

# Thoughts about Network Protocols and the Internet

Radia Perlman  
Dell EMC

[Radia.perlman@dell.com](mailto:Radia.perlman@dell.com)

Most important point I'll make

# Most important point I'll make

- Not everything you read, or hear is true

# How networking tends to be taught

- Memorize these standards documents, or the arcane details of some implementation that got deployed
- Nothing else ever existed
- Except possibly to make vague, nontechnical, snide comments about other stuff

# My philosophy on teaching (and books)

- Look at each conceptual problem, like how to autoconfigure an address
- Talk about a bunch of approaches to that, with tradeoffs
- Then mention how various protocols (e.g., IPv4, IPv6, Appletalk, IPX, DECnet, ...) solve it

## But some professors say...

- Why is there stuff in here that my students don't “need to know”?

# Where does confusion come from?

- Hype
- People repeating stuff
- Buzzwords with no clear definition
  - Or persons A and B have a clear definition in mind, but different from each other
- Or the world changing, so something that used to be true is no longer true

# Things are so confusing

- Comparing technology A vs B
  - Nobody knows both of them
  - Somebody mumbles some vague marketing thing, and everyone repeats it
  - Both A and B are moving targets



# Standards Bodies...

# What about “facts”?

- What if you measure A vs B?

# What about “facts”?

- What if you measure A vs B?
- What are you actually measuring?...one implementation of A vs one implementation of B

# What about “facts”?

- What if you measure A vs B?
- What are you actually measuring?...one implementation of A vs one implementation of B
- *So don't believe something unless you can figure out a plausible property of the two protocols that would make that true*

# Fostering and Practicing Critical Thinking

- Don't believe something (and certainly don't repeat it!) unless you understand something intrinsic that makes it true
- Encourage “naïve” questions
  - Delight in teaching what “everyone knows”
  - Cherish the chance to question your basic assumptions
  - Be a role model by asking questions yourself

# How to understand Network Protocols

# How to understand Network Protocols

- Nobody would have designed what we have
- The only way to understand it is to look at the history

# Why do we have both Ethernet and IP?

- Most network people will say “because Ethernet is layer 2 and IP is layer 3”
- I claim that is wrong...the answer is subtle



# The story of Ethernet

- What is Ethernet?
- How does it compare/work with IP?
- People talk about “layer 2 solutions” vs “layer 3 solutions”.  
What’s that about?

# So, first we need to review network “layers”

- ISO credited with naming the layers
- It’s just a way of thinking about networks

# Perlman's View of ISO Layers

- 1: Physical

# Perlman's View of ISO Layers

- 1: Physical
- 2: Data link: (neighbor to neighbor)

# Perlman's View of ISO Layers

- 1: Physical
- 2: Data link: (neighbor to neighbor)
- 3: Network: create path, forward data (e.g., IP)

# Perlman's View of ISO Layers

- 1: Physical
- 2: Data link: (neighbor to neighbor)
- 3: Network: create path, forward data (e.g., IP)
- 4: Transport: end-to end (e.g., TCP, UDP)

# Perlman's View of ISO Layers

- 1: Physical
- 2: Data link: (neighbor to neighbor)
- 3: Network: create path, forward data (e.g., IP)
- 4: Transport: end-to end (e.g., TCP, UDP)
- 5 and above:

# Perlman's View of ISO Layers

- 1: Physical
- 2: Data link: (neighbor to neighbor)
- 3: Network: create path, forward data (e.g., IP)
- 4: Transport: end-to end (e.g., TCP, UDP)
- 5 and above: ..... boring



# So...why are we forwarding Ethernet packets?

- Ethernet was intended to be layer 2
- Just between neighbors – not forwarded

# So...why are we forwarding Ethernet packets?

- Ethernet was intended to be layer 2
- Just between neighbors – not forwarded
- What exactly is Ethernet?

# So...why are we forwarding Ethernet packets?

- Ethernet was intended to be layer 2
- Just between neighbors – not forwarded
- What exactly is Ethernet?
- No way to understand it without seeing the history

# Back then...

- I was the designer of layer 3 of DECnet
  - the routing protocol I designed was adopted by ISO and renamed IS-IS
- Layer 3 calculates paths, and forwards packets
- Layer 2 just marked beginning and end of packet, and checksum (links between two nodes)

# What designing “layer 3” meant

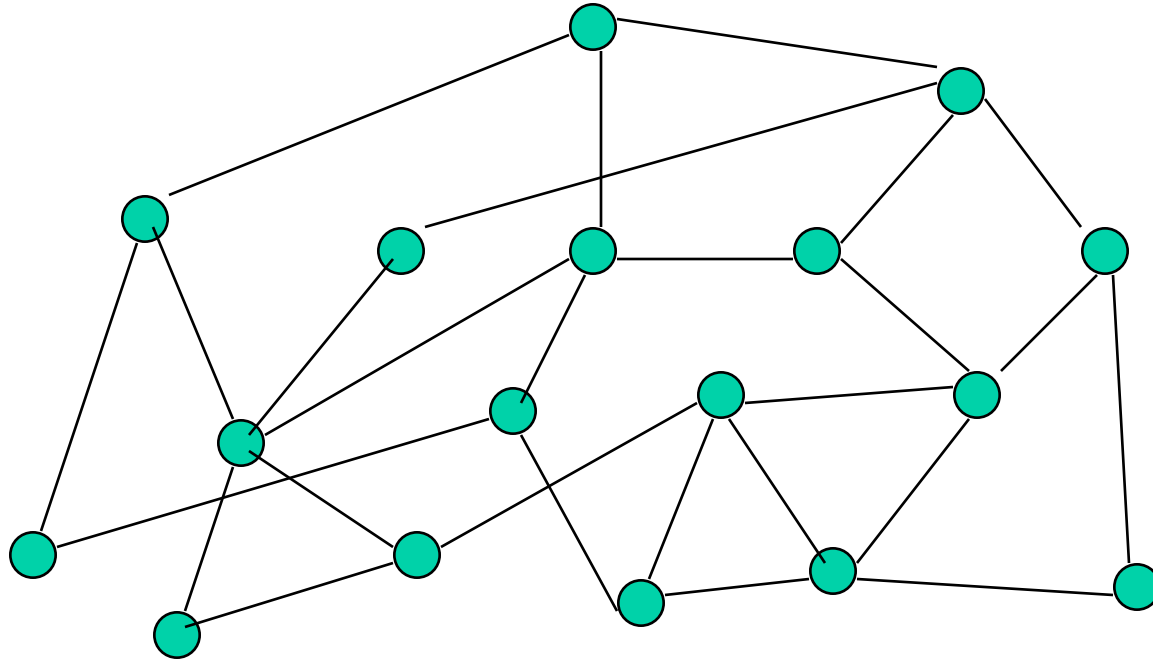
- Layer 3 addresses
- Layer 3 packet format (IP, DECnet)
  - Source, destination, hop count, ...
- A routing algorithm
  - Exchange information with your neighbors
  - Collectively compute routes with all rtrs
  - Compute a forwarding table

# Computing the Forwarding Table

# Computing the Forwarding Table

- Could be done with a central node
  - ATM, Infiniband, ...
- Or with a distributed algorithm

# Distributed Routing Algorithms



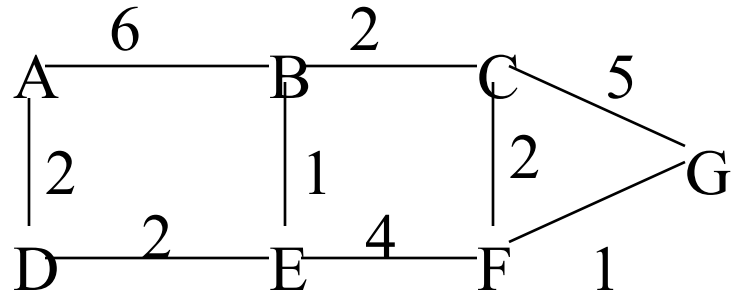


# Distributed Routing Protocols

- Rtrs exchange info
- Use it to calculate forwarding table

# Link State Routing

- meet nbrs
- Construct Link State Packet (LSP)
  - who you are
  - list of (nbr, cost) pairs
- Broadcast LSPs to all rtrs
- Store latest LSP from each rtr
- Compute Routes (breadth first, i.e., “shortest path” first—well known and efficient algorithm)



A
B/6
D/2

B
A/6
C/2
E/1

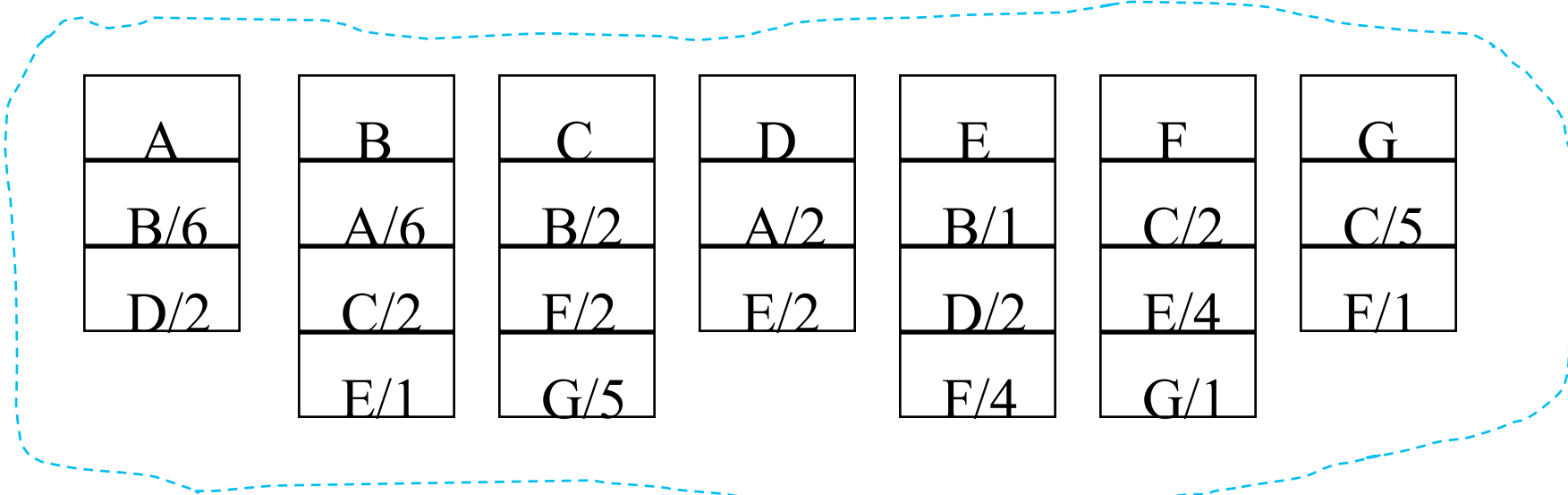
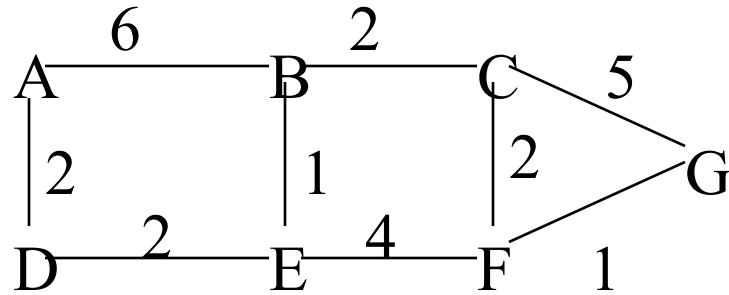
C
B/2
F/2
G/5

D
A/2
E/2

E
B/1
D/2
F/4

F
C/2
E/4
G/1

G
C/5
F/1



Each router has this same database  
 Gives enough info to compute paths

# Back to history

- I was doing layer 3
- Then along came Ethernet

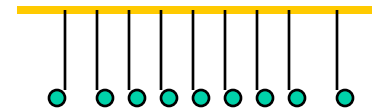
# The story of Ethernet

- CSMA/CD
- Spanning Tree
- TRILL/overlay networks

# CSMA/CD Ethernet

- CSMA/CD...shared bus, peers, no master

- CS: carrier sense (don't interrupt)
- MA: multiple access (you're sharing the air!)
- CD: listen while talking, for collision



- Lots of papers about goodput under load only about 60% or so because of collisions
- Limited in # of nodes (maybe 1000), distance (kilometer or so)

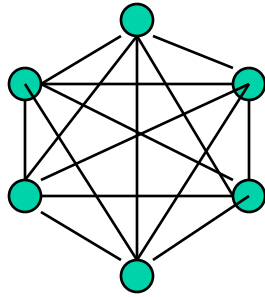
# I saw Ethernet as a new type of link

- I had to modify the routing protocol to accommodate this type of link
- For instance, the concept of “pseudonodes” and “designated routers” so that instead of  $n^2$  links, it's  $n$  links with  $n+1$  nodes

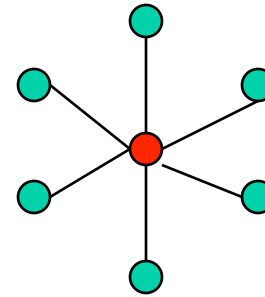


# Pseudonodes

Instead of:



Use pseudonode



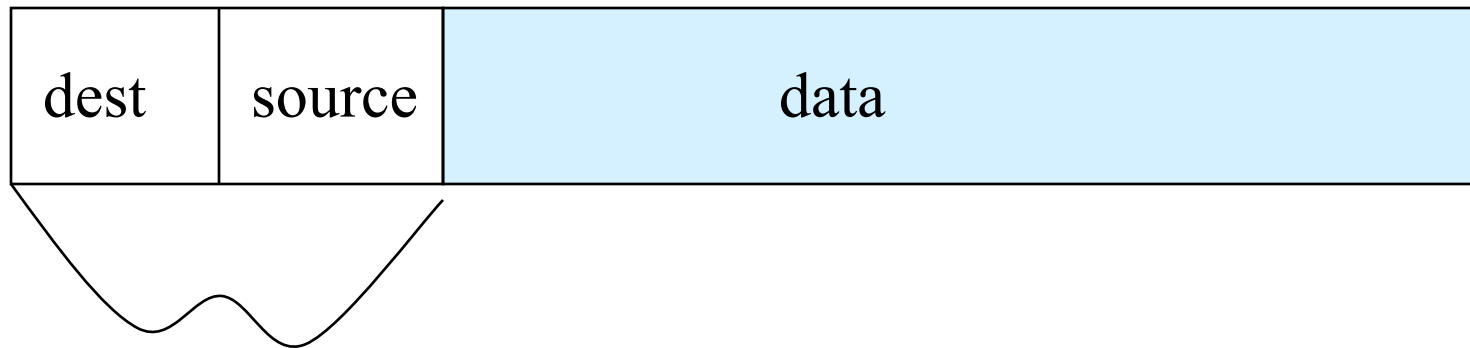
But Ethernet was a link in a network, not a network

- I wish they'd called it “Etherlink”

# Original Invention

- A way of cheaply hooking together lots of nodes on a single link
- Everyone could directly talk to everyone
- No forwarding

# Ethernet packet



# Layer 3 Packet



# It's easy to confuse Ethernet with layer 3

- It looks sort of the same
- No hop count field...
- Flat addresses (no way to summarize a bunch of addresses in a forwarding table)
- But it never occurred to the Ethernet inventors that anyone would be forwarding an Ethernet packet

So...why are we forwarding Ethernet packets?

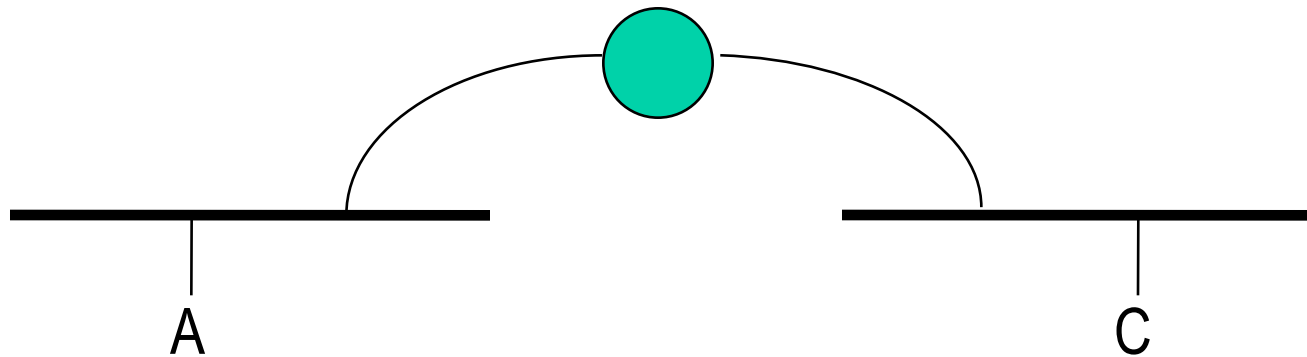
# How Ethernet evolved from CSMA/CD to spanning tree

- People got confused, and thought Ethernet was a network (layer 3) instead of a link (layer 2)
  - Link (layer 2) = nbr-nbr
  - Network (layer 3) = forward along a path
- Built apps on Ethernet, with no layer 3
- Router can't forward without the right envelope
- I tried to argue...



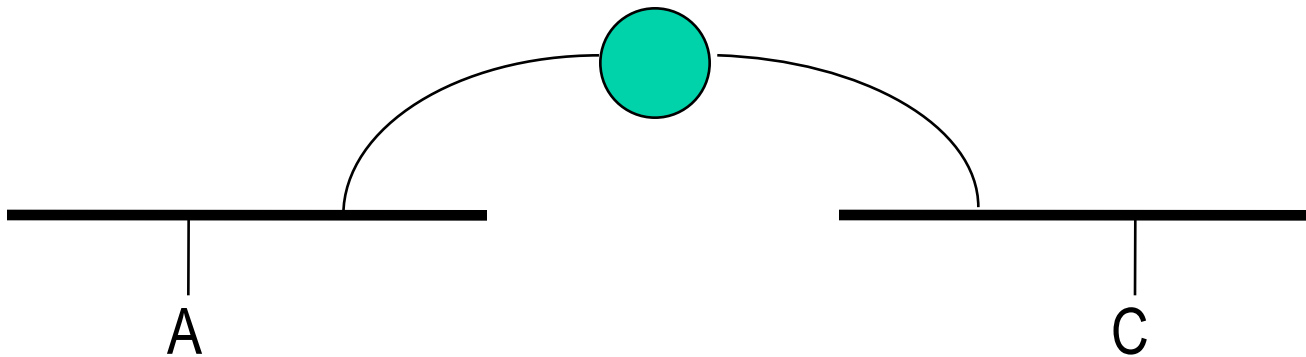
# Problem Statement (from about 1983)

*Need something that will sit between two Ethernets, and let a station on one Ethernet talk to another*



# Problem Statement (from about 1983)

*Need something that will sit between two Ethernets, and let a station on one Ethernet talk to another*

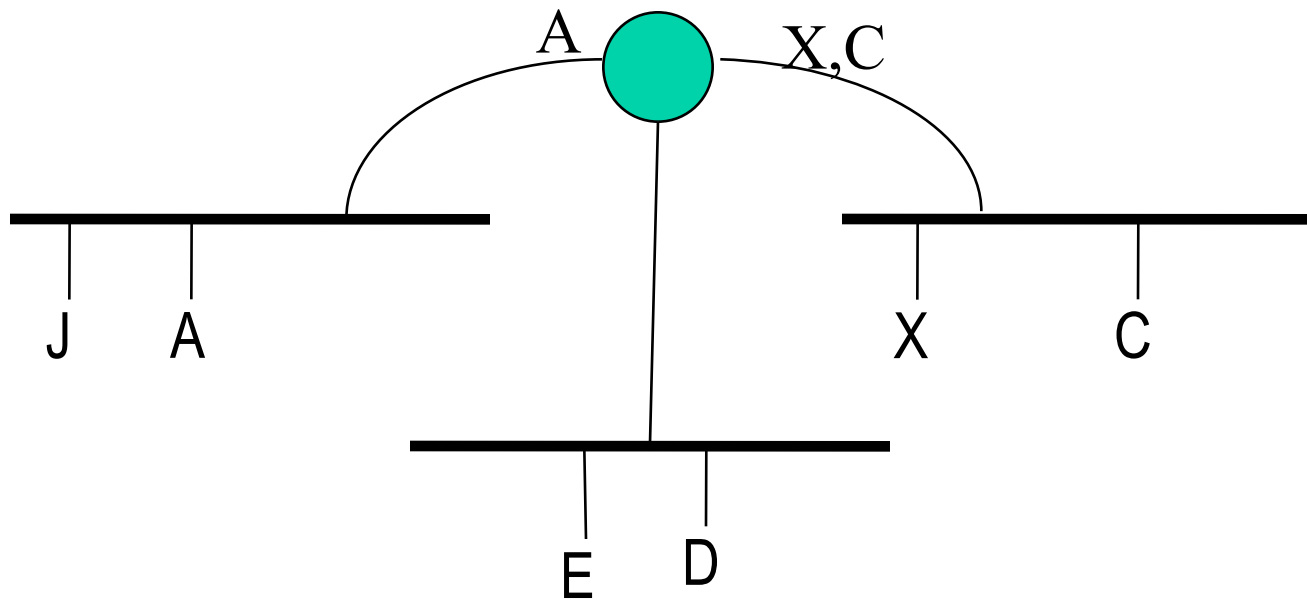


***Without modifying the endnode, or Ethernet packet, in any way!***

# The basic concept

- Bridge just listens “promiscuously”, and forwards to each other port(s) when the ether is free
- Learn (Source=S, input port). Once learned, if see a packet with destination=S, know where to forward it (rather than “all the ports”)
- This requires a topology with only one path between any pair of nodes

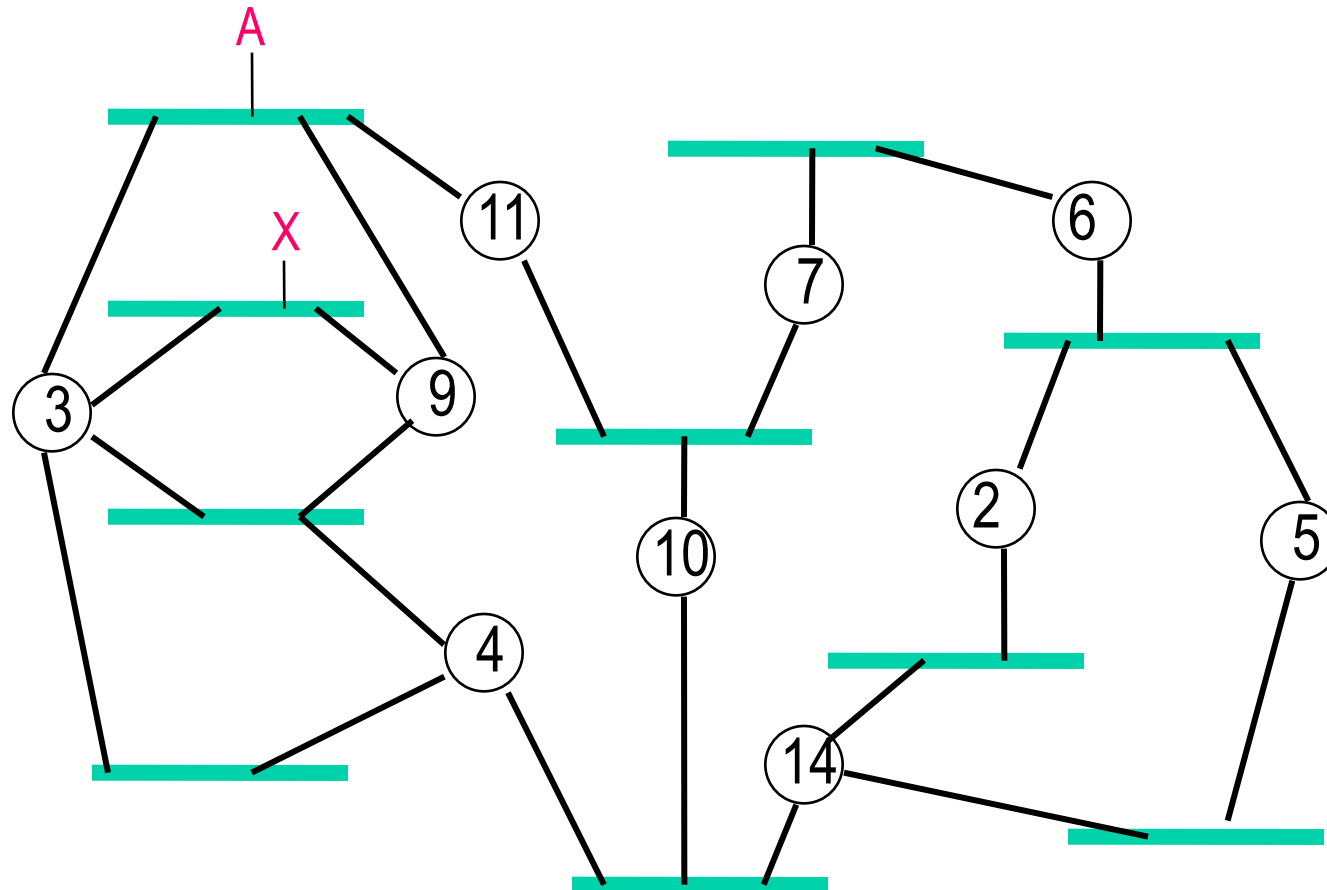
# Basic concept



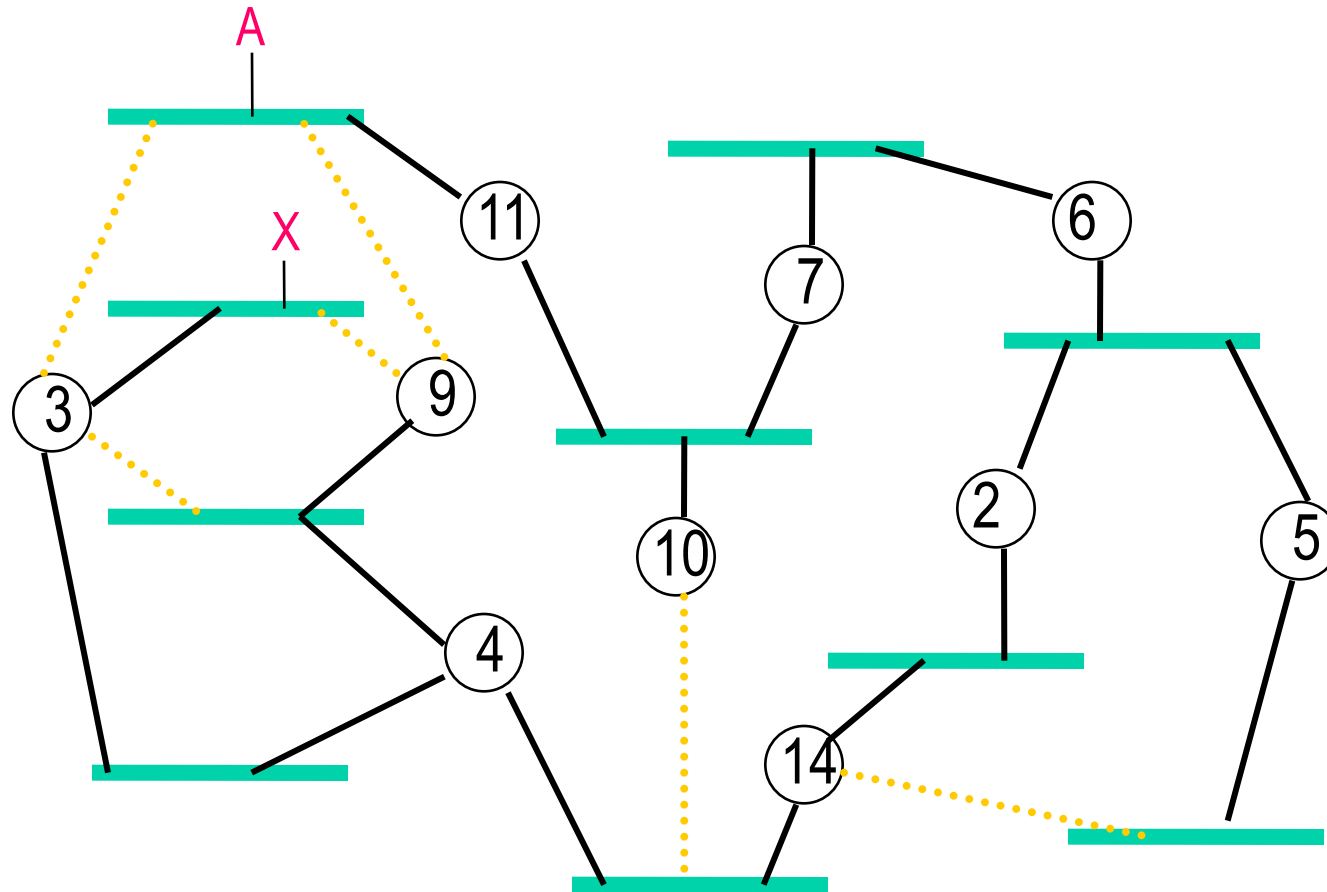
# How about require physical tree topology?

- What about miscabling?
- What about backup paths?
- So...spanning tree algorithm
  - Allowing any physical topology
  - Pruning to a loop-free topology for sending data

# Physical Topology



# Pruned to Tree



# Algorhyme

*I think that I shall never see  
A graph more lovely than a tree.  
A tree whose crucial property  
Is loop-free connectivity.  
A tree which must be sure to span  
So packets can reach every LAN.  
First the root must be selected,  
By ID it is elected.  
Least cost paths from root are traced,  
In the tree these paths are placed.  
A mesh is made by folks like me.  
Then bridges find a spanning tree.*

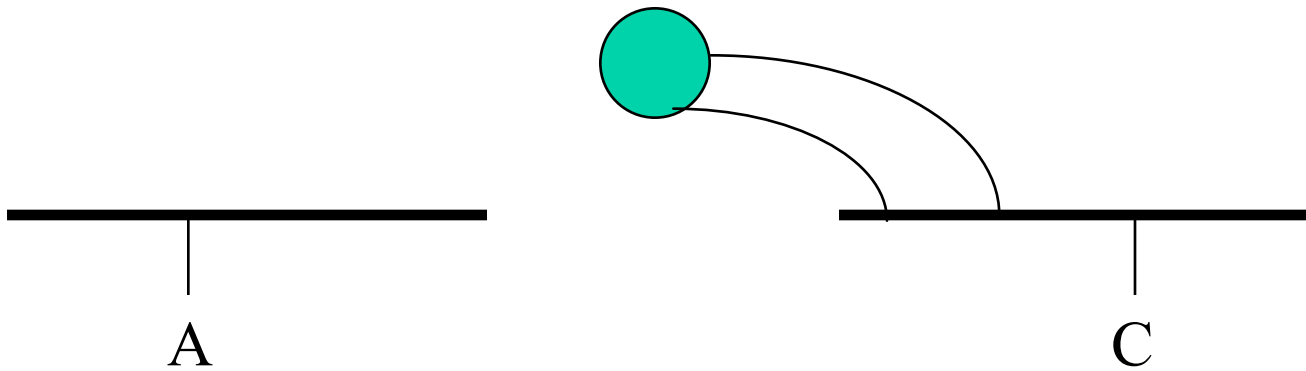
*Radia Perlman*



# Bother with spanning tree?

- Maybe just tell customers “don’t do loops”
- First bridge sold...

# First Bridge Sold



# CSMA/CD died long ago

- A variant is used on wireless links
- But wired Ethernet quickly became spanning tree
- So “Ethernet” today has nothing to do with all the papers about CSMA/CD

# Next Topic

- Why do we need both IP and Ethernet?

# Why not get rid of Ethernet and use only IP?

- World has converged to IP as layer 3, and it's in the network stacks

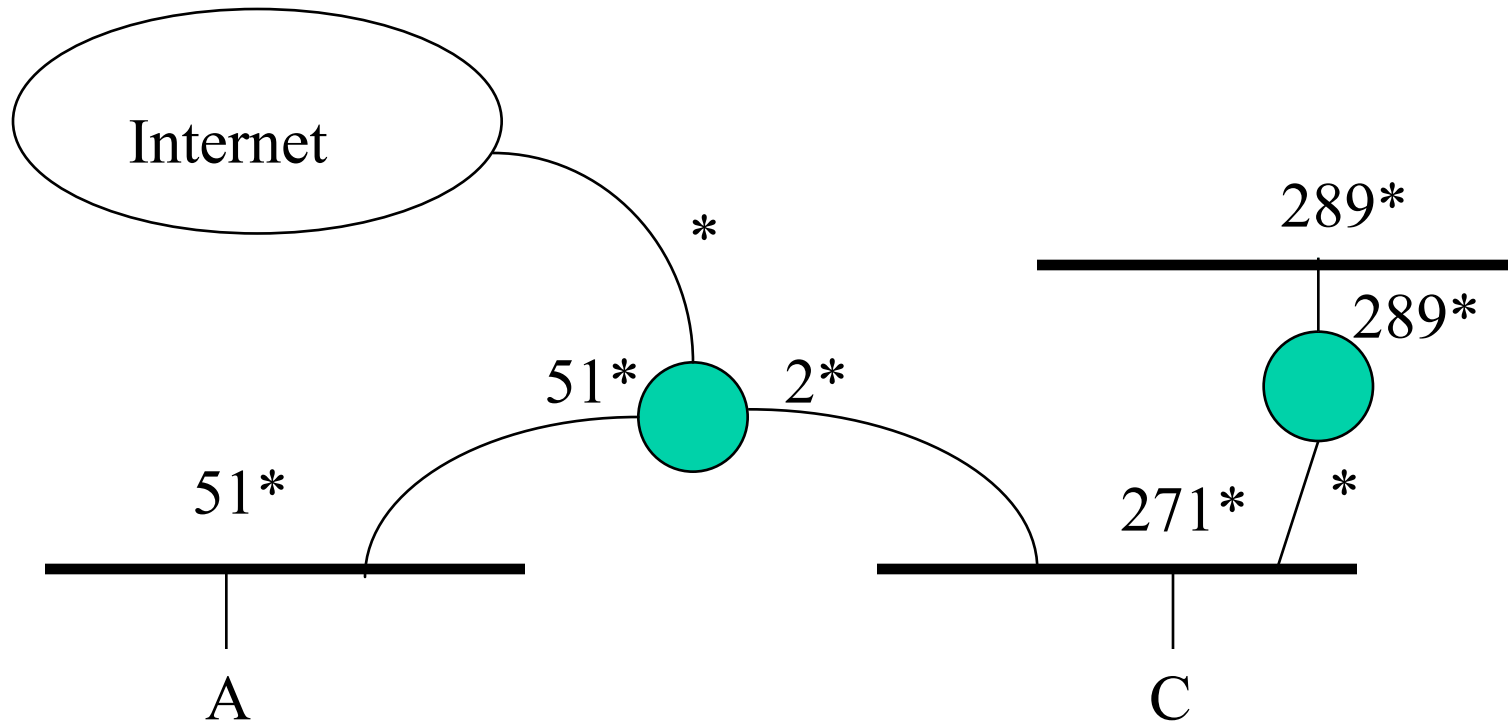
# Why not get rid of Ethernet and use only IP?

- World has converged to IP as layer 3, and it's in the network stacks
- If IP were designed slightly differently, we wouldn't need Ethernet anymore
- Just put your data in a layer 3 envelope!

# What's wrong with IP?

- IP is configuration intensive, moving VMs disruptive
  - IP protocol requires every link to have a unique block of addresses
  - Routers need to be configured with which addresses are on which ports
  - If something moves, its address changes

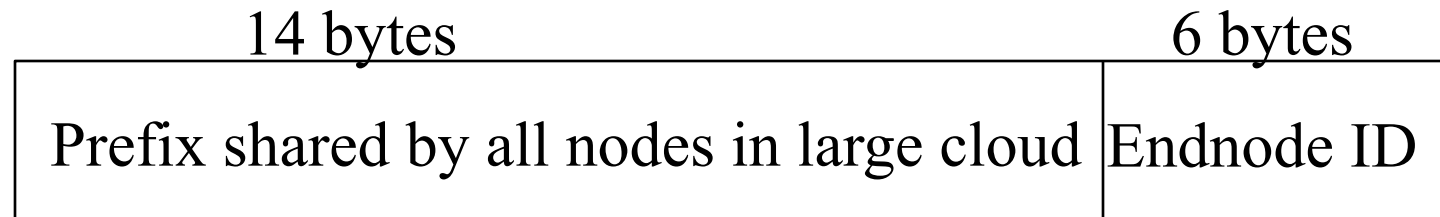
# Each Link Different Address Block



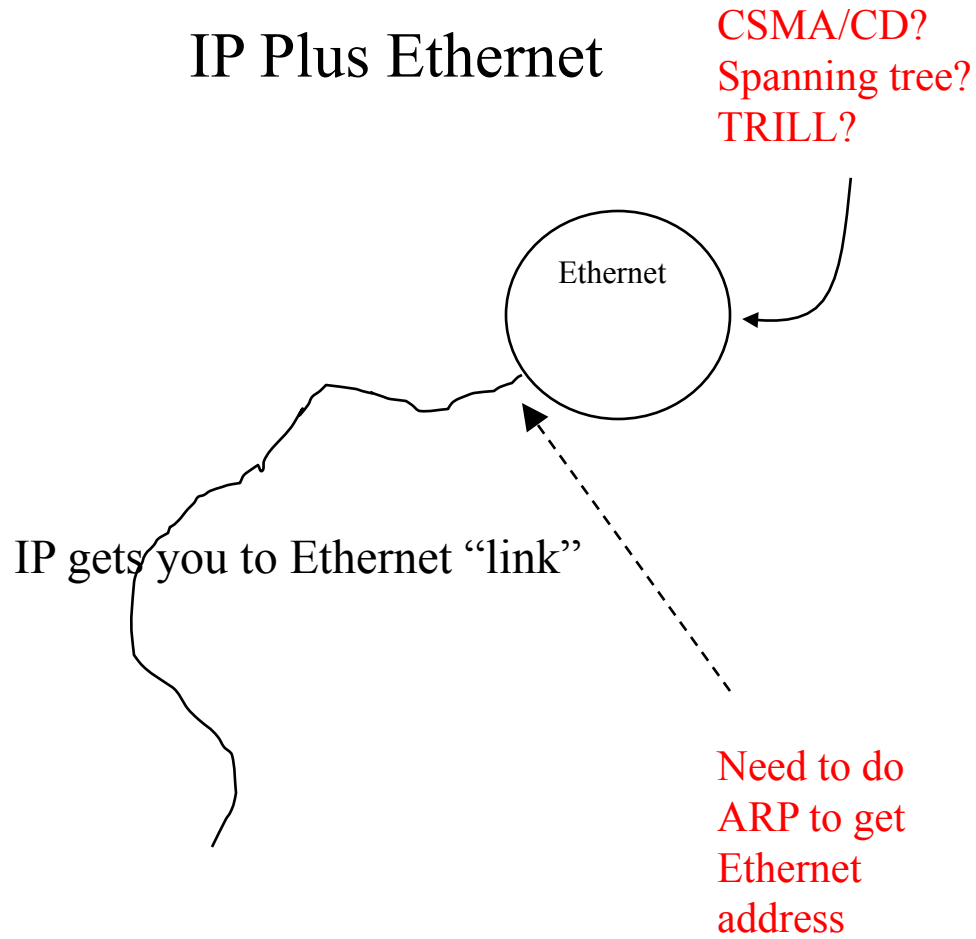


# Layer 3 doesn't have to work that way!

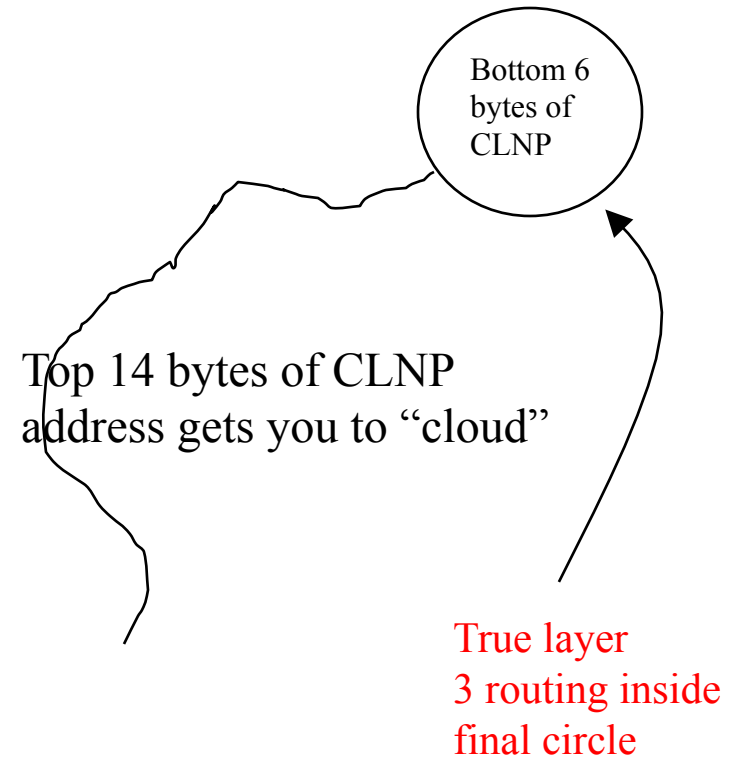
- CLNP / DECnet...20 byte address
  - Bottom level of routing is a whole cloud with the same 14-byte prefix
  - Routing is to 6 byte ID inside the cloud
  - Enabled by “ES-IS” protocol, where endnodes periodically announce themselves to the routers



## IP Plus Ethernet

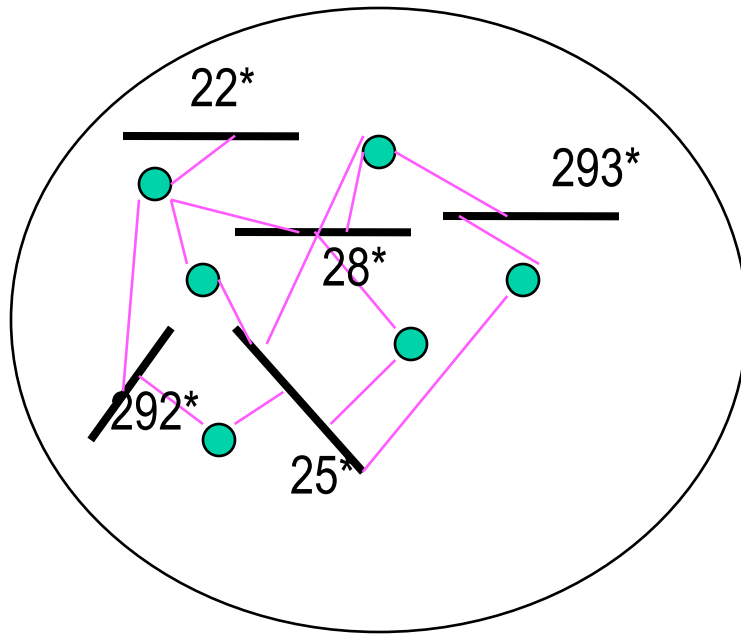


## CLNP



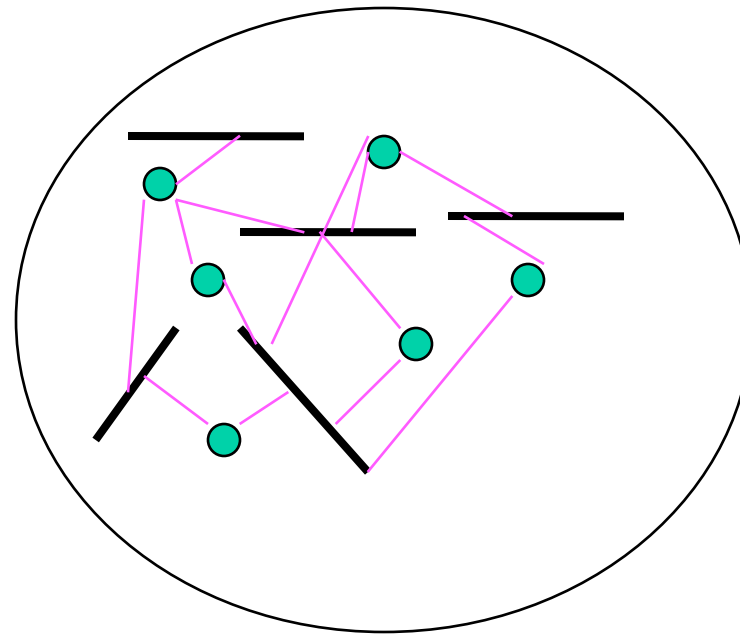
# Hierarchy

One prefix per link (like IP)



2\*

One prefix per campus



2\*

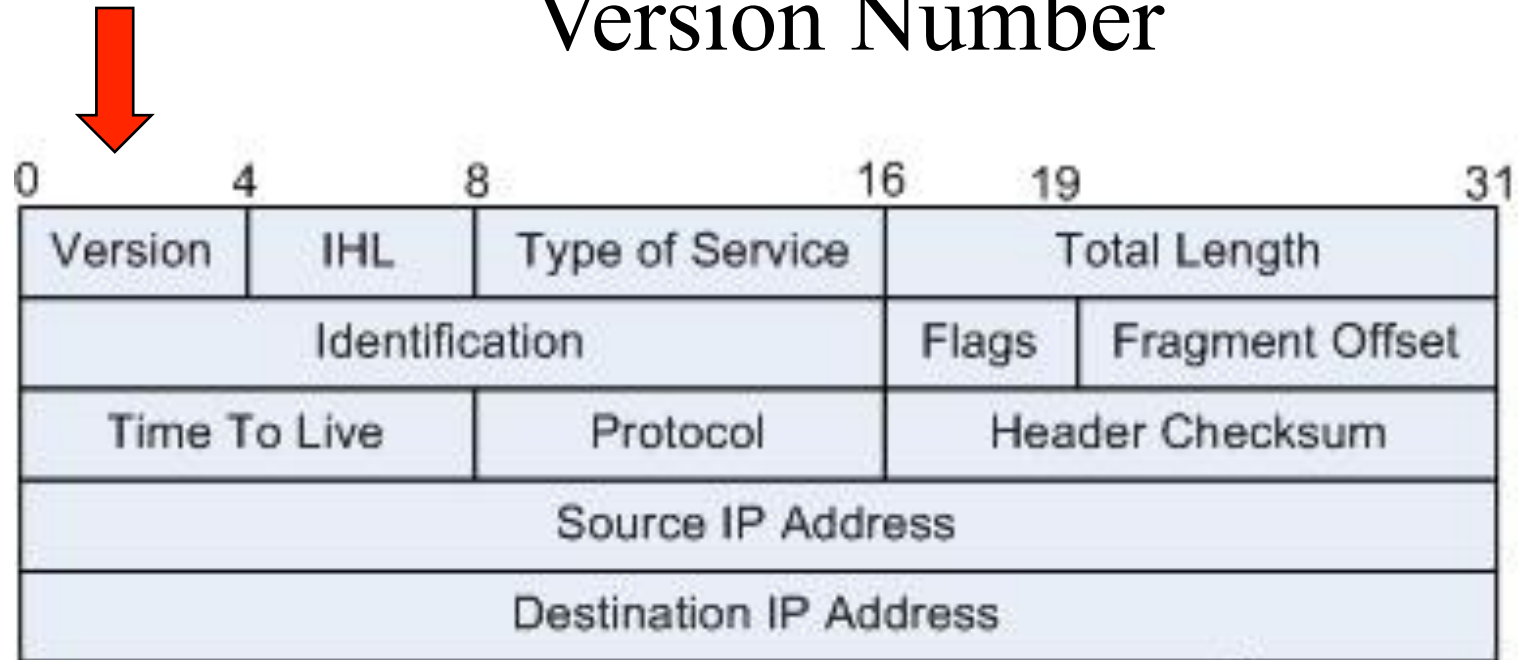
# Worst decision ever

- 1992...Internet could have adopted CLNP
- Easier to move to a new layer 3 back then
  - Internet smaller
  - Not so mission critical
  - IP hadn't yet (out of necessity) invented DHCP, NAT, so CLNP gave understandable advantages
- CLNP much cleaner than IP; wouldn't need ARP, wouldn't need Ethernet/spanning tree
- IPv6 still inferior to CLNP! (IPv6 also routes to a link, so will require Ethernet clouds, and ARP-like thing)

# Protocol Folklore

- Obvious stuff everyone gets wrong

# Version Number



# What's a Version Number?

- Version number
- What is the purpose?
- Philosophical question:
  - what is “new version” vs “new protocol”?

# What I think makes sense

- Envelope says what the protocol is (how to parse the packet)
  - Ethernet: Ethertype
  - IP: Protocol Type
  - TCP/UDP: port



# What I think makes sense

- Envelope says what the protocol is (how to parse the packet)
- If differentiate based on protocol type, then it's a new protocol
- If differentiate based on version number, then it's a new version of the same protocol

# If differentiate based on version number

- You can't just say "write this value into this field"
- You have to say "Look at the version number, and if it's not your version, then drop the packet"!

# Version #

- Nobody seems to do this right
- IP, IKEv1, SSL don't say what to do if version # different.  
Most implementations ignore version number field
- SSL v3 moved version field!

# Parameters

- Minimize these:
  - someone has to document it
  - customer has to read documentation and understand it
- How to avoid
  - architectural constants if possible
  - automatically configure if possible

# Settable Parameters

- Make sure they can't be set incompatibly across nodes, across layers, etc. (e.g., hello time and dead timer)
- Make sure they can be set at nodes one at a time and the net can stay running

# Example: Hello Timer

- IS-IS
  - pairwise parameters reported in “hellos”
  - So you know what to expect from that neighbor
- OSPF
  - Kind of copied IS-IS, but decided...

# Example: Hello Timer

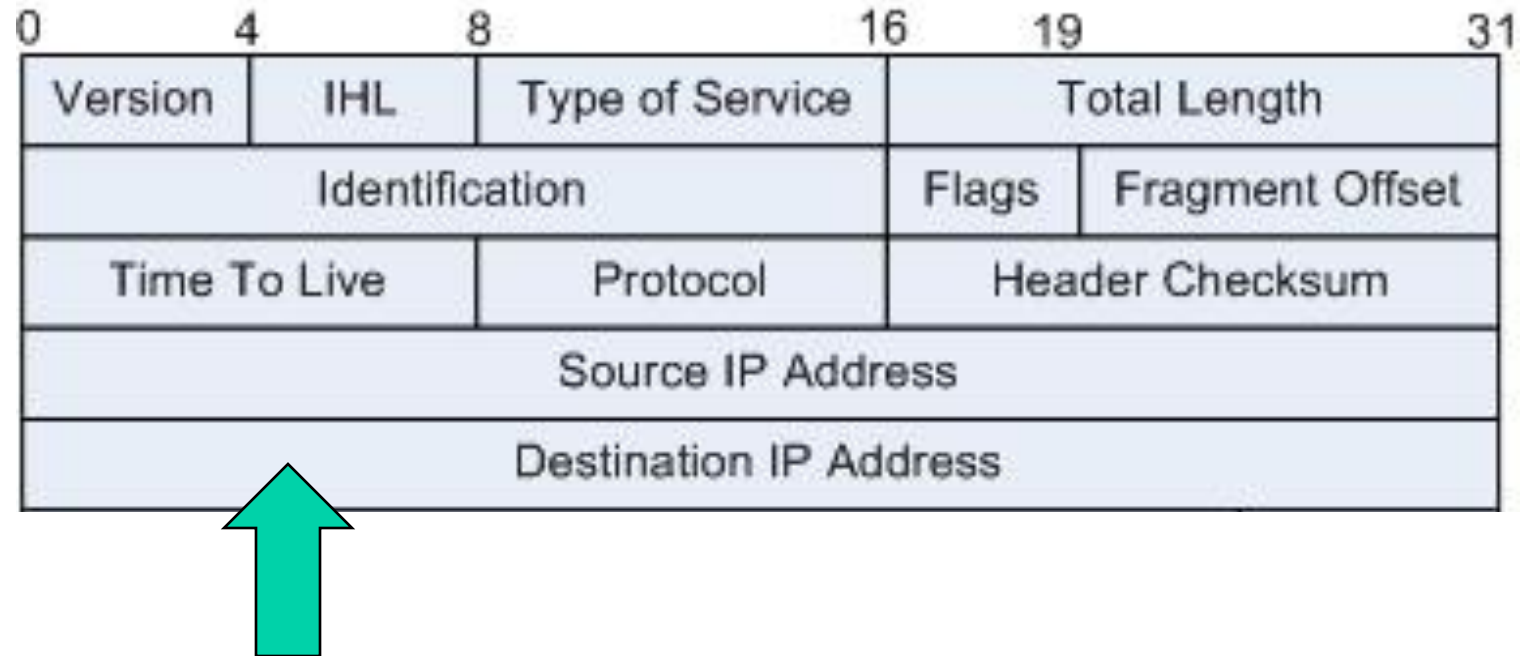
- IS-IS
  - pairwise parameters reported in “hellos”
  - So you know what to expect from that neighbor
- OSPF
  - Kind of copied IS-IS, but decided...
  - Refuse to talk if timers not identical with neighbor's!

# Latency

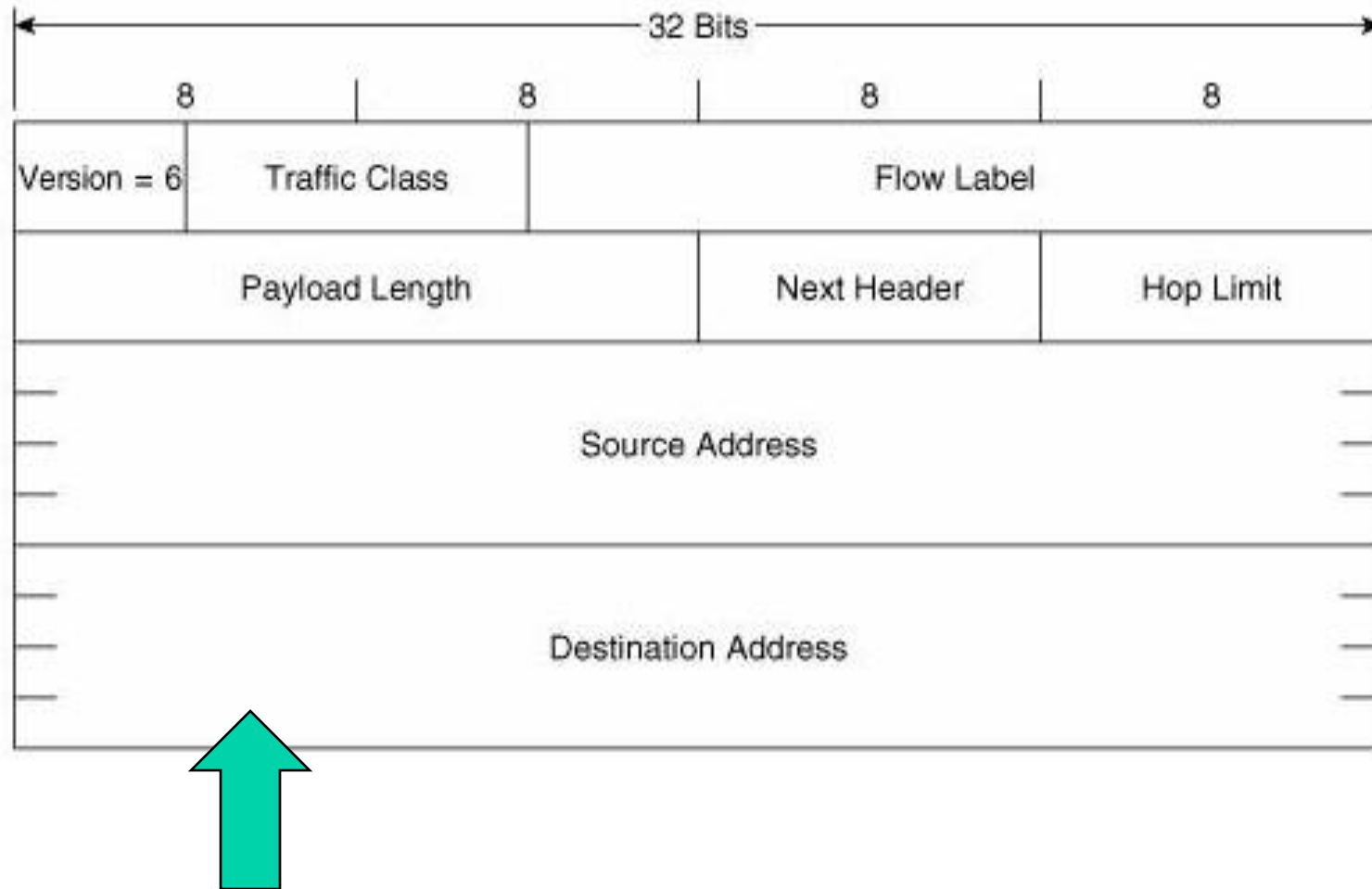
- Store-and-forward vs cut-through
- Cut through can start after the forwarding decision is made
- What field do you need to see for forwarding decision?



# IPv4 header



# IPv6 header



And now for something different

# What I focused on, at the start of my career (and the early days of computer networks)

- The foundation...plugging the network together and letting it figure out how to move data. My contributions:
  - Robustness
    - You can reboot a PC, but not a network
    - Self-stabilization and later, Byzantine robustness
  - Scalability: innovations to allow much larger networks
  - Manageability/Usability
    - Reducing the need for configuration
    - Ensuring if there are parameters, that they will interoperate

# What I didn't worry about, because it seemed solvable

- Knowing who is sending you data
- Knowing the data has not been corrupted along the way
- Theory sounds great:
  - Service gets a DNS name and a certificate
  - Service proves it owns that DNS name
  - Use a protocol like SSL/TLS to protect the conversation
  - And user authenticates with username/password or something...

# How it works

User searches for  
“MyBank”, clicks on  
URL

Sends to DNS name in URL



Sends certificate showing (DNS name, public key)



Crypto, and protocol, server proves it owns the DNS name



# In reality

- DNS names don't really mean anything
- Example: I fell for a scam recently...
- I wanted to renew my Washington state driver's license
- I knew it could be done online
- I did an Internet search for “renew Washington State driver's license”

The top search result looked legitimate



## Washington License Renewal | Renew WA DMV Drivers License

Ad [www.washington-information.org/Drivers-License/Renewal](http://www.washington-information.org/Drivers-License/Renewal) ▼

Find All the Information You Need to **Renew** Your **Driver's License** Here! Up to date information and assistance with all the necessary steps to **renew** your **license**. Car Registration. Categories: Community Service, Government, Recreation, Business.

I didn't notice that it said "Ad"

Everything was as I expected...



Renew Driver License



New Driver's License



Replace Driver's License



ID Card



Change of Name



Change of Address

# I typed in my information, including credit card

- I got suspicious when afterwards it presented me with a bunch of “offers” (which I knew were scams)
- I then looked more carefully at the site...it just claims to “give information” about how to get a license, not actually getting a license
- They initially charged me \$3.99, then started charging more things over the next couple of days...\$9.99, \$19.99.
- Fraud department of my bank called me, asking if these were fraudulent. I explained, and they changed my credit card number and denied all the charges
- This scam is so lucrative, that multiple organizations are doing it, in all 50 states

# Don't blame the user!

- These are scam sites (which appear first in the search, because they pay the search engine companies to put them first, or because they understand the ranking algorithm and game the system:

**[washingtondriverslicense.org](http://washingtondriverslicense.org)**

**[www.washington-information.org](http://www.washington-information.org)**

**[www.dmv.org](http://www.dmv.org)**

- The site I should have gone to was

**[www.dol.wa.gov](http://www.dol.wa.gov)**

# New Topic: Hype

# Getting confused by hype

- People get taken in by hype
- It's hard to believe, when you hear something from so many places, that it could possibly be wrong
- Don't start with a technology and figure out how to use it
- Instead start with “what problem am I solving”?

# Example: “Blockchain”

- People assume
  - Distributed = “bad”
  - Centralized = “good”
  - Distributed = “blockchain”
- Not even clear what “distributed” means?
  - Storing data in multiple locations?
  - Having multiple servers to respond?
  - Distributed trust....



# Distributed Trust

- Makes sense not to totally depend on one organization
- Example of how to do it without “Byzantine consensus”
  - US credit rating organizations
  - Each has its own algorithm, and source of data
  - Your credit rating is likely to be quick similar at the different organizations
  - Someone who wants to check can check whichever they trust most, or look at several and compare

# Privacy

- I don't post on social media
- So I'm safe...right?
- Recently, I made a hotel reservation on the phone, getting an email confirmation
- I searched on Internet maps for the location of the hotel
- It showed the hotel and said "your reservation is for October 9"
- The company that makes my browser, and for which I use email...reads my email, my browser history, etc....all for my convenience I'm sure, but still...
- The Internet knows more about me than I know about myself...

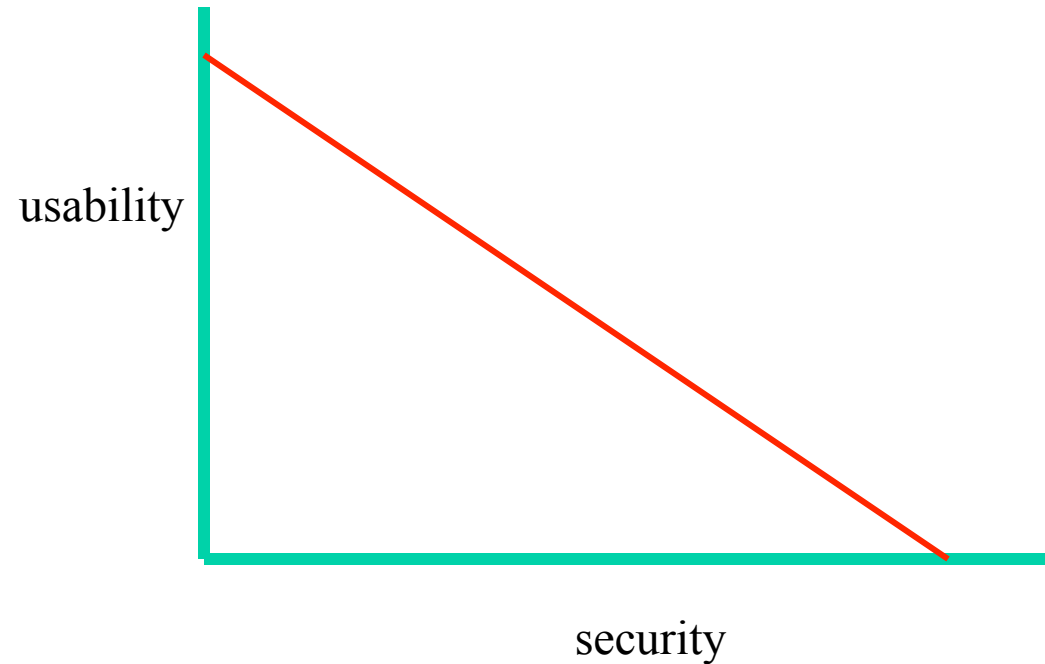
# Truth

- Is it the end of truth?
- It is possible to make fake photos, audios, and videos that a human can't distinguish from real...possibly even professionals can't...
- Then the Internet can spread it virally
- This is disastrously polarizing society

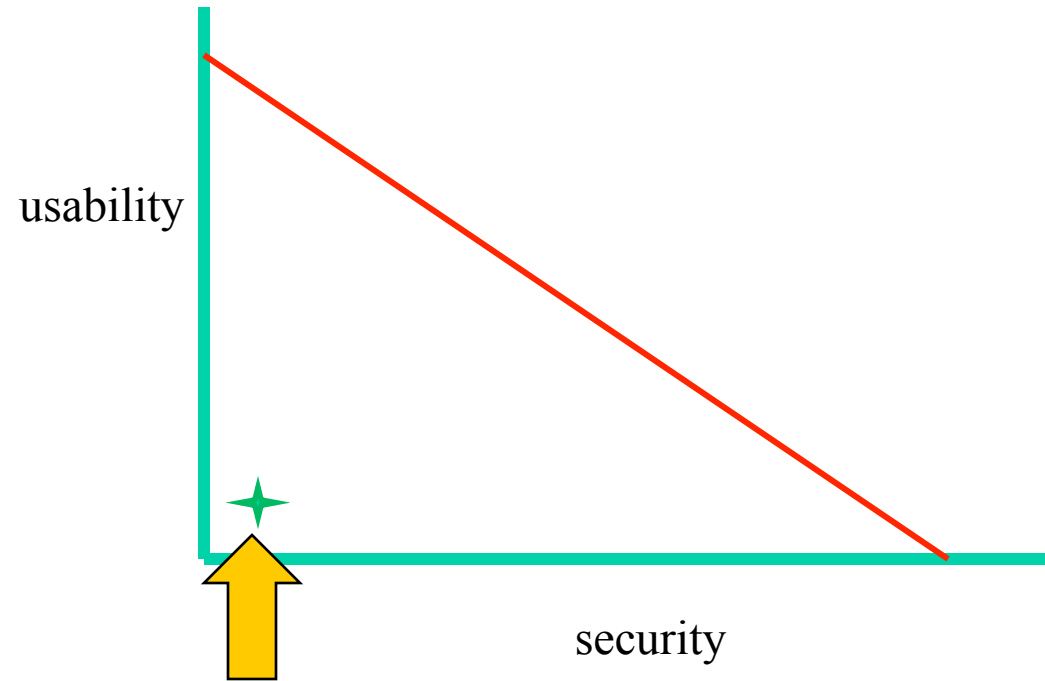
# User Authentication

It's common to have to tradeoff usability vs  
security

It's common to have to tradeoff usability vs security



# We are here!



# User authentication

- Every site has different rules for usernames and passwords
  - At least  $n$  characters, no more than  $x$  characters, must have at least one letter, one number, one special character, must not contain anything but letters and numbers, ....



# User authentication

*“Sorry, but your password must contain an uppercase letter, a number, a haiku, a gang sign, a hieroglyph, and the blood of a virgin.”* .....(unknown author)

# Security questions

- Who comes up with these?
  - Father's middle name
  - 2<sup>nd</sup> grade teacher's name
  - Veterinarian's name
  - Favorite sports team
  - My middle name

# User authentication

- I do not want to hear...

# User authentication

- I do not want to hear...  
“We need better user training”

# User authentication

- I do not want to hear...  
“We need better user training”  
Or things like “people shouldn’t click on suspicious links”

# Humans

*“Humans are incapable of securely storing high-quality cryptographic keys, and they have unacceptable speed and accuracy when performing cryptographic operations. They are also large, expensive to maintain, difficult to manage, and they pollute the environment. It is astonishing that these devices continue to be manufactured and deployed, but they are sufficiently pervasive that we must design our systems around their limitations.”*

- (me), in “Network Security: Private Communication in a Public World”

# Parting Thoughts

- Don't blame the user!
- Be skeptical about what you read and hear
- Know what problem you are solving