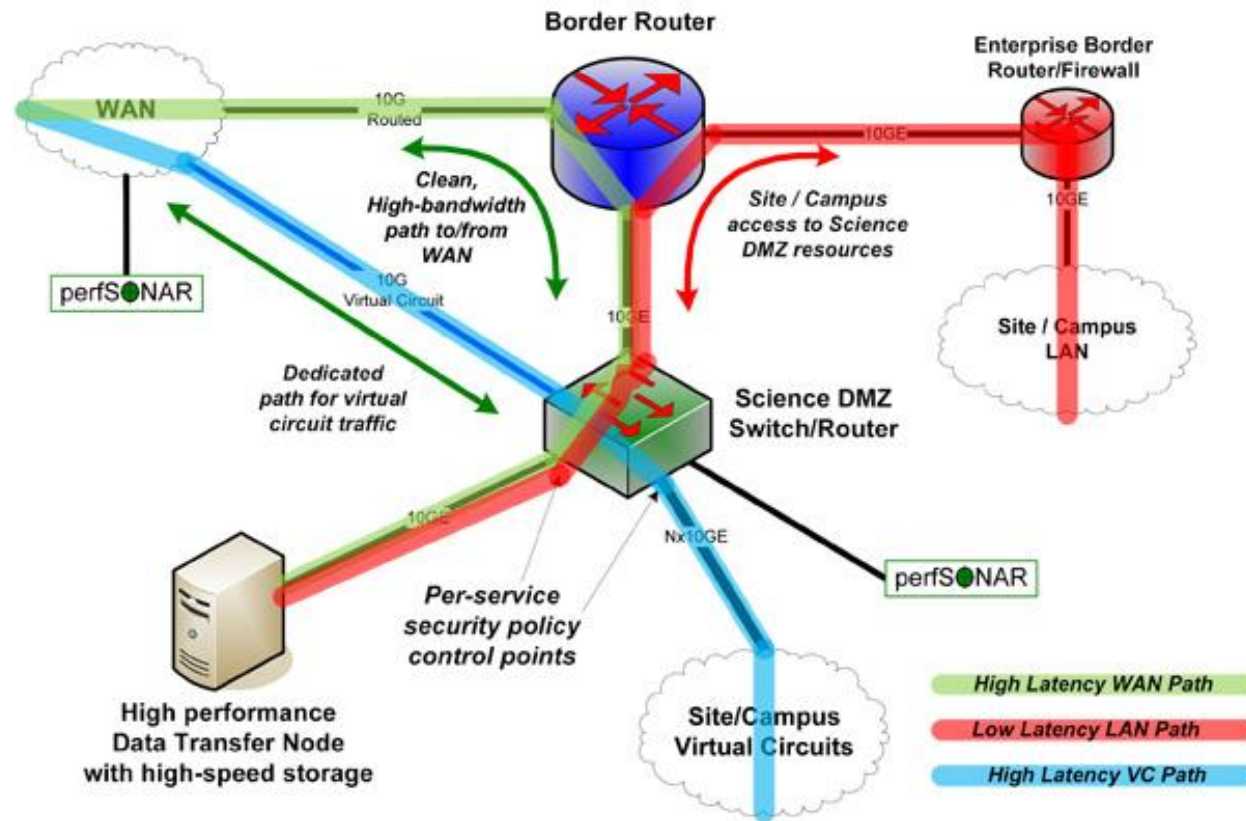
When Computational Science Meets Traditional Networks

But how do we overcome this? *I can't stop my research* just because **the network can't keep up!** Being able to collaborate is absolutely necessary!



Specialty networks to the rescue!

- Both internally to your organization and externally
- Science DMZ is an example of a specialty network

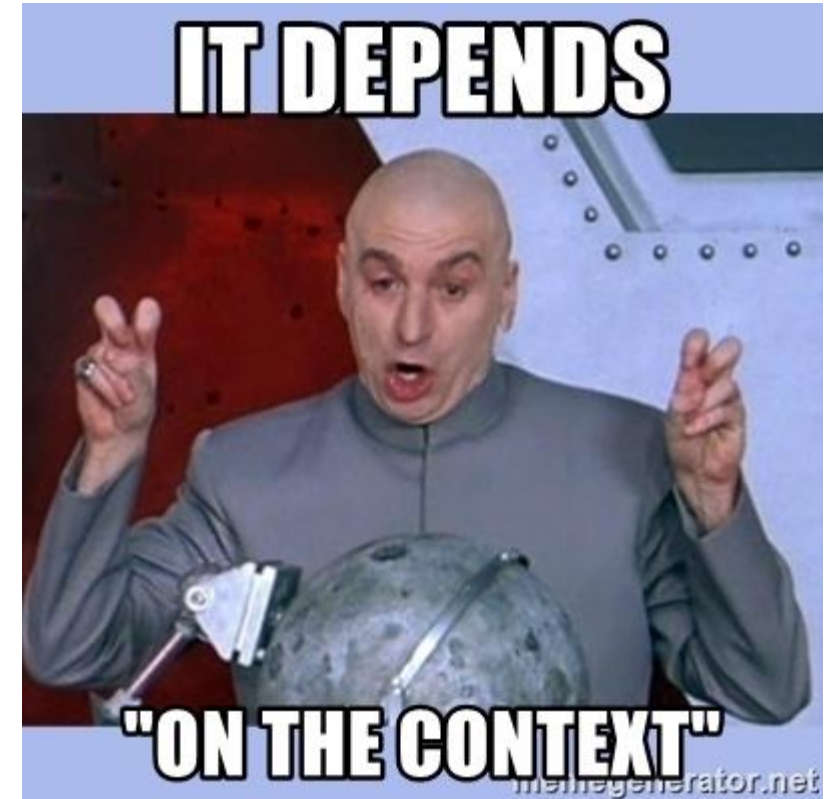


Does everyone need one?



Different SciDMZ architectures

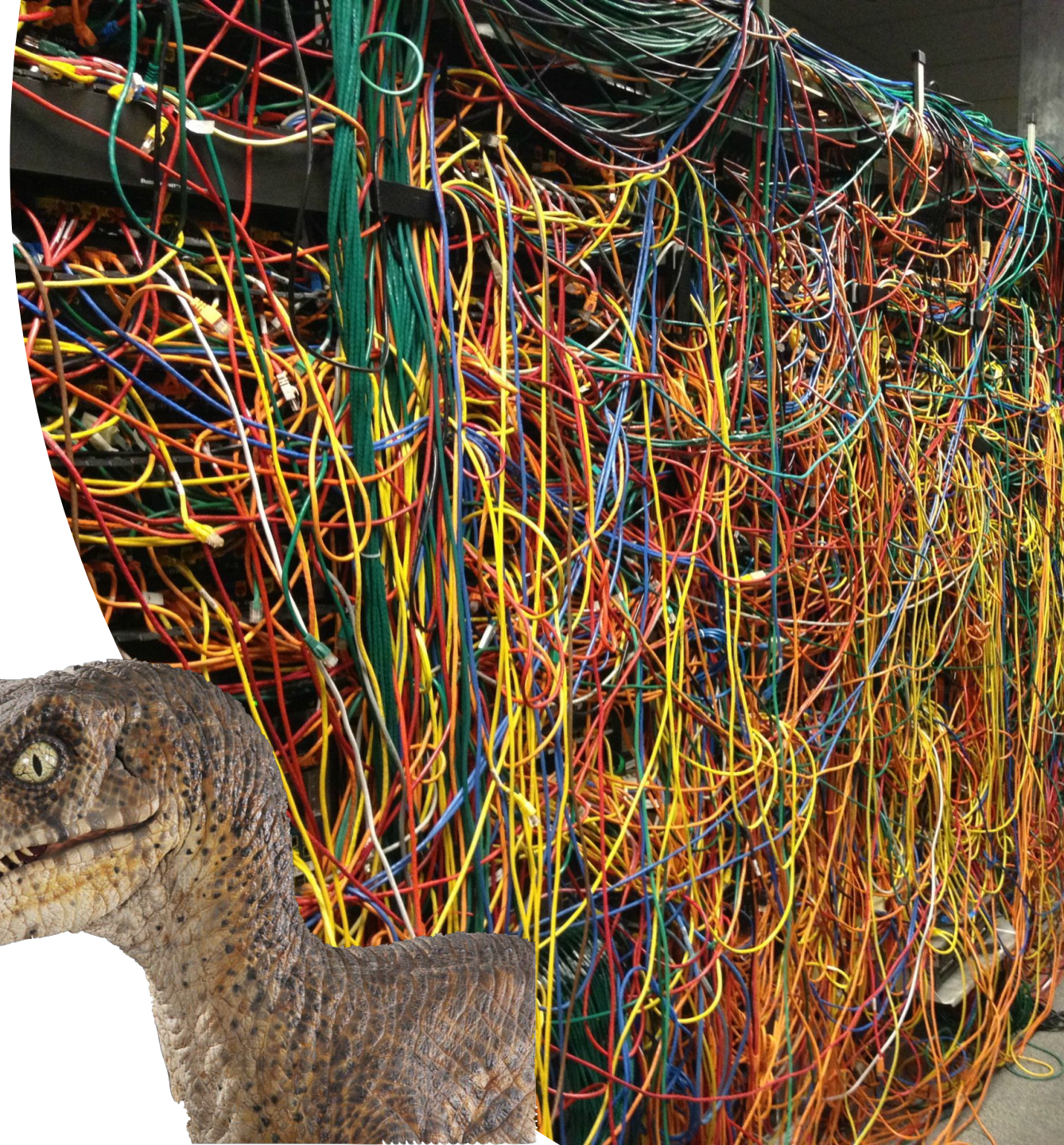
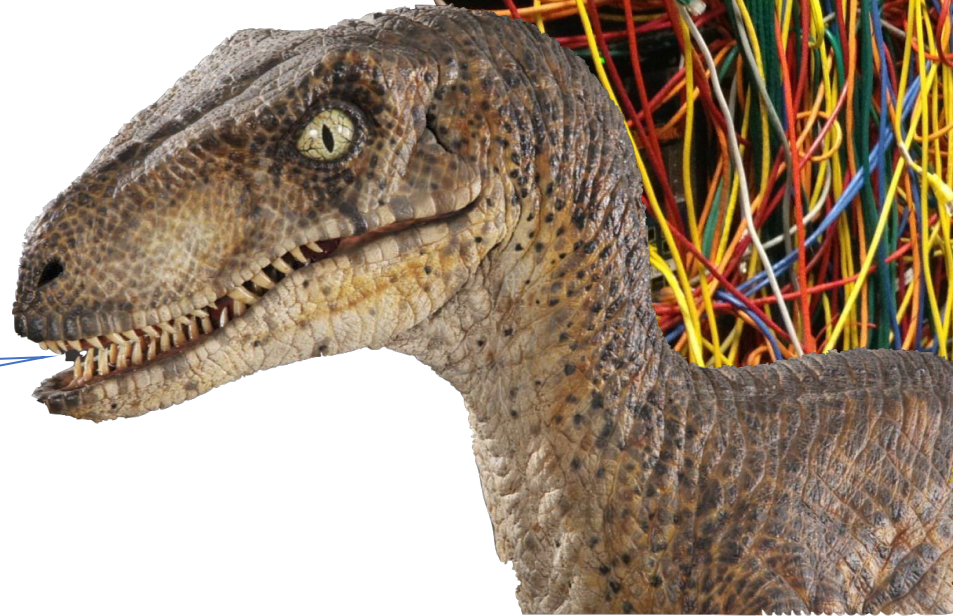
- Brute force method
- Apply money method
- Help me help you method
- Get fancy method
- Special snowflake method
- Come to me method



Brute force method...


- Fiber to the end user
- Build to demand
- Sustainability issues
- Bang for buck issues
- Looong lead time issues

You expect me to wait for how long?



Apply money...

- Full hardware designs
 - Sometimes just to the buildings
 - Sometimes all the way to the jacks
-
- Sustainability issues
 - Bang for buck issues
 - Loop anyone?

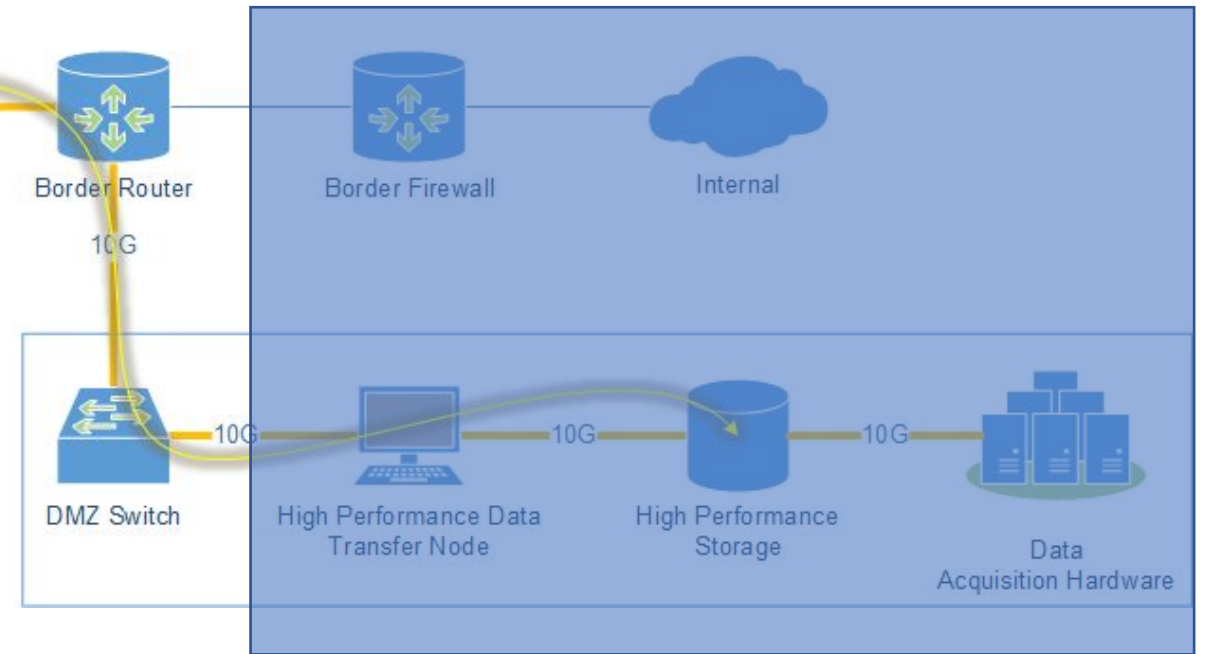
A realistic-looking dinosaur head, possibly a T-Rex, is shown in profile, facing left. It has a textured, scaly skin in shades of brown and grey. A blue speech bubble originates from its mouth, containing text. In the background, there is a green plant with large, palmate leaves in a red pot.

You expect me to fund what? Some kind of tubes?

Help me help you...

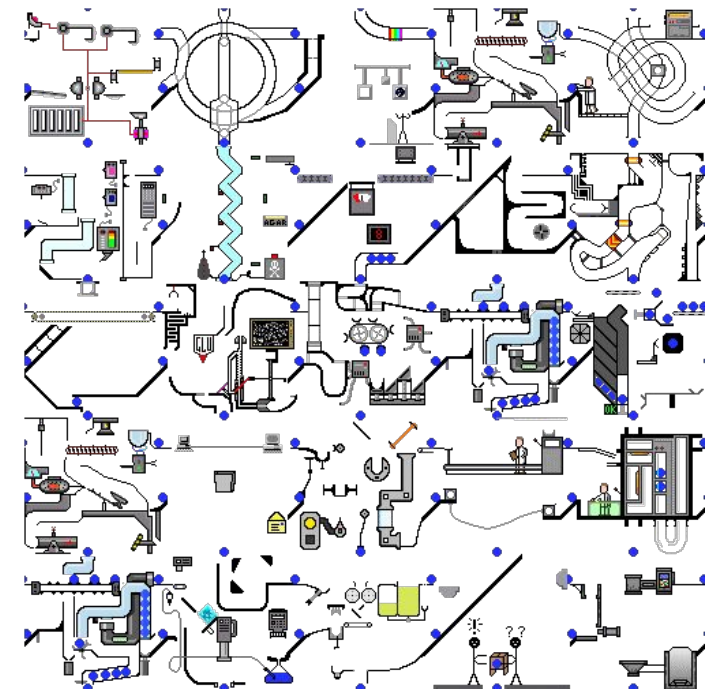
- Replacement of border routers with more powerful boxes
- Often a L2 switch in a central location
- Good starting point

- Is the science driver located where the DMZ is?
- What about the labs?
- Only a start



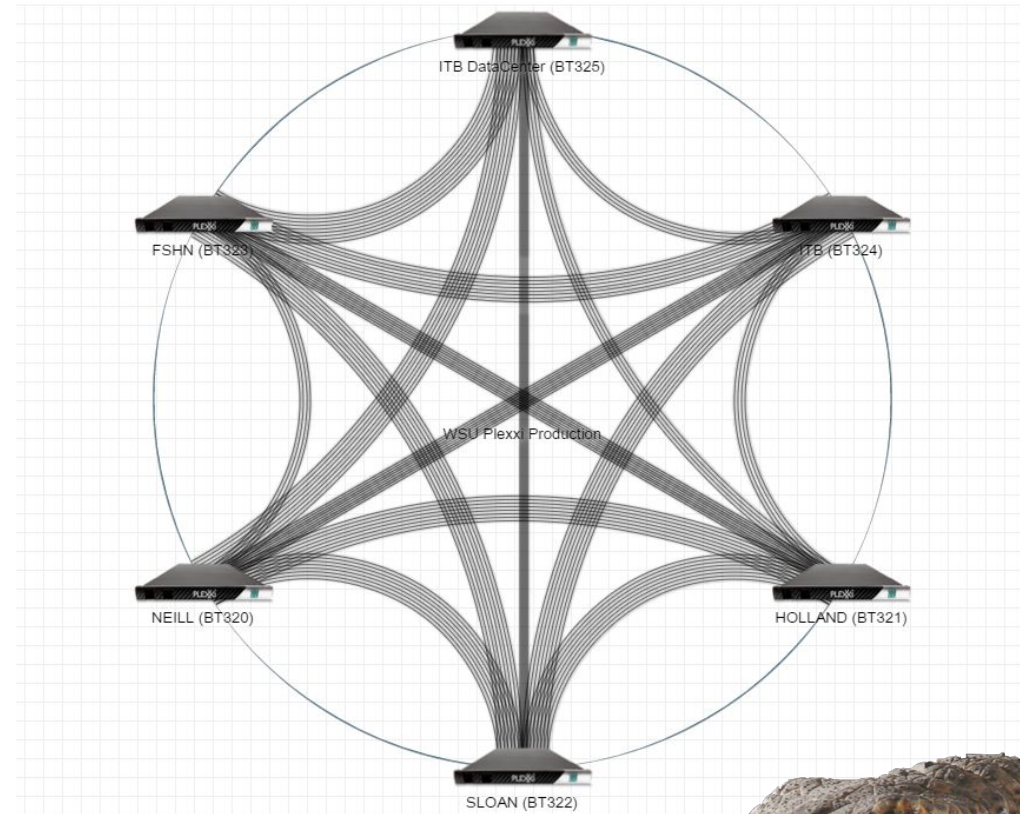
Get fancy...

- Virtualize all the things
- VRF or other such segmentation of existing network gear
- At its most basic – VLANs
- Low cost – as sustainable as the network is
- ~~Can be~~ is complex
- Will IT set this up in a reasonable amount of time?
- Congestion still an issue for large flows
- More readily sustainable



Special snowflake method...

- Unique design
- Can be overly complex
- Will IT set this up in a reasonable amount of time?
- Is it sustainable?
- Supporting research or doing research?

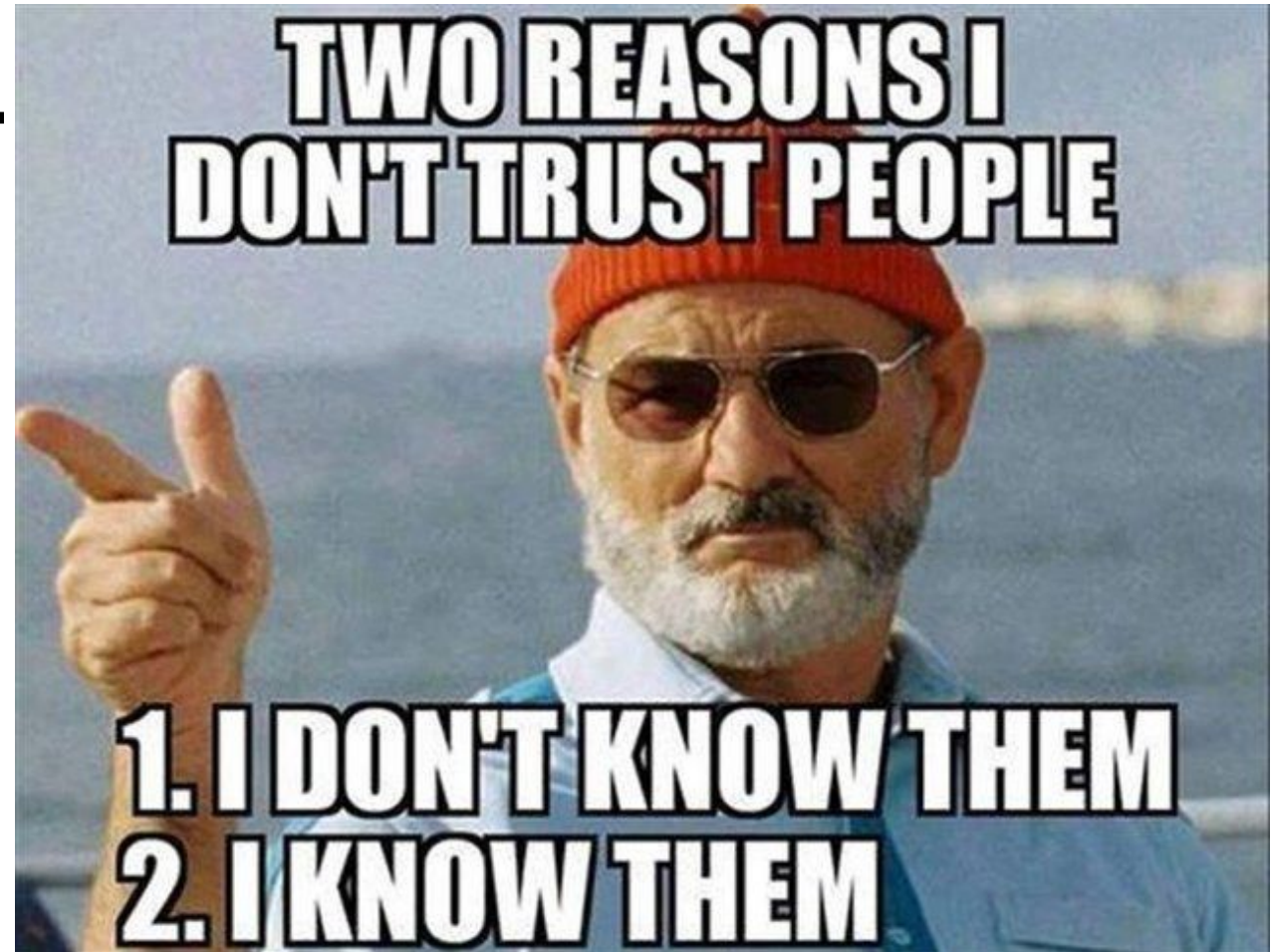


Ice ages have not gone well for my kind...

Come to me method...

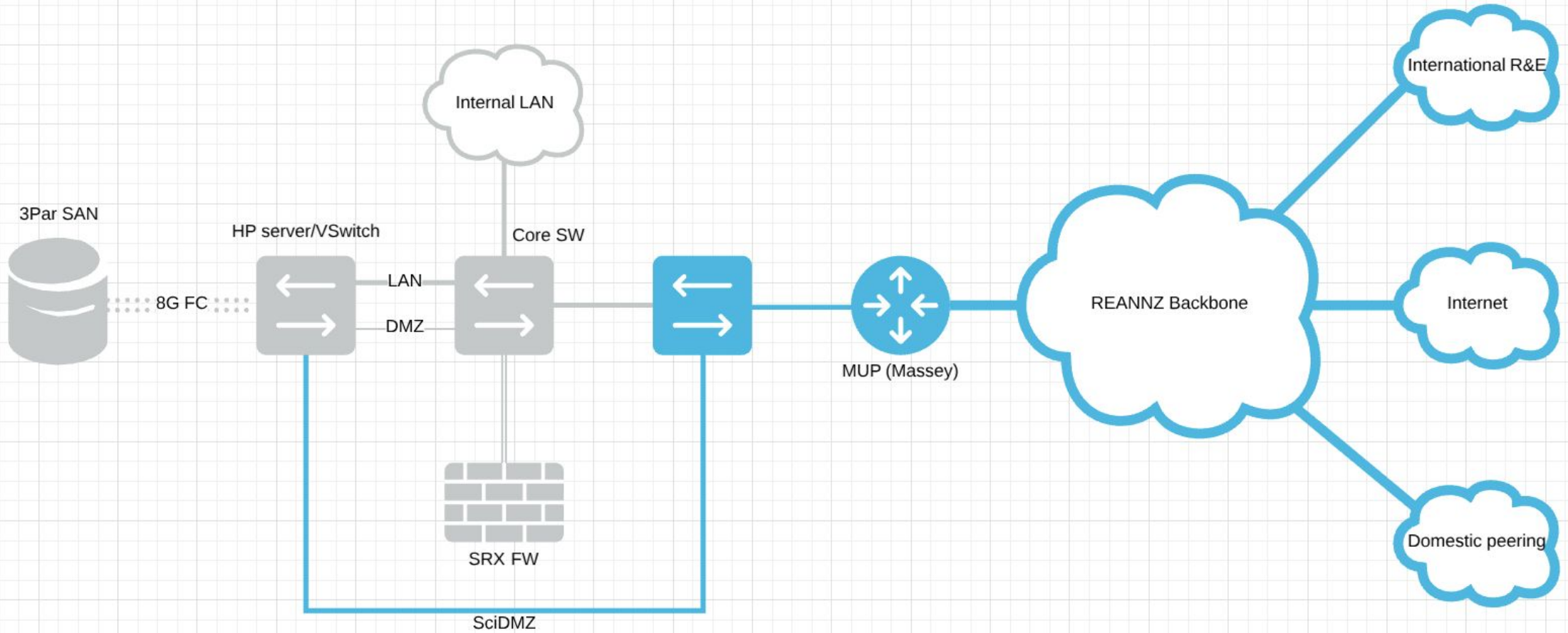
- Centralization!
- Less hardware and complexity

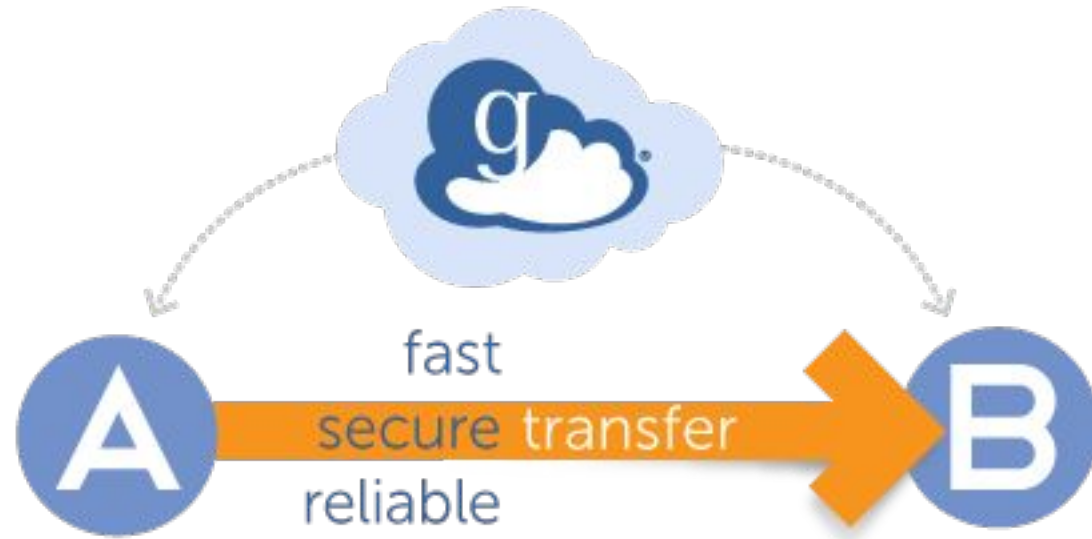
- Must have trust
- Will IT set this up in a reasonable amount of time?
- Is it sustainable?
- Start with one





Manaaki Whenua
Landcare Research







FILE MANAGER



BOOKMARKS



ACTIVITY



COLLECTIONS



GROUPS



CONSOLE



FLows



ACCOUNT



LOGOUT



HELP

Collection

Path

Start

Transfer & Timer Options

Start

Navigation bar with icons for home, back, refresh, settings, and menu.

	100G.dat	4/28/2016, 06:00 AM	100 GB
	100M.dat	4/28/2016, 06:00 AM	100 MB
	10G.dat	4/28/2016, 06:00 AM	10 GB
	10M.dat	4/28/2016, 06:00 AM	10 MB
	1G.dat	4/28/2016, 06:00 AM	1 GB
	1M.dat	4/28/2016, 06:00 AM	1 MB
	500G.dat	4/28/2016, 06:08 AM	500 GB
	500GB-in-large-files		

- Share
- Transfer or Sync to...
- New Folder
- Rename
- Delete Selected
- Download
- Open
- Upload
- Get Link
- Show Hidden Items
- Manage Activation

Search for a collection to begin

Get started by taking a short tour.

ESnet Fasterdata Knowledge Base

An Expert Guide for End-to-End Performance Tuning, Tools and Techniques

The Fasterdata Knowledge Base provides proven, operationally sound methods for troubleshooting and solving performance issues. Since 1986 ESnet has operated an advanced research network with the goal of enabling the highest levels of performance for the Department of Energy (DOE) scientific community. During this time, our engineers have identified a common set of issues that hinder performance. We share our experiences and findings in this knowledge base.

Our solutions fall into five categories:

- Network Architecture, including the [Science DMZ](#) model
- [Host Tuning](#)
- [Network Tuning](#)
- [Data Transfer Tools](#)
- [Network Performance Testing](#)

How long should it take to transfer a TeraByte of data across your network? It is probably less time than you think. Check out our [performance expectations guide](#).

Want to contribute material or know of something that should be on the site? Send an email to fasterdata@es.net.

Linux Tuning

- 100G Host Tuning
- TCP Tuning

Network Troubleshooting

- Tools
- Techniques

Science DMZ

- Architecture
- Data Transfer Nodes

Data Transfer Tools

- Data Transfer Tools Overview
- Globus

NSF Info

NSF International Research and Education Network

Weak links break file transfers...

- Know your provider and leverage them for help!
- Use <https://fasterdata.es.net/>
- Networks come in many shapes and sizes
- Networks interconnect to make more networks
- Networks get exponentially complex the more connections you have

If your data has to transit it - its “your” network!

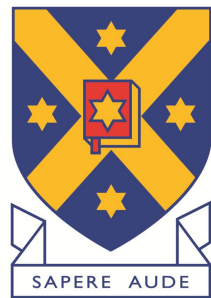
You need to know who to go to for help!

Thank you for your time today!

Wallace A. Chase

Head of Department, ITS
wallace.chase@otago.ac.nz

@bmtfr



UNIVERSITY
of
OTAGO

Te Whare Wānanga o Otāgo
NEW ZEALAND

