Internetworking networks

Gentian Jakllari
Département Sciences du Numérique
jakllari@enseeiht.fr
http://jakllari.perso.enseeiht.fr/

The internetworking idea (Robert Kahn, 1972)

- Build a single network (an interconnected set of networks, or *internetwork*, or *internet*) out of a large collection of separate networks

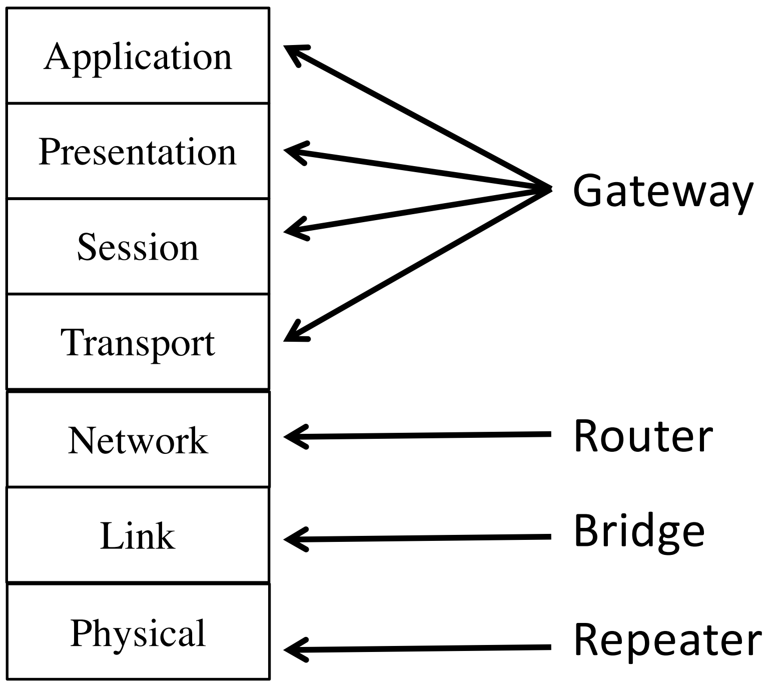
- Four ground rules[1]
  1. Each distinct network must stand on its own, with no internal changes required to connect to the internet.
  2. Communications should be on a best-effort basis.
  3. ***"Black boxes" should be used to connect the networks.***
  4. No global control at the operations level.

[1]Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff. 2009. A brief history of the internet. *SIGCOMM Comput. Commun. Rev.* 39, 5 (October 2009)

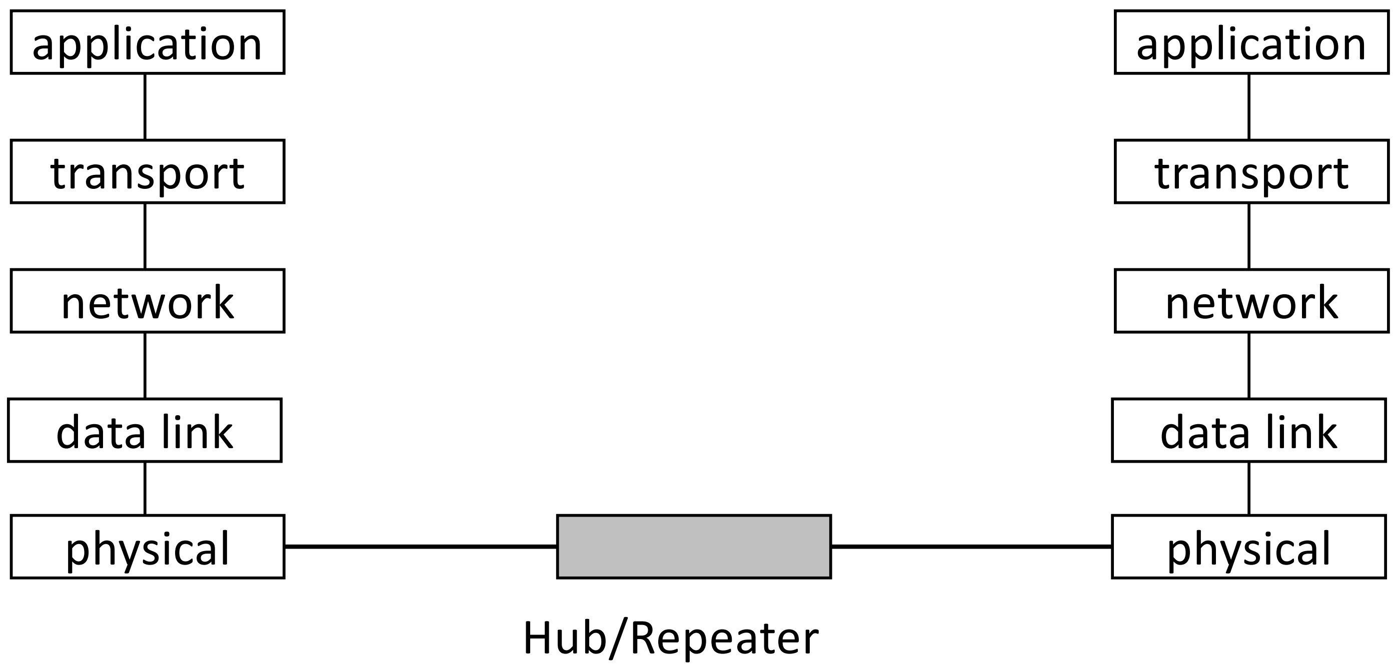
“Black Boxes” challenges

- Heterogeneity
  - Lots of different kinds of networks (Ethernet, FDDI, ATM, WiFi, point-to-point)
  - How to unify this hodgepodge?
- Scale
  - how to keep together potentially billions of nodes?

Different kinds of “Black Boxes”



Internetworking with repeaters

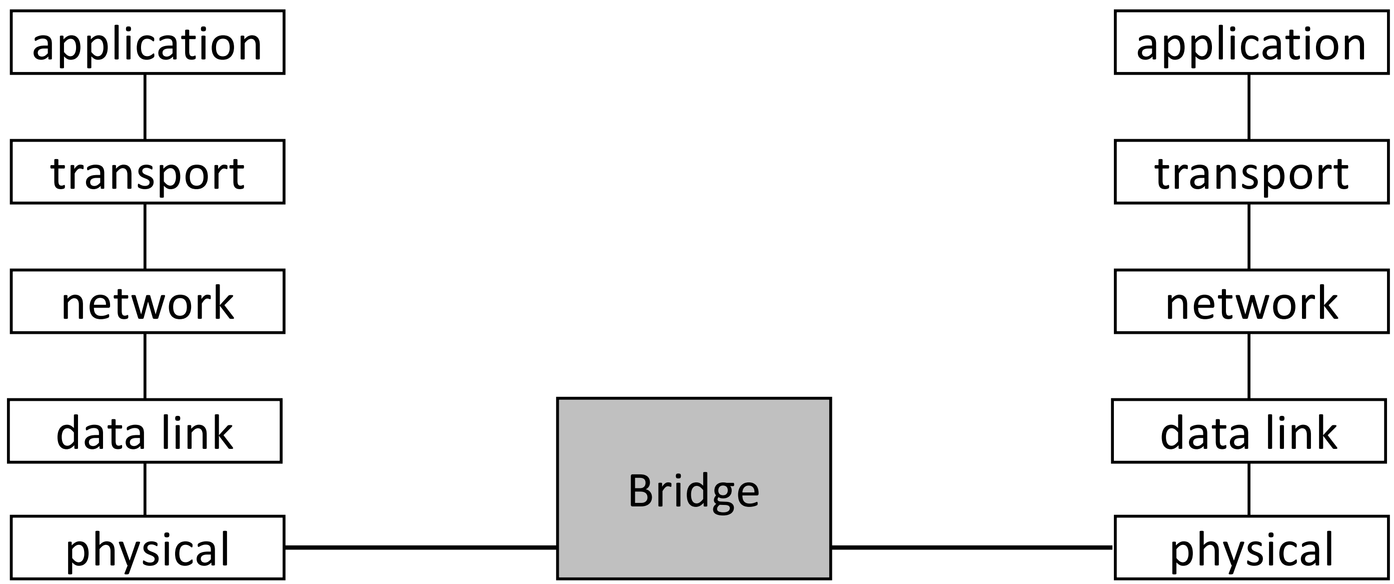


Vote on live.voxvote.com   PIN: 91758

- Can an internetwork scale to billion of nodes using repeaters only ?
  1. Yes, we can build as many repeaters as necessary
  2. No, building this many repeaters would be prohibitively costly
  3. No, there wouldn’t be enough communication bandwidth to handle all the traffic
  4. Good question

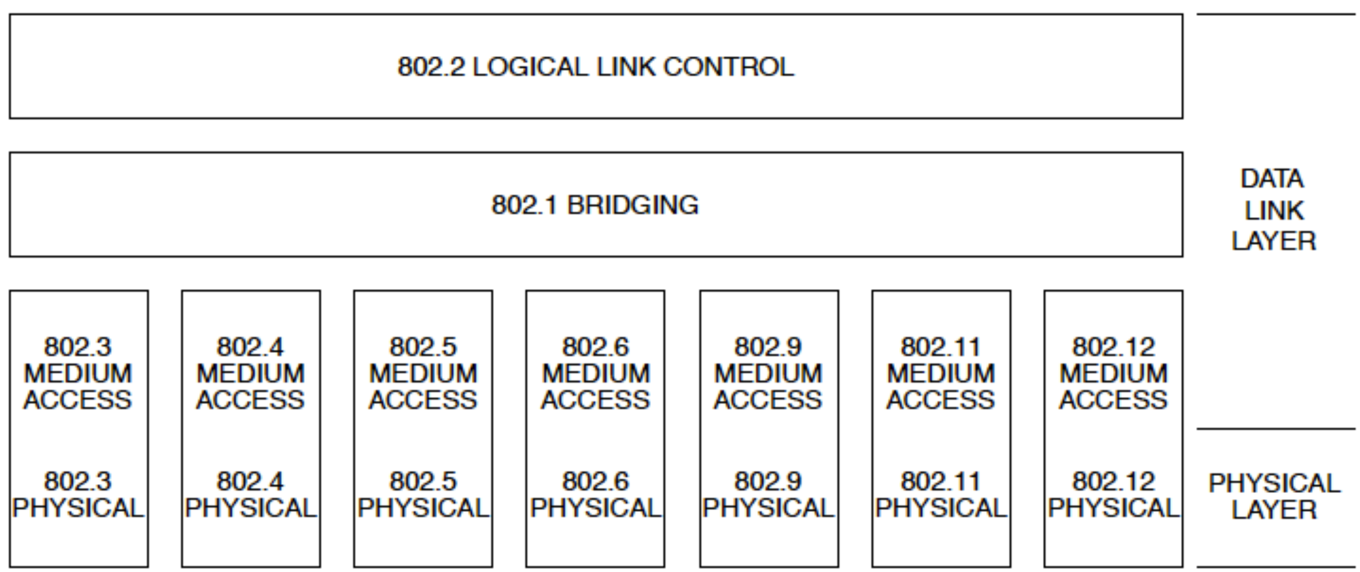
- Repeaters can handle heterogeneity:
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree

Internetworking with bridges

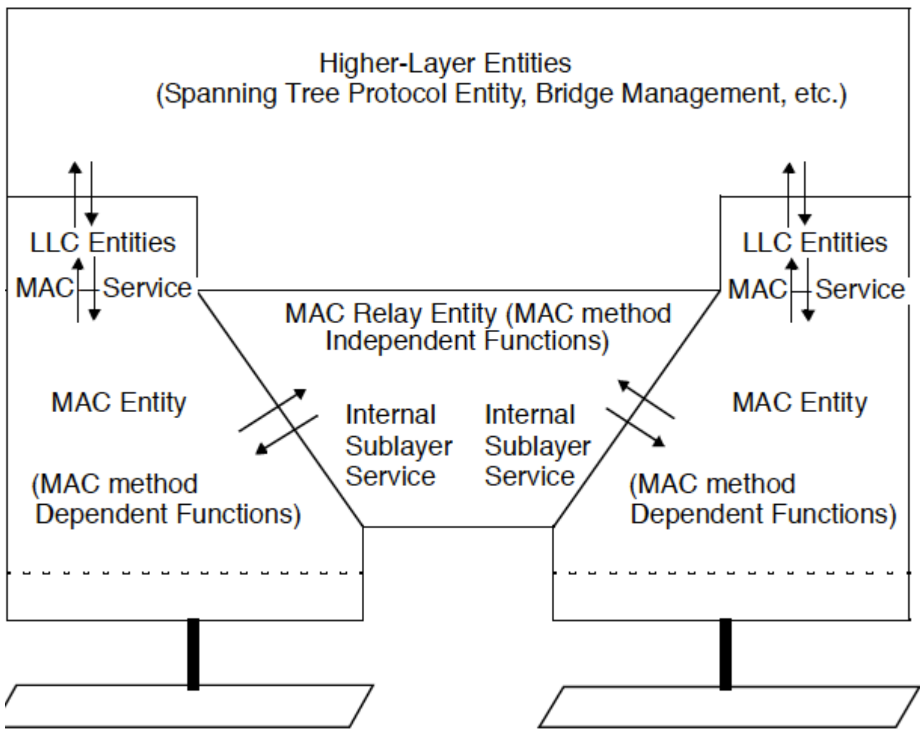


Les adresses mac n'ont pas de système de localisation contrairement à ip donc les bridges ne peuvent pas passer à l'échelle. Par exemple à l'n7 on a les adresses 147.27.X.X/16 du coup par exemple on peut mettre dans un routeur à paris une route vers ce réseau pour tout message destiné à l'n7 et un routeur à l'n7 s'occupe de les router vers son destinataire final. Avec les adresse mac il faudrait faire pareil pour toutes les machines individuellement (dans les tables des bridges). Les tables de routage des routeurs ip sont de ce fait plus simplifié.
Un bridge avec de la surcharge se comporterait comme un hub. Les bridges comprennent le niveau 2 donc peut faire de la traduction (faire le lien entre dux technologies).

Bridge based internetworking



Bridge Architecture



- MAC relay entity: relaying frames between ports, filtering, filtering learning
- Each bridge port transmits and receives frames to and from the LAN to which it is attached
- MAC Entity handles Media Access

MAC Relay Entity

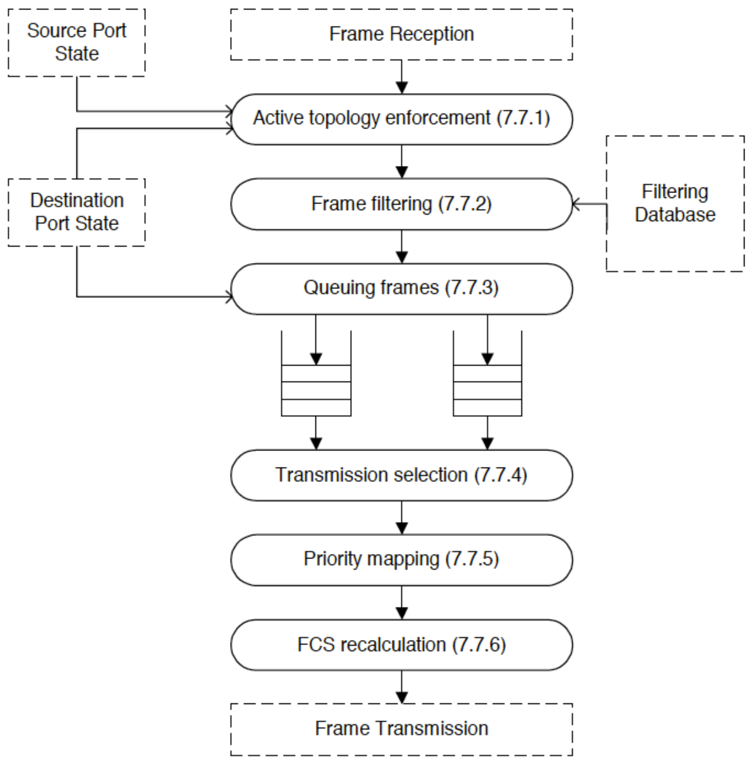
- **Forwarding**
- Learning
- Filtering

Question

- What is the difference between forwarding and routing?

  - *Write your answer on a post-it*

Forwarding



- Active topology enforcement: only allow transmissions on ports selected by the spanning tree
- Filtering: Based on the destination MAC address carried in a received frame and Filtering Database
- Queuing: Up to eight traffic classes are supported by the traffic class tables

Forwarding: User priorities and traffic classes

Number of queues	Traffic types
1	{*Best Effort*, Excellent effort, Background, Voice, Controlled Load, Video, Network Control}
2	{*Best Effort*, Excellent effort, Background} {*Voice*, Controlled Load, Video, Network Control}
3	{*Best Effort*, Excellent effort, Background} {*Controlled Load*, Video} {*Voice*, Network Control}
4	{*Background*} {*Best Effort*, Excellent effort} {*Controlled Load*, Video} {*Voice*, Network Control}
5	{*Background*} {*Best Effort*, Excellent effort} {*Controlled Load*} {*Video*} {*Voice*, Network Control}
6	{*Background*} {*Best Effort* {*Excellent effort*} {*Controlled Load*} {*Video*} {*Voice*, Network Control}
7	{*Background*} {*Best Effort*} {*Excellent effort*} {*Controlled Load*} {*Video*} {*Voice*} {*Network Control*}

- Network control – maintain network infrastructure
- Voice – less than 10 ms delay
- Video – less than 100 ms delay

Forwarding: Transmission selection

- Frames are selected from a queue for transmission only if all queues corresponding to numerically higher values of traffic class supported by the Port are empty

- A frame queued for transmission on a Port is dropped if that is necessary to ensure that the maximum bridge transit delay
  - Recommended value: 1.0 second
  - Absolute maximum: 4.0 seconds

MAC Relay Entity

- Forwarding
- **Learning**
- Filtering

The learning process

- The bridge listens promiscuously
- For each packet received, the bridge stores the source address field in the Filtering Database together with the port on which the packet was received
- For each packet received the bridge looks through its stations cache for the address listed in the packet’s s destination address field to decide on which port to forward
- The bridge ages each entry
  - Recommended default value: 300.0 seconds
  - Range: 10.0-1000000 seconds

MAC Relay Entity

- Forwarding
- **Learning**
- Filtering

Addressing – End Stations

- All MAC Entities communicating across a Bridged Local Area Network use 48-bit addresses
- Frames transmitted between end stations carry the MAC Address of the source and destination peer end stations
  - The address of a Bridge is not carried in frames transmitted between peer users for the purpose of frame relay

Addressing – Bridge

- The individual MAC Entity associated with each Bridge Port has a separate individual MAC Address
- A unique 48-bit Universally Administered MAC Address, termed the Bridge Address, shall be assigned to each Bridge
- The Bridge Address may be the individual MAC Address of a Bridge Port, in which case, use of the address of the lowest numbered Bridge Port (Port 1) is recommended

- Can a network scale to billion of nodes connected through bridges ?
  1. Yes, we can build as many bridges as necessary
  2. No, bridges create loops
  3. No, there wouldn’t be enough communication bandwidth to handle all the traffic
  4. Yes, with bridges there is enough bandwidth to handle all the traffic

- Bridges can handle heterogeneity:
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree

Bridges vs. Hubs

- Who does internetworking best?

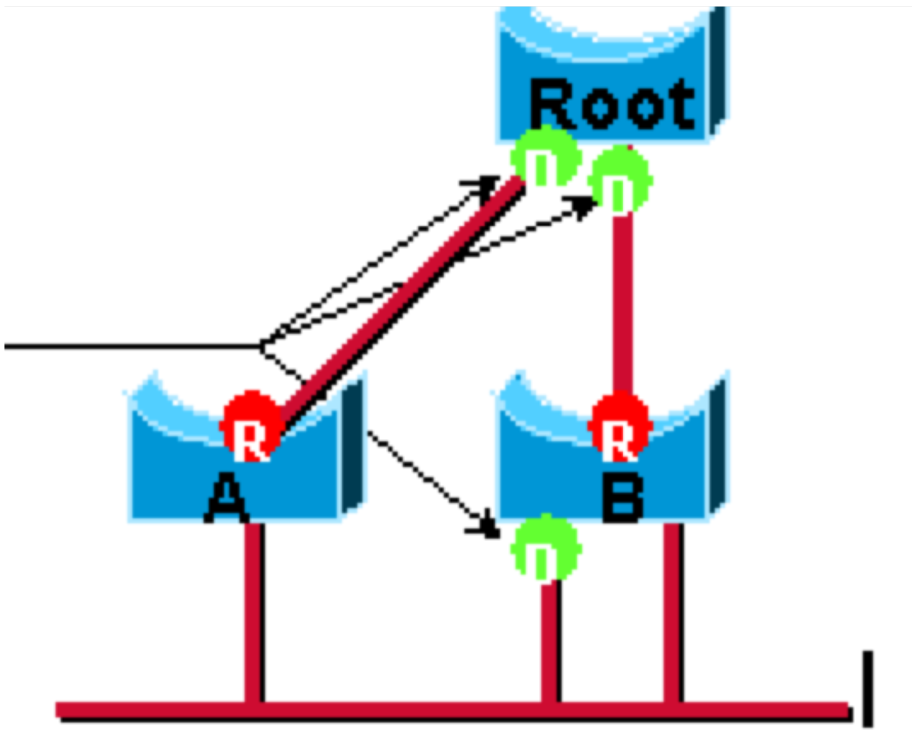
- We can replace bridges with routers
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree

Rapid Spanning Tree Protocol

- Spanning Tree Protocol was designed at a time when the recovery of connectivity after an outage within a minute or so was considered adequate performance

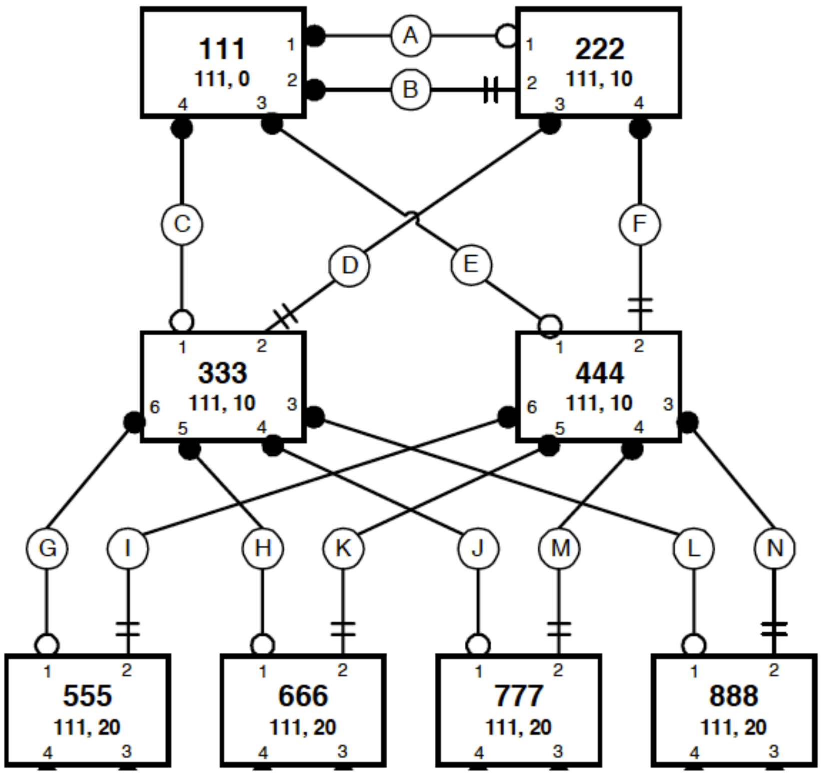
- With the advent of Layer 3 switching in LAN environments, bridging now competes with routed solutions where protocols, such as OSPF recover faster

Spanning Tree Protocol



- Identify root, root ports and designated ports
  - Using Bridge Protocol Data Units (BPDUs) packets
- The rest of the ports are blocked to avoid loops

Spanning Tree Protocol



Port Role	Port State	Legend
Designated	Discarding	
	Learning	
	Forwarding	
& operEdge	Forwarding	
Root Port	Discarding	
	Learning	
	Forwarding	

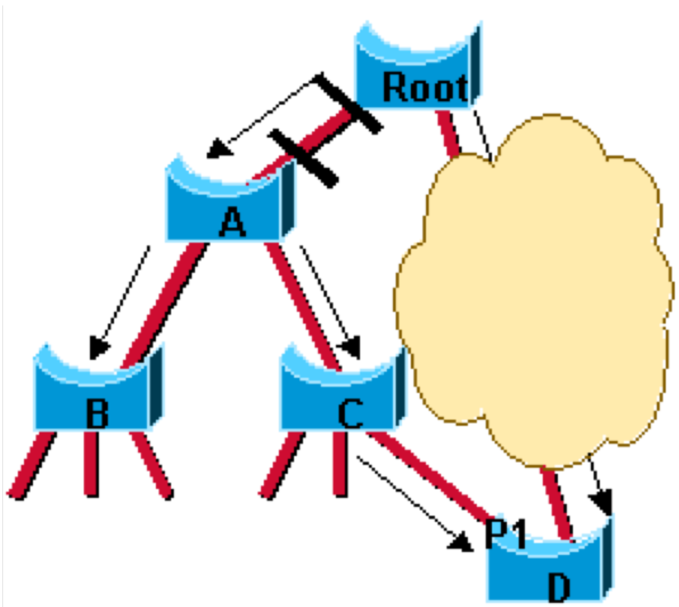
Accelerating STP - New BPDU Handling

- BPDU are Sent Every Hello-Time (2 s by default)
  - In legacy STP, a non-root bridge only generates BPDUs when it receives one on the root port
- Faster Aging of Information
  - A bridge considers that it loses connectivity to its direct neighbor root or designated bridge if it misses three BPDUs in a row

Accelerating STP - Rapid Transition to Forwarding State

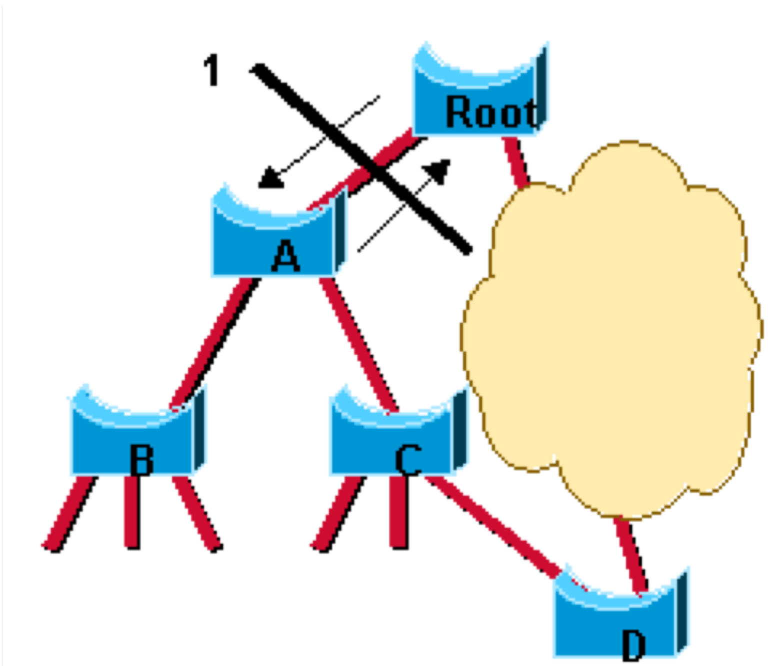
- The legacy STP passively waited for the network to converge before it turned a port into the forwarding state
- Edge ports: all ports directly connected to end stations
  – Cannot create bridging loops
  – Directly transition to the forwarding state, skipping the listening and learning stages
- Point-to-point links
  – A port that operates in full-duplex is assumed to be point-to-point and can transition to forwarding state
  – In switched networks today, most links operate in full-duplex

Legacy STP – Adding a new link



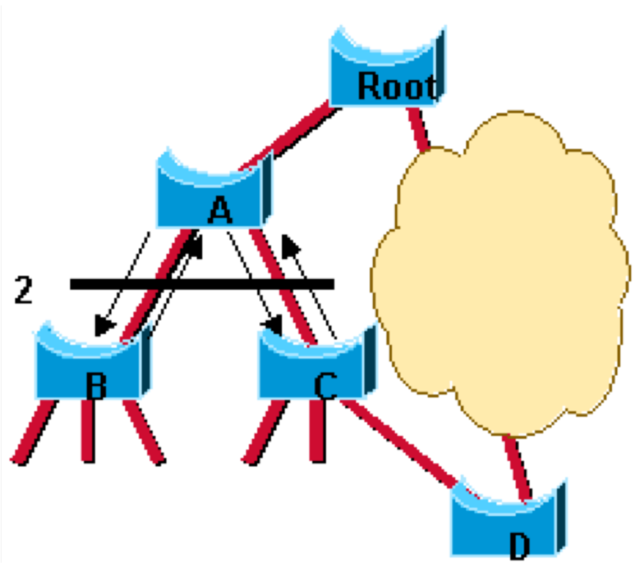
- New link between Root and A is added
- The respective ports will be set in listening waits twice the forward delay seconds (2x15 by default) before they can switch to forwarding
- D finds out fast and blocks P1 to avoid a loop, leaving A, B, C isolated for 30 s.

Rapid STP – Adding a new link



- As soon as A receives the BPDU of the root, it blocks the edges to B and C
- Explicitly authorizes the root bridge to put its port in the forwarding state
- Root and switch A can start immediately to exchange data

Rapid STP – Adding a new link



- The cut travels down the tree along with the new BPDUs originated by the root through Switch A
- The rest of the switches can function normally

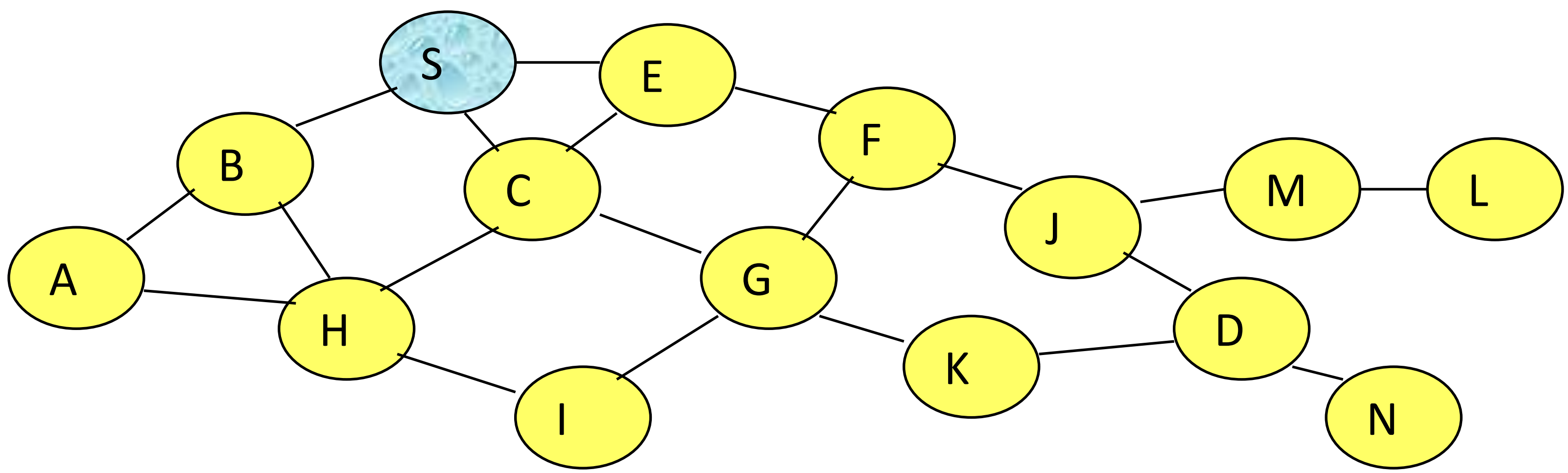
Source routing bridges

- Basic idea: the packet header contains a route and the route is inserted by the source station

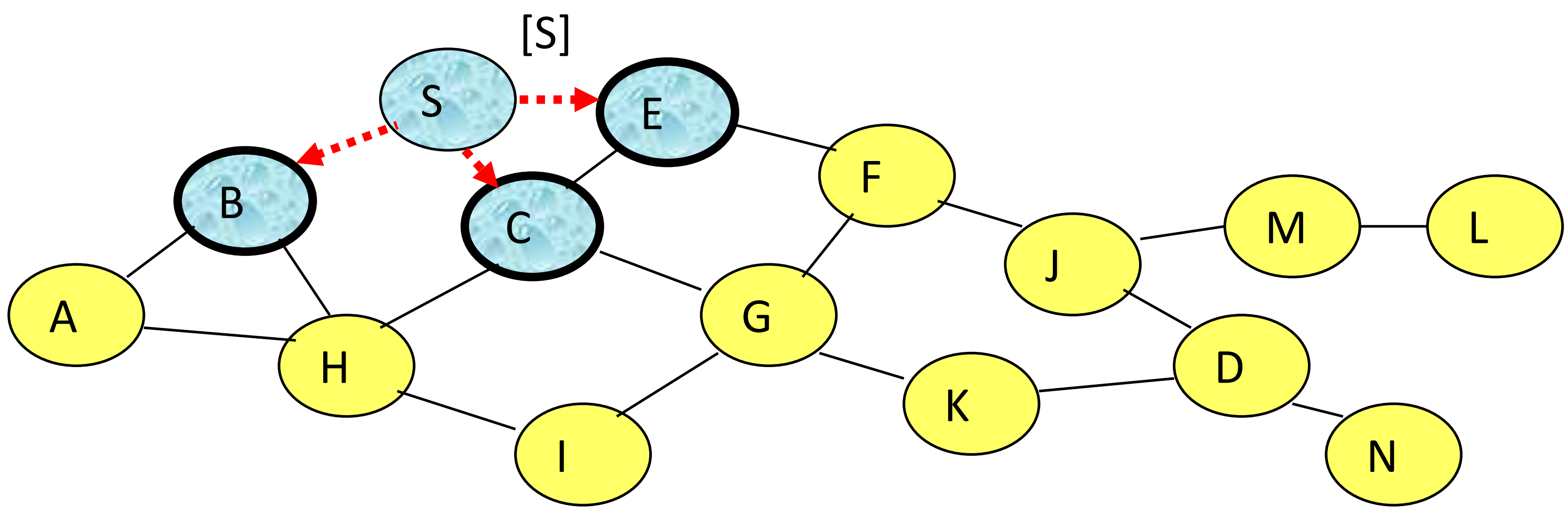
- Stations must discover routes by transmitting a special kind of packet that replicates itself as it reaches other stations

- Each copy collects a diary of its travels so that when the copies reach the destination station a route can be selected

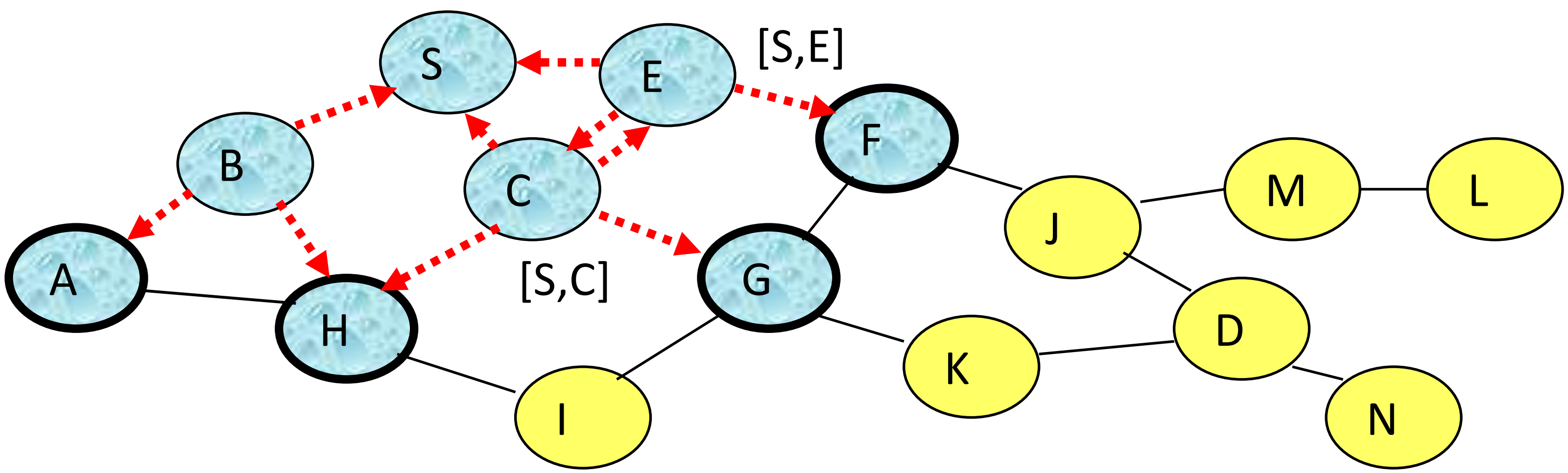
Route discovery in source routing



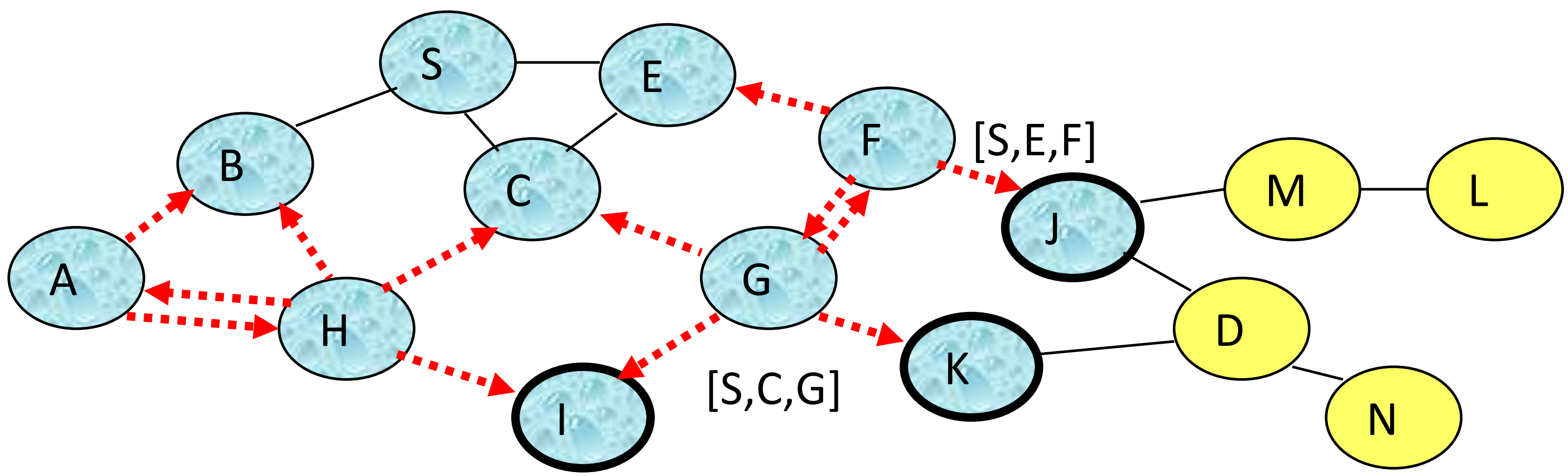
Route discovery in source routing



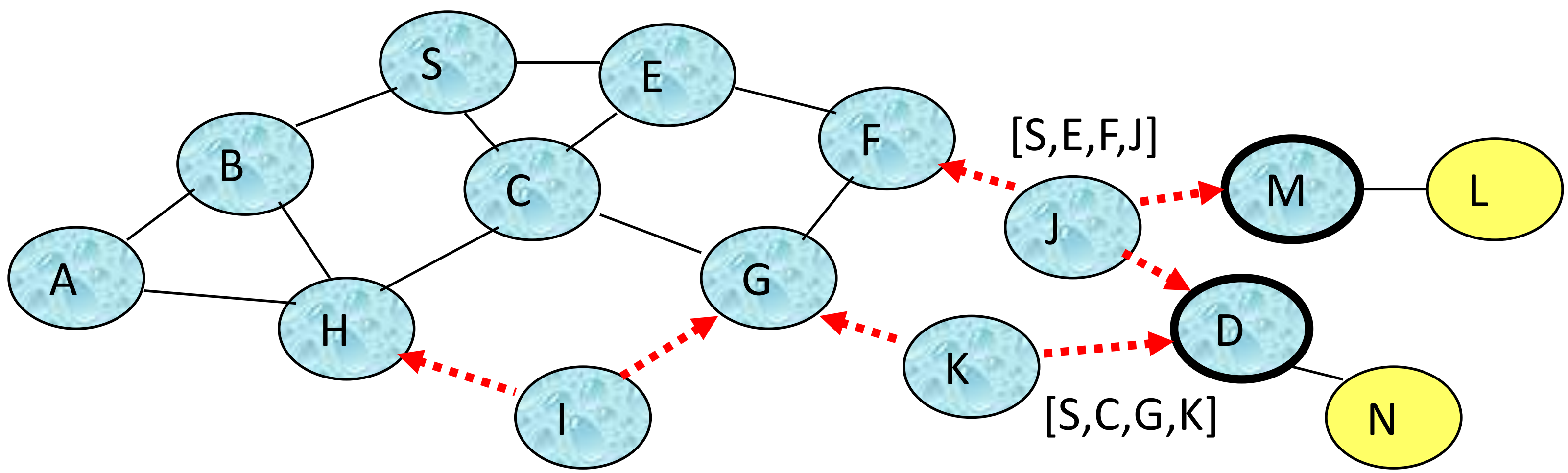
Route discovery in source routing



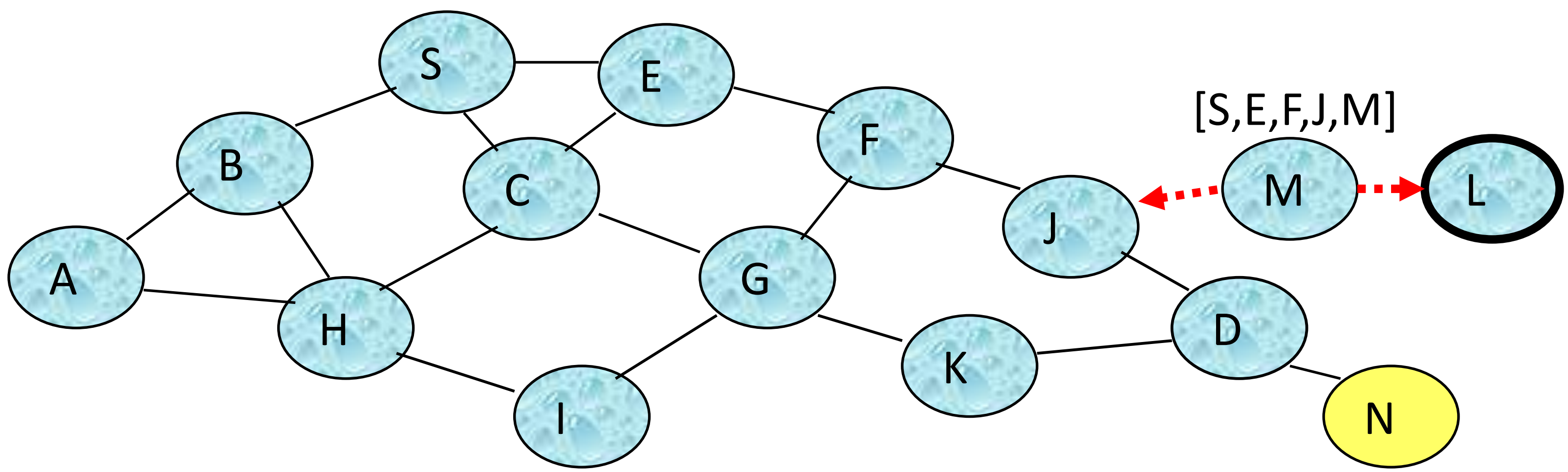
Route discovery in source routing



Route discovery in source routing



Route discovery in source routing



Source routed packet

destination	source	RI	data

- Type
  1. Specifically routed (route in the header)
  2. All paths explorer
  3. Spanning tree explorer (cross only spanning tree)
- Length: number of bytes in the RI field
- Route: a sequence of 2-byte-long fields, route designators, each of which consists of a 12-bit LAN number followed by a 4-bit bridge number

LAN	bridge

- We can replace routers with bridges
  - Strongly agree
  - Agree
  - Disagree
  - Strongly disagree

Bridges

- How many networks can we internetwork with bridges?