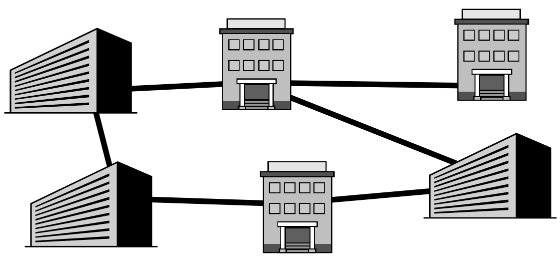
Packet Switching Techniques

Problem

- Aim: Build larger networks connecting more users
  - also spanning different network technologies



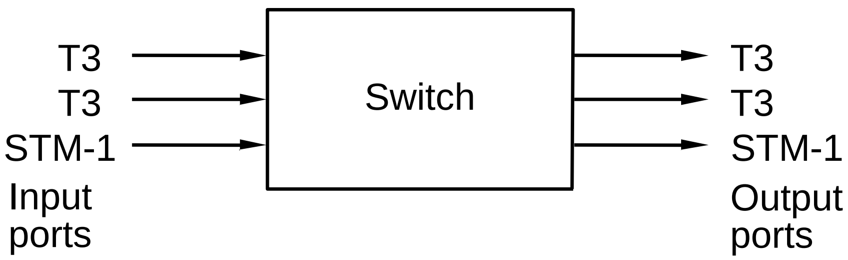
- Shared media networks
  - limited number of stations that can be connected
  - limited geographical coverage
- Solution: switched networks
  - review general switching technologies
  - extending LANs through LAN switches (bridges, Ethernet switches)

Outline

- General switching techniques
- Bridges and LAN switches
- ATM technology
- ATM in LAN

Scalable networks

- Switch
  - provides star topology
  - forwards (switches) packets from input port to output port
    - ports terminate link layer protocols (T1, T3, STM-n)
  - port selected based on address in packet header



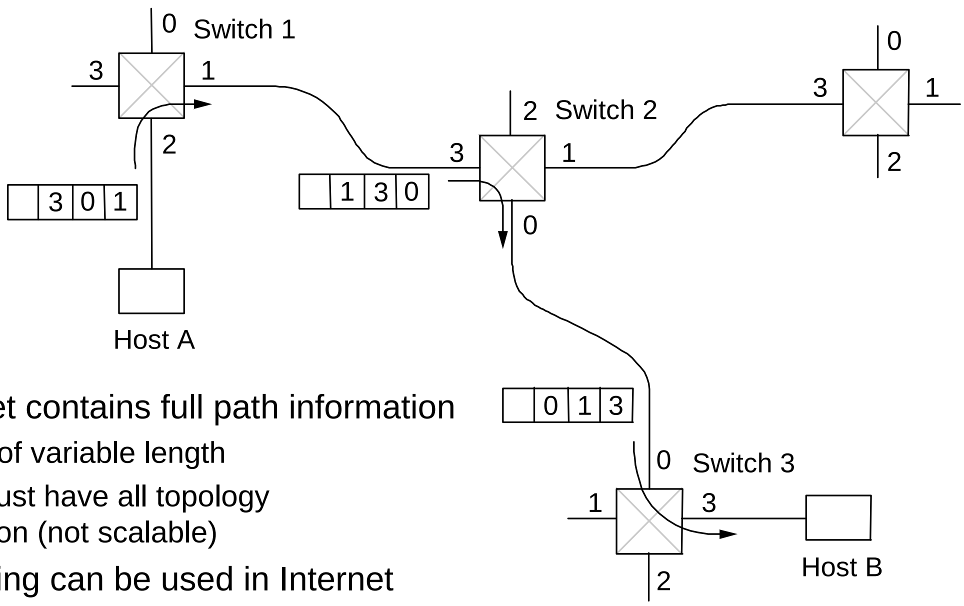
- Advantages
  - can cover large geographic area (tolerate latency)
  - can support large numbers of hosts (scalable bandwidth)
    - in Ethernet adding new hosts decreases bandwidth available to users

Switching methods and addressing

- Switching/forwarding
  - systematic process of determining to which output port a packet is sent based on packet's header
  - different techniques to achieve this
    - source routing model
    - connectionless (datagram) model
    - connection oriented (virtual circuit) model

- Requires method to identify end hosts
  - done by using addressing
  - here it is assumed hosts have globally unique addresses (telephone number, Ethernet)
  - in Internet addresses can be "reused" (discussed in lectures on routing)

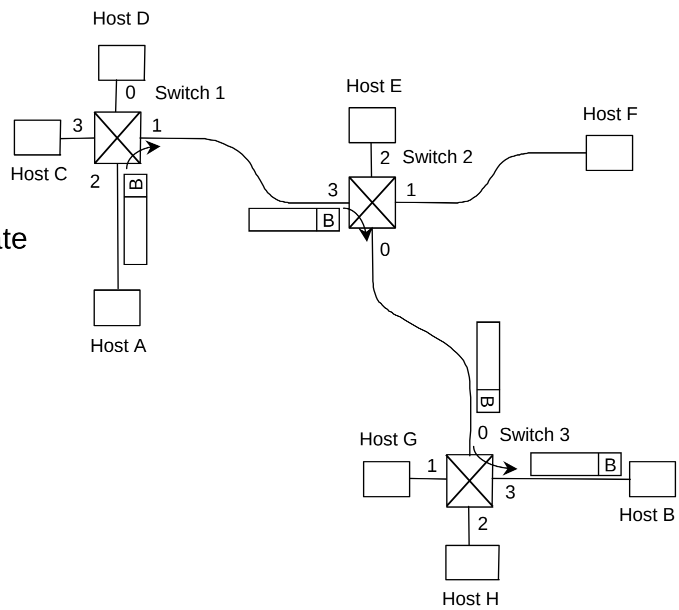
Source routing



- Every packet contains full path information
  - headers of variable length
  - nodes must have all topology information (not scalable)
- Source routing can be used in Internet

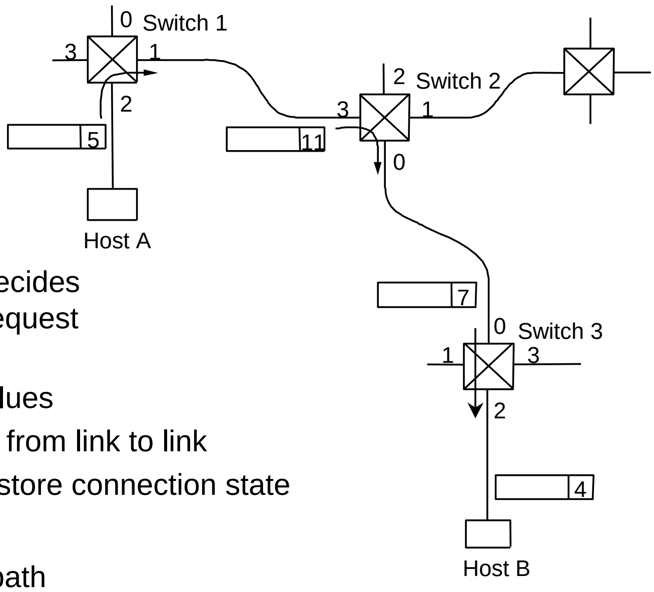
Datagram switching

- Each packet contains complete destination address
  - Analogy: postal system
- No notion of connections
  - network does not store any state
- Switches (routers) maintain routing tables
  - packets routed independently
  - constructing routing tables difficult for complex networks



Virtual circuit (VC) switching

- Connection setup phase before information transfer
- VCs can be
  - permanent (nw mgmt)
  - on-demand (signaling)
- Connection set up (on-demand)
  - along forward path, each switch decides VC identifier (VCI) and forwards request
  - along reverse path, nodes inform upstream nodes of chosen VCI values
  - VCIs local identifiers, may change from link to link
  - switches maintain VCI tables and store connection state
- Information transfer
  - subsequent packets follow same path
- Connection tear down after transfer over
- Analogy: telephone call



Features of VC switching

- Source must wait full RTT before sending any data (VC set up delay)
- While the connection request contains the full address for destination, each data packet contains only a small identifier
  - per-packet header overhead smaller than in datagram switching
- If a switch or a link in a connection fails, the connection is broken and a new one needs to be established.
- Connection setup provides an opportunity to reserve resources
  - in POTS/cellular networks it is checked if time slot is available or not
  - In general, it can be checked if enough bandwidth/buffer resources exist and if users' delay requirements can be met
- Examples: POTS network, cellular networks, Frame Relay, ATM

Features of datagram switching

- No connection set up delay
  - host can send data as soon as it is ready
  - network does not need to store connection state
- As a result, source can not know
  - if the network is capable of delivering a packet
  - if the destination host is even up
- Packets are treated independently
  - it is possible to route around link and node failures
  - important e.g. in military environment (ARPANET developed for military)
- Since every packet must carry the full address of the destination, the overhead per packet is higher than for the connection-oriented model
  - forwarding based on global addresses more resource consuming

Traffic control and switching techniques

- Congestion
  - packets arrive at a rate greater than link speed $\Rightarrow$ buffer fills up
- Preventive traffic control
  - network attempts to ensure that congestion does not occur by allocating resources to VCs
  - circuit switching model
  - may result in poor resource utilization (if VCs exhibit random behavior)
- Reactive traffic control
  - network tries to recover from congestion as fast as possible
  - datagram switching model
  - achieves better utilization since resources are not dedicated to any particular connection, resources shared among all (statistical multiplexing)

Outline

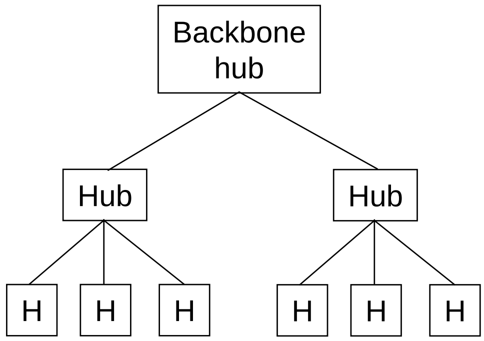
- General switching techniques
- Bridges and LAN switches
- ATM technology
- ATM in LAN

Extending LANs

- A single LAN
  - all stations on the LAN share bandwidth
  - limited geographical coverage (Ethernet 2500 m)
  - = single collision domain
  - nof stations limited
- Techniques:
  - hubs
  - bridges
  - Ethernet switches

Hubs

- Layer 1 device
  - repeats received bits on one interface to all other interfaces (repeaters)
- Hubs can be arranged in hierarchy
  - provides star topology
- Each connected LAN referred to as LAN segment
- Hubs do not isolate collision domains
  - collisions may happen with any node on any segment



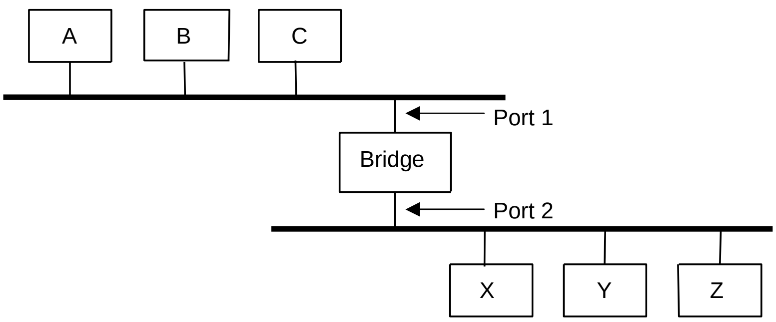
Note! Still a single LAN and collision domain

Hubs: advantages and limitations

- Hub advantages:
  - simple, inexpensive device
  - hierarchy provides graceful degradation: portions of the LAN continue to operate if one hub malfunctions
  - extends maximum distance between node pairs (100m per Hub)
- Hub limitations:
  - single collision domain results in no increase in max throughput
  - multi-tier throughput same as single segment throughput
  - individual LAN restrictions pose limits on number of nodes in same collision domain and on total allowed geographical coverage
  - cannot connect different Ethernet types (e.g., 10BaseT and 100baseT)

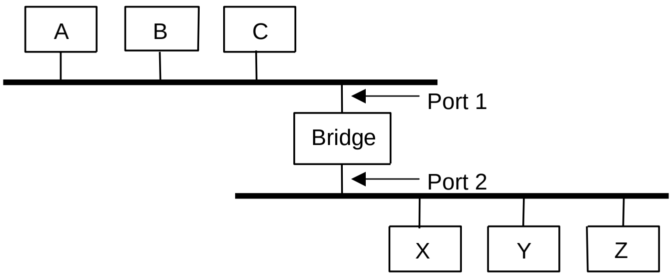
Simple bridges

- Connect two or more LANs with a bridge
- Simple bridge
  - all packets from a particular port forwarded on all other ports
  - level 2 forwarding/switching (does not add packet header)
  - isolates collision domains
  - Ethernet bridge uses CSMA/CD to transmit packets onto connected LANs



Learning bridges

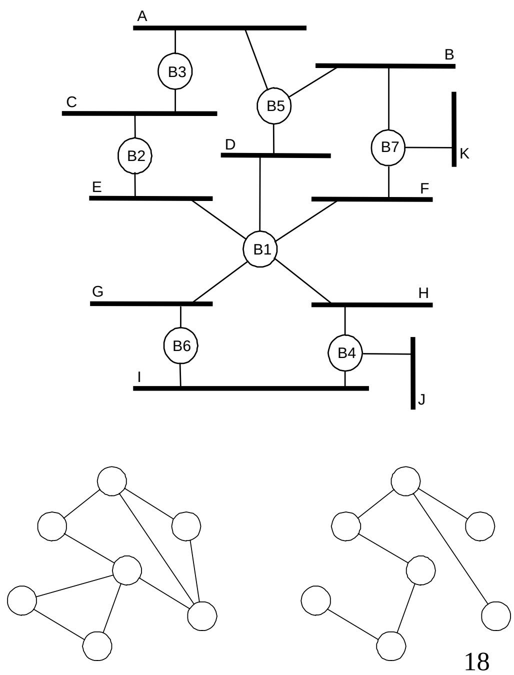
- Idea: forward only when necessary
  - bridges maintain forwarding tables
  - datagram switching on layer 2
- Dynamic algorithm
  - bridge examines source address of each packet seen on a port
  - addresses saved in a table
  - table entries have time outs (if host moves from one segment to another)
  - broadcast frames always forwarded
- Table is an optimization; need not be complete



Host	Port
A	1
B	1
C	1
X	2
Y	2
Z	2

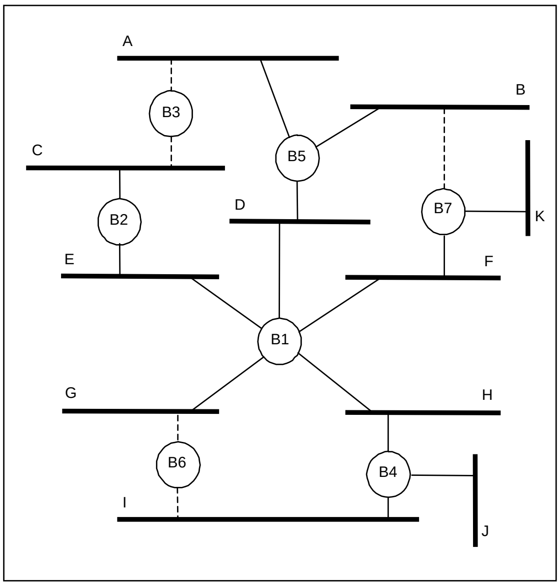
Spanning tree algorithm

- Problem: loops in topology
  - loops used to provide redundancy in case of link failure
- Bridges run a distributed spanning tree algorithm
  - selects which bridges actively forward traffic
  - subset of network graph representing a tree that spans all nodes
  - dynamic algorithm: tree reconfigures when topology changes
  - developed by Radia Perlman
  - now IEEE 802.1 specification



Algorithm overview

- Bridges have unique ids (B1, B2, etc.)
- Bridge with smallest id is root
  - root forwards all traffic onto all ports
- Select designated bridge on each LAN
  - LAN may be connected to many bridges
  - bridge “closest” (=min nof hops) to root selected as designated bridge
  - id used to break ties
- Each bridge forwards frames over each LAN for which it is the designated bridge
- Dynamic (survives link failures), but not able to utilize multiple paths during congestion



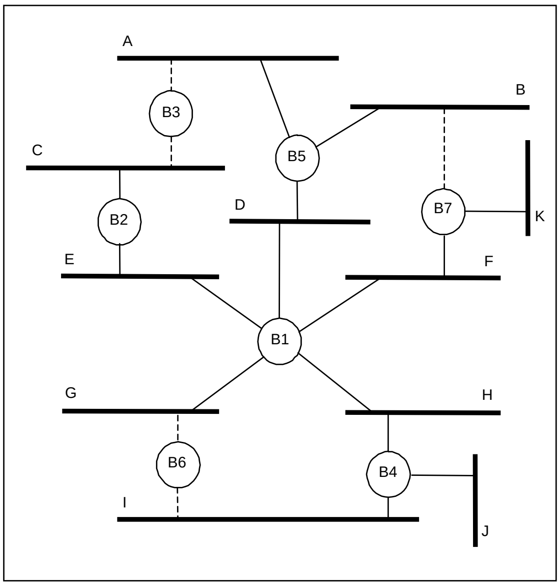
Algorithm details

- Bridges exchange configuration messages containing
  - id for bridge sending the message
  - id for what the sending bridge believes to be root bridge
  - distance (hops) from sending bridge to root bridge
- Each bridge records current best configuration message for each port
  - identifies root with smaller id
  - identifies root with same id but with shorter distance
  - root id and distance are same but sending bridge has smaller id
- Initially, each bridge believes it is the root
- When learn not root, stop generating config messages
  - in steady state, only root generates configuration messages
- When learn not designated bridge, stop forwarding config messages
  - in steady state, only designated bridges forward config messages
- Root continues to periodically send config messages
  - If any bridge does not receive config message after a period of time, it starts generating config messages claiming to be the root

Algorithm example

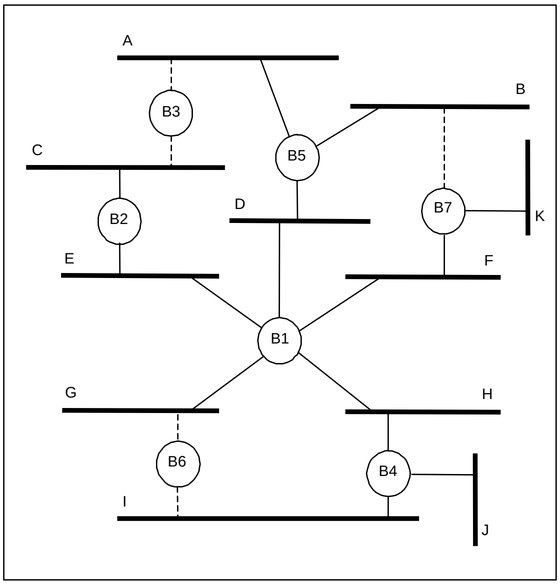
- (Y, *d*, X): Message from X, at distance *d* from root Y
- Consider node B3:

1 B3 receives (B2, 0, B2)
2 2 < 3, B3 accepts B2 as root
3 B3 increments d and sends (B2, 1, B3) towards B5
4 B2 accepts B1 as root (lower id) and sends (B1, 1, B2) towards B3
5 B5 accepts B1 as root and sends (B1, 1, B5) towards B3
6 B3 accepts B1 as root, notes that B2 and B5 are closer to root $\Rightarrow$ stop



Broadcast and Multicast

- Forward all broadcast/multicast frames
  - current practice
- Possible optimization for multicast:
  - learn when no group members downstream (similarly as in learning bridges)
  - typically group members are not sending any traffic
  - accomplished by having each member of group G send a frame to bridge multicast address with G in source field
  - not widely deployed

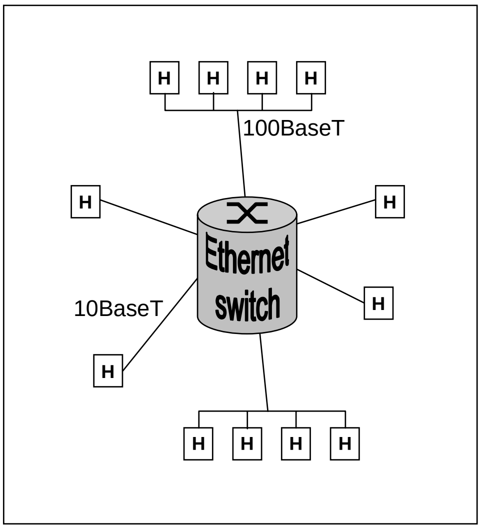


Bridges: advantages and limitations

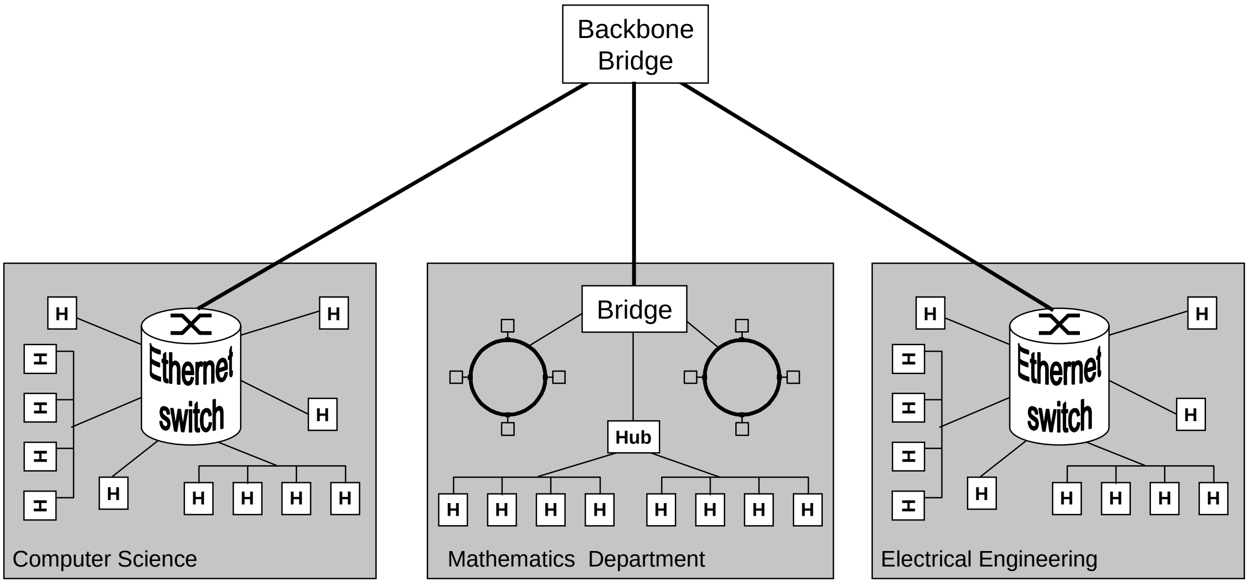
- Advantages:
  - isolates collision domains (higher throughput than when using hubs)
  - can connect multiple LAN types (different Ethernets, Token Rings)
  - transparent: no need for any changes in end hosts
  - used to build network of “tens” of LAN segments within e.g. a campus area
- Limitations:
  - scalability:
    - spanning tree algorithm does not scale
    - broadcast does not scale (VLANs can be used to alleviate)
  - heterogeneity: not all network technologies use 48 bit addresses
- Caution: beware of transparency
  - in an extended LAN, there may be congestion, larger delays, ...
  - end host applications should not assume that all is behind a single LAN

Ethernet switches

- Layer 2 forwarding and filtering using LAN addresses
- Uses switching (frames can be sent in parallel between multiple ports)
- Can accommodate large number of interfaces
  - mix of 10/100/1000 Mbit Ethernets
  - shared (multiple hosts) or dedicated (single host) Ethernets
- Common configuration
  - star topology: hosts connected to switch
  - Ethernet, but no collisions!



Building a campus area network



Outline

- General switching techniques
- Bridges and LAN switches
- ATM technology
- ATM in LAN

History

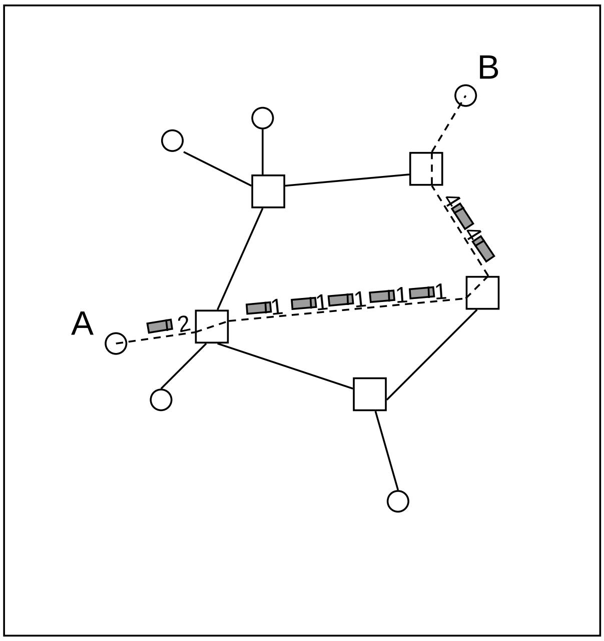
- Traditionally: dedicated networks for different services
  - For example: telephone, telex, data, broadcast networks
  - Optimized for the corresponding service
- Need for integration of all services into a single ubiquitous network
  - "One policy, one system, universal service" (T. Vail, AT&T's first president)
  - Early 80's: Research on Fast Packet Switching started
- Answer from the "Telecom World" : **B-ISDN**
  - 1985: B-ISDN specification started by Study Group SGXVIII of CCITT
  - 1988: Approval of the first B-ISDN recommendation (I.121) by CCITT
  - idea: replace old PSTN network with new one
- Chosen implementation method: **ATM**
  - 1990: ATM chosen for the final transfer mode for B-ISDN by CCITT
  - 1991: ATM Forum founded
    - to accelerate development of ATM standards
    - to take into account needs of the "Computer World"

Design objectives

- Flexibility
  - integration of different types of traffic (like voice, data and video) into a single network
  - supporting any transmission rate (even variable)
  - combination of best properties of circuit and packet switching
- Guaranteed Quality of Service (QoS)
  - end-to-end delay, delay variation, cell loss ratio
- Scalability
  - suitable for both WAN and LAN
- Fastness
  - both in switching and transmission $\Rightarrow$ broadband
- Supports multiple physical layers
  - typically used over SDH (SONET)
- Minimal error correction (only header protected by check sum)
  - very low bit error rates on optical links

ATM

- Connection oriented:
  - virtual connections set up end-to-end before information transfer
  - Q.2931 signaling protocol
  - resource reservation possible but not mandatory
- All information carried in short, fixed-length packets (**cells**)
  - along the route chosen for that virtual connection
  - statistical multiplexing at nodes
  - identifier label (local address) at cell header
  - no error detection/recovery for the information field



Technical choices

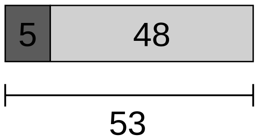
- Connection oriented
  ⇒ resource reservation possible
  ⇒ guaranteed QoS possible

- Packet switching
  ⇒ statistical multiplexing
  ⇒ flexibility (any bit rate possible) and efficiency (high utilization)

- Packets small and fixed-length (cells)
  ⇒ cell switching
  ⇒ fast switching, easier to implement in hardware

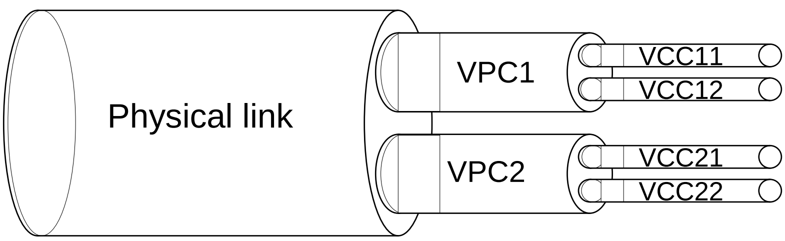
Cell

- **Cell = short, fixed-length packet**
  - Total length = **53 bytes** (octets)
  - **Header**: 5 bytes
    - **GFC**, generic flow control (4 [0] bits at UNI [NNI]), (not used)
    - **VPI**, virtual path identifier (8 [12] bits ⇒ 256 [4096] values)
    - **VCI**, virtual channel identifier (16 bits ⇒ 65,536 values)
    - **PT**, payload type (3 bits)
      - 4 values (1xy) used for management functions
      - user data = 0xy, x used for EFCI (ABR), y=delineate AAL5 frames
    - **CLP**, cell loss priority (1 bit)
    - **HEC**, CRC-8 header error control (8 bits)
  - **Information field**: 48 bytes
    - carried transparently, without error detection/recovery



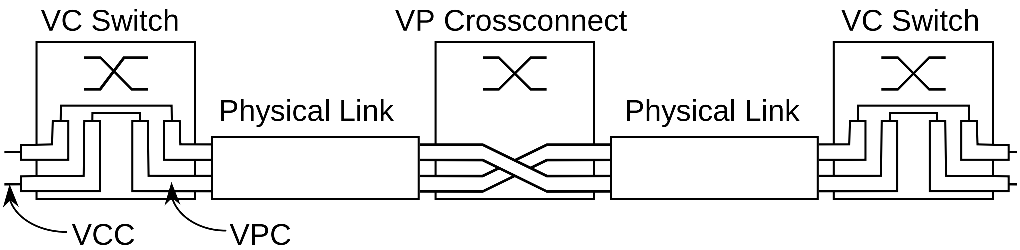
Virtual connections

- Basic connection: **Virtual Channel Connection** (**VCC**)
  - identified by the **VPI/VCI** pair (24 [28] bits) in **each** cell header
    - max. 16,777,216 [268,435,456] VCCs per physical link
  - VPI and VCI fields are **local** labels
    ⇒ reuse possible in different physical links ⇒ scalability
- Aggregated connection: **Virtual Path Connection** (**VPC**)
  - identified by the **VPI** field (8 [12] bits) in **each** cell header
    - max. 256 [4096] VPCs per physical link
  - consists of the VCCs with the same VPI (can be switched together!)



Virtual paths

- Pros
  - faster connection establishment
  - easier network management
  - differentiated QoS possible
  - virtual networks possible
- Cons
  - reduced statistical multiplexing gain



Variable vs. fixed length packets

- No optimal length
  - if small: high header-to-data overhead
    - but e.g. traditional telephone calls generate small sized packets
    - ATM intended to carry all traditional voice traffic effectively
  - if large: low utilization for small messages
    - but, in computer networking packets are typically large
- Fixed length packets easier to switch in hardware
  - in 80’s a lot of effort put into designing hw for switching fixed size packets
  - simpler to implement than variable size packet switching
  - can utilize parallelism in switch design (switches operate based on “slotted” time)

Big vs. small packets (1)

- Small improves queue behavior
  - finer-grained pre-emption point for scheduling link
    - maximum packet = 4KB
    - link speed = 100Mbps
    - transmission time = 4096 x 8/100 = 327.68us
    - high priority packet may sit in the queue 327.68us
    - in contrast, 53 x 8/100 = 4.24us for ATM
  - near cut-through behavior
    - two 4KB packets arrive at same time
    - link idle for 327.68us while both arrive
    - at end of 327.68us, still have 8KB to transmit
    - in contrast, can transmit first cell after 4.24us
    - at end of 327.68us, just over 4KB left in queue

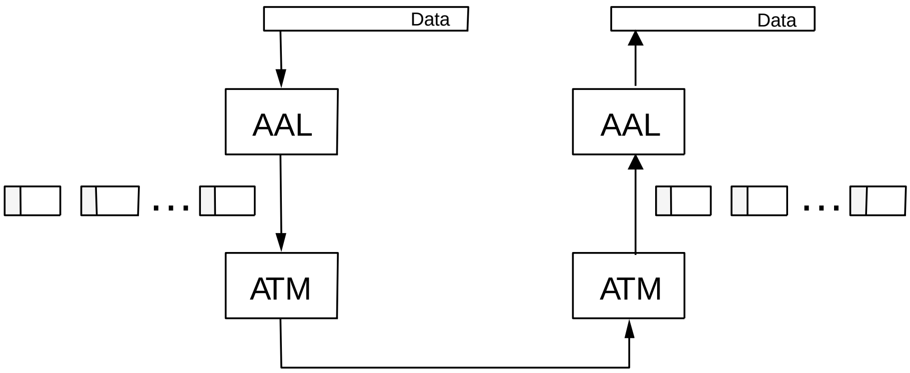
Big vs. small packets (2)

- Small improves latency (for voice)
  - voice digitally encoded at 64KBps (8-bit samples at 8KHz)
  - need full cell's worth of samples before sending cell
  - example: 1000-byte cells implies 125ms per cell (too long)
  - smaller latency implies no need e.g. for echo cancellors

- ATM compromise: 48 bytes = (32+64)/2
  - Europe: 32 bytes
  - USA: 64 bytes

Segmentation and reassembly (SAR)

- Problem: higher layer frames longer than 48 bytes
  - frames need to fragmented and reassembled
- ATM Adaptation Layer (AAL)
  - AAL 1 and 2 designed for applications that need guaranteed rate (e.g., voice, video)
  - AAL 3/4 designed for packet data
  - AAL 5 is an alternative standard for packet data



AAL5

- Problems with AAL 3/4
  - complex, too much protocol overhead
  - 44 bytes of 48 available for splitting AAL 3/4 frame into cells
  - 4 bytes used for segmentation header (enables multiplexing, CRC-10)
- AAL 5
  - all 48 bytes available for AAL 5 frame segmentation, no multiplexing
  - in practice, AAL 5 is used (AAL 3/4 not)
  - CS-PDU format (CS-PDU = AAL 5 frame, CS=Convergence Sublayer)

< 64 KB	0– 47 bytes	16	16	32
Data	Pad	Reserved	Len	CRC-32

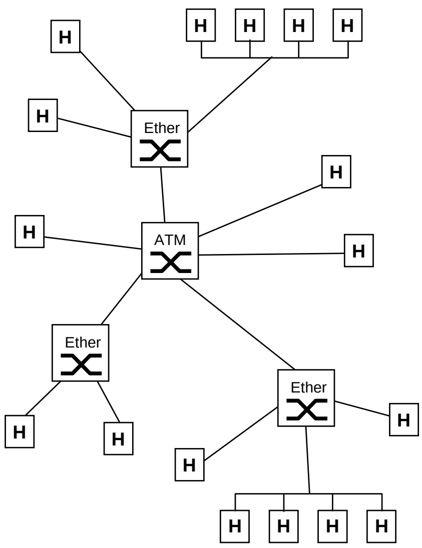
    - pad so trailer always falls at end of ATM cell
    - length: size of PDU (data only)
    - CRC-32 (detects missing or misordered cells)
  - Cell Format
    - end-of-PDU bit in Type field of ATM header

Outline

- General switching techniques
- Bridges and LAN switches
- ATM technology
- ATM in LAN

Historical perspective

- ATM also intended for LAN environment
  - switched technology offers better scalability
  - aimed to operate at link speeds of 155 Mbps and above (faster than 10BaseT)
  - idea: ATM replaces Ethernet and Token ring
  - …but 100BaseT Ethernet and Ethernet switches appeared (almost same speed as ATM)



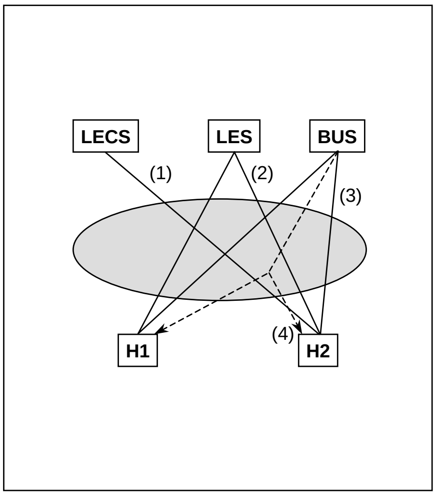
- ATM in the LAN backbone
  - does not have distance limitations
  - link speeds above 622 Mbps
  - hosts connected to Ethernet switches, ATM in backbone
  - … but now Gigabit Ethernet competes with ATM in backbone

Problems with ATM in LAN

- Broadcast (and multicast) crucial for LAN operation
  - easily implemented in shared media LANs (Ethernet)
  - but, ATM is connection oriented and switched in nature ...
  - broadcast important for ARP (Address Resolution Protocol)
- Two approaches
  - ATMARP: does not make assumptions on existence of broadcast
  - ATM LANE (LAN Emulation): makes ATM behave like shared media LAN
    - adds number of functional entities to network to make an illusion of shared medium, does not add functionality to ATM switches
    - LEC = LAN Emulation Client (host connected to ATM LAN)
    - Servers: LECS (LANE Configuration Server), LES (LANE Server), BUS (Broadcast and Unknown Server)
- Problem with addresses:
  - ATM uses NSAP or E.164 addresses, different than 48 bit MAC addresses
  - to make ATM work in LAN environment each ATM device needs a MAC address also

Overview of LANE (1)

- Joining the LAN and setting up broadcast
  - new client (H2) contacts LECS using “a well known“ VC (1)
  - H2 provides LECS with its address and receives configuration information about the LAN and learns ATM address of LES
  - H2 contacts LES and provides its ATM address and MAC address, H2 learns ATM address of BUS (2)
  - BUS maintains single point-to-multipoint connection connecting all registered clients
  - H2 signals to BUS (3)
  - H2 is added to the multipoint connection (4)



Overview of LANE (2)

- Delivery of unicast traffic to a particular MAC address
  - H2 wants to send to H1
  - H2 does not know ATM address of H1
  - first packets sent directly to BUS (1)
  - H2 sends request to LES, LES returns ATM address corresponding to the MAC address (2)
  - H2 signals a VC directly to H1 (3)
  - old unused VCs time out and are disconnected automatically