Managing the Bandwidth Explosion

Scott Bradner
Harvard University
sob@harvard.edu

September 23rd, 1997

Schedule

Gartner Group	08:00 - 10:00
Coffee	10:00 - 10:30
Scott Bradner	10:30 - 12:00
Lunch	12:00 - 13:00
Scott Bradner	13:00 - 14:30
Coffee	14:30 - 15:00
Scott Bradner	15:00 - 16:30
Closing	16:30

Syllabus

- introduction
- the networking future
- performance requirements in today's LANs
- QoS Implications
- flattening vs. routing
- network technology
- interconnect technology
- ATM vs. *
- summary
- Q & (hopefully) A

Introduction

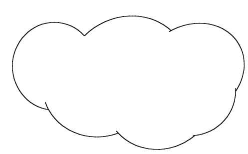
- network managers live in interesting times
- managing expectations
  - boss reads trade press
  - final solution of the week
    - passé a week later
- plan for the future
  - without knowing what demand will be?
  - who predicted the web?
- will corporate networks need to support IVD?
  - what does it mean if the answer is yes?

Near Future

- ever more traffic
- ever more critical to organization
- ever higher expectations
- ever more constrained budget

Further Future

- only a fool would put down money on a
  - 5 year contract for long distance telephone
  - 3 year contract for Internet service
  - N year contract for LAN hardware
- only broad strokes predictable
  - general data network
    - not separate private nets
  - open protocol
    - will be called IP
  - cycles of distributed vs. centralized control
  - no single ubiquitous LAN technology
  - no single ubiquitous WAN technology



IP

What are the Future Scale Problems?

- what problems need to be solved?
  - overall Internet scale
  - management complexity
  - Quality of Service
  - unrestricted usage growth
- what technology to try with?

Internet Scale

- "running out of IP addresses"
- yes - once, not now
  - NAT, CIDR & registry allocation procedures
  - was 2008 +-3, now don't see when
- IPng driver was IP address scarcity

Internet Scale - II

- "running out of routing table space"
- multiple connectivity means multiple paths
  - 10 or more per destination
- yes, once, not now
  - CIDR
  - memory cost

Internet Scale - III

- "running out of route processing time"
- yes, once and now
- changes in routing tables are constant
  - > 100/sec peaks
- recalculating forwarding table is expensive
- use route damping?

Management Complexity

- are you running the network?
  - or is it the other way around?
- guru factor in management products
- basic measure of network quality is user response time
  - many components, network and server
- how monitor user response time with NMS?

Unrestricted Usage Growth

- ISPs have only one "product"
  - best effort delivery
- if flat rate price
  - no reason to limit usage
- if variable-rate price
  - no way to get good feedback
- class-based pricing would help
  - pay extra for what you "need", best effort for rest

Performance Requirements of Today's Local Area Networks

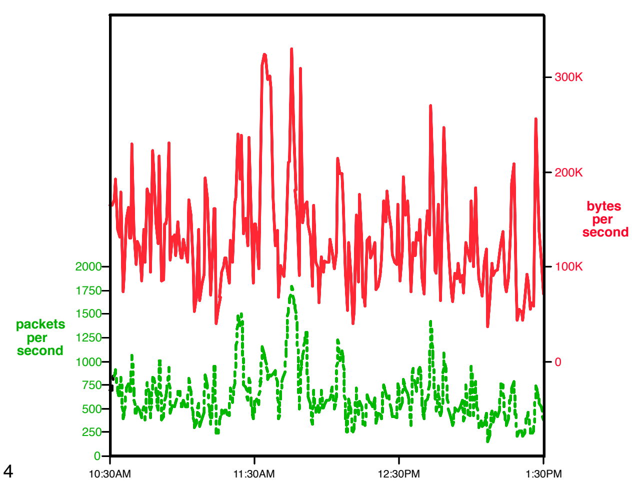
- hype vs. reality

  do you *need* the fastest LAN on the block?

  is this a case of bragging rights, or actual need
- note it's *system* response not link bandwidth that makes a difference to users
- does 30% utilization on a 10 Mb Ethernet mean 100 Mb is needed now?
- does "high" utilization mean high pps?

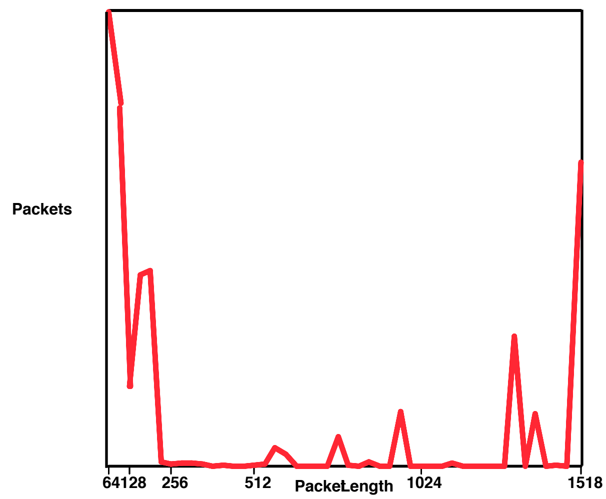
Life on a real world enterprise backbone

- load on a Harvard backbone net



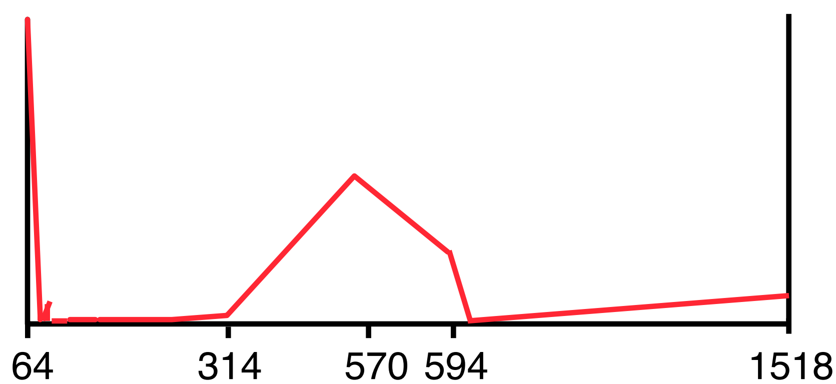
Life on a real world network

- packet size distribution on a Harvard subnet



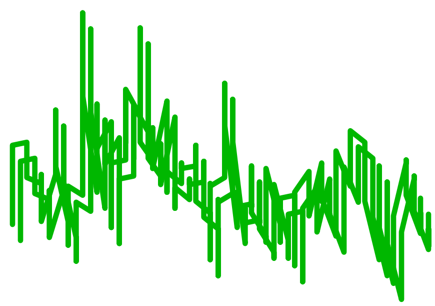
Life on the big world network

- packet size distribution on the Internet backbone



Data Traffic Patterns

- non-poisson distribution
  - thumper.bellcore.com/pub/dvw/sigcom.93
- tendency to synchronize peaks
- buffering can help or hurt
  - smooth out bursts
  - add latency



How Do You Tell?

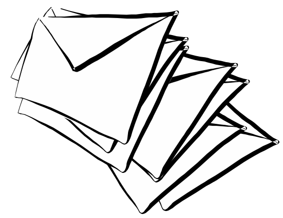
- how do you tell if network is overloaded?
  - user complaints per hour
  - monitoring loads (utilization)
    - links, servers
  - monitoring errors
    - dropped packets, collisions
  - note time of day/month factor can be important

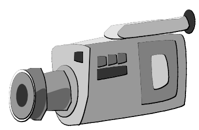
October 1997

S	M	Tu	W	Th	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Applications Driving Network Traffic Loads

- more power on desktop
- centralized servers
- corporate www services
- graphics and binaries in email
- someday?
  - voice and video





Traffic

- traffic load
  - WWW very heavy load
    - the web is the answer (what was your question?)
  - excess graphics
  - will be shipped with new (all) services
  - high performance apps
- traffic pattern
  - no logical pattern in accessing servers
  - picture from next door or 3,000 miles away

Dreams

- dream of IVD
  - integrated Voice Data and Video
  - glue headset to PC
- dream of video on demand
  - Terminator 34 at your command
  - more interesting than the web?
- dream of the Network Computer
  - terminals on steroids
  - download all apps

Quality Of Service (QoS)

- people say they "need" QoS!
  - but what is it?
  - where do you need it?
  - can you get QoS in datagram network?
- types of QoS approaches

What is QoS

- predictable system behavior
  - note not just network
- e.g. predictable latency, loss rate or bandwidth
- real-time interactive vs big batch processing
- SNA as ideal
  - much manual configuration
  - SLA for expected response time and procedures

Where do you need QoS?

- QoS needed where there are constraints
  - i.e fighting bandwidth explosion
  - no constraints = no QoS controls needed
  - if never have packets queued for output or if total queue time is OK then no need to control order of packet transmission
- LAN - easy to throw bandwidth at problem
- WAN - a real issue

QoS in a Datagram Network

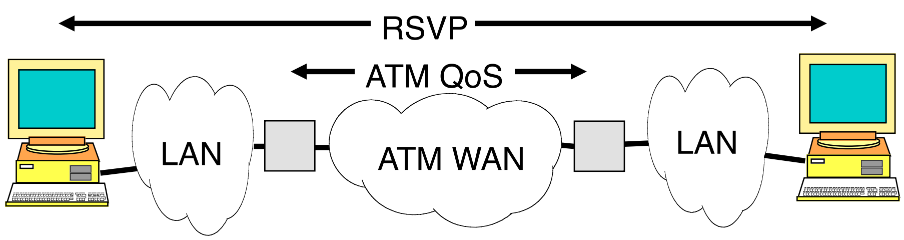
- very old argument, at least since 1964
- "need session-oriented communication to be able to control QoS"
- RSVP installs "soft state" to mimic sessions

Types of QoS Approaches

- probabilistic QoS
  - architecting current datagram networks for predicted load
- instance of application QoS
  - RSVP & ATM QoS
- non-flow-based QoS
  - future?

Mixed QoS

- since only sure end-to-end technology is IP
  (for now)
- use IP signaling (like RSVP) to control link-level QoS (like ATM) when present
- can map classes of traffic into separate VCs

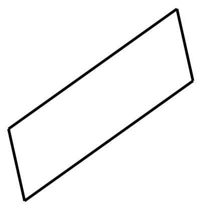


QoS Issues

- scaling issues
- authorization issues
- accounting issues
- advanced reservations *very* hard

Flattening vs. Routing

- switch vendors often advocate flattening networks
  - i.e., reduce subnets
- remember why subnet created before removing it
- product limitations or architectural purity?



Breaking up Broadcast Domains

- routers act as firewalls for broadcast packets
- some “intelligent” switches on the way
  - confine broadcasts to VLANs or learn ARPs
  - what about new protocols?
- some switches limit broadcasts
  - all broadcasts or from over active source
- subnets to remove broadcasts can reduce system latency by reducing LAN load

Manageability Implications

- Protocol dependent
  can control what is on LANs
- Network not seen as a single LAN
  local problems do not become global
- Filter at protocol level
  can do protocol and application access control

Manageability, contd.

- Isolates broadcast/multicast
  creates firewalls against broadcast traffic
- Isolates administrative zones
  can delegate control
- Same difficulty of management in normal and abnormal conditions

Network Technology

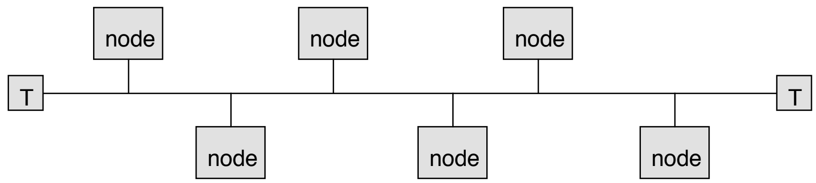
- Ethernet
- token ring
- FDDI
- frame relay
- SONET
- ATM
- xDSL

Etherneten

technology	bit time	slot time	max RT	coll dia	min packet	max packet
10 Mb half-duplex	100ns	51.2us	51.2us	4KM	64B	1518B
10 Mb full-duplex	100ns	51.2us	-	-	64B	1518B
100 Mb half-duplex	10ns	5.12us	5.12us	400M	64B	1518B
100 Mb full-duplex	10ns	5.12us	-	-	64B	1518B
1Gb half-duplex	1ns	4.09us	4.09us	300M	512B*	1518B
1Gb full-duplex	1ns	4.09us	-	-	64B	1518B

*smaller packets must be padded out to 512 bytes using carrier extension

Half-Duplex Ethernet & 802.3



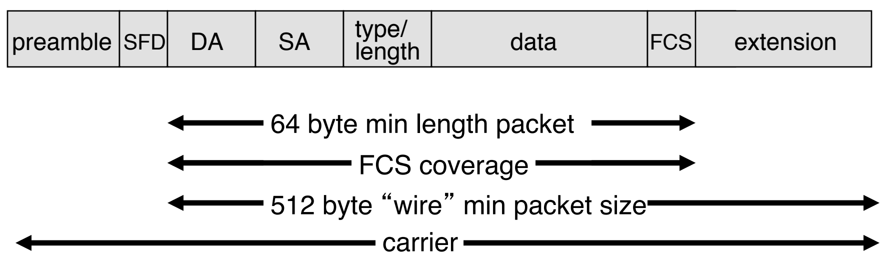
- Provides multi drop LAN service
  - Carrier Sense Multiple Access with Collision Detection (CSMA/CD) procedure
    - a/ wait for no carrier on cable
    - b/ start transmitting frame
    - c/ if collision detected:
      - send 32-bit jam signal to ensure bad FCS
      - wait for “random” interval (“longer” interval for each retry)
        - *Binary Exponental Backoff (BEB)*
      - go to "a" (limit on retries of 16)

Full-Duplex Ethernet & 802.3

- removes collision zone
- getting more common
- full bandwidth in each direction

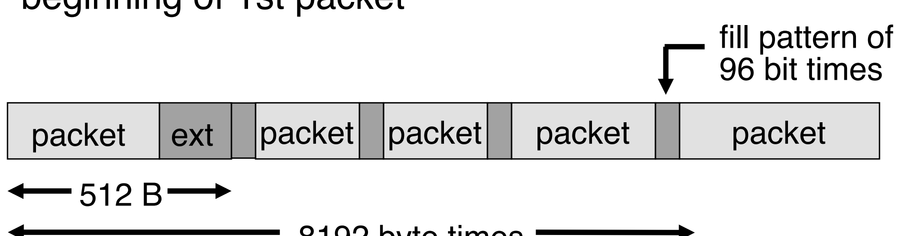
Gigabit Ethernet

- uses Fiber Channel interfaces
- 1Gb data rate
- includes carrier extension on half duplex
  - 512 B minimum packet duration on wire
    - to expand collision diameter
- 8B/10B encoding



Gigabit Ethernet, Burst Mode

- burst mode (half duplex only)
- to reduce effect of extension
- packets can be attached to end of extension or any packet longer than 512 bytes
- added packets must start within 8192 bytes of the beginning of 1st packet



802.3x Flow Control

- help small memory switches not drop packets
- designed to work on different 802.3 LANs
- link-based, not end-to-end
- does not “fix” long-term congestion

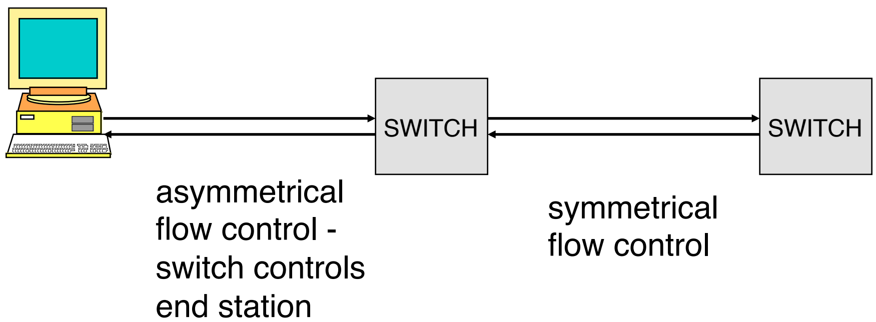
DA = 01-C2-80-00-00-01
SA
TYPE=8808
OPCODE=0001
PAUSE TIME
PAD

802.3x Flow Control, contd.

- send *PAUSE* frame to tell other end to hold off transmission
- Pause *PAUSE TIME* slot times
- new *PAUSE* frames override current
- note must send new *PAUSE* frames periodically to stop transmission
- negotiated at startup or manually configured
- can be symmetric or asymmetric

Flow Control Symmetry

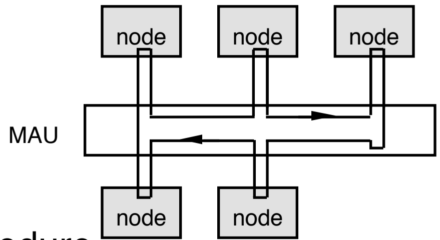
- should throttle packet source
  - normally an end station
  - do not want end station to throttle network
  - no “source” between switches



Ethernet Myths

- collisions are evil
  - normal part of csma/cd - > 20% an issue
- utilization should not exceed 40%
  - 60%-70% utilization common and fine
- Ethernet will collapse under high load
  - tests show that this is not true
- Ethernet is non-deterministic
  - simulation shows it to be quite deterministic
- no such questions in full duplex switched Ethernet

802.5 Token-Ring

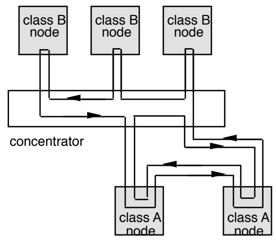


- Procedure
  - a/ wait for token on "input" link
  - b/ send a frame (only one per IBM, >1 per 802.5)
  - c/ wait for frame to return
  - d/ strip own frames
  - e/ transmit token when done
    - early token release option (default on IBM 16Mb)
      - send token after transmitting frames (don't wait for frame)

Token Ring

- very wide use in some areas
  - e.g. IBM mainframe shops
- 4 & 16 Mbps
- seen as “better” than shared Ethernet
  - less of an issue with switched Ethernet
- limited upgrade paths
  - can use FDDI for backbone
  - new effort for “fast token ring” starting

Fiber Distributed Data Interface (FDDI) (ANSI X3T9.5)

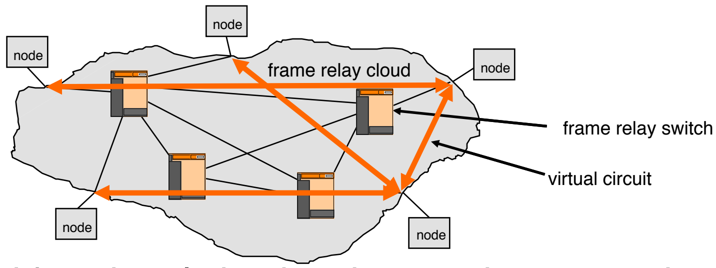


- Procedure
  - a/ wait for token on "input" link
  - b/ send one or more frames
  - c/ transmit token when done
  - d/ wait for frame to return - strip own frames
- Maximum ring size
  - 200 km

FDDI, contd.

- 100 Mbps
- supports long distances over private fiber
- transparent redundancy
- not seen as a “future” technology

Frame Relay



- Provides virtual circuit point to point connections
- no delivery guarantees
- Data Link Connection Identifier (DLCI)
  - identifies logical connection
  - hop-level significance - not end-to-end
  - 10 bits - ~1K simultaneous sessions per link
- 56Kb to T3 (OC3 & OC12 in future?)

Frame Relay - Flow Control

- includes information about network congestion
  - Forward Explicit Congestion Notification (FECN)
    - congestion seen in path from source to destination
  - Backward Explicit Congestion Notification (BECN)
    - congestion seen in path from destination to source
  - information for higher-level protocol's flow control
- note - generally not sent to end node
- simple priority indication
  - Discard Eligibility Indicator (DE)
    - discard this frame if congestion found

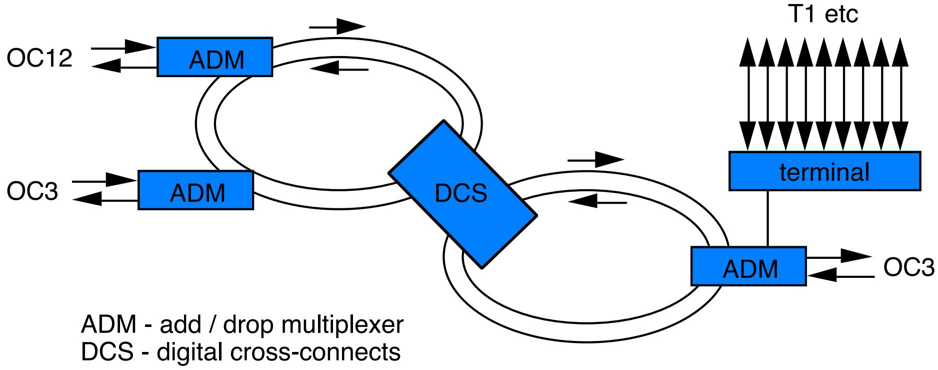
Frame Relay - CIR

- traffic shaping parameters
  - T_c - time slot
  - B_c - # bytes that are committed to be delivered
  - B_e - excess bytes that will be accepted but DE bit will be set
    - can be discarded if network is congested
- Committed Information Rate (CIR)
  - $CIR = B_c / T_c$
- can be less than tail circuit speed
  - sometimes 0

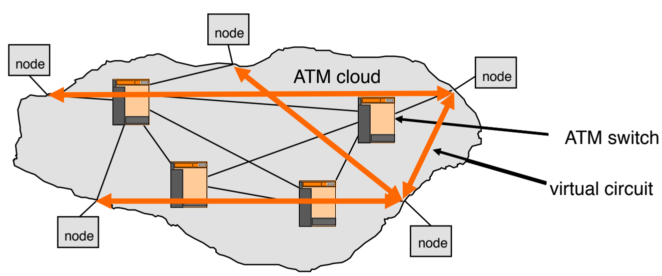
Synchronous Optical Network (SONET)

- time-division multiplexing of data streams
- ANSI & ITU-T standard
- central clock source
- base data rate
  - US/Japan - STS-1 51.48 Mbps
  - Europe - STS-3 - 155.52 Mbps
- can use dual ring architecture for redundancy
  - like FDDI
- uses *add/drop multiplexer* (ADM)
  - drop one incoming stream and replace with another
- mixed availability

SONET Network

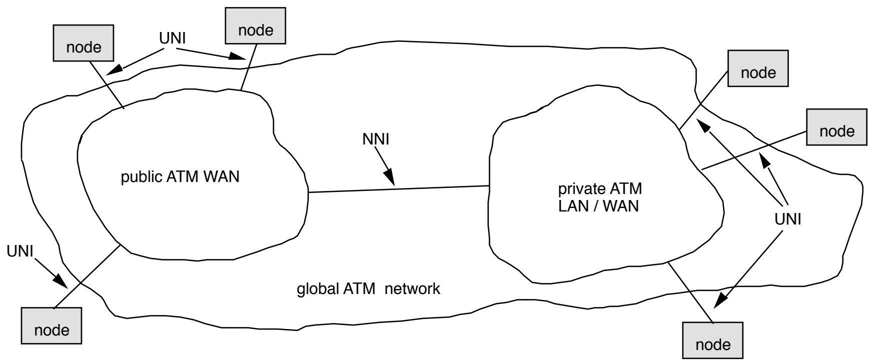


Asynchronous Transfer Mode (ATM)



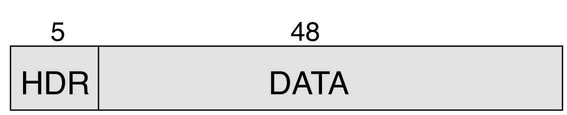
- Provides virtual point to point connections through network mesh
- Originally for core of telephone system
- can have per connection QoS support
- hop-local virtual circuit ID - ~64K per link

ATM - Conceptional Network



ATM - Connections

- cell-based, i.e. no SRC & dest address in cell
- Virtual Path Identifier (VPI) and Virtual Channel Identifier (VCI) used to tag cells for routing
- connection-oriented
- path must be set up before data can flow
- path set up in switches
- path released when done
- quality of service (QoS) requested at set up time



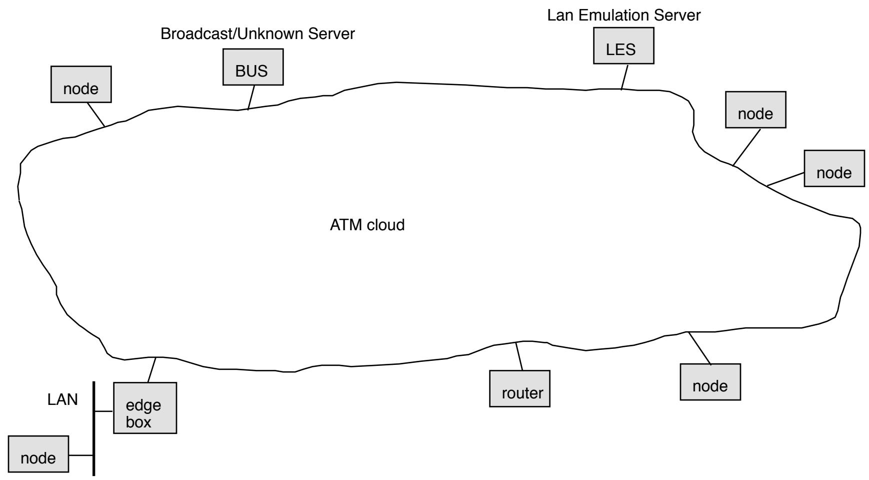
ATM - Service Types

- ATM service types
  - Constant Bit Rate (CBR) service
    - fixed data rate service, e.g. uncompressed voice or video
  - Variable Bit Rate (VBR) service
    - variable data rate service with a maximum permitted rate
    - e.g. compressed voice or video
    - RT-VBR (real-time) & NRT-VPR (non-real-time)
      - amount of delivery time variation allowed
  - Available Bit Rate (ABR) service
    - lower priority service, e.g. file transfer, e-mail
    - source required to control usage rate
  - Unspecified Bit Rate (UBR) service
    - what's left over, cells discarded on overload

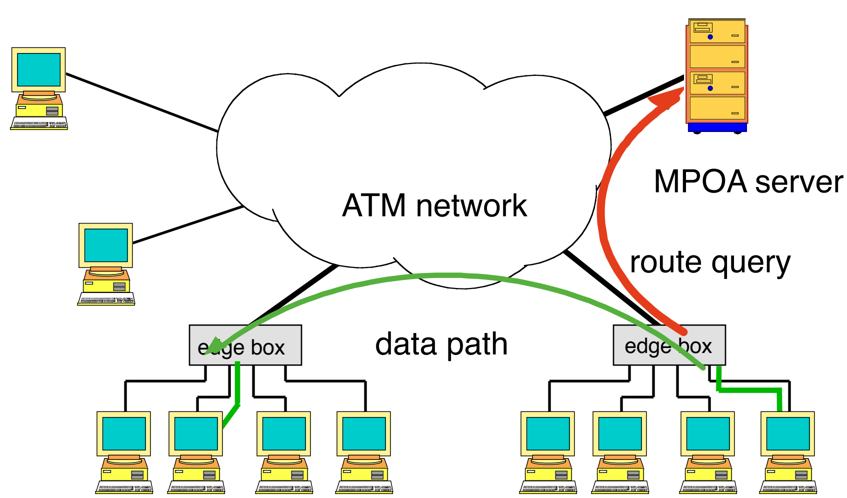
ATM Virtual LANs

- pretends to be separate LANs
  - individual nodes configured in switch to be on LAN
  - “logical subnets” connected with routers
  - e.g.,
    - Classical IP and ARP over ATM - RFC 1577
    - LANE - ATM Forum LAN emulation
    - MPOA - ATM Forum Multi-Protocol Over ATM
- ATM is not a broadcast media
  - need to support broadcast functions like ARP
  - uses servers to emulate broadcast

ATM LAN Emulation



MPOA, Model



xDSL

- DSL = Digital Subscriber Line
- digital signaling over twisted pair “phone” lines
- requires sophisticated electronics on each of lines
  and expensive
- limited distances
- many types

Interconnect Technology

- hub
- switch
- router
- level 3 switch

Hub

- share bandwidth between attached nodes
- Ethernet
  - single collision domain
  - packets go everywhere
  - single broadcast domain
- token ring & FDDI
  - extend ring to additional nodes

Switch

- dedicated bandwidth to each node
- traffic only goes where it needs to
  - security advantage
  - performance advantage
- single broadcast domain
- Ethernet, token ring, FDDI, ATM

Router

- separates LANs
- multiple broadcast domains
- ideal LAN size a discussion topic
  - can vary based on protocols in use
    - e.g. Novell can generate a lot of broadcast traffic

Level 3 Switch

- two types
  - just a router
  - does something special with 1st packet in flow
    - may not match flow characteristics of Internet well
- much marketing hype
- "tag switching" & Ipsilon
  - move routing function to edge of network

ATM vs. *

- ATM was once seen as the future network
- picture far more complicated now
- “lost” desktop - Ethernet 1000 to 1 (at least)
- ATM is generally under attack

ATM under attack

- by competing technology
- by need
- by zealots
- by complexity
- by standards process
- by ATM

Competing Technology

- switched Ethernet
- 100 Mb Ethernet
- Gb Ethernet
- RSVP

Need for ATM

- QoS is a major driver
- is it needed?
- LAN vs. WAN

Zealots

"and there is just no ATM there!
we are talking about _real_ 155MB, no fake"

Complexity

- much session-dependent state in the net
- routing
- QoS scheduling (time-share again?)
- in comparison to alternatives

Standards Process

- rather big
- rather political
- rather commercial
- rather confusing
- slow?

ATM

- the ATM dream
- was it the original mission?
- could it have been real?

ATM's future

- one of the gang

Summary

- bandwidth requirements are expanding
  - on LAN, campus backbone, and WAN
- IP is present & future layer 3 technology
- routers are your friends (and even L3 switches)
  - level 2 switches have their places
- QoS needed where there are constraints
  - most WAN links will have constraints
  - IP QoS is future direction from end system point of view

Summary, contd.

- current most popular LAN technology
  - Ethernet (10 -> 1000 Mbps)
- campus backbone harder
  - Ethernet, FDDI, ATM
- WAN is a tariff and service availability issue
  - leased lines, frame relay, ATM, SONET
- ATM evaluation as much political as technical

Q & A