# CS484 Lecture 6
# The Network Layer

Fall 2022
Joshua Reynolds

# Layer 3 - Networking vs Switching

Layer 1: Physical connections and signalling

Layer 2: Links between nodes

Layer 3: Finding the best path across the world to specific machines

1. Data “Plane” – Sending Data
2. Control “Plane” – Planning Routes for forwarding Data towards its destination

# Routers

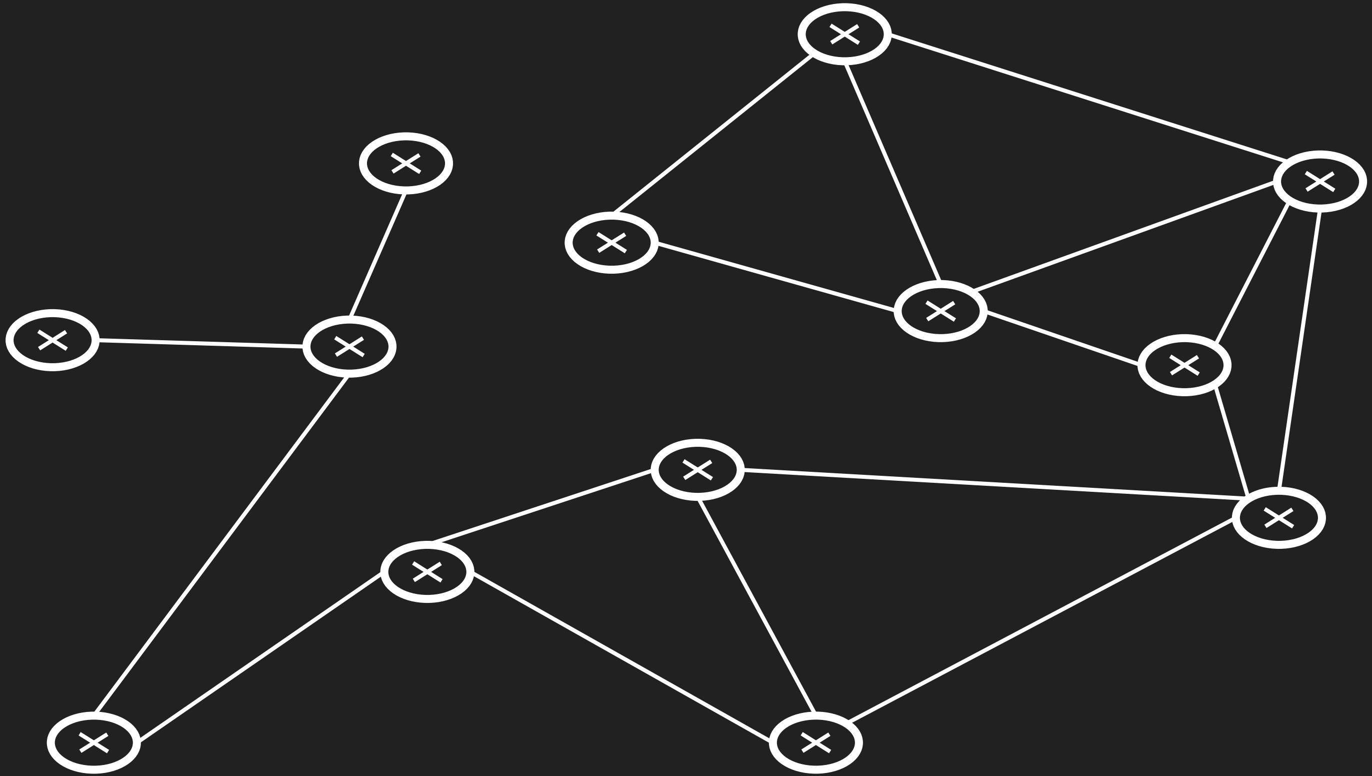
Relays in the internet

Bridges between layer 2 networks

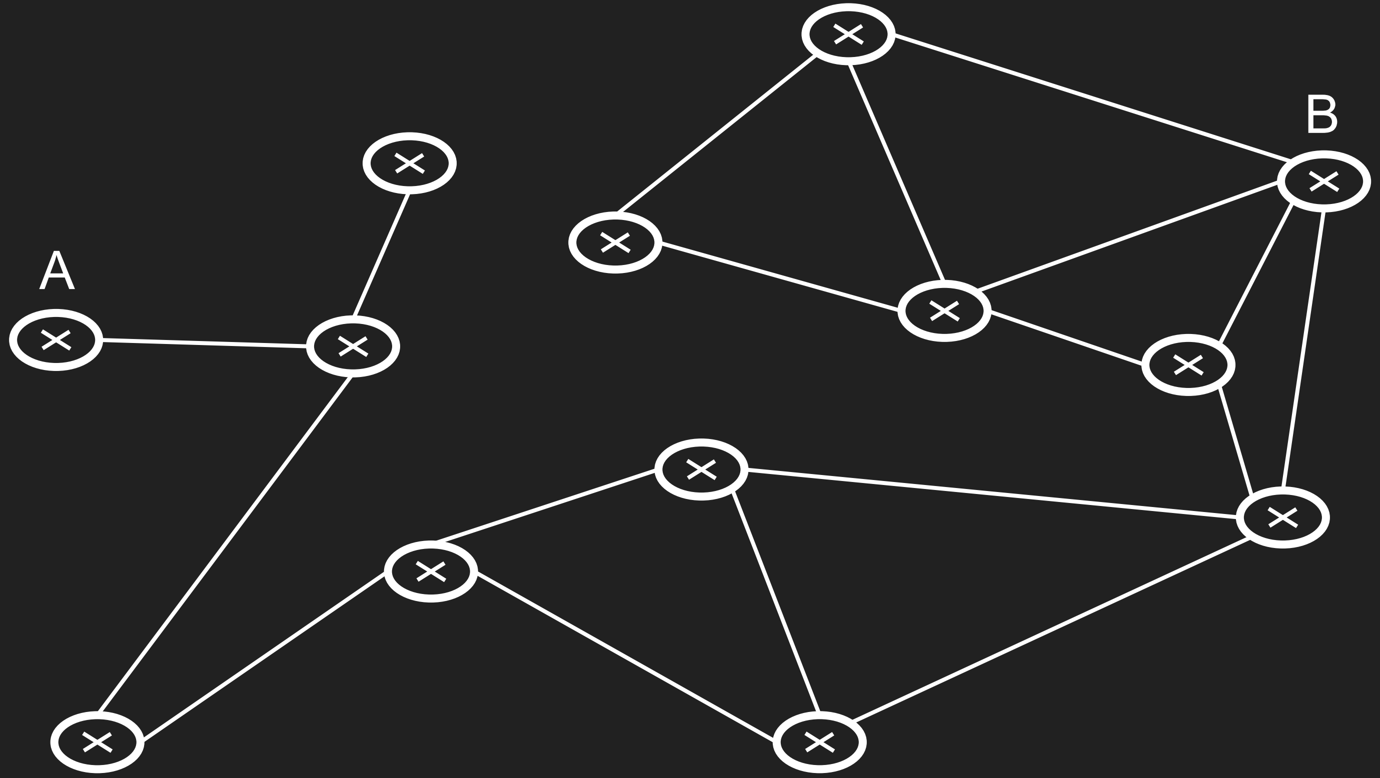
Need to be able to forward data to whatever destination address is on the data

Needs to evaluate which layer 2 link is the best one to send data on so it eventually arrives at its destination as efficiently as possible.

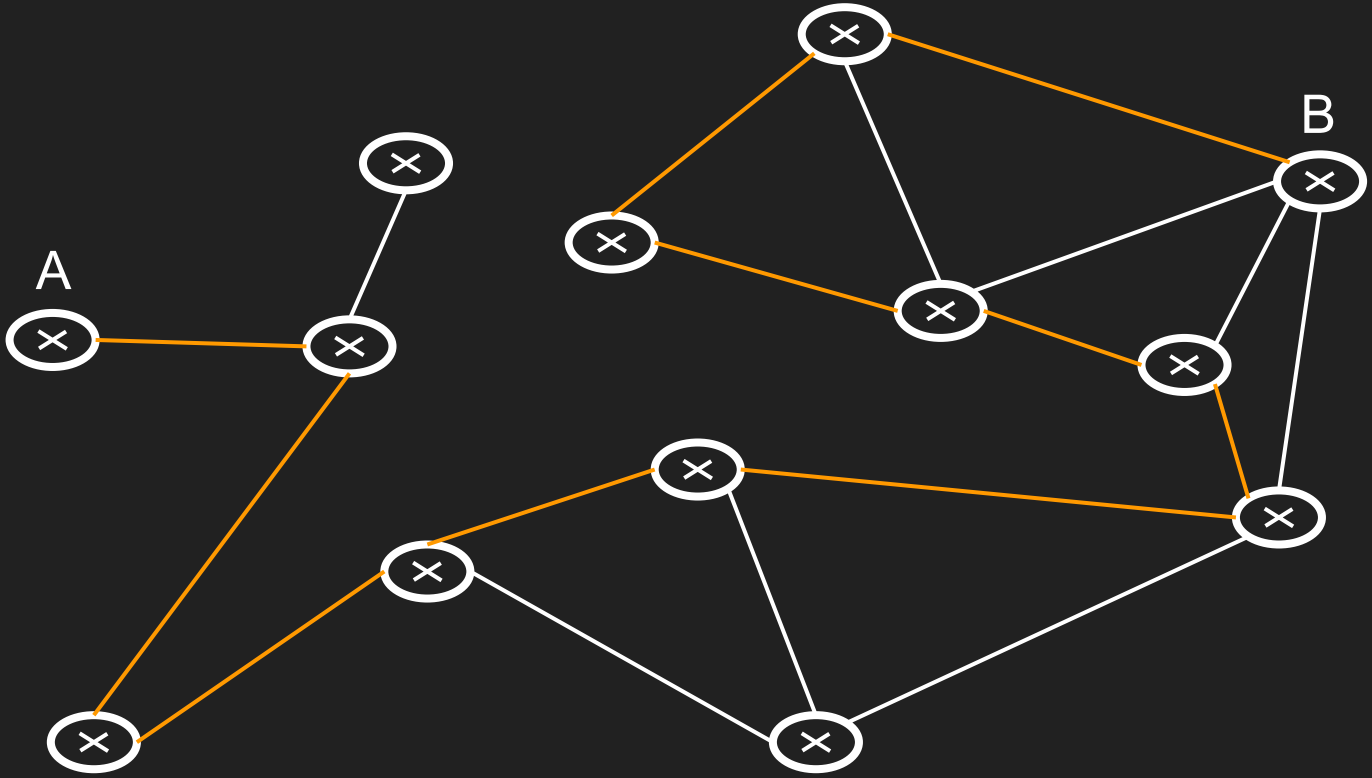
# Networks as Graphs



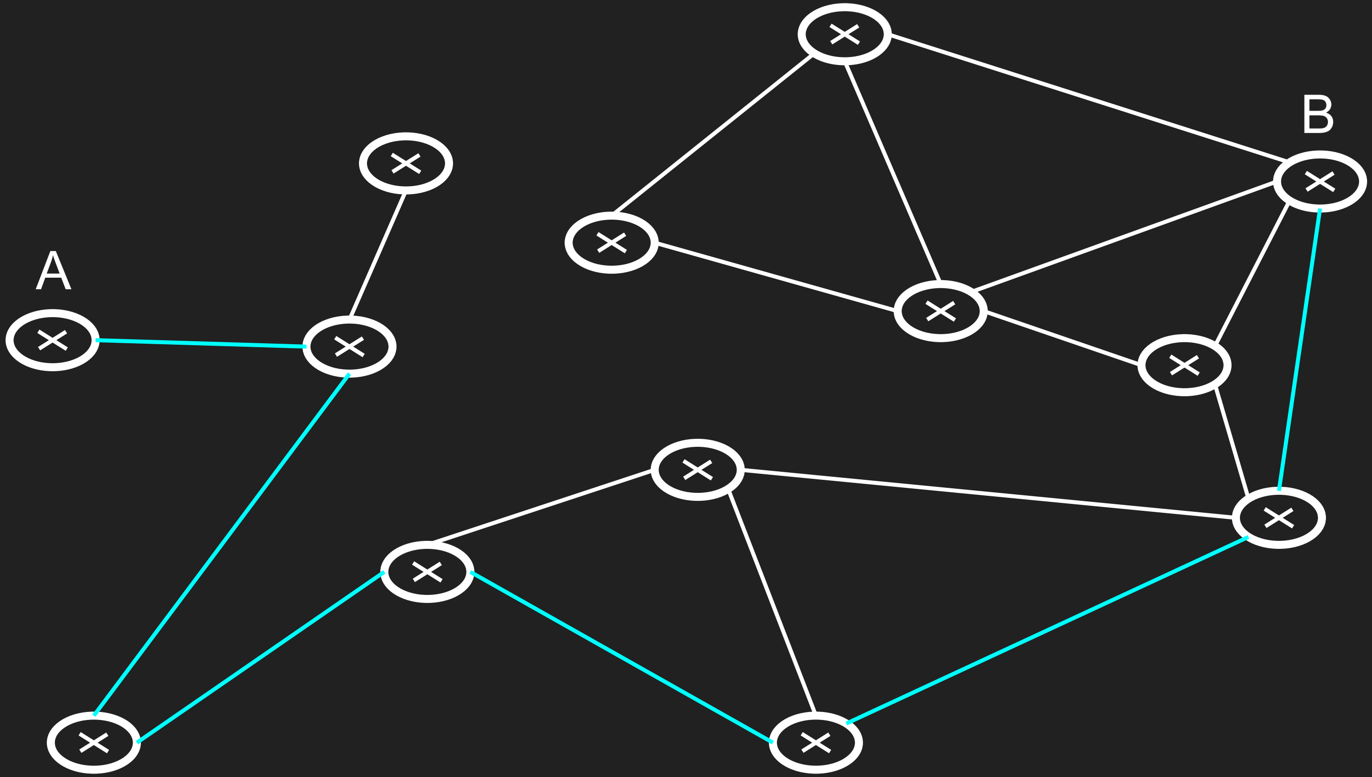
# Networks as Graphs



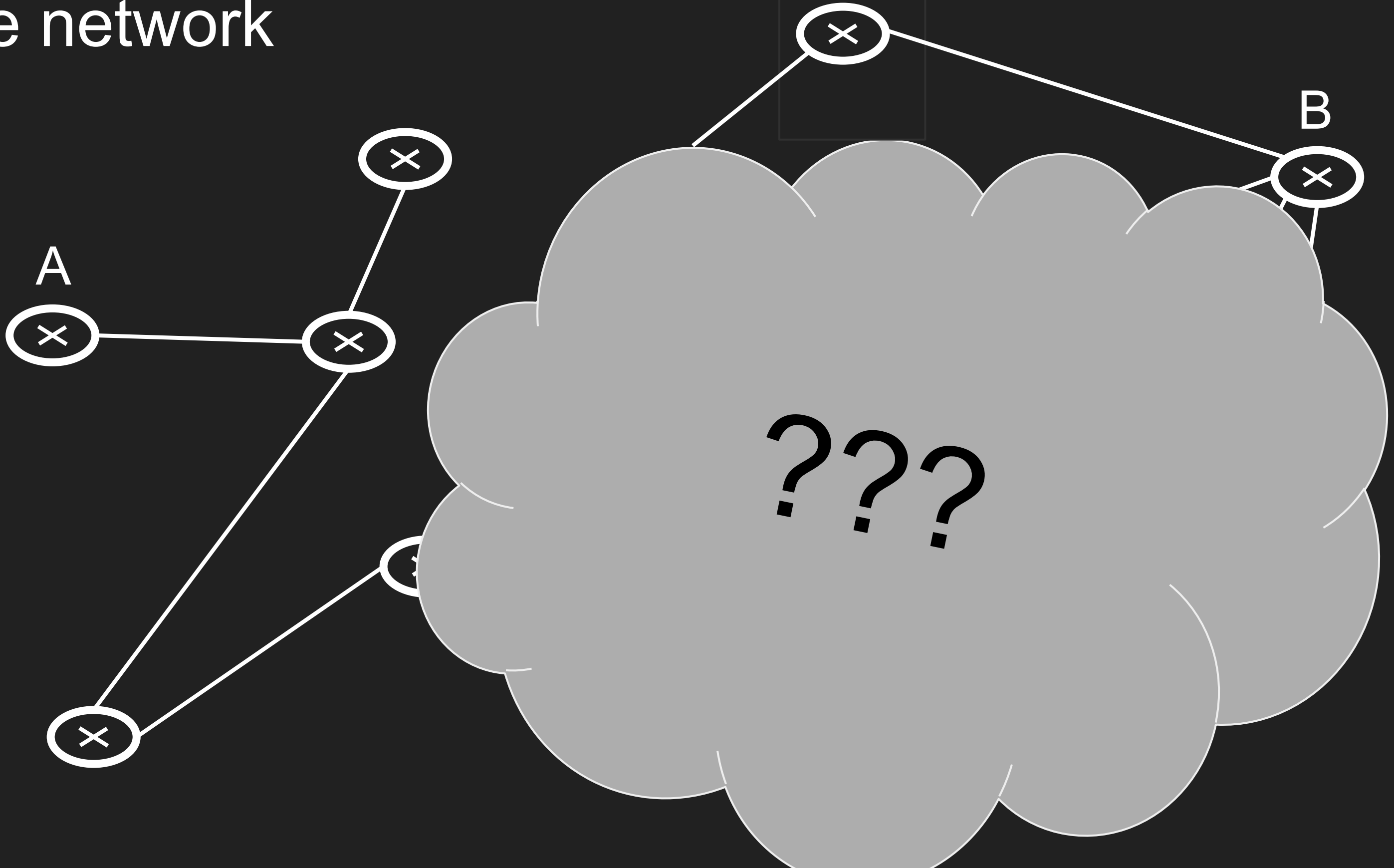
# Networks as Graphs: Inefficient



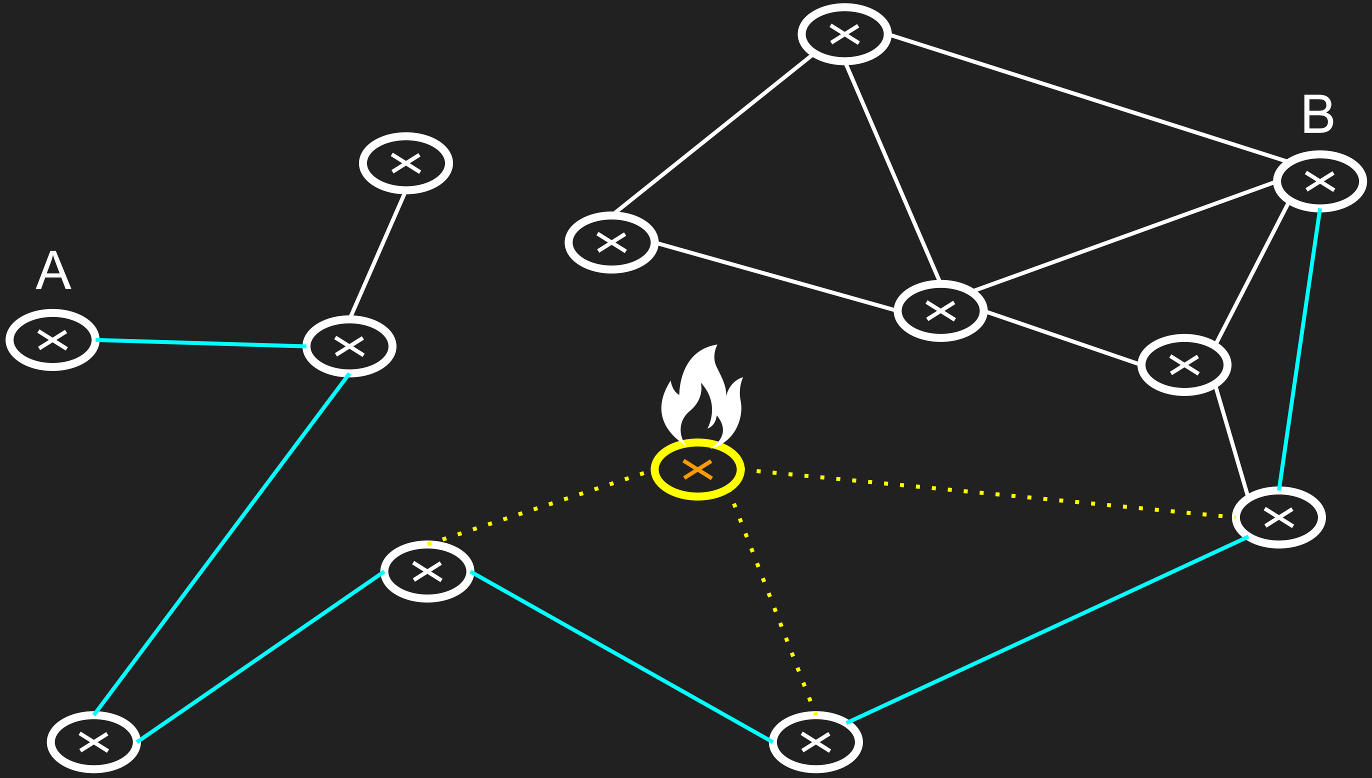
# Networks as Graphs: Efficient



Need to find efficient routes even when A cannot see the whole network



# Needs to handle losses of network components



# Networking Algorithms: Math Terms

We call the network graph “**G**”

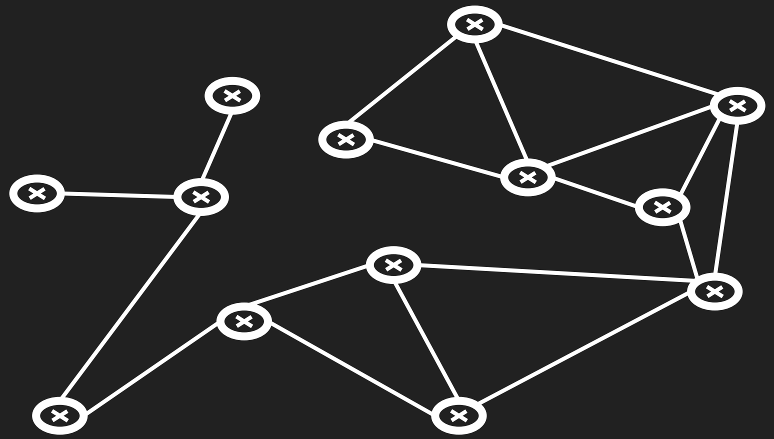
Networks are made up of **Nodes** and **Edges**

We call the set of all nodes “**N**”

We call the  set of all edges “**E**”

All together the network graph **G** can be mathematically described as:

**G = (N, E)**

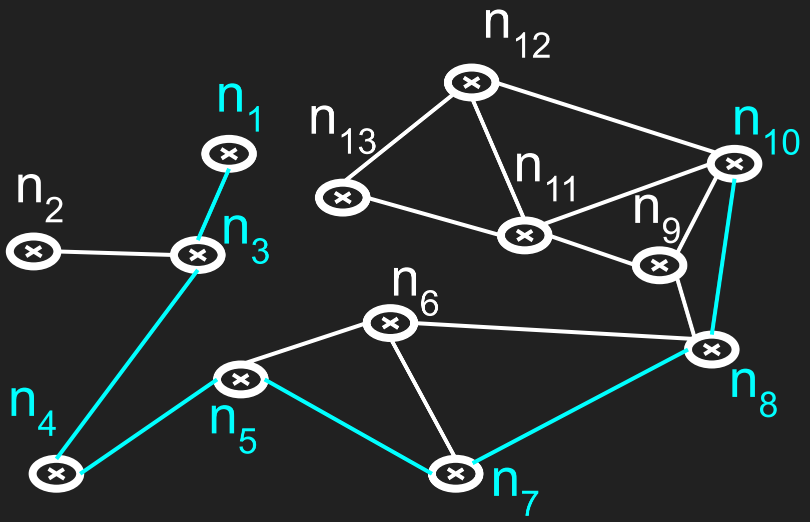


# Networking Algorithms: Math Terms

**$G = (N, E)$**

A **path** is an ordered list of nodes $\in N$

$P = (n_1, n_3, n_4, n_5, n_7, n_8, n_{10})$



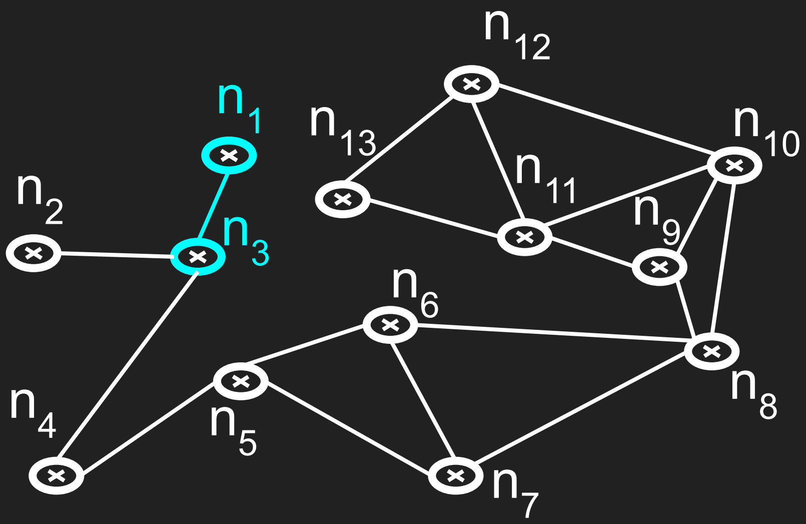
# Networking Algorithms: Math Terms

**G = (N, E)**

A **path** is an ordered list of nodes ∊ N

$P = (n_1, n_3, n_4, n_5, n_7, n_8, n_{10})$

Every edge can be described by its two endpoints $(n_1, n_3)$



# Networking Algorithms: Math Terms

**G = (N, E)**

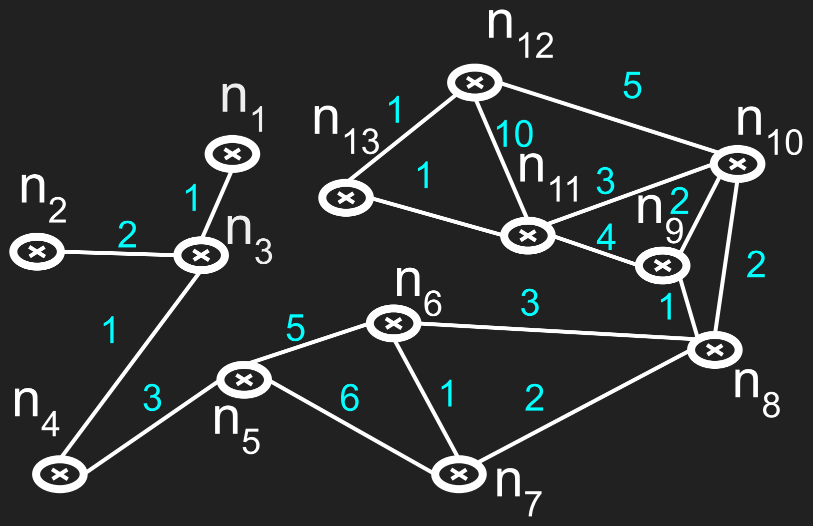
A **path** is an ordered list of nodes $\in$ N

$P = (n_1, n_3, n_4, n_5, n_7, n_8, n_{10})$

Every edge can be described by its two endpoints $(n_1, n_3)$

Each edge has a “cost”. This cost is an abstraction of latency, congestion, or error frequency.

We write this edge cost as $c(n_1,n_3)$



# Networking Goals:  Shortest Path

Shortest Path – it is more complicated with costs considered

# Networking Goals: Lowest-Cost Path

Shortest Path

Lowest-Cost Path

# Networking Goals: Lowest-Cost Path

Shortest Path

Lowest-Cost Path – But people join, leave, and rearrange the network!

# Networking Goals: Lowest-Cost Path

Shortest Path

Lowest-Cost Path – But people join, leave, and rearrange the network!

We call this problem “churn”

# Networking Goals: Adaptive Lowest-Cost Path

Shortest Path

Lowest-Cost Path – But people join, leave, and rearrange the network!

We call this problem “churn”

We need to constantly adapt and keep finding the lowest-cost path from our source to our destination

# Networking Goals: Adaptive Lowest-Cost Path

Shortest Path

Lowest-Cost Path – But people join, leave, and rearrange the network!

We call this problem “churn”

We need to constantly adapt and keep finding the lowest-cost path from our source to our destination

# Centralized vs Decentralized Routing

**Centralized:**

One central system knows the whole network topology and assigns routes.

Works in datacenters where one entity controls anything.

**Decentralized:**

No system sees the whole network topology.

Works in the Internet where people who do not trust each other can work together without sharing information.

No one group can control the Internet for everyone else

Resilient to network damage and churn - Cold War roots

# Inter-Network Network History: ARPANET

The Advanced Research Projects Agency NETwork
Late 1960’s through 1980’s

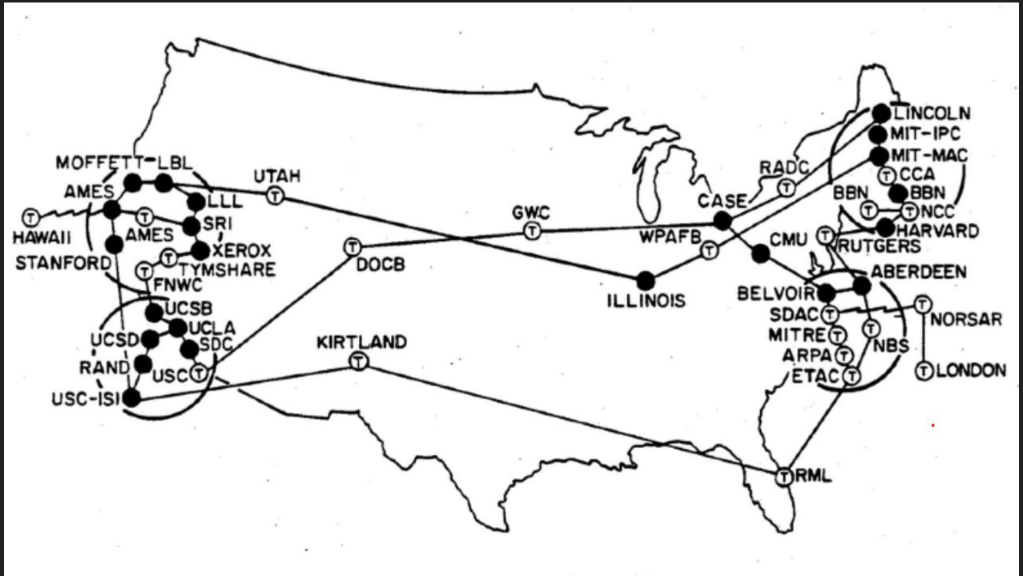


Image by Semaforo GMS
CC BY-SA 4.0

# 1974: Arthur C. Clarke Interview on the “future” (2001)

Science Fiction Author of screenplay for movie: “2001, A Space Odyssey”

https://en.wikipedia.org/wiki/File:ABC_Clarke_predicts_internet_and_PC.ogv

# Trace Route Activity

```
traceroute 79.171.97.215
```

noc.checs.net/

conterra.com/wp-content/uploads/2018/09/South-West-Region-El-Paso.pdf

he.net/3d-map/

www.siminn.is/

basis.is

# Autonomous Systems (ASes)

A useful network that can help you get to other networks

New Mexico CHECS is AS #3912 as assigned by the IANA

Google is AS #15169

Comcast is AS #7922

Each of these subnets got assigned AS numbers because they have the capacity and popularity to be a major player in the Internet

They have Layer 2 connections to each other at Internet Exchange Points (IXPs)

# BGP Gossip Demo

# BGP Attack Gone Wrong

“...in 2008, a Pakistani ISP attempted to use a BGP route to block Pakistani users from visiting YouTube. The ISP then accidentally advertised these routes with its neighboring ASes and the route quickly spread across the Internet’s BGP network. This route sent users trying to access YouTube to a dead end, which resulted in YouTube’s being inaccessible for several hours.”

https://www.cloudflare.com/learning/security/glossary/what-is-bgp/

# HW1, Q4 - Design your own 1D PAM4 Line Encoding Rules

# Internet’s “Best-Effort Service”

Guaranteed Delivery

Guaranteed Delivery Time

In-Order Delivery

Guaranteed No Packets Dropped @ Minimum Bitrate

End-to-End Encryption Between Layer 3 Source and Destination

# Internet’s “Best-Effort Service”

~~Guaranteed Delivery~~

~~Guaranteed Delivery Time~~

~~In-Order Delivery~~

~~Guaranteed No Packets Dropped @ Minimum Bitrate~~

~~End-to-End Encryption Between Layer 3 Source and Destination~~

No Guarantees ¯_(ツ)_/¯

# Internet Protocol (IP) Address

Version 4:

4 bytes long: 143.23.53.254

Version 6:

16 bytes long: 2001:0db8:0000:0000:0000:ff00:0042:8329

You can omit writing sections with zeros: 2001:db8::ff00:42

# Internet Protocol (IP) Address

Version 4:

4 bytes long: 143.23.53.254 (4,294,967,296 options)

Version 6:

16 bytes long: 2001:0db8:0000:0000:0000:ff00:0042:8329

You can omit writing sections with zeros: 2001:db8::ff00:42

($3.8*10^{38}$ options)

# Internet Protocol (IP) Address

Version 4:

4 bytes long: 143.23.53.254 (4,294,967,296 options)

Version 6:

16 bytes long: 2001:0db8:0000:0000:0000:ff00:0042:8329

You can omit writing sections with zeros: 2001:db8::ff00:42

($3.8*10^{38}$ options)

**PROBLEM:** There are more devices in the world than IPv4 addresses

# Subnet Masks and CIDR Notation

Methods of describing a subnet (small part of the network) that exists within an IP address range.

Subnet Mask:

All bits set in the mask are going to be the same on this subnet

ex: 256.256.256.0

Classless Inter-Domain Routing CIDR

Adds a forward slash to say how many bits of the total 32 are fixed.

For example:

128.123.63.0/24 - the range from 128.123.63.0 thru 128.123.63.255

# IPv4 Class A, B, and C networks

Class A - 8 bits set, 24 bits free (huge address space)

*OG internet orgs USPS, US DOD, GE, AT&T (2013 caida.org)*

Class B - 16 bits set, 16 bits free (some ISPs need this many)

*At $50 per address, these are worth millions of dollars*

Class C - 24 bits set, 8 bits free

*NMSU CS Department has 128.123.64.0/24 and 128.123.63.0/24*

*Assigned by New Mexico Council for Higher Education Computing/Communication Services (CHECS)*

https://www.caida.org/archive/id-consumption/census-map/images/2013-hilbert-plot.png

# ipconfig/ifconfig vs. “what’s my IP?” Demo

What is your gateway’s IP address?

Is it shown as a local IP or a global IP?

# Network Address Translation (NAT)

A way to “stretch” IPv4 address space

Many router clients share their gateway router’s IP address externally.

Internally, the router gives every machine a local IP address

The router keeps track of who starts outgoing connections, and makes sure the replies go to the right place.

Keeping track of these conversations is actually a layer 4 task

# Special IP Address Ranges to Know

127.0.0.1/32 – Reserved for IPv4 lookback interfaces

10.0.0.0/8 – Class A sized space for local networks

192.168.0.0/16 - Class B sized space for smaller networks

0.0.0.0 - Me, or IDK, or ERROR

fc00::/7 - IPv6 local networks - Unique Local Addresses

::1 - IPv6 loopback address

:: - Me, IDK, or ERROR

# Forwarding Strategies

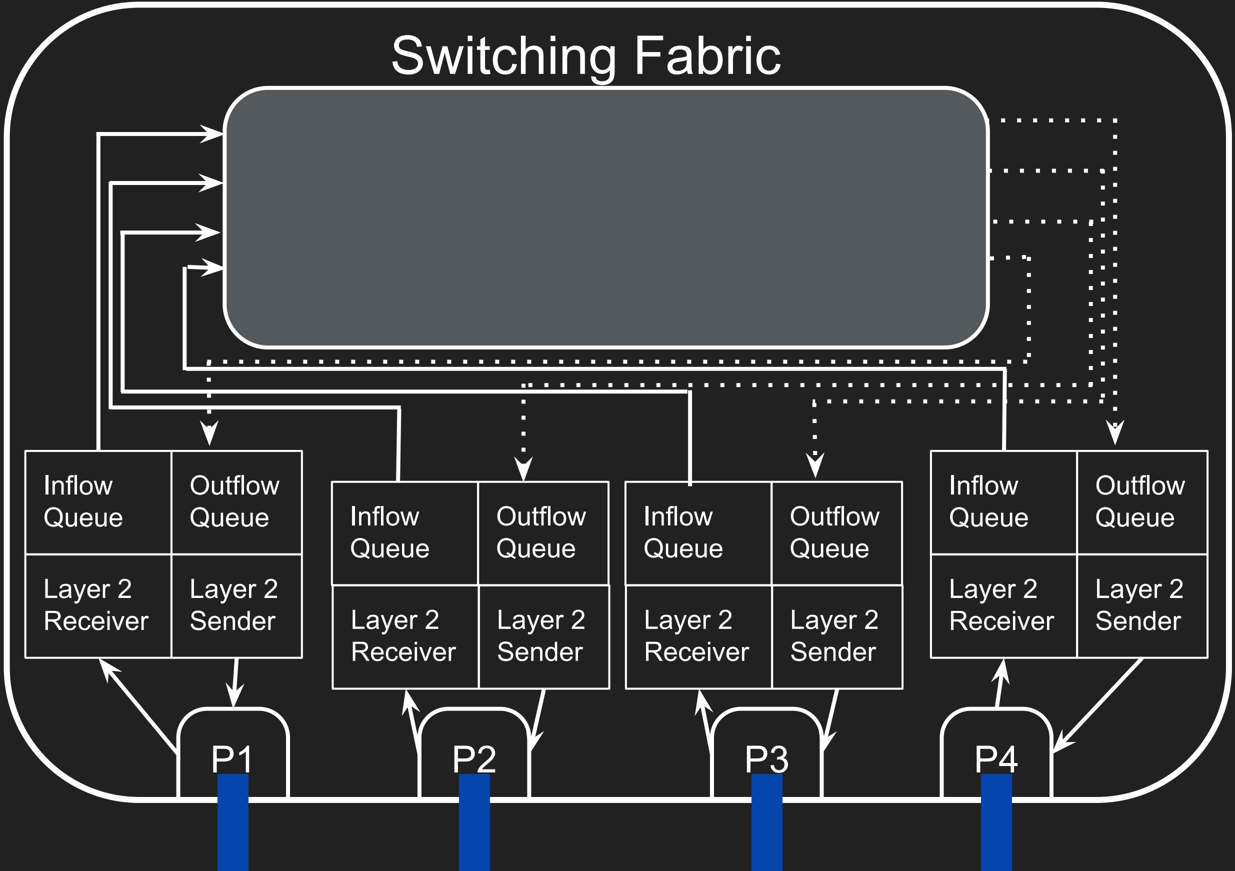
**Generalized Forwarding**

- Brainless rules like input on port 5 should be sent out on port 7
- Faster than consulting a processor with a routing table

**Destination-Based Forwarding**

- Requires the router’s processor to look up where to forward data to
- Uses a routing table
  - Uses ranges to avoid needing a huge table.

# How Data Flows in a Router



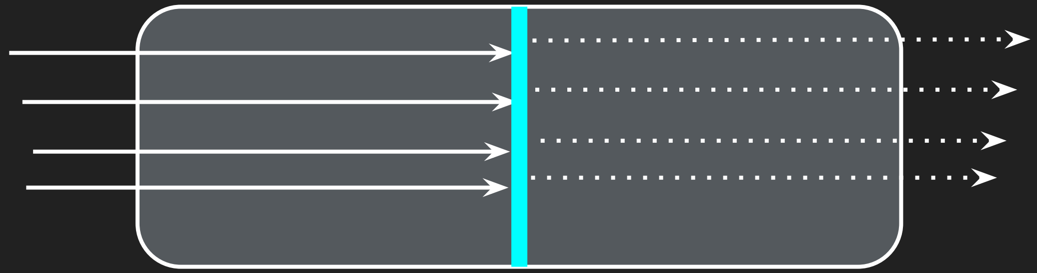
# Bus Switching Fabric

Every port’s input and output are connected on a single bus

Messages are heard by every output port, but are labeled with an output port

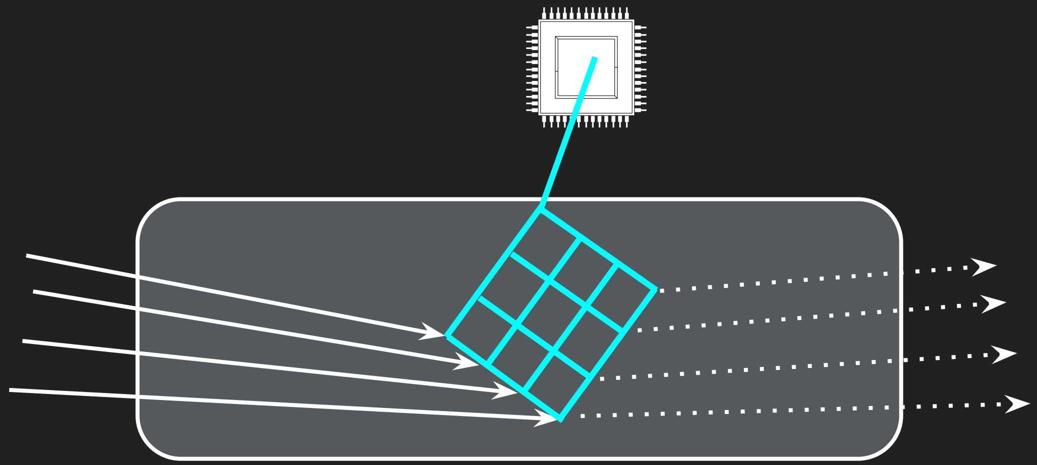
Must do collision avoidance on the bus

Low-capacity



# Crossbar Switching Fabric

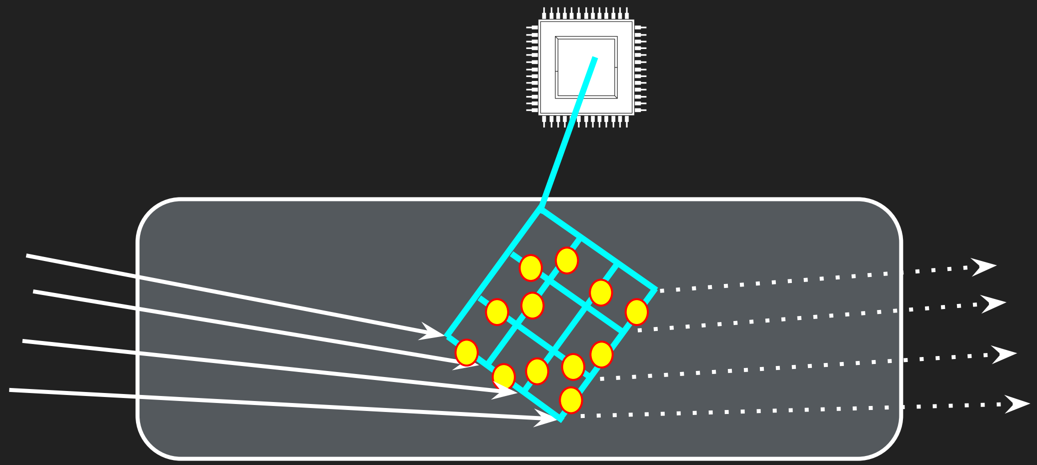
A series of busses that can be connected and disconnected at each junction by the router CPU



# Crossbar Switching Fabric

A series of busses that can be connected and disconnected at each junction by the router CPU

May include several crossbars operating in parallel

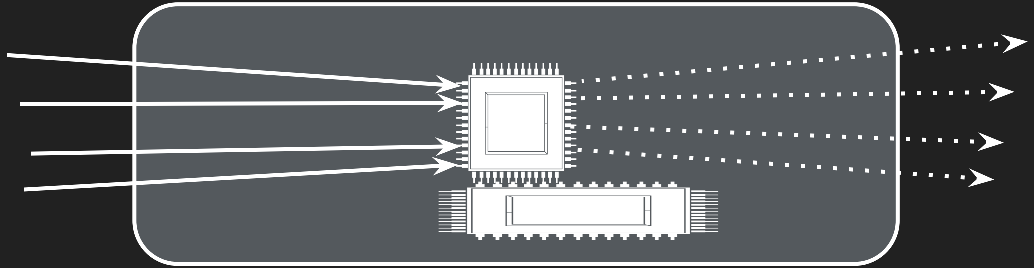


# Memory Switching Fabric

The switch or router CPU processes and forwards all messages using a piece of shared memory (to eliminate copying overhead)

Adds extra overhead on networks where the same paths are often repeated

Very similar to how a motherboard with multiple CPUs works



# Queueing

If all incoming ports data flows need to go out one single egress/outbound port, there is a bottleneck.

Messages will wait in queues.

Queues have finite space.

Packets get dropped when queues are full

# FIFO Packet Scheduling

Memory and Crossbar routers/switches need to decide how to prioritize which packets get the best routes soonest through the switching fabric

First In, First Out priority scheduling is the simplest

It is first come, first served

Causes ***head-of-line blocking***

# Head-of-Line Blocking

Because of different Switching Fabric designs, the fabric may be busy even though the input and output ports are idle

A switch/router can be overloaded even when only a little over 50% capacity

For example:

1. Consider a router/switch using FIFO scheduling
2. The first packet can’t go out because of collisions on its outbound port
3. Any subsequent packets are blocked by the packet at the **head of** the **line**

# Priority Queue Packet Scheduling

Packets are marked with a priority level.

Higher priority packets can skip ahead in line - either always or at a designated rate.

Otherwise, FIFO for same-priority packets.

# Round Robin Packet Scheduling

Give each network flow a specified chunk of time

Basically taking turns getting priority and everyone gets the same chunk of time.

No head of line blocking!

# Weighted Fair Queue Packet Scheduling

Like round robin, but with unfair “turn” lengths

Priority flows get longer turns

Also no head-of-line blocking

Requires a mechanism to assign priorities

Proprietary router “speed-up” mechanisms

# Link State Routing Algorithm (Dijkstra's Algorithm)

https://www.youtube.com/watch?v=_lHSawdgXpI

# Distance Vector Routing Algorithm (Bellman-Ford)

https://www.youtube.com/watch?v=obWXjtg0L64

# Organizing the Network Layer: Path Finding Algorithms

Both have problems when link costs change

Link-State Routing Algorithm

- Every node needs to hear about every change

Distance-Vector Routing Algorithm

- Can get stuck routing traffic in loops

# Open Shortest Path First OSPF

Adaptive Routing for Interior Networks

1. All routers have:
   a. 32-bit Router ID
   b. Neighbor Table: Who do they have Layer 2 connections to?
   c. Topology Table: How is the whole AS connected
   d. Routing Table: After Dijkstra's Alg, what are all the best routes
2. Requires all routers to keep watching their neighbors and gossipping about what they see

# Routing Tables

Maps IP address ranges to which Layer 2 connection to use (or maybe a Layer 3 address to be looked up with ARP)

| Address Range | Forward To |
|---|---|
| 10.0.0.0/16 | 10.0.0.1 |
| 78.45.32.0/22 | 54.34.56.25 |
| 78.45.32.16/28 | 55.55.55.57 |
| 8.0.0.0/8 | 54.34.56.25 |
| 0.0.0.0/0 | 67.122.233.12 |

# Routing Table Rule Precedence

Precedence goes to the rule with the longest matching “prefix”. Specific rules beat general rules.

| **Address Range** | **Forward To** |
|---|---|
| 10.0.0.0/16 | 10.0.0.1 |
| 78.45.32.0/22 | 54.34.56.25 |
| 78.45.32.16/28 | 55.55.55.57 |
| 8.0.0.0/8 | 54.34.56.25 |
| 0.0.0.0/0 | 67.122.233.12 |

# ICMP - Internet Control Message Protocol

Used to troubleshoot layer 3 routing problems

“ping” uses ICMP packets to check connectivity

“traceroute” uses ICMP packets to check routes

Both require routers to be willing to answer

# Dynamic Host Configuration Protocol (DHCP)

1. SRC 0.0.0.0, DST 255.255.255.255 - **DHCPDISCOVER**
2. SRC (Router IP),  DST 255.255.255.255 - **DHCPOFFER**
   a. IP address
   b. Subnet Mask
   c. Gateway
   d. Available DNS (For a Higher Layer)
3. **DHCPREQUEST**
4. **DHCPACK**

# Network Game: Classroom Routing

Everyone gets an IP address

Everyone should sit with their AS (first three octets match)

Create IXPs

Do BGP

Route Packets

# Router Config Exploration Demo

# Want total control of your router?    openWRT

openwrt.org

# Networking Field Trip

# Image Credits

Fire by YANDI RS from Noun Project

computer chip by ilCactusBlu from Noun Project