



ICT-onderzoek Platform Nederland

Radia Perlman

(Sun Microsystems)

*Mythology and Folklore of
Computer Network Protocol Design*



Mythology and Folklore of Network Protocol Design

Radia Perlman
radia.perlman@sun.com

What are protocols?

- How things communicate
- There are protocols all around us
 - Meetings: Channel acquisition
 - 2-party
 - My favorite protocol:
 - Take turns talking and listen when the other one speaks

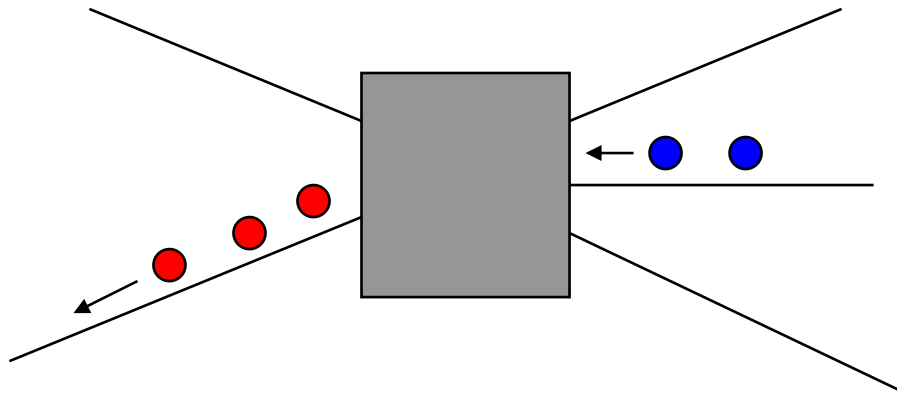
Some things I've learned

- Standards bodies are mostly political
- The technology that “wins” is not necessarily the best
- It's almost random whether a paper gets accepted or rejected
- Networking is seldom taught as a science
- A lot of what is “known” is false

Example of how confused
networking is

Topic 1: Bridges, Routers, and Switches, Oh My!

- These all move data from one port to another



Topic 1: Bridges, Routers, and Switches, Oh My!

- These all move data from one port to another
- Everyone is confused about what the differences are
- Especially those that think they know...

First let's start with the OSI Model of Networking

- ISO (a standards body) came up with a way of thinking about networking, and a way to break into committees, so that pieces could be independently designed
- “ISO Reference Model”
- Divide the problem into “layers”, where each layer uses the layer below

ISO Reference Model

- Layer 1: Physical (move bits)
- Layer 2: Data Link (neighbor-to-neighbor)
 - Mark beginning and end of packets, perhaps number messages and do hop-by-hop acks

ISO Reference Model

- Layer 1: Physical (move bits)
- Layer 2: Data Link (neighbor-to-neighbor)
- Layer 3: Network
- Layer 4: “Transport” (end-to-end, e.g., TCP numbering messages, sending acks, and retransmitting)

ISO Reference Model

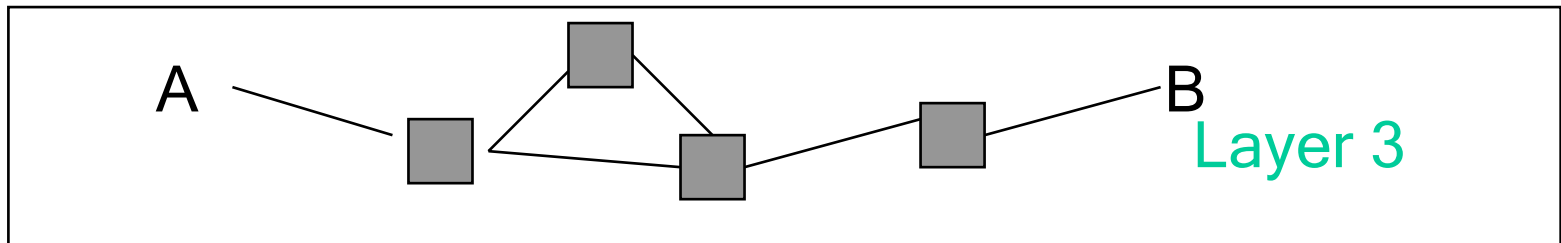
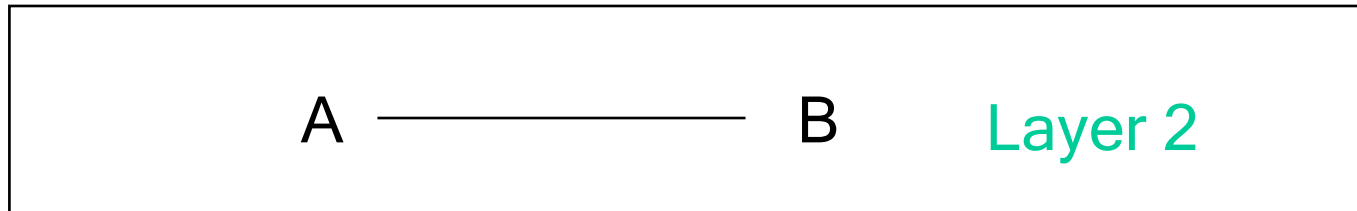
- Layer 1: Physical (move bits)
- Layer 2: Data Link (neighbor-to-neighbor)
- Layer 3: Network
- Layer 4: “Transport”
- Layers 5 and above: boring...

Now we can define bridges vs routers

- Router: layer 3 relay
- Bridge: layer 2 relay
- Wait....what's a "layer 2 relay"? Isn't layer 2 neighbor-neighbor?

If I ran the world

- Definition of layer 2: neighbor-neighbor.
No forwarding
- Forwarding is what layer 3 does



But I don't run the world, so

- True definition of a layer 2 protocol
 - *Anything defined by a committee whose charter is to design a layer 2 protocol*

Focus on layer 3

- Put your data in an “envelope” with a source and destination address
- Lots of layer 3 protocols
 - IPv4, IPv6, Appletalk, DECnet, IPX
- All of those are similar: just add header with source, destination, and hop count
- Main difference is size of the addresses

What's a hop count?

- News about a topology change (link or router going up or down) can't take effect simultaneously everywhere
- So you might have temporary loops
- Hop count kills off looping packets

But back to “what are
bridges?”

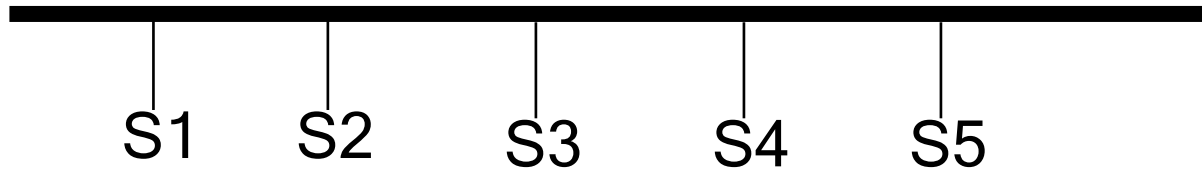
Early 1980's

- I was DECnet layer 3 architect
- Routing algorithms
- Originally DECnet addresses were 2 bytes
- DEC decided to adopt ISO's packet format (CLNP), with 20 byte addresses
- ISO adopted DECnet's routing algorithm, and renamed it IS-IS

Then along came “Ethernet”

- Myth: Ethernet is a successful technology
- Actually: what was designed (CSMA/CD) is now dead
- CSMA/CD: “contention” one big cable, when anyone talks, everyone hears, collisions mean nobody can understand anyone else
- Note: No packet forwarding

Original Ethernet



Ethernet had implications for layer 3

- Routing algorithm overhead is proportional to the number of neighbors
- If you consider a link with 1000 nodes as 1000^2 links, routing alg doesn't scale
- So I adapted DECnet to use this new form of link

But the world got confused

- Thought of Ethernet as a *competitor* to layer 3, rather than a link in a network
- But the Ethernet header was not intended to be an end-to-end, forwardable header (e.g., no hop count)
- I tried to argue with a group that was leaving out layer 3

Arguing...

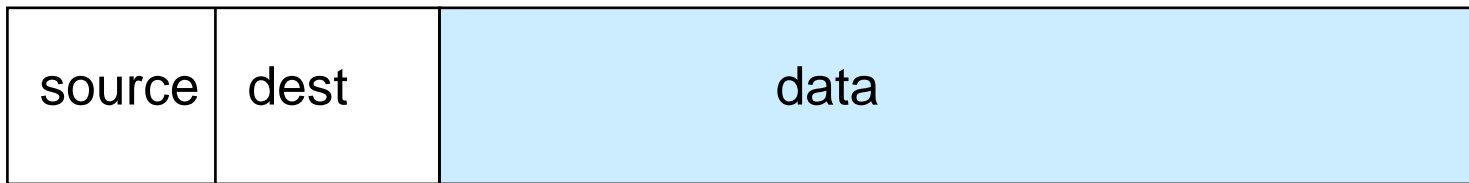
- Me: please don't leave out layer 3
- Them: you're just upset that your layer is a dinosaur and we don't need it anymore
- Me: But you may want to talk from one Ethernet to another
- Them: Our customers would never want to do that

Layer 3 packet



Layer 3 header

Ethernet packet



Ethernet header

Notice hop count is missing

Another difference

- Layer 3 addresses are “hierarchical”
 - Like phone or postal: country, state, city, ...
- Ethernet addresses are “flat”
 - Like social security numbers
- Flat addresses give no hint about where something is: routing has to keep track of every individual

It's easy to confuse “Ethernet” with “network”

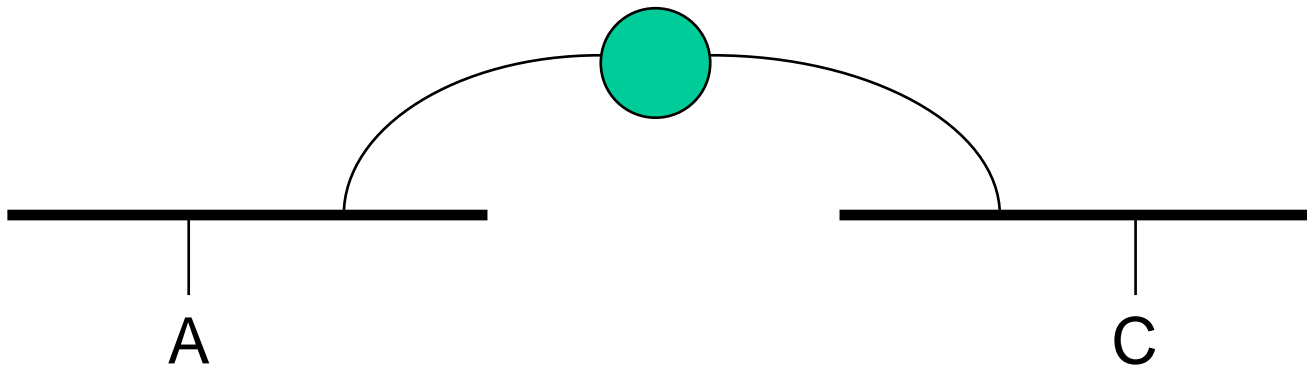
- Both are multiaccess clouds
- Why can't Ethernet replace IP?
 - Flat addresses
 - No hop count
 - (as originally conceived) limited in physical size, number of nodes, total amount of traffic

Why did the world need something other than routers?

- People built protocol stacks leaving out layer 3
 - Routers only work if endnodes implement the router's layer 3 protocol
- There were lots of layer 3 protocols (IP, IPX, Appletalk, CLNP), and few multi-protocol routers
 - So you'd need a router for each protocol

Problem Statement

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



Basic bridge idea

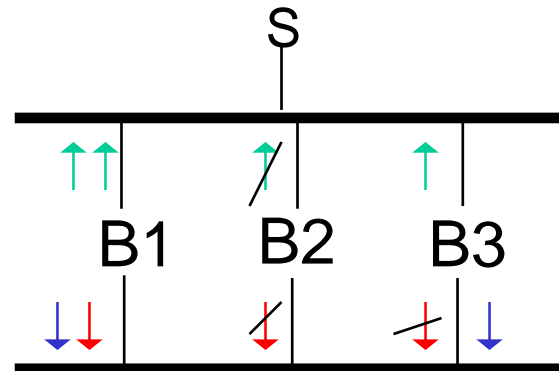
- Listen promiscuously
- Learn location of source
- Forward based on Ethernet header,
based on learned location of destination

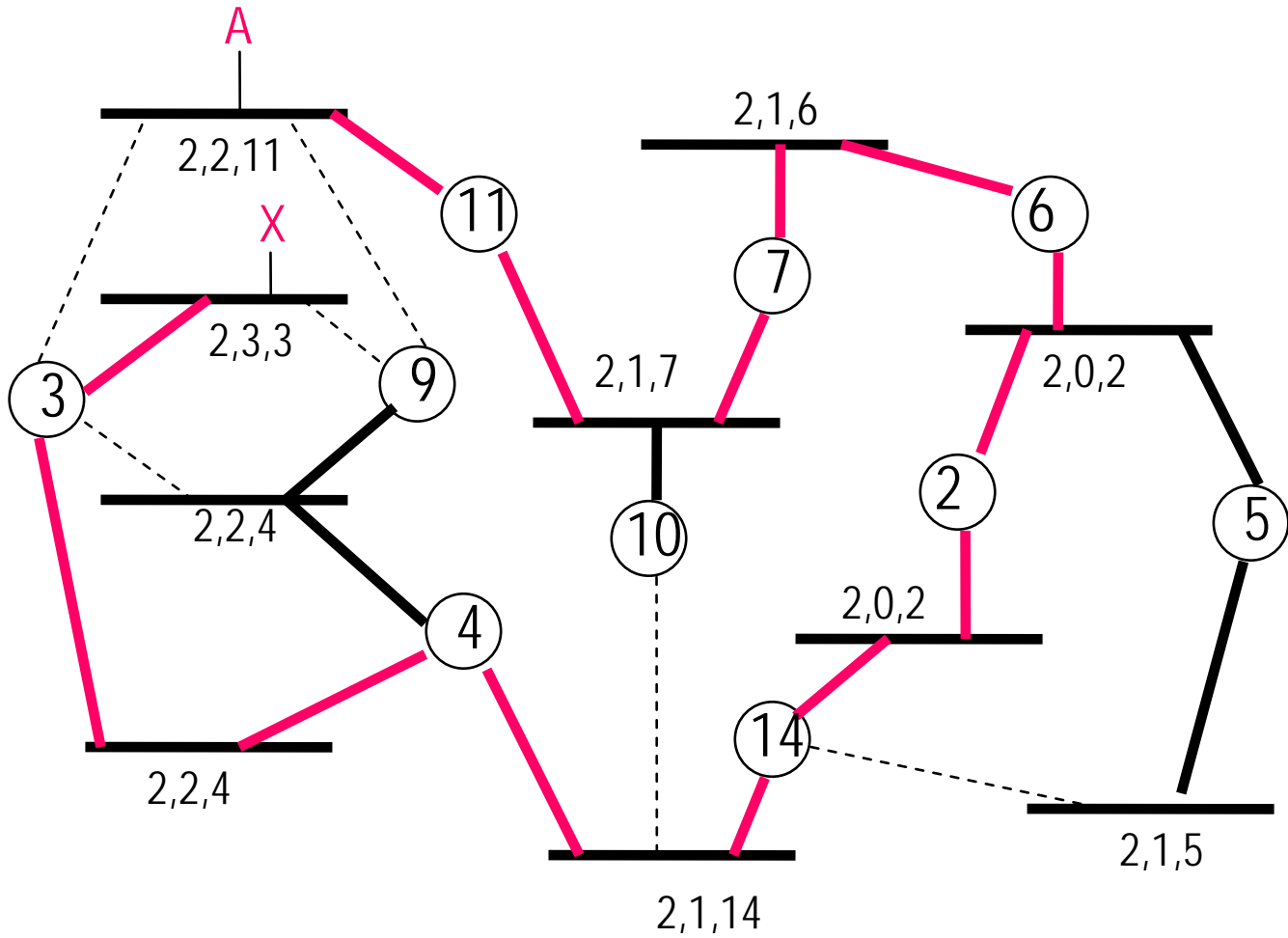
Need loop-free topology

- Can't learn location of source if it's in multiple directions
- loops are a disaster (no hop count, exponential proliferation)

Bridge Loops

- No hop count
- Exponential proliferation





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.*

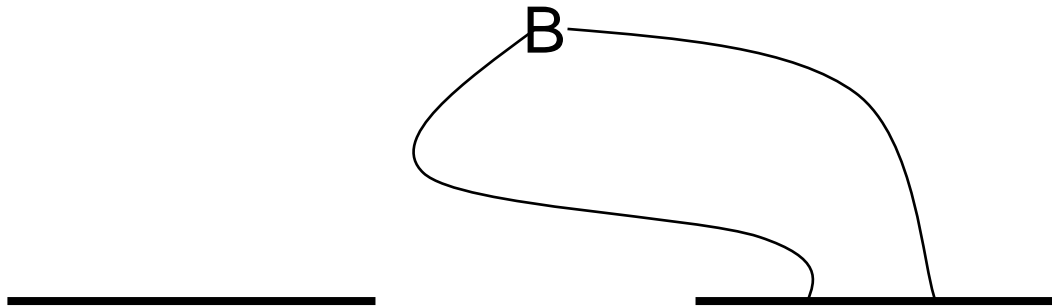
*Least-cost paths from Root are traced.
In the tree these paths are placed.*

*A mesh is made by folks like me.
Then bridges find a spanning tree.*

Bridges without Spanning Tree?

- Implementers wanted bridges as simple as possible. “Don’t allow loops”
- Felt a little bad about forcing them to do STP
- ... Until, first customer site

First customer site



What's wrong with bridges?

- Suboptimal routing
- Traffic concentration
- Temporary loops real dangerous (no hop count, exponential proliferation)
- Fragile
 - If lose packets (congestion?), turn **on** port

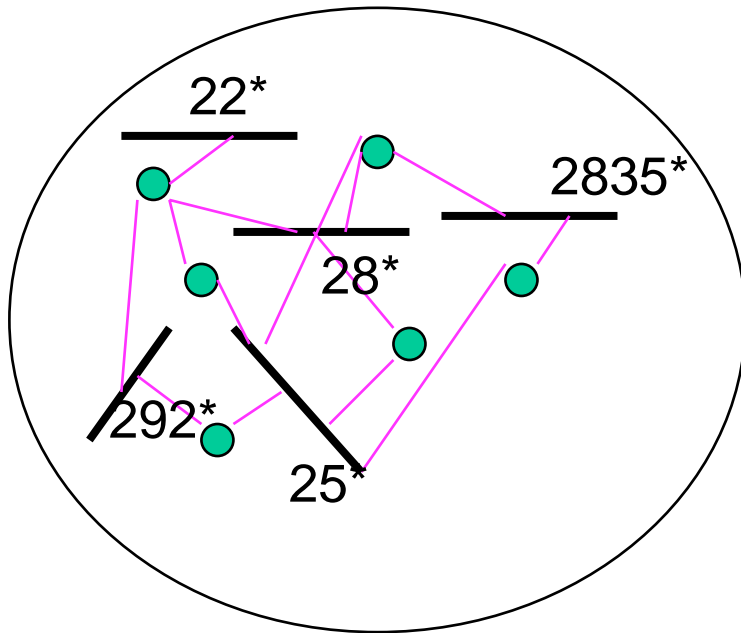
Bridge success

- Simple, fast, reliable, plug-and-play
- Routers have gotten better. Why still bridges?
- Subtle reason: IP needs address per link.
- Layer 3 doesn't have to work that way
 - CLNP and DECnet don't work that way:
 - Bottom level of routing is a whole cloud with the same prefix
 - Routing is to endnodes inside the cloud

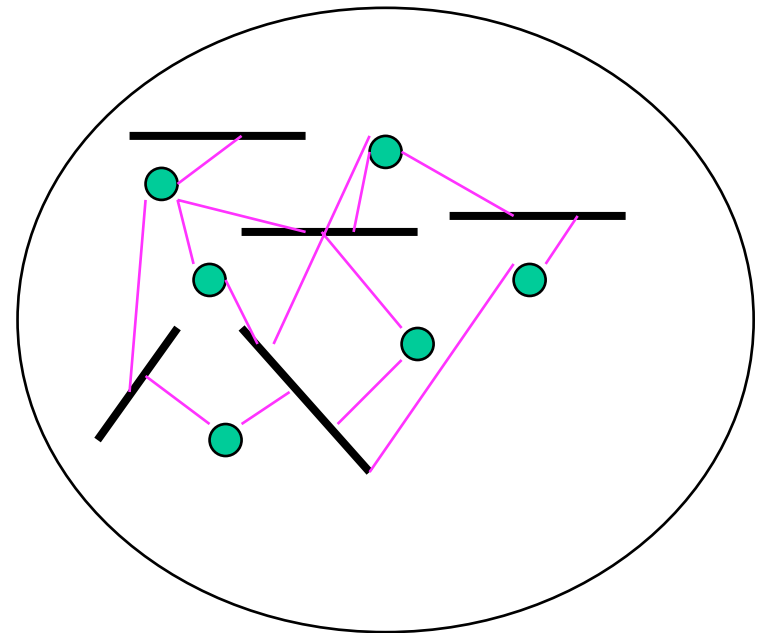
Hierarchy

One prefix per link (IPv4 and IPv6)

One prefix per campus



2*



2*

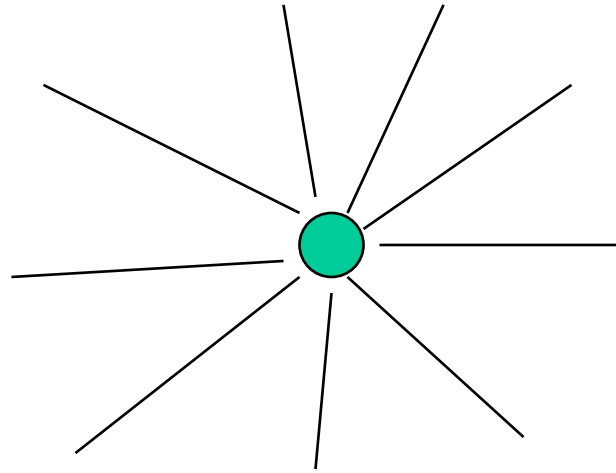
Advantages of lots of links in a prefix

- Zero configuration of routers inside campus
- Nodes can move within campus and keep address
- Address space doesn't need to be chopped up

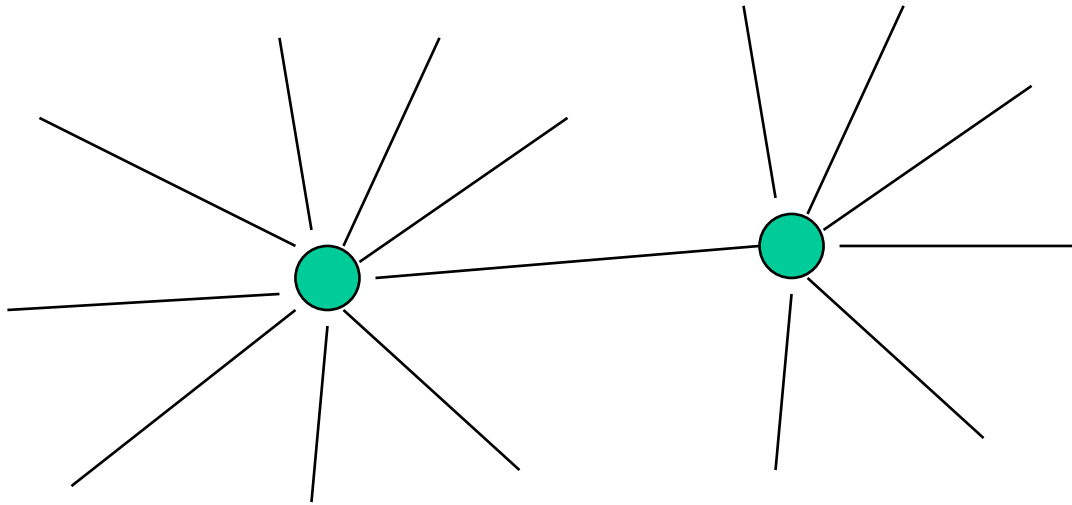
What are switches?

- Myth: Ethernet is wildly successful
- Reality: Ethernet (CSMA/CD) doesn't exist anymore!
- Panel: bus (Ethernet) vs ring (vs star)

Stars



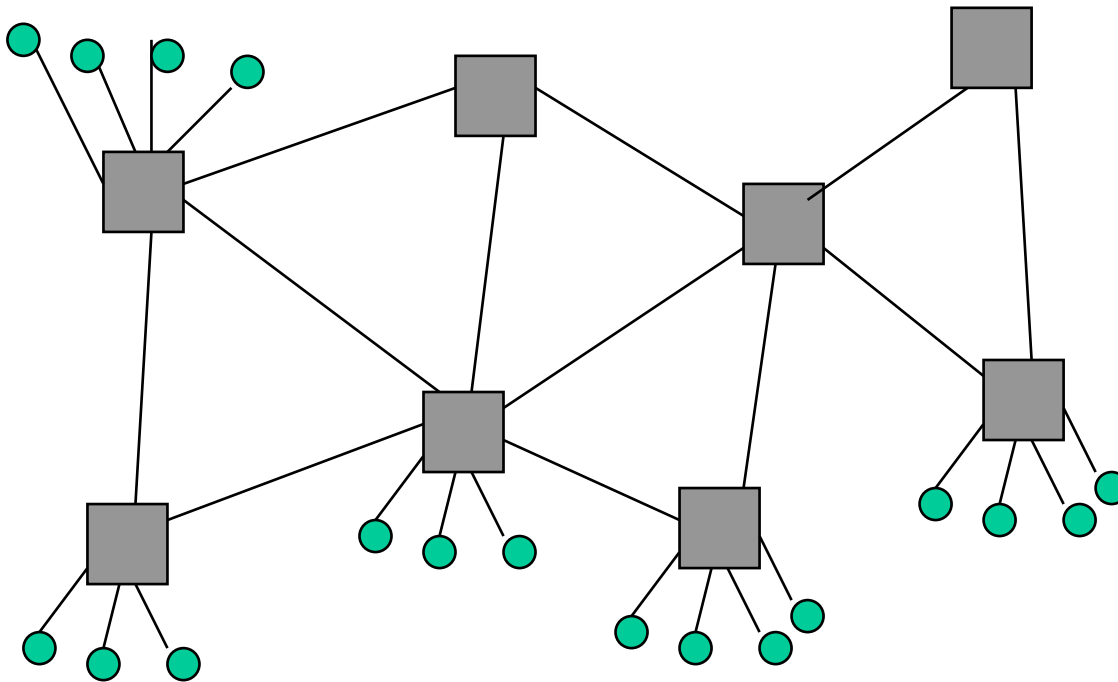
Plug star into another star



Stars

- Start with multi-port repeater
- Then notice that hub can be smarter. Store and forward, learn location of sources.
- Then notice can cascade them.
- Should run spanning tree.
- We've reinvented the bridge!
- "Switched Ethernet": pt-to-pt links with bridges!

Current Ethernet



Result

- Switches=bridges
- Once people figured that out, they started calling routers “layer 3 switches”

New topic: What's with IPv6?

- Why should it have taken over a decade to design?
- What's really new?
- IAB, in 1992, realized IP addresses were too small, and said we should replace IP with CLNP, ISO's "Connectionless Network Protocol"
- At the time, CLNP fully implemented, mature standard, enthusiasm in Europe

What's CLNP?

- ISO's connectionless packet format
- 20 byte addresses
- Lots of links in a prefix (so corporate networks could be zero configuration, with optimal routes, ...)
- In 1992 all the major vendors had implemented it, and supported replacing IP with it

IPv6 reality

- Much harder to change Internet to bigger addresses now than in 1992
 - Internet much larger
 - More “mission-critical”
- Less incentive now
 - delay necessitated inventions like DHCP, NAT
- Per-link prefixes, like IPv4, so people will still want bridges

Result

- We are probably stuck with IPv4 forever
- Or at least, migration will be much more painful than it would have been
- And IPv6 is not as good technically as CLNP would have been

New topic: simple things people screw up

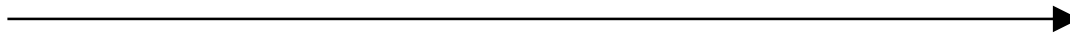
- Parameters. Need to set in running net!
 - better if plug and play
 - Sometimes parameters must be compatible:
 - hello timer vs dead timer

IS-IS Strategy

Alice

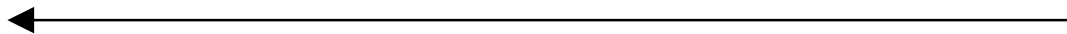
Bob

Hello...I send Hellos every 22 secs



Oh. So I should wait 70 secs before panicking

Hello...I send Hellos every 8 secs



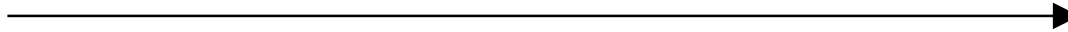
Oh. So I should wait 25 secs before panicking

OSPF Strategy

Alice

Bob

Hello...I send Hellos every 22 secs



Oh no! It's not the same as my Hello timer. I refuse to talk to her!

Forward compatibility

- TLV encoding “Type”, “Length”, “Value”
- For optional fields that it’s OK to skip if you don’t implement them
- Example: priority might be encoded as:
 - Type=17
 - Length=2 (bytes)
 - Value

How can you mess up TLV?

- L has to be always expressed in the same units!

Version number

- A lot of protocols have a field called “version number”
- What’s it for?

Version number

- What's the difference between a new protocol and a new version of a protocol?
- For instance, is IPv6 a new version of IP, whereas CLNP would have been “replacing IP with a different protocol”?

What's a “different” protocol vs a new “version”?

- Reasonable definition of “new protocol”
 - It has a different Ethertype (or port)
- New version
 - share the same Ethertype (or port), disambiguate based on version number
- If you ever make the format incompatible, change the version number

Which means...

- That a spec has to say “set the version number as x , and if a received packet has a version number you don't understand, don't try to parse it”

Version numbers

- Lots of protocols made (and continue to make) this mistake
- Implementations don't check the version field. They just set it.
- So can't do "new version"
- Need a different "Ethertype" (or port)
- Which means it's a different protocol

Version numbers

- IPv4
 - Just says “set the field to 4”
 - So you can’t send an IPv6 packet to an IPv4 node (with same Ethertype)
 - Because it would totally misparse it
 - So IPv6 is a *new protocol*, not a new *version* of IPv4

So you'd assume they've learned
their lesson for IPv6

- No...the spec says “fill in this field as 6”

SSL

- Version 3 completely changed the format from version 2
- And even moved the version field!

Things that shouldn't work but do

- Web search
- Wikipedia

Things that can't possibly be hard

- Safely reading email and surfing the web
- User interface
- Authentication

Rant about authentication

- Should be based on public keys
- Why are we still just using passwords?

Returning user

- Scenario: Buy something from a merchant you haven't bought from recently
- All prepared with your info, credit card, etc.
- It asks you for your email address...

You're a returning user!

- Type your username and password
- Of course you can't remember it, so...
 - you manage to find “recover username”
 - suddenly you are in a Monty Python movie
 - Answer the following questions three:
 - Telephone number
 - Address
 - Mother's maiden name

New Rule

- It should be no more onerous to be a returning user than a new user

Security questions for password/username recovery

- Favorite sports team
- 2nd grade teacher's name
- Pet's name
- Father's middle name
- My middle name

New Rule

- Security questions must be specifiable by the user
- I'd say "or selectable from a very large list", but I'm sure they can come up with an arbitrarily long list of questions I can't answer

Security question in comedy routine

- Question: “Are you wearing underwear”?
- Answer: “I don’t think that’s an appropriate question”

Keeping customer information

- I do not want to do “single click ordering”
- I do not mind typing in my address
- I do not mind typing in my credit card number
- Merchants insist on keeping all of this information
- And eventually this information gets stolen

New Rule

- After a merchant is paid, any subset of information about a customer (including all of the information) must be expunged by the merchant at the customer's request

Summary

- We need to have protocol designers humble enough to learn from previous protocols
- If things aren't simple they won't work
- Know what problem you're trying to solve before you try to solve it!



ICT-onderzoek Platform Nederland

Andy Tanenbaum (VU)

Questions?

