Computer Networks Design and Control – Class intro and review

Computer Networks Design and Control

Class intro and review

Andrea Bianco
Telecommunication Networks Group
firstname.lastname@polito.it
http://www.telematica.polito.it/

1

Class goals

- Describe, mainly in a qualitative way, techniques and algorithms
  - to design networks, with the aim of offering quality of service to users
  - to ease network management and control
- Focus on
  - Algorithms for QoS
  - Protocol Standardization
    - Internet
    - Ethernet
    - Broadband forum and 5G networks (including Frame-relay and ATM)
  - Management and control
    - SNMP, MIB
    - Software Defined Networking
    - NFV (Network Function Virtualization), Network slicing, Edge computing

3

Course syllabus

- Technology: review
  - SDH, WDM, Frame-Relay, ATM, Ethernet, Internet
- Introduction to quality of service and traffic characterization
- Quality of service standardization efforts
  - Frame Relay and ATM
  - Internet
    - Intserv
    - Diffserv
  - Ethernet
  - Wireless networks *("5G and next-generation mobile networks")*

4

Course syllabus

- Algorithms
  - Policing / shaping
  - CAC: Connection Admission Control
  - Scheduling and buffer management
  - Congestion control
  - Network protection and restoration
- SNMP and network management
- Capacity planning
- Network measurements
- SDN: Software Defined Networking
- CDNs, Data center, data distribution and application taxonomy

5

Other info*

- Class web site (to be updated)
  - https://www.telematica.polito.it/course/computer-network-design-and-control/
  - Linked from the Politecnico portal https://didattica.polito.it/
- Use of
  - Mail for announcements (delayed or cancelled or added lessons,…)
  - Messages on the portale della didattica
- Teaching material
  - On Dropbox
  - Pay attention in class and take notes!
- Oral examination
  - Contact the teacher via e-mail (andrea.bianco@polito.it) to fix the examination date.
    - Provide tentative dates, a phone number and a skype id
- * In slide title means important slide
  - Fundamental to understand
  - Not the only ones that are fundamental

6

Review and basic concepts

- Topologies
- Channel sharing: Multiplexing and multiple access
- Node sharing: Switching techniques
- Protocol and layering
  - Physical layer: SDH and WDM
- Error/loss management
- Old technologies
  - ISDN and Frame Relay
  - B-ISDN (ATM)
- Ethernet
- Internet (TCP/IP, UDP)
- "Low" layers (Frame relay , ATM and Ethernet, "high" layers in Internet

7

Networks

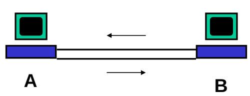
- Focus on modern telecommunication and computer networks
  - Telephone (for comparison) and Internet
- One possible network definition (service oriented view)
  - Infrastructure that provides services to applications:
    - Web, VoIP, email, games, e-commerce, social nets,
    - phone calls, fax, …
  - Internet
    - provides programming interface to apps program to "connect" to Internet
    - provides service options

Networks

- Another network definition:
  - A set of **nodes** and **channels** that offer a connection among two or more points to make telecommunication possible among **users** (i.e. move **infos (data/flows)** among nodes)
  - Represented as a topology (graph)
    - Different levels of detail
  - Key issue in networks is **resource sharing**
- Node is the point where
  - Multiplexing /multiple access (link sharing) and switching (node sharing) occurs

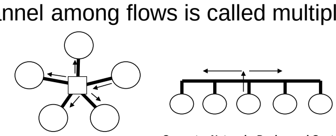
Type of channels

- Point-to-point channel
- Only two nodes connected to the channel
- The channel is used by both nodes (often in the same fashion)
- Sharing the channel among flows is called multiplexing



Type of channels

- Broadcast channel
  - Many Tx and many Rx
    - Sometimes one (many) Tx and many (one) Rx
- Single communication channel shared by all nodes
  - This room!
- The information sent by one node is received by all other nodes (with some delay)
- Sharing the channel among flows is called multiple access



Channel properties

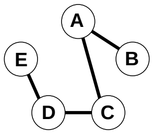
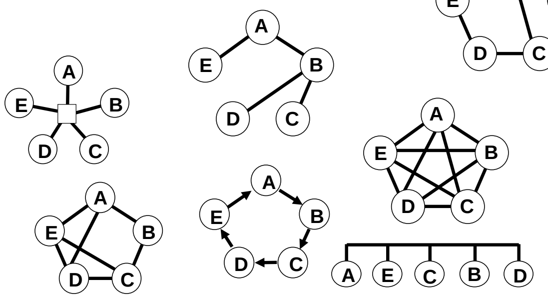
- Many quality indices
  - Attenuation, robusteness to mechanical stress, ease of installation, robusteness to interference, cost, etc,
- Mainly interested in
  - bit rate [bit/s]
    - Also named bandwidth, throughput, with slightly different meaning
  - delay [s]
    - propagation delay
    - depends on the channel length
  - Bandwidth x delay [bit]
    - channel ''size''
    - how much we can ''push'' on the channel

Topologies in TLC networks

- The network topology is defined by the relative position of nodes and channels
- A network topology is a graph G=(V,A)
  - V = set of vertices (represented as circles - nodes)
  - A = set of edges (represented as segments - channels)
- Edges may be:
  - direct (directed segments (arrow) – unidirectional channels)
  - undirect (non directed segments – bidirectional channels)
- Abstraction of real networks
  - Several levels of abstraction are possible

Many topologies

- Full mesh, mesh, tree, bus, star, …

14

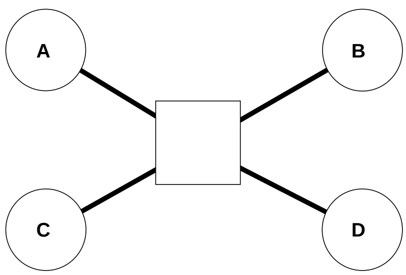
Physical and logical topology *

- It is important to distinguish between the physical and the logical topology
  - Logical topology: logical interconnection among nodes via logical channels
  - Physical topology: takes into account transmission media constraints

15

Physical topology

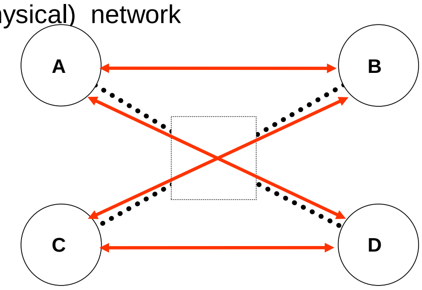
- E.g., satellite network



16

Logical topology

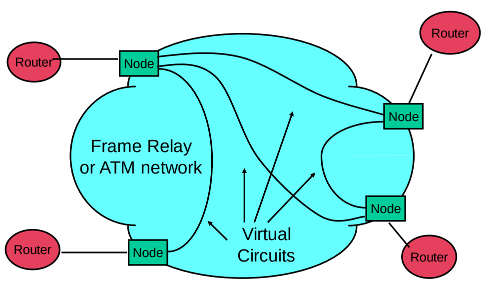
- Private (logical) network over satellite (physical) network



17

Logical topology

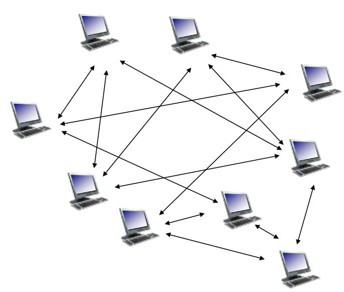
- Router interconnection for an ISP



18

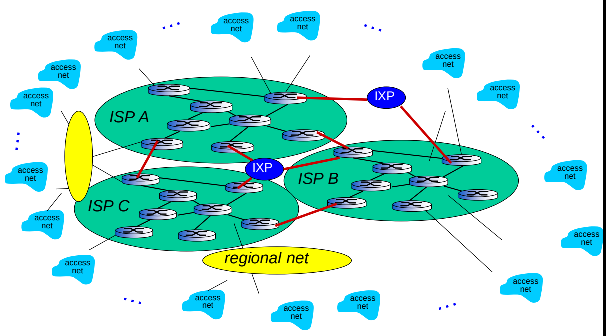
Logical topology

- Overlay among peers in a P2P network



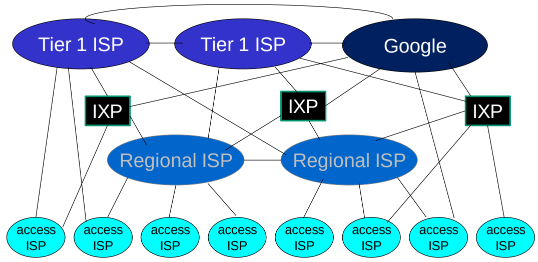
19

Logical topology



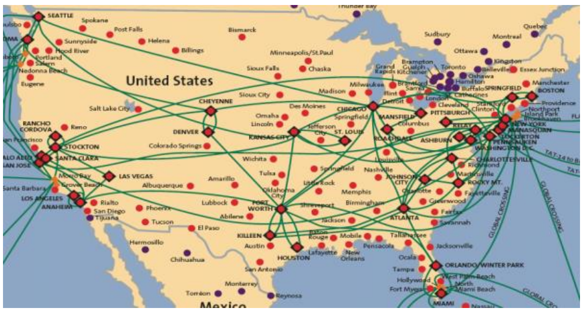
20

Logical topology



21

Tier-1 ISP: e.g., Sprint



22

Topologies and performance *

- The amount of traffic that can be succesfully transferred (throughput) in a network is
  - for a given available capacity
  - inversely proportional to the average distance among node pairs
  - weighted by the amount of traffic exchanged between the two nodes
- For uniform traffic and regular topologies the average distance on the topology determines the throughput

23

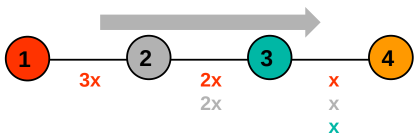
Topologies and performance

- Exercise
  - Comparison among topologies, with the same number of nodes (4) and (almost) the same number of channels
- Uniform traffic
  - Every node pair exchange x bit/s. Total generated traffic is 12x.
- Every unidirectional channel has capacity B bit/s.
- Compute: average distance, network capacity (maximum throughput), maximum channel load,maximum node load

24

Topologies and performance

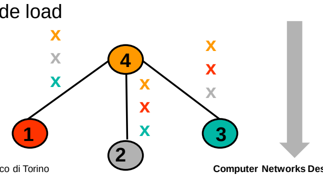
- Capacity: 3x2B=6B
- Average distance:
  - (2(1+2+3) + 2(1+1+2))/12 = 20/12 = 1.66
- Consider only traffic from left to right
  - right to left similar due to simmetry
  - maximum channel load is 4x. Thus, x <= B/4
- Node 3 (or 2) must handle 7B/4 of traffic unit
- Uniform traffic, non regular topology, unbalanced channel load, unbalanced node load



25

Topologies and performance

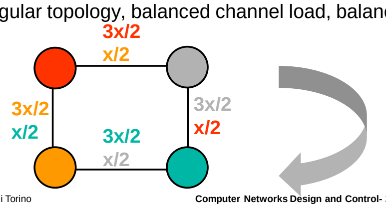
- Capacity: 3x2B=6B
- Average distance: 1.5
- Considering the traffic from node 4 downstream
  - Maximum load on (all) channel is 3x. Thus x <=B/3
  - The same holds for the other, upstream, direction
- Node 4 must handle 3B of traffic unit
- Uniform traffic, non regular topology, balanced channel load, unbalanced node load



26

Topologies and performance

- Capacity: 4x2B=8B
- Average distance: 1.33
- For clockwise traffic the maximum channel load is 2x. Thus x <= B/2.
  - The same holds for counter clock wise traffic
- Each node must handle 2B unit of traffic
- Uniform traffic, regular topology, balanced channel load, balanced node load



27

Channel sharing
Multiplexing and multiple access

28

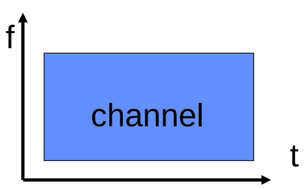
Sharing channel resources *

- Sharing of channel resources among data flows comes in two different flavours
  - Multiplexing
    - All flows access the channel from a single point
    - Single transmitter scenario
    - Centralized problem
    - A radio access from an antenna (base station in a cellular network, access point in a WI-FI network, satellite transmission), an output link in a switch or a router
  - Multiple-access
    - Flows access the channel from different access points
    - Many transmitters are active
    - Distributed problems
    - Local area networks (if not switched), mobile phones in a cellulare network, PC accessing via a Wi-FI hot-spot

29

Channel sharing techniques

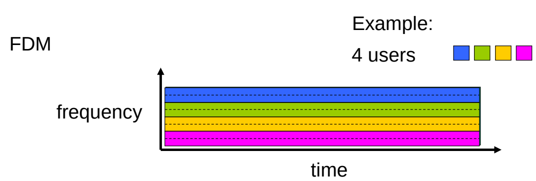
- Channel is a set of frequencies available for tx/rx (to a target distance)
- More frequencies imply higher bit rates
- Division techniques
  - Frequency (FDM - FDMA)
  - Time (TDM - TDMA)
  - Code (CDM - CDMA)
  - Space



30

Frequency division (FDM – FDMA)

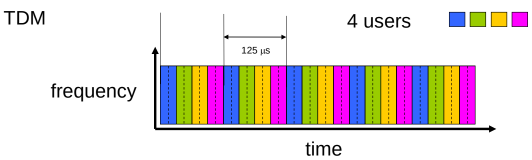
- Each flow is transmitted using different frequency bands
- Overlapping in time
- Need for band guard



31

Time division (TDM-TDMA)

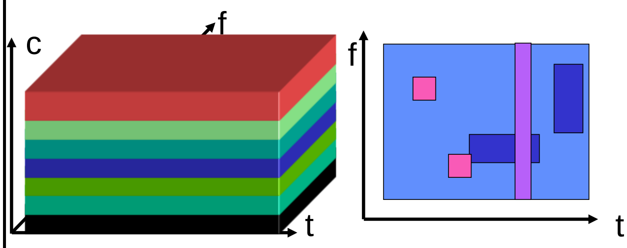
- Each flow exploits different time intervals (slots)
- Overlapping in frequency
- Need to define frame over which slot allocations are repeated
  - 125μs frame duration



32

Code division (CDM – CDMA)

- Each flow exploits a different code (waveform with higher frequency than the bit tx rate)



33

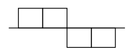
Code division (CDM-CDMA)

- Flow separation through different codes
  - Neither time nor frequency
  - Need for orthogonal codes
  - Codes assigned to tx (need to know at the rx)
- Transmission and reception imply multiplication of information bit and the given code
  - Equivalent to a scalar product among vectors

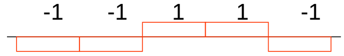
34

Code division

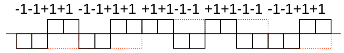
- Example



  - Code word used by user i: +1 +1 -1 -1
  - Coded sequence = information bit x code word
  - Information bit: -1 -1 1 1 -1



  - Coded sequence: -1-1+1+1 -1-1+1+1 +1+1-1-1 +1+1-1-1 -1-1+1+1



35

Code multiplexing

- Example
  - Code word for user 1: +1 +1 -1 -1
  - Code word for user 2: +1 +1 +1 +1
  - Code word for user 3: +1 -1 +1 -1
  - Code word for user 4: +1 -1 -1 +1
- Hp
  - User 1, 2 and 3 are active
  - We are interested in receiving transmissions of user 1
- Over the channel, transmitted signals sum up (need to equalize power)
  - Transmissions of 1+2+3: +3 +1 +1 -1
    - Everything we get
  - Transmissions of 2+3: +2 0 +2 0
    - What we are not interested in

36

Code multiplexing

- Reception = correlation with code words
- Reception of user 1 = scalar product of the received sequence with the code word +1 +1 -1 -1
- Transmissions of 1+2+3: +3 +1 +1 -1
- Correlation with +1 +1 - 1 -1 = 4
- Transmissions of 2+3: +2 0 +2 0
- Correlation with +1 +1 -1 -1 = 0
  - No contribution form the other users!

37

Space multiplexing

- Networks exploit also space multiplexing
- First idea is to use multiple parallel wires
- Routing techniques may also try to exploit space multiplexing to increase network capacity
- Cell in wireless access are another example of space multiplexing (reuse)

38

Multiplexing or multiple access *

- Time, frequency, code and space (multiple wires) are all equivalent alternatives
  - Given a channel capacity we can choose one among the above techniques depending on technological constraints
- Code division permits to “increase” channel capacity (by allowing more users) if using pseudo-orthogonal codes but degrading the signal to noise ratio at the receiver (increase the bit error rate)

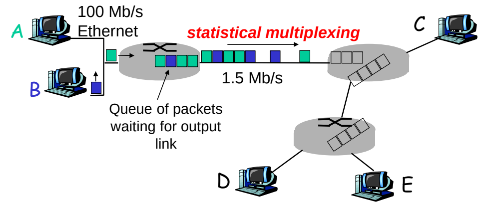
39

Statistical multiplexing *

- Multiplexing can be
  - deterministic, fixed in time, on the basis of requirements determined at connection setup
  - statistical, variable in time, to adapt to instantaneous traffic requirements
  - easy to obtain (see next slide)
- Statistical multiple access more tricky

40

Statistical multiplexing

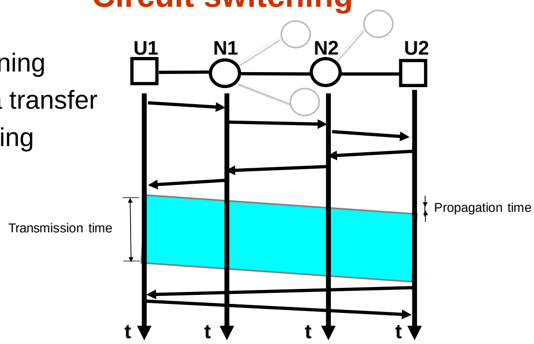


- Sequence of A & B packets does not have fixed pattern
- Dynamic TDM scheme
- Bandwidth shared on demand

41

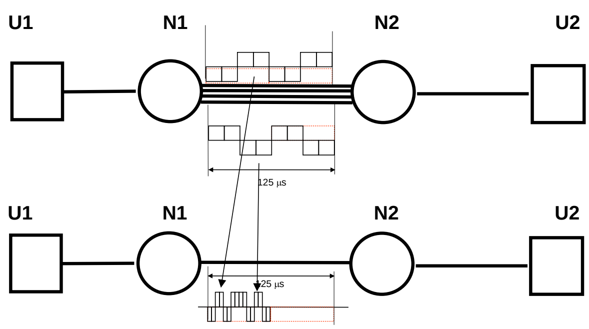
Circuit switching

- Opening
- Data transfer
- Closing



42

Space vs time switching

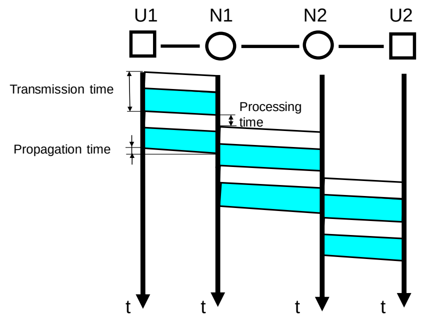


43

Switching techniques *

- **Circuit switching**
  - Resources allocated uniquely to a circuit
    - Physical channel, time-slot in TDM frame
  - Connection oriented
    - Need to open (and close) the circuit prior (after) data transmission
    - Store state information on each circuit (stateful approach)
  - Address (unique for each user in the network) used only when opening the circuit, not carried in data
  - Data unit identified by position
  - Routing (choice of the best route) performed only when opening the circuit
    - Done through routing table lookup
  - Data forwarding
    - Through forwarding table look-up (one entry for each active circuit)
    - Static (always the same scheduling, unless circuits are closed or opened)
    - **With no delay!**

Packet switching



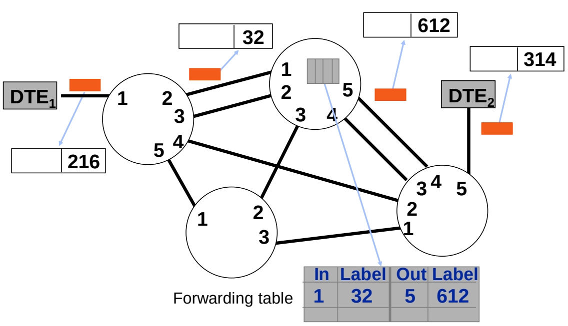
Switching techniques *

- **Packet switching, with datagram service**
  - Shared resources
    - Ideally the full network is available to a single user
    - Resources are shared with all other users
  - Connectionless
    - Free to send data when available, no need to check network or user availability
    - Stateless approach
  - Each packet must carry the destination (and source) address
  - Data unit identified through source and destination addresses (unique for each pair of users in the network)
  - Routing and forwarding performed independently over each packet
    - Through routing table look-up
    - **Buffering and delays**

Switching techniques *

- **Packet switching, with virtual circuit service**
  - Shared resources
    - Resources are shared with all virtual circuits sharing the same link
  - Connection oriented
    - Need to open (and close) the virtual circuit prior (after) data transmission
      - Permanent virtual circuits available
    - Store state information on each virtual circuit (stateful approach)
  - Address (unique for each user in the network) used only when opening the virtual circuit, not carried in data
  - Data unit identified through a label (unique for each existing virtual circuit on each link in the network)
    - Label is unique on each link, but has a local scope, i.e. the value assumed is different on each link for simplicity
  - Routing (choice of the best route) performed only when opening the virtual circuit
    - Done through routing table lookup
  - Data forwarding
    - Through forwarding table look-up (one entry for each active virtual circuit)
    - Re-labelling needed

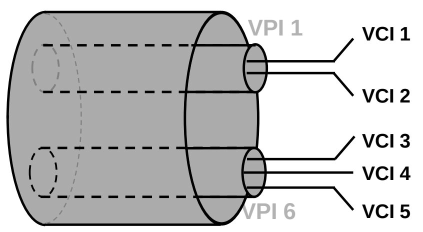
Data transfer over virtual circuits



Grouping virtual circuits

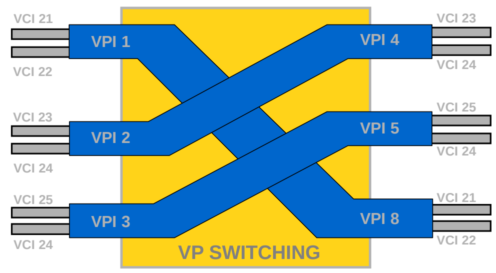
- A virtual circuit is logically identified by a label
- Label = often a pair of identifiers (VCI-VPI in ATM, LCN-LGN in ISDN)
- The grouping allows flow aggregation
  - Eases network management
  - Increases scalability
- Possible use
  - LAN inter-connection to crete a VPN (Virtual Private Network)
  - Multimedia flows (video, audio, data)

Virtual circuits and paths (ATM)



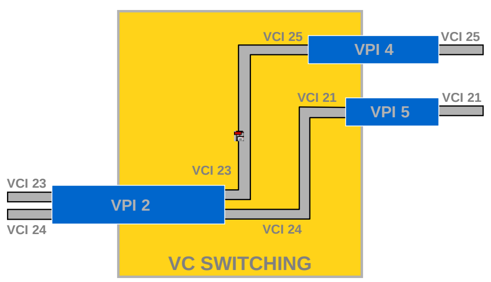
50

ATM: VP switching



51

ATM: VC switching



52

Virtual circuits

- Switched virtual circuit (SVC)
  - Established on-demand, through signaling, in real-time
  - Three phases
    - Virtual circuit opening
    - Data transfer
    - Virtual circuit closing
  - Users (and network) exchange signaling packets (over dedicated VCI/VPI) to establish a virtual circuit; then, data transfer can occur
- Permanent virtual circuit (PVC)
  - Established through agreement among user and network provider
    - Off-line, through management procedures
  - Define a semi-static network
    - Logical topology
  - Users can immediately exchange data, with no delay

53

Fundamentals of packet switching *

- Data sent as packets
- Nodes operate in store&forward (almost always)
  - Buffers
  - Delay
- Many operations on data in the network (not in circuit switching)
  - Error detection, error recovery, flow control, routing, forwarding, congestion control, packet inspection ……..
  - Need to define a network architecture to organize functionalities

54

Packet size *

- Packet size P
  - Measured in bit
- Packet size in time T_{TX}
  - Transmission time measured in s
  - Different on every link
  - $T_{TX} = P/V_{TX}$ where V_{TX} is the link bit rate
- Packet size in meter M on a given link
  - M = Speed of light x T_{TX}

55

Layering

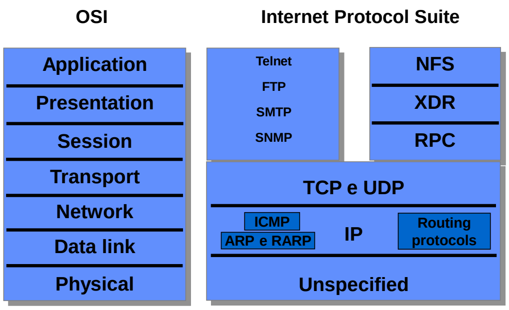
56

Protocol architectures

- In networks, many functions beside data transmission are needed (error detection, loss and error recovery, addressing, routing, …)
- Need to organize these functions
- Need to have formal and standardized rules
- Layered architectures
  - separation among functions
  - simple design
  - simple management
  - simple standardization
- Definition of protocols!

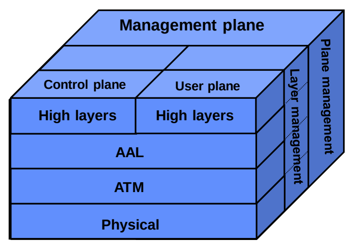
57

OSI and Internet



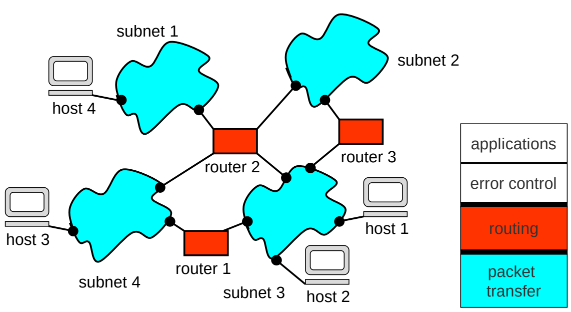
58

Several plane interact



59

Not all devices run all protocols *



60

Protocols

- Definition
  - formal definition of the procedures adopted to guarantee the communication between two or more objects on the same hierarchical level
- Protocols permit to exchange messages
  - composed of bits
  - organized in fields
- Protocols are set of
  - semantic rules
    - field meaning (commands, answers,…)
  - syntactic rules
    - message formats
  - timing

61

SDH and WDM
A look at the physical layer

Physical layer (not only in telephone networks) *

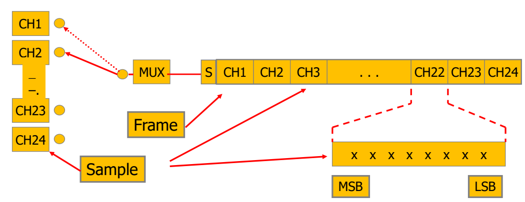
- TDM based scheme
- No store and forward in nodes
- Two technologies
  - Plesiouchronous Digital Hierarchy (PDH)
    - No global network synchronization
    - T and E hierarchies
  - Synchronous Digital Hierarchy (SDH)
    - Global network synchronization
    - Fiber based (optical) transmission

PDH: T and E hierarchies

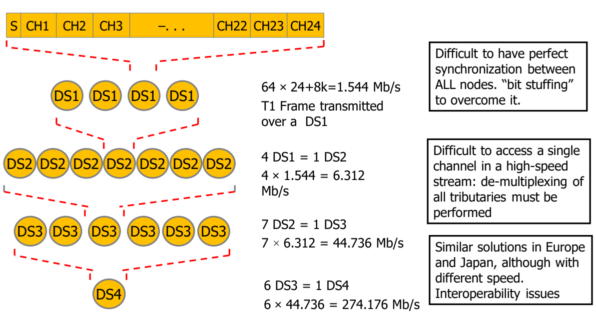
Level	USA (T-)	Europe (E-)	Japan
0	0.064 Mb/s	0.064 Mb/s	0.064 Mb/s
1	1.544 Mb/s	2.048 Mb/s	1.544 Mb/s
2	6.312 Mb/s	8.488 Mb/s	6.312 Mb/s
3	44.736 Mb/s	34.368 Mb/s	32.064 Mb/s
4	274.176 Mb/s	139.264 Mb/s	97.928 Mb/s

USA: T-1 carrier system

- Frame duration 125µs
- One sample (8bit) per channel every 125µs
- 24 TDM PCM channels
- Transmission rate (24*8+1)*8000=1.544Mb/s
- 1 signalling channel (1 bit per frame)
- No support for management functions



T- and DS- hierarchy



PDH: synchronization

- Each device has its own clock (no network wide global synchronization)
- Local clocks would lead to synchronization errors
- To solve it
  - TXs faster than receivers
  - Short buffer to store in transit bits
  - Insert bits through bit stuffing at the end of a frames
- Stuffed bits must be signalled to the other end to permit bit removal

PDH limitations

- Lack of efficiency
  - Difficult to extract slower tributaries from high speed aggregates
- Lack of flexibility
  - No monitoring standard
  - No management standard
- No common physical standard
  - Every manufacturer goes alone
  - No NNI standard

68

SONET/SDH

- Exploits network-wide synchronizaton
- Hierarchies
  - SONET - Synchronous Optical NETwork
    - Used in USA
    - Based on fiber transmission
    - Base signal at 51.84 Mbit/s
  - SDH - Synchronous Digital Hierarchy
    - International and European
    - Base signal at 155.52 Mbit/s
  - STS - Synchronous Transport Signal
    - Labels to identify electrical signals
- Introduction of management, signalling, protection

69

SONET/SDH Goals

- Main goals of SONET/SDH:
  - Fault tolerance as required by telecom providers
    - 99.999%, or five nines availability
  - Interoperability among different manufacturers
  - Flexibility of upper layer formats to adapt to different source (not only voice)
  - Complex monitoring capabilities of performance and traffic
    - 50 ms of recovery time after failure thanks to double ring architecture

70

SONET/SDH speed

OC level	STS level	SDH level	Mbit /s
OC-1	STS-1		51.84
OC-3	STS-3	STM-1	155.52
OC-12	STS-12	STM-4	622.08
OC-24	STS-24	STM-8	1244.16
OC-48	STS-48	STM-16	2488.32
OC-192	STS-192	STM-64	9953.28
OC-768	STS-768	STM-256	39813.12
OC-3072	STS-3072	STM-1024	159252.48

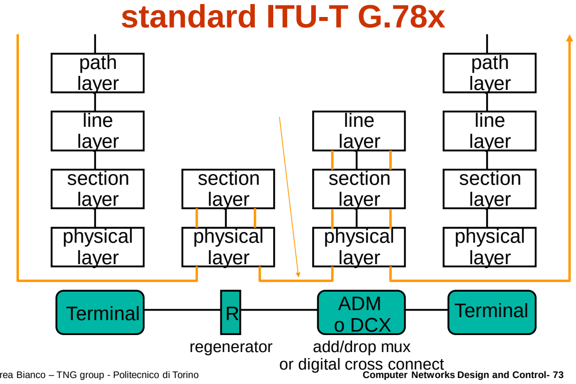
71

SONET/SDH framing

- Continuous bit transmission
- Complex TDM scheme
  - Designed to permit a very efficient VLSI implementation
- Each frame includes a PCI (Protocol Control Information) or overhead which includes
  - Synchronization infos
  - Voice channels for OAM services
  - Support for complex fault/error management procedures
- Layered architecture
  - Path, Line, Section (each including overhead infos)

72

Layering in SONET/SDH standard ITU-T G.78x

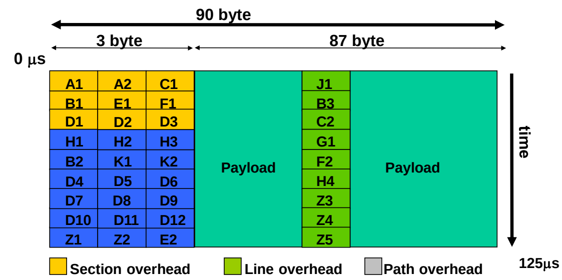


73

SONET/SDH reference model

- Path layer (similar to OSI layer 3)
  - Manages end-to-end connections
  - Monitoring and management of user connection
- Line Layer (similar to OSI layer 2)
  - Multiplexing of several path-layer connection among nodes
  - Protection and Fault Management
- Section Layer (similar to OSI layer 2-1)
  - Frame alignment
  - Define regenerator functions
- Photonic/Physical Layer (same as OSI layer 1)
  - Defines all the transmission requirements of signals.

SONET: Framing STS-1

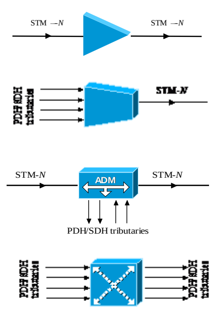


SONET/SDH: overhead

- Section
  - Framing
  - Parity check
  - Voice channel for operators
  - OAM and protection
- Line
  - Pointers to indentify payload
  - Parity check
  - Performance monitoring
  - Automatic protection switching(signalling to face fault management)
  - Support for restoration
  - Synchronization info distribution
  - Voice channel for operators
- Path
  - End to end monitoring

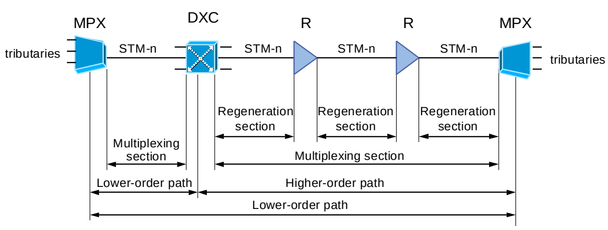
SONET/SDH equipment

- Regenerators
  - 3-R functions
  - SOH processing and regeneration
- Two multiplexer types
  - Terminal multiplexer (TM)
    - PDH and SDH tributaries
  - Add-drop multiplexer (ADM)
    - PDH and SDH tributaries
    - Used in ring topologies
    - Transit traffic forwarded transparently
- Digital cross-connect
  - PDH and SDH tributaries
  - Operate links with different speed



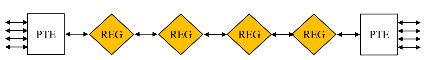
SONET/SDH transport scheme

- Transport of digital tributaries through SDH equipment



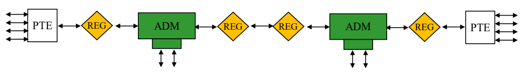
SONET/SDH Network Configurations

- Point-to-point topology
  - Simplest topology
  - The point-to-point start and end on a PTE (Path Terminating Equipment), which manages the mux/demux of tributaries
  - No routing, and no demux along the path
  - Regenerators may be used to cope with transmission problems



SONET/SDH Network Configurations

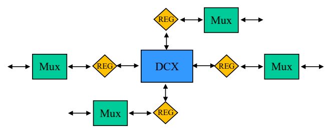
- Linear add-drop topology
  - Still a linear topology
  - ADM (and regen) along the line
  - ADM allow to add and drop tributaries along the path
  - ADM are designed to work in this kind of simple topologies, which often translates to rings
    - there is no need to mux and remux in transit tributaries



80

SONET/SDH Network Configurations

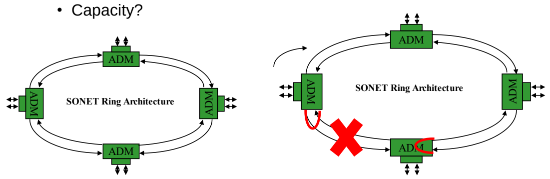
- Hub network setup
  - Typically on large aggregation point
  - Adopt Digital Cross connect (DCX) working at high rate
  - DCXs are much more complex that ADMs
    - they have to manage both single tributary and SONET stream
- Meshed topologies
  - Often seen as interconnected rings



81

SONET Network Configurations

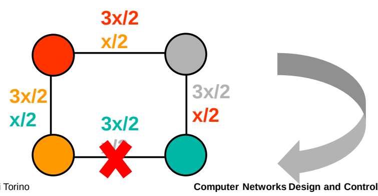
- SONET Rings
  - The most used topology. Can use two or four fibers and an ADM at each node
  - Bidirectional topology
  - Simple protection functions against single failure
    - Bidirectional ring becomes unidirectional ring
    - Capacity?



82

Performance after fault? *

- Capacity 8B
- Average distance: 1.33
- Maximum channel load: 2x.
  - Thus, x <= B/2.
- Each node must handle 2B unit of traffic
- Uniform traffic, regular topology, balanced channel load, balanced node load

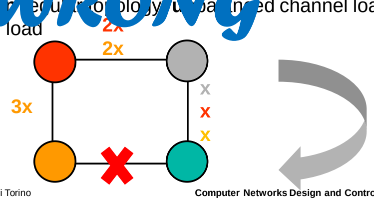


83

Performance after fault? *

- Capacity 6B (was 8B)
- Average distance: 20/12=1,6 (was 1,33)
- Maximum channel load: 4x (was 2x)
  - Thus, x <= B/4 (was x<=B/2)
- Node handle different load (was «each node must handle 2B unit of traffic»)
- Uniform traffic, **n**on regular topology, unbalanced channel load, **un**balanced node load

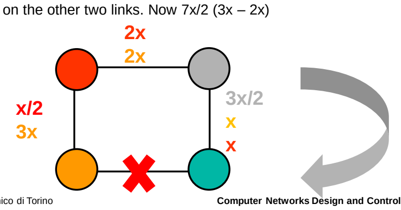
WRONG



84

Performance after fault? *

- Should consider the routing imposed by automatic reconfiguration
- Capacity: 6B (8B – 6B)
- Average distance (due to routing): 1.83 (1.6 – 1.3)
  - (2(1+2+3) + 2(1+1+**(2+4)/2**))/12
- Maximum channel load is 4x (2x – 4x)
  - Thus x <= B/2 (B/4 – B/2)
- Same botlleneck link
  - But, higher load on the other two links. Now 7x/2 (3x – 2x)



85

WDM: Optical transmission (and more?)

- Fibers can carry huge bandwidth
- Signals are generated in the electronic domain
  - Limited ability to exploit the optical bandwidth
  - Today 100Gbit/s transmission systems are the standard commercial for high speed transmission
  - 400Gbit/s are the next step
  - Still a huge bandwidth gap
- WDM (a FDM technique well suited to the optical domain)
  - Many wavelengths on a single fiber
  - Each wavelength transport an independent electronic signal
  - 128 × 2.5 Gbit/s or 32 × 10 Gbit/s

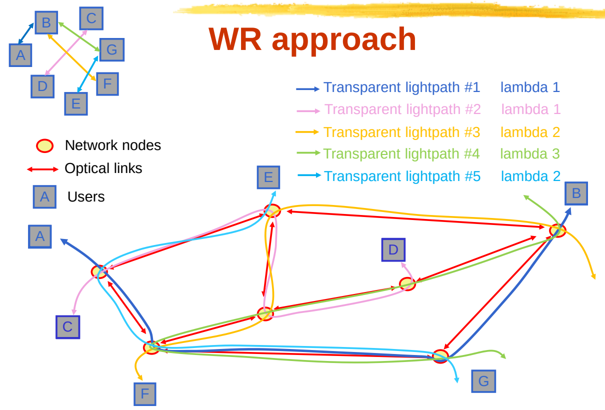
WDM for optical networks

- Which functions can be performed in optics?
  - Transmission
  - Switching/routing
    - Implies buffering?
  - Management
- First generation optical networks
  - Optics for transmission only
- Second generation optical networks
  - Perform also switching/routing in the optical domain
  - Wavelength routing approach

Wavelength Routing (WR) networks

- WDM is exploited to route and switch information in the optical domain using wavelengths
- Transparent or opaque optical circuits, called lightpaths, are used to connect network nodes.
- Lightpaths are optical circuits that may span one or more hops in the physical topology, and may cross switching elements in the optical layer
- Traffic carried by a lightpath may be
  - packet-based, e.g., IP datagrams
  - circuit oriented, e.g., telephone streams.
- The optical network is not aware of traffic formats

WR approach



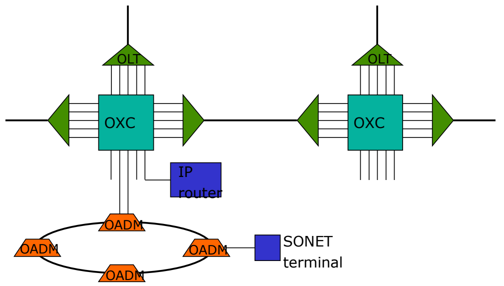
WR approach *

- Building a logical topology
  - using optical technology for logical links!
- Building a ligthpath similar to
  - Building a circuit in circuit switching
  - Building a virtual circuit in packet switching with virtual circuit service
  - Lambdas – labels – time/frequency/space slots
- BUT
  - lightpath transparently bypass nodes
  - no electronic processing required in nodes!

WR devices

- In a wavelength routed network
  - Optical Line Terminal (OLT): line termination, taking care of physical functions, signal regeneration, wavelength adaptation, amplification, traffic multiplexing/demultiplexing
  - Optical Add-Drop Multiplexer (OADM): it allows to add and drop traffic carried by one or more wavelengths in a (bidirectional) WDM link
  - Optical Cross Connect (OXC): switches incoming wavelengths to multiple outgoing fibers
- These devices are similar to the equivalent first-generation SONET/SDH devices

WR network architecture



- Class: "Open Optical Networks"

92

Layer 2 – Data link
Error management

93

Data link layer functions

- Frame delineation
  - Explicit delimiters (flag) before and after packet transmission
  - Lenght indicator
  - Fixed length
  - Silence between packets
- Multiplexing (of higher layer protocols)
- Addressing (local to the link)
- Error detection
- Window protocol
  - Flow control
  - Sequence and error control through retransmission
- Multiple access protocols for shared media

94

Techniques to protect against transmission errors

- FEC (Forward Error Correction)
  - Try to correct errors at the receiver
- ARQ (Automatic Retransmission reQuest)
  - Tries to detect errors at the receiver
  - Receiver ask transmitter to retransmit lost PDUs
  - Can be used aldo for **packet loss**
- Both need additional bits in the packet header
- FEC requires more bits because correcting is more difficult than detecting
- Both have limited capability (if too many bit errors, nothing can be done)

95

Dealing with bit errors

- Additional bits in the packet header required
- No error control
  - Typically checking for errors on packet header
- Error detection
  - Ability to detect errors in transmitted bits (data and header)
  - Packets received with errors are discarded
  - No assumptions can be made on the received packet
    - Equivalent to "nothing happened"
- Error recovery
  - Correction at the receiver
  - Error detection and retransmission
    - Requires the ability to identify packets (sequence number)
    - Implies sequence control
    - Slower than error correction (one round trip time needed) but requires less overhead
    - Can be implemented with less overhead (bits in the header) that error correction

96

Detection and retransmission *

- Window (ARQ) protocols
- Basic idea: send and wait for ACK
- If ACK not received, resend after a timeout
- Need packet (and ACK) numbering and error check field
- Timeout establishment (related to RTT, round trip time)
- How many packet to send before stopping and waiting for ACK?
  - One packet: stop and wait
    - Performance limitation
  - More packets: go back N and selective repeat protocols
    - Higher complexity: more memory, store packets out of order, etc
- Performance (bit rate) control
  - May be used for flow (and congestion control)

97

Window protocols throughput

- In absence of errors, the throughput of a window protocol is ruled by

$$\min\left\{\frac{transmission\_window}{RTT}, link\_speed\right\}$$

- For a given RTT there is an optimal window size that maximize throughput
- Larger windows translate in a ficticious RTT increase
- “Short” connections get higher throughput, for a given window size
- To control TX throughput, it is possible to act on
  - RTT (delaying ACK transmission)
    - May induce retransmissions
  - Window size

Integrated Networks ISDN and B-ISDN

ISDN

- ISDN: **Integrated** Services **Digital** Network
- Main goals
  - Extend telecommunication services of traditional telephone (POTS) network architectures
- Integrated Services: different services are provided to users using the same network resources
  - Not a dedicated network, rather an integrated network
- Digital: data transferred in digital format (bits or symbols), independently of their original nature, up to the user terminal
  - Take advantage of digital transmission
  - Get rid of original nature of data
    - Everything bits!

Integrated vs dedicated networks *

- Dedicated networks
  - Telecommunications networks were traditionally defined and designed to provide a specific service
  - one network paradigm for each service
    - Telephone network for the interactive human voice transport service
    - Internet for data exchange among computers
    - TV or radio distribution for the TV or radio systems
- Integrated networks
  - one network for any service
    - narrowband ISDN o N-ISDN
    - broadband ISDN o B-ISDN

Integrated vs dedicated networks *

- Dedicated networks
  - Easier to optimize for the specific service
  - “Optimal” engineering solutions for the specific requirements of the service
- Integrated networks advantages
  - No need to create an independent infrastructure for each service
  - Supporting different requirements implies sub/optimal choices
- Integrated networks trade flexibility and infrastructure cost reduction with perfomance and increased control complexity

ISDN: Main characteristics

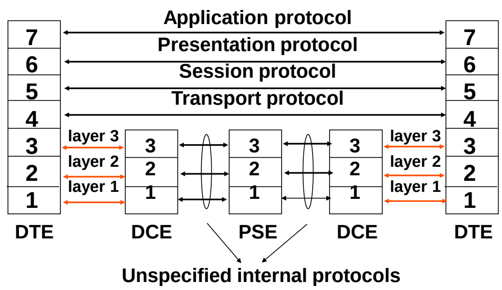
- Derived as an extension to the telephone network architecture
  - Connection oriented
    - Time-based billing
  - TDM frame at the physical layer
- Digital end-to-end
  - Also from user to first network node
  - POTS (old phones) supported through D/A conversion at user premises
- Offers both circuit and packet services (phone calls, fax, data transmission) but on a circuit-switched based network
- Standardized by CCITT (now ITU-T) from 1975 to 1980
  - Commercial services available to users starting from late 80s

ISDN-B channel

- Recommendation that describe the first three (lower) layers in data public networks
  - Similarly for the signalling D channel
- Packet transfer, connection oriented
- Packet switched network with virtual circuit service
- Specifies an “interface” between:
  - DTE (user terminal, computer, concentrator, multiplexer)
  - DCE (network device)
- “Interface” = protocols of layers 1, 2 and 3 in the OSI model

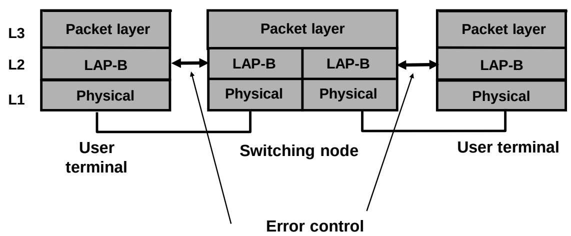
104

Network architecture



105

Layered architecture: ISDN B channel



106

ISDN B channel layers

- Physical layer
- Data link layer: LAPB (derived from HDLC)
  - Packet delineation
  - Addressing (one address for historical reasons)
  - Flow and sequence control with error recovery
- Packet layer:
  - Defines
    - the use of virtual circuits
    - data unit format
  - Flow error and sequence control (per virtual circuit)

107

Flow and sequence control

- Window (ARQ) protocol independent for each VC
- Transmitter window W is negotiated (default W = 2)
  - The transmitter can send up to W packets before receiving an ACK
- Cumulative ACKs
- An out-of-sequence (loss or duplication) causes a VC RESET
- An ACK out of the transmitter window causes a VC RESET

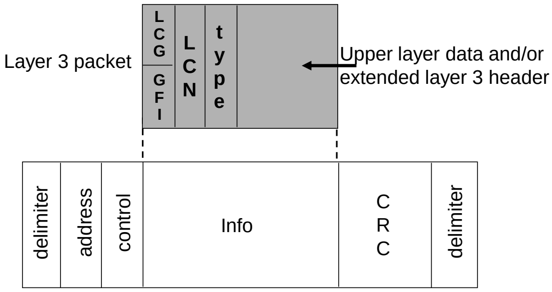
108

ISDN B channel layer 2

- Deals with the reliable data transfer on the link connection the DTE and the DCE
- Layer 3 packets are encapsulated in layer 2 packets
- Variable size packets, maximum size is negotiated and can reach 4096 byte
- The layer 2 protocol adopted is a variant of the ISO HDLC (High-Level Data Link Control) named LAPB (Link Access Procedure Balanced)

109

Encapsulation



110

Data format (of many layer 2 protocols)

- PDU format:

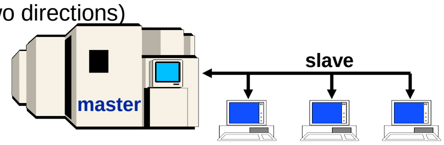
01111110	address	control	data	CRC	01111110
8	8	8/16	>=0	16	8

- Bit oriented protocol, with bit-stuffing to ensure data transparency (the flag fiels 01111110 must not appear in other fields)
- Address field is derived from multi-point (master/slave) configuration
- Control field differentiates the PDU type

111

Master/Slave Configuration

- Protocol used to manage the communication among master and slaves (multiplexing and multiple access in the two directions)
- PDUs sent from the master carru the slave (destination) address. PDUs send from slaves carry the slave (source) address.



112

HDLC: three types of PDUs

- Control field
  – Differentiates three types of PDUs

Information

0	N(S)	P/F	N(R)

Supervision

1	0	S	S	P/F	N(R)

Unnumbered

1	1	M	M	P/F	M	M	M

- P/F = poll/final bit

113

HDLC: three types of PDUs

- Information
  – Data sent after connection opening
  – Packet and ACK numner
- Supervision
  – ACKs (positive and negative)
- Unnumbered
  – Used for control
    - Link management
    - Setting up operational modes
  – Data sent in connectionless mode

114

Frame relay

115

Frame Relay

- Standard to create packet networks based on virtual circuits (normally permanent VCs) on a wide area
- The standard was originally proposed within the ISDN framework
- Today used (see later)
  - to create VPN (Virtual Private Networks) for companies
  - to interconnect LANs
  - to build logical topologies to interconnect Internet routers for ISP
- Bit rate ranging from 64 kb/s to 2 Mb/s
- Variable size packets (well suited to data traffic)
  - Maximum size 4096 byte
- http://www.frforum.com

Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 116**

116

Frame Relay

- Similarities with X.25
  - DCE-DTE "interface" standard
  - Multiplexing of different virtual circuits over the same transmission line
- Dissimilarities from X.25
  - Only defines layer 2 (and 1) protocols
  - Avoids link-by-link error control (wired transmission lines with negligible transmission errors)
    - core-and-edge approach
  - Defines a Frame Relay "network"
    - how is it possible without a layer 3, needed for routing purposes?

Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 117**

117

Logical topology design

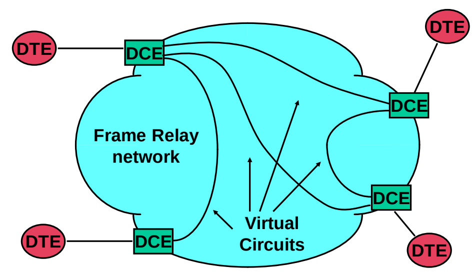
- **Recall**
- Need to distinguish between
  - Logical topology: interconnections among nodes (e.g. routers) via logical channels
  - Physical topology: physical layout of nodes and transmission channels
- Properties of a network depend on the logical topology
  - The physical topology imposes constraints on how logical topologies can be designed, due to capacity limitations

Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 118**

118

Frame Relay

- Operates on Permanent Virtual Circuit (although signaling protocols to deal with SVC are defined)



Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 119**

119

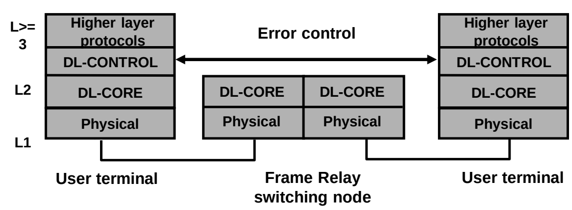
LAPF

- Frame Relay defines the LAPF protocol (Link Access Procedure to Frame mode bearer services)
- LAPF is divided in two parts:
  - DL-CORE (reccomendation I.233)
    - Used in all network nodes
  - DL-CONTROL
    - Optionally used only by end users (today, mainly IP routers)
    - In most applications, it is not used

Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 120**

120

Frame relay core and edge approach



Andrea Bianco – TNG group - Politecnico di Torino — **Computer Networks Design and Control- 121**

121

LAPF packet

- Packet delimitation through *flag* and *bitstuffing* to guarantee data transparency

Flag	Address	Control	Information	CRC	Flag

- DL-CORE
- DL-CONTROL (like HDLC with extended numbering)

122

LAPF packet

- ADDRESS field contains
  - the DLCI (Data Link Connection Identifier),the virtual circuit identifier
  - some additional bits for congestion control and traffic policing

FLAG
ADDR (MSB)
ADDR (LSB)
control
information
CRC (MSB)
CRC (LSB)
FLAG

- DL-CORE
- DL-CONTROL

123

ADDRESS field

- DLCI: Data Link Connection Identifier
- FECN/BECN: forward/backward explicit congestion notification
- DE: discard eligibility
- C/R: command/response
- D/C: DLCI or DL-CORE
- EA: extension bit

default format (2 byte)

upper DLCI			C/R	EA 0
lower DLCI	FECN	BECN	DE	EA 1

format 3 byte

upper DLCI			C/R	EA 0
DLCI	FECN	BECN	DE	EA 0
lower DLCI or DL-CORE control			D/C	EA 1

format 4 byte

upper DLCI			C/R	EA 0
DLCI	FECN	BECN	DE	EA 0
DLCI				EA 0
lower DLCI or DL-CORE control			D/C	EA 1

124

B-ISDN

125

B-ISDN

- First real attempt to design an integrated network to provide any service
  - Not only ISDN evolution
  - Support all type of services, with different transmission speeds and quality of service requirements over the same network infrastructure
- Private and public networks
- Standardized by ITU-T and ATM Forum
- Exploit ATM as a transport, multiplexing and switching technique
- Rec. I.121,1991: B-ISDN supports switched, semi-permanent and permanent, point-to-point and point-to-multipoint connections and provides on demand, reserved and permanent services. Connections in B-ISDN support both circuit mode and packet mode services of a mono and/or multi-media type and of a connectionless or connection-oriented nature and in a bidirectional or unidirectional configuration.

126

B-ISDN requirements

- Broadband network
  - Large area coverage
  - Large number of user
  - High speed
- Integrated network
  - Heterogeneous traffic over a single infrastructure
  - At least efficient support for voice and data
- QoS (Quality of Service) guarantee
  - Different guarantees for each connection (virtual circuit)
    - Negotiated between user and network
    - a priori control of delay
    - a priori control of bandwidth
    - a priori control of loss probability
- End/to/end guarantees
  - Must reach the end user

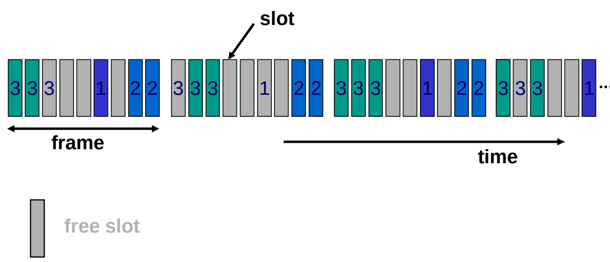
127

ATM: Asynchronous?

- ATM (Asynchronous Transfer Mode) versus STM (Synchronous Transfer Mode)
- STM transport
  - The link is organized in frames of fixed temporal duration (0.125 ms)
  - Each frame is subdivided in slots, data unit of fixed size
- ATM transport
  - No fixed size frame
  - Data unit of fixed size

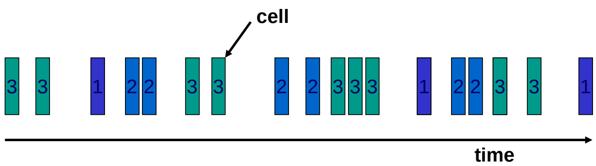
128

STM transport



129

ATM transport



130

STM and ATM switching and multiplexing

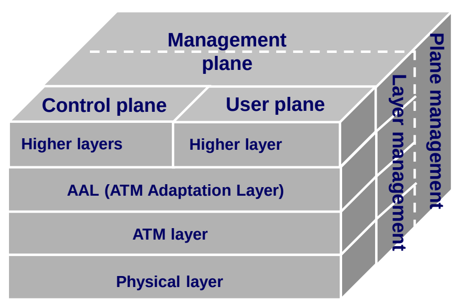
- STM
  - A fixed number of slots per frame is assigned to each connection
  - No need to identify explicitly each connection ⇒ positional switching
  - Deterministic multiplexing
  - Suited to fixed bit-rate circuit services
- ATM
  - No assignment is done, or a given number of cells per time is assigned
  - Need to explicitely identify the connection ⇒ label switching
  - Statistical multiplexing
  - Suited to both circuit and packet services

131

B-ISDN: Protocol architecture

132

B-ISDN: reference model

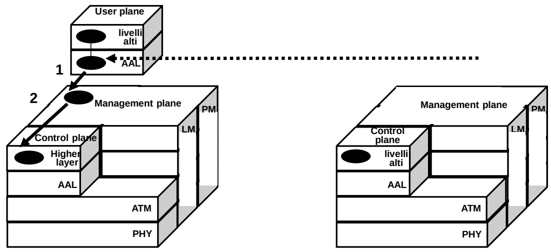


133

Planes

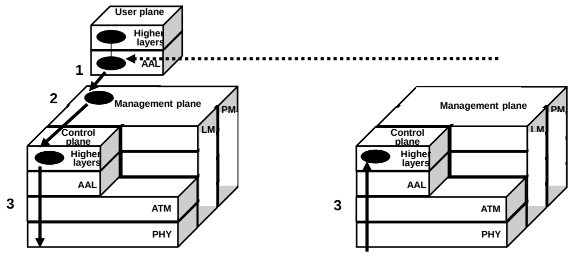
- Data plane
  - Dealing with user data transport
- Control plane
  - Dealing with signaling
  - Open/close virtual circuits
- Management plane
  - Coordination among data and control plane
  - Measurements
  - Monitoring
  - Charging
  - Partly layered and partly not layered

Opening a SVC



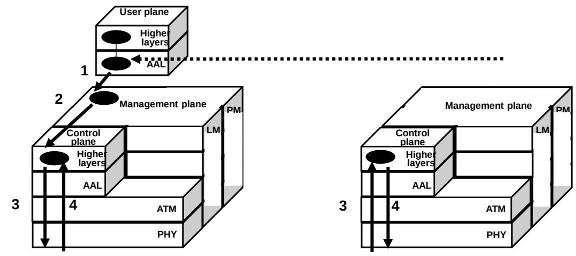
PM = plane management
LM = layer management

Opening a SVC



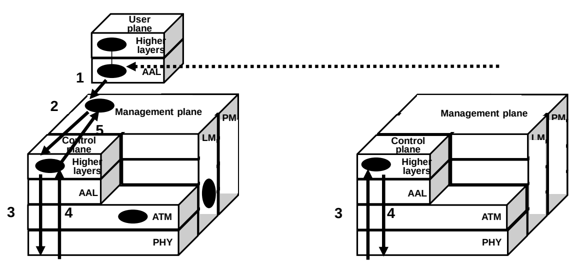
PM = plane management
LM = layer management

Opening a SVC



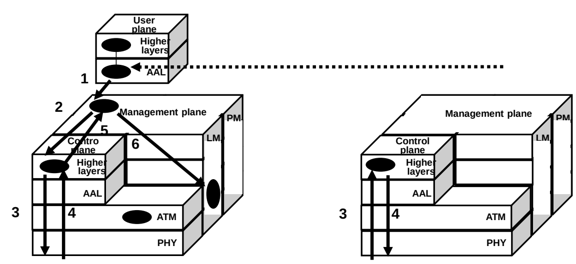
PM = plane management
LM = layer management

Opening a SVC



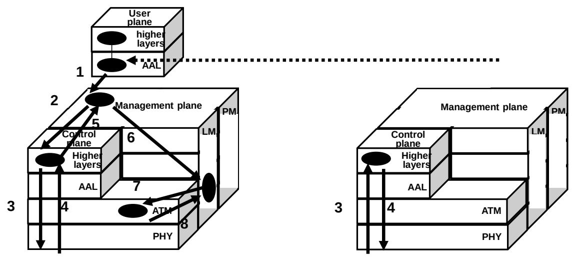
PM = plane management
LM = layer management

Opening a SVC



PM = plane management
LM = layer management

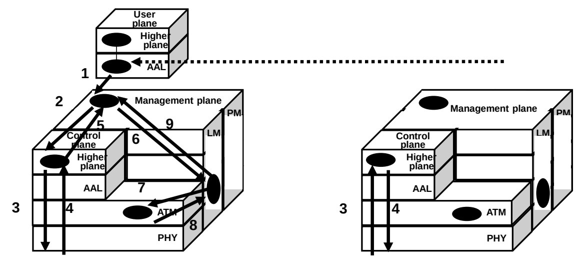
Opening a SVC



PM = plane management
LM = layer management

140

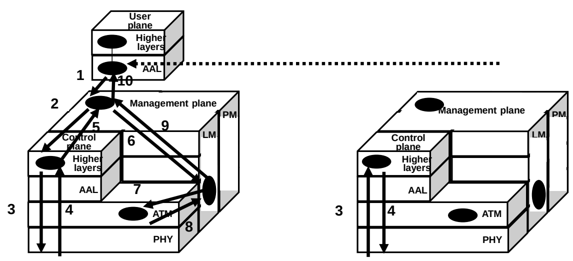
Opening a SVC



PM = plane management
LM = layer management

141

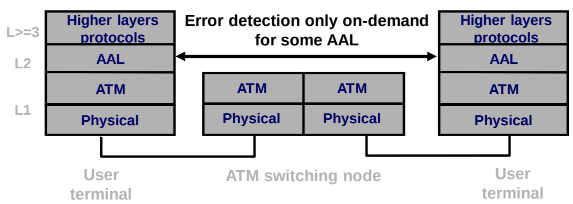
Opening a SVC



PM = plane management
LM = layer management

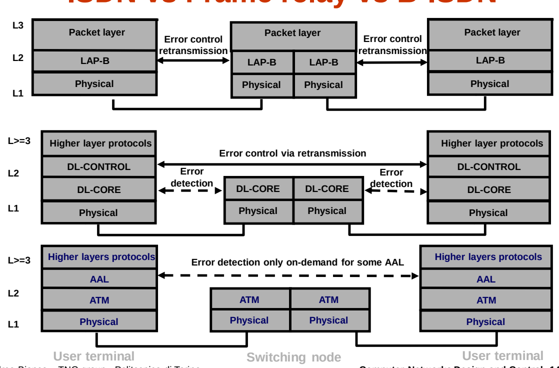
142

B-ISDN
Core and edge approach (user plane)



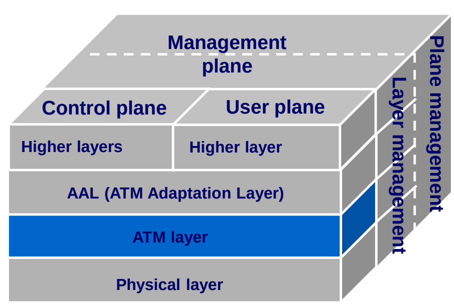
143

ISDN vs Frame relay vs B-ISDN *



144

B-ISDN: reference model

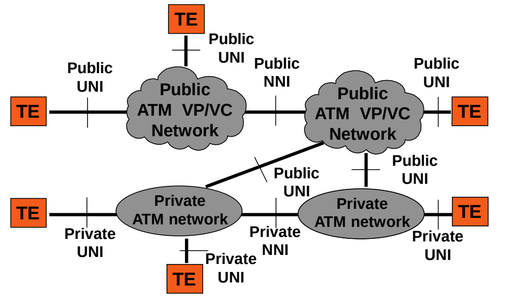


145

ATM cell format

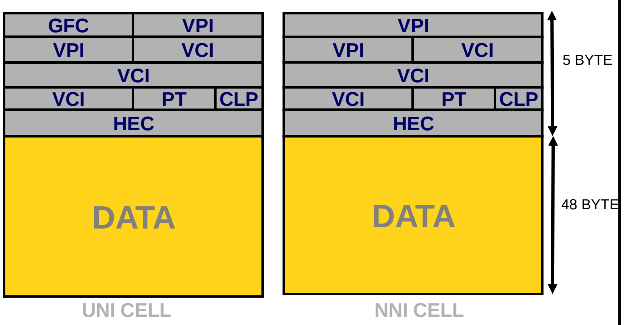
- Header (5 bytes) + payload (48 bytes)
- Fixed size cell
  - To ease the switching task at high speed (synchronous switching)
- Small cell size
  - Reduced latency (can be obtained by increasing transmission speed)
  - Small packetization delay for interactive voice services
  - Segmentation overhead
- Not only data but also connection management and rate control

B-ISDN interfaces



UNI: User to Network Interface NNI: Network to Network Interface

ATM cell format



ATM cell format

- ATM cell header (5 bytes = 40bit)
  - GFC (4 bit): Generic Flow Control
  - VPI (8-12 bit): Virtual Path Identifier
  - VCI (16 bit): Virtual Circuit Identifier
  - PT (3 bit): Payload Type
  - CLP (1 bit): Cell Loss Priority
  - HEC (8 bit): Header Error Code

ATM cell format

- GFC - Generic Flow Control
  - Only at the UNI interface.
  - The network issues information to user on the number of cells that can be sent
  - Two control algorithms:
    - ON-OFF
    - Credit based

ATM cell fromat

- VPI - Virtual Path Identifier
  - Variable length:
    - 8 bit at the UNI (256 VP's)
    - 12 bit at the NNI (4096 VP's)
  - Some VPIs are reserved to network management functions and to signalling

ATM cell format

- VCI: Virtual Circuit Identifier
  - Identifies a single virtual circuit within a given VPI.
  - 65536 VC's are available for each VP.
  - Example: link at 2,4 Gb/s, 1 VP, all VCs with the same capacity ⇒ 36Kb/s for each VC.

152

ATM cell format

- PT - Payload Type
  - Classifies the payload information type.
  - It contains an identifier named Payload Type Identifier (PTI).
  - Among the eight possible codes,
    - four are reserved to network functions
    - four to user function

153

PT field (Payload Type)

PT			MEANING
0	0	0	*User cell* EFCI No congestion AAL 5 indication=0
0	0	1	*User cell* EFCI No congestion AAL 5 indication=1
0	1	0	*User cell* EFCI Congestion AAL 5 indication=0
0	1	1	*User cell* EFCI Congestion AAL 5 indication=1

154

PT field (Payload Type)

PT			SIGNIFICATO
1	0	0	OAM cell (Operation and Maintenance)
1	0	1	OAM cell (Operation and Maintenance)
1	1	0	RM cell (Resource Management)
1	1	1	Not used Reserved for future use

155

ATM cell format

- CLP - Cell Loss Priority
  - Two priority levels at the ATM layer (within each VC)
  - In ATM switches, it permits to selectively discard cells in case of buffer congestion
  - CLP=0 indicates a high priority cell

156

ATM cell format

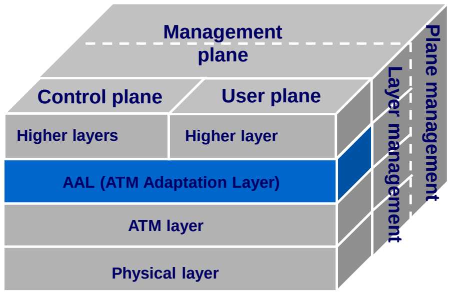
- HEC - Header Error Code
  - It permits to check the correctness of the ATM cell header only
  - No error detection on paylod!
  - Single errors are corrected
  - Two errors are detected
    - SEC/DED: Single errore correction/ Double Error Detection

157

ATM layer functions

- Connection opening and closure
  - Label assignment
- Cell header generation and extraction
  - 48 byte + 5 byte = 53 byte
- Switching and multiplexing
- Label swapping
- Performance monitoring at the ATM layer
  - Delay management
  - CLP bit management
    - Selective discarding
  - User parameter control
  - ECN (Explicit Congestion Notification)
  - Cell type discrimination
    - User, OAM, Control

B-ISDN: reference model



AAL: ATM Adaptation Layer

- Integrates ATM transport to offer service to users
- Service dependent layer
  - Four AALs
- Examples of AAL functions:
  - **Segmentation and reassembly**
  - Error detection and management
  - Cell loss management
  - Flow control
  - Synchronization
  - Timestamping
  - Sequence numbering

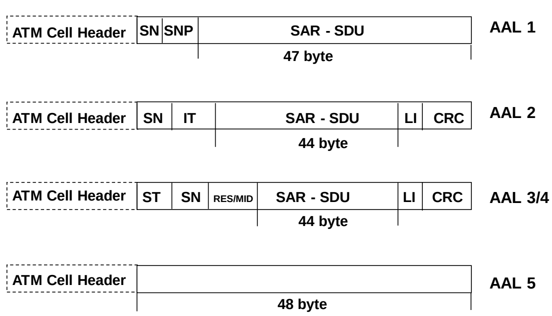
AAL service classes

	Class A	Class B	Class C	Class D
Synchronism required between source and dest	required		not required	
Speed	costant (CBR)	variable (VBR)		
Connection type	Connection oriented			connection less
AAL type	AAL 1	AAL 2	AAL 3/4 - 5	
Possible applications	voice 64kbit/s video CBR	video/audio VBR	data	data

AAL layer: architecture

- The AAL layer is subdivided into two sub-layers:
  - convergence sublayer (CS):
    - Multiplexing
    - Error detection
    - Synchronism recovery
    - Sequence numbering
    - Timestamp
  - segmentation and reassembly (SAR):
    - Segmentation in transmission, reassembly in reception of CS PDUs

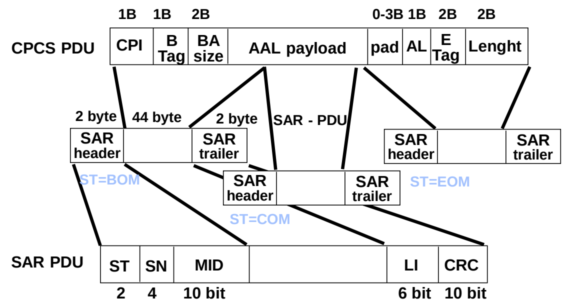
AAL format



AAL

- AAL1 and AAL2 for muktimedia
  - Packetization
  - Adaptive source clock recovery
  - Timing information transfer
  - Sequence counter
  - Counter error recovery
  - Cell loss notification

AAL 3/4



CS AAL 3/4 main header fields

- Packet delineation via BTag e ETag
  - Assume the same value (BTag=ETag)
- PAD: padding field, to align the PDU size to a multiple of 4 byte
- Length: PDU length measured according to the CPI field

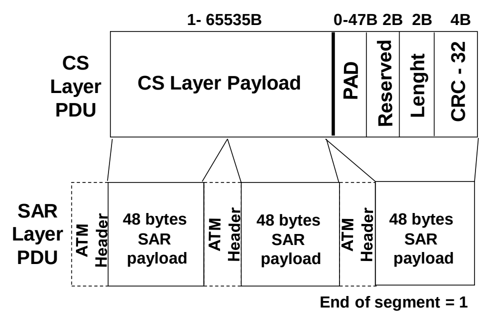
AAL 3/4 main SAR header field

- ST (Segment Type):
  - BOM (Begin of Message), COM (Continuation), EOM (End), SSM (Single Segment)
- SN (Sequence Number): increasing number
- LI (Lenght Indicator): PDU length (in byte)
  - Equal to 44 for BOM, SSM and COM cells
- CRC: error control on data

AAL 3/4 SAR function

- When transmitting:
  - Data segmentation, ST and SN management
  - CS-PDU multiplexing by using different MIDs
- When receiving:
  - Length verification through the LI field
  - CRC verification
  - Data re-assembly
  - Dropping incomplete or not correct CS-PDUs

AAL 5



AAL 5

- No CS layer
- SAR layer exploit all 48 byte payload
- Last cell created by the segmentation process has the third bit in the PT field of the ATM header set to 1
  - Layer separation principle violated!
- Error control over the full CS-PDU

AAL 5 vs AAL 3-4

- Advantages Simplicity
  - Efficiency
  - Improved reliability (CRC - 32)
- Disadvantages
  - Uses the third bit of the PT field in the ATM header!
  - Loss of the cell with the PT bit set =1 implies that two full CS-PDUs are lost

B-ISDN today *

- Did B-ISDN reaches the original goals?
  - No
- Is it still used?
  - Yes, mainly as an alternative to Frame Relay to create logical topologies
- From the performance point of view for **data transfer**, is there any benefit in using ATM with respect to Frame Relay?
  - No, the segmentation process required by ATM may only worsen performance
    - More losses
    - «Useless» traffic

LANs (Ethernet) : Protocol architecture

LANs

- Small geographical extension
- Shared transmission medium (originally) $\Rightarrow$ only one node can transmit at a time
  - Multiple access problem
  - Motivation: bursty traffic
    - Dedicated channel would be wasted
    - When sending, each node would like a high tx speed
  - Useful for broadcast-multicast transmission
    - See next slide
    - Need to use address to identify node for unicast traffic
- Many topologies
  - bus,ring, star

Multicast in meshed topologies *

- More complex than in broadcast channel
- If treating multicast traffic with a group size of k as k unicast connections
  - Scalability issue at the source node
  - Lot of resources required in the network
    - k flows from source to destination
- Better solution: create a multicast tree
  - Optimal tree definition is NP (broadcast is polynomial, spanning tree)
  - Requires network support
    - Nodes must create packet copies
    - K flows still generated within the network (task distributed)
- Multicast groups may be dynamic

Solutions for medium access *

- Static channel division
  - Fixed assignement of portion of channels
    - Time Division
    - Frequency Division
    - Code Division
- Not suited to bursty traffic
  - (N queues and servers at speed C are worse that 1 queue and server at speed NC)
- Could extend to a dynamic assignment scenario
  - Suppose a centralized controller
  - Need to collect node tx needs (according to which access scheme?)
  - Need to send allocation decision to nodes (according to which access scheme?)
  - Complexity and increase in delay
- Solution: rely on distributed, access protocols
  - Goal: to emulate statistical multiplexing

176

Access protocols for LANs: taxonomy

- Three main families:
  - Random access (CSMA/CD, Ethernet)
  - Ordered access (Token Ring, Token Bus, FDDI)
  - Slotted, with reservation (DQDB)
- How to evaluate LAN access protocols performance
  - Throughput
  - Fairness
  - Access delay
  - Number of nodes, network size, reliability, ease of deployment

177

Random access protocols

- Free access
  - Each node send at the channel speed R
  - No coordination among nodes
- If two concurrent transmissions ⇒ collision
- MAC (Medium Access Control) random access protocols specify:
  - How to detect a collision
  - How to recover after a collision has been detected
- ALOHA: random transmission. If collision is detected, retransmit after a random delay

178

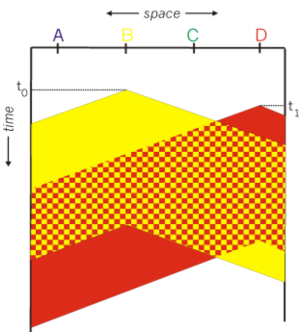
CSMA: Carrier Sense Multiple Access

- Sense the channel before transmission
  - If the channel is **sensed** free, transmit a packet
  - If the channel is busy, defer transmission to avoid collision
    - 1-persistent CSMA: retry transmission as soon as channel sensed free
    - 0-persistent CSMA : retry transmission after a random time
    - p-persistent CSMA: with p behave as 1-persistent, with probability (1-p) behave as 0-persistent

179

CSMA: collisions?

- Collisons occur due to propagation delay
- If a collision occurs, a full packet transmission time is wasted
- The propagation delay (distance) plays a fundamental role in collision probability
  - Vulnerability period depends on propagation delay



180

CSMA/CD (Collision Detection)

- CSMA/CD adds to CSMA
  - If a collision is (quickly) detected, packet transmission is suspended
  - Reduce the waste due to useless transmission
- Collision detection:
  - Compare the tx signal with the rx signal
  - Easy in wired LANs:
  - Almost impossible in wireless LANs: half duplex (when tx the rx is disbled)

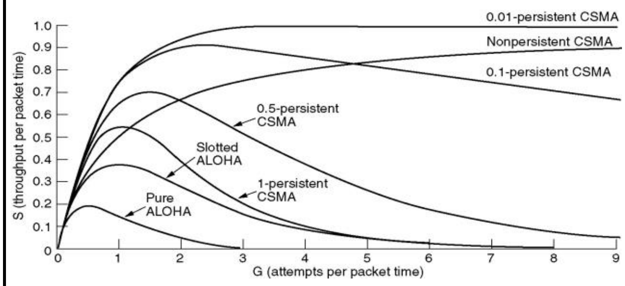
181

CSMA/CD: performance

- Throughput performance strongly depend on the end to end propagation delay
  - More precisely, on the ratio between packet transmission time and the propagation delay
- Very good throughput performance on small size networks (with respect to packet size) and with relatively small transmission speed
- Large packets, much larger than network size!
- Constraint on the minimum packet size to detect collisions (a node must transmit when detecting a collision)

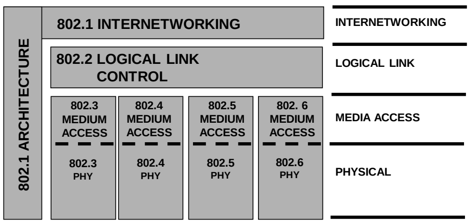
182

Random access protocols performance



183

Standard IEEE 802



- Many other committees
  - 802.11: Wireless Networks
  - 802.13: 100 base X
  - 802.15: Bluetooth

184

MAC Address

- Identifies each NIC (Network Interface Card) on a local area network
- 6 byte
- MAC addresses can be:
  - single or unicast: data for a single access node
  - multicast: data for a group of station
  - broadcast (FF FF FF FF FF FF): data for all stations

185

MAC addresses

- When a MAC NIC receives a correct packet
  - If the MAC unicast destination address is the NIC address, accept the packet
  - If the MAC destination address is broadcast, accept the packet
  - If the MAC destination address is multicast, accept the packet if the multicast group has been (via software) enabled
- Promiscuos mode bypass any control

186

Ethernet and IEEE 802.3

- Packet switching with datagram service
- Ethernet: commercial standard developed by Digital, Intel e Xerox (DIX) in the '70s
  - Ethernet 2.0 specification defined by DIX in 1982
- IEEE defines the 802.3 standard, based on Ethernet (1985)
- Ethernet and IEEE 802.3 have minor differences
  - Etehernet and 802.3 NICs co-exhist in the same LAN
- Protocol
  - CSMA/CD 1 persistent
  - No ACK is sent to confirm packet reception

187

Ethernet: packet format

Field	BYTES
Preamble = 101010……..	7
SFD = 10101011	1
MAC Destination Address	6
MAC Source Address	6
Protocol type > 1500	2
DATA	46 - 1500
FCS	4
Inter Packet GAP (silence))	12 bytes time

188

IEEE 802.3: packet format

Field	BYTES
Preamble = 101010…	7
SFD = 10101011	1
MAC Destination address	6
MAC Source address	6
Lenght (<1500)	2
DATA	0 - 1500
Padding	0-46
FCS	4
Inter Packet GAP (silence)	12 bytes time

189

LAN Interconnection

- Needed to
  - Extend LAN physical size
  - Increase the total number of access nodes
    - Each portion (cable) of a LAN has limitations on the maximum number of access nodes
- Requirement
  - Should be a transparent extension
  - Avoid to modify protocol architecture
    - Keep the same NIC

190

Interconnecting devices

- Repeater or Hub (layer 1)
  - **Not an interconnecting device**
  - Permit to extend cable lenghts
  - Equivalent to amplifiers
- Bridge or Switch (layer 2)
  - Simple routing algorithms
  - Work only on loop free topologies
- Router (layer 3)
  - Complex routing algorithms
  - Any topology
- Gateway (layer 4-7)
  - Useful to interconnect networks with different layering structure

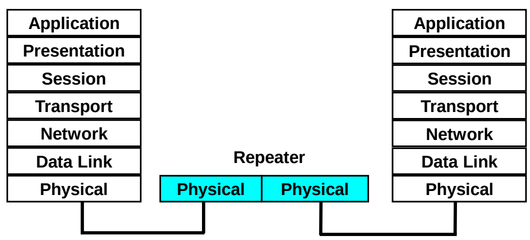
191

LAN interconnection *

- Bridge, switches, routers, gateways operate at different protocol layers
  - Different device complexity
- Key element: store and forward behavior of the interconnecting device
  - Move from a single channel scenario to a topology
- May increase LAN throughput performance
  - Increase the number of links in the topology
    - Increase network capacity
    - More space diversity
  - Need to exploit traffic locality to gain advantage

192

Repeater or Hub

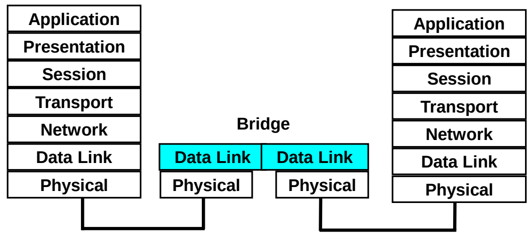


193

Repeater or Hub

- Multi-port device
- Operates at the bit level (layer one)
- Extend the cable lenght
  - No increase in network capacity
- Regenerates strings of bit and forwards them on all the ports
- Shared bandwidth on all ports
- 3R: re-generation, re-shaping, re-timing
  - May introduce delays
- Repeaters
  - On coaxial cable
  - Tree-like topology (interconnected buses)
- Hubs
  - Structured cabling (ease cabling and maintenance)
  - On twisted-pair or fiber
  - Star based topology

Bridge or Switch



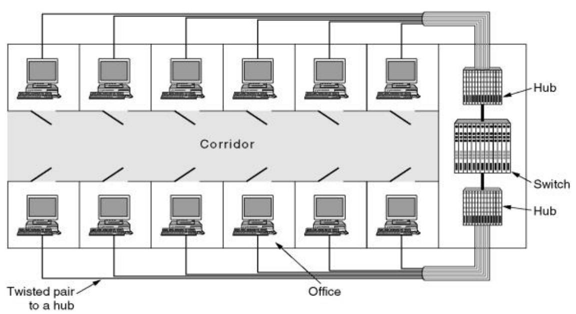
Bridge/Switches

- Layer 2 devices
  - Operate on layer 2 addresses
- From one segment LAN to extended LANs
  - Interconnect segments of LANs
- Enable to increase the network size
- Store and forward devices
- Dedicated bandwidth per port
- Transparent to users (same behaviour with or withouth bridge/switch)
- Do not modify packet content
- Limited routing capability
  - Backward learning algorithm (see later)

Bridge/Switches

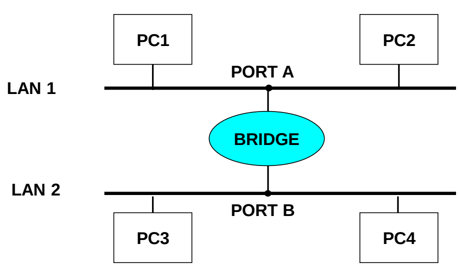
- Bridge
  - Operates on coaxial cable
  - Interconnect LANs, possibly with different MAC
  - Run the spanning tree protocol (see later)
- Switches
  - Operates on twisted pair
  - Interconnect LANs (or single users) with the same MAC
  - Support VLANs
  - Sometimes do not run the spanning tree protocol (see later)

Modern LANs



Bridge/Switch

- Packets received on LAN 1 are transmitted on LAN 2 only when needed



Bridge/switch operations

- Focus on transparent bridging
- Each bridge/switch has a unique ID
- Each bridge/switch port has a unique id
- Forwarding tables are initially empty!
- Three fundamental functions:
  - address learning: to dynamically create a routing (forwarding) table at the MAC layer (MAC Address, port_id)
  - frame forwarding: forward packets depending on the outcome of the routing table look-up
  - spanning tree algorithm execution to operate on a loop-free (tree) topology

200

Address learning

- Exploits the Backward learning algorithm
- For each received packet
  - Read the **source** MAC address MAC_S to associate the address with the port PORT_X from which the packet has been received
  - Update timer associated to the entry (MAC_S, PORT_X)
  - Will later use PORT_X to forward packets to MAC_S
- Timer needed to automatically adapt to topology variations and to keep the table size small

201

Frame forwarding

- When a correct packet (wrong packets are dropped) with a unicast MAC_D destination address is received on PORT_X
  - Look for MAC-D in the table
  - If found and associated to PORT_X, drop the packet
  - If found and associated to port_Y, forward to PORT_Y
  - If not found, forward to any other output port except PORT_X
- If the packet has a multicast/broadcast address
  - Forward to any port except PORT_X

202

Spanning tree

- Needed to avoid loops
  - Build a logical tree topology among bridges/switches by activating/de-activating ports
- Some switches may not support the spanning tree
  - Need to interconnect in a loop-free physical topology

203

Bridge/Switch/Router properties

- From a multiple-access network to a multiplexed network
  - Reduce collision probability by partitioning the network in independent segments
- For a full duplex fully switched network
  - Ethernet becomes a framing and transmission technique alternative to LAP-B, LAP-F, ATM
  - The MAC layer becomes useless
  - Physical distance limitations induced only by the media transmission properties, not by the MAC
- Ease security and management
  - Traffic separation

204

Bridge/Switch properties *

- Throughput performance may increase
  - More space diversity (higher capacity)
  - Need to exploit traffic locality
- Introduce store and forward (and queueing) delays
  - Worse delays than hubs
  - Store and forward delay significant with respect to propagation delay
    - Transmission time of a minimum packet size at least twice of the propagation delay
- Potential packet losses when queues are filled-up
- Unfairness in resource access

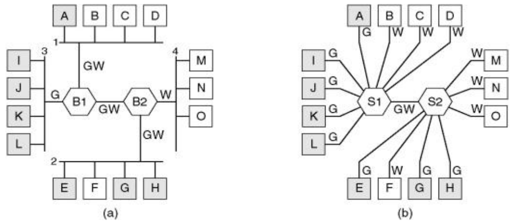
205

VLAN (Virtual LAN)

- Host are physically connected to the same network segment, but logically separated
- Broadcast/multicast packets forwarded only on ports belonging to the VLAN
- Need to extend the PCI MAC to identify packets as belonging to a specific VLAN
- Hosts belonging to separate VLANs cannot directly exchange packets

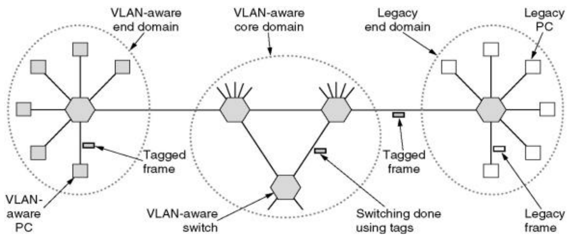
Virtual LANs

- (a) 4 LAN segments organized as 2 VLANs (white and grey) through two bridges
- (b) similar scenario with two switches



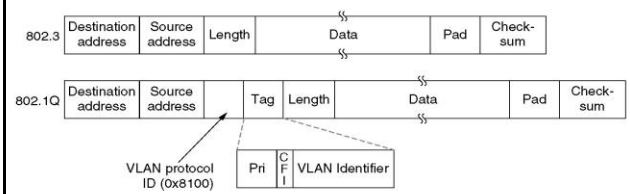
The IEEE 802.1Q Standard

- From legacy Ethernet to Ethernet with VLANs

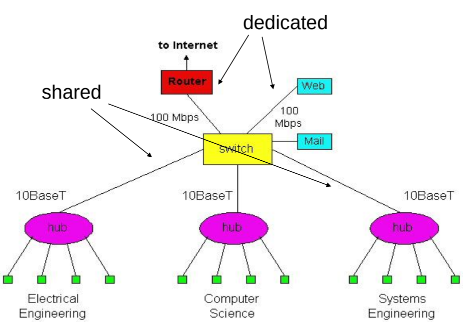


IEEE 802.1Q

- 802.3 Packet format (legacy) e 802.1Q.



Hierarchical LAN organization



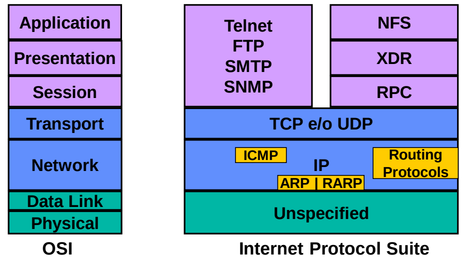
Layer 2 protocol comparison *

Protocol	Packet delimitation	Layer 3 protocol multiplexing	Error detection	Error correction (window protocol)	QoS support
LAPB + Layer 3	Flag	Through VC at layer 3	YES in both layers	Yes in both layers	Through VCs
LAPF core + LAPF control	Flag	Through VC	YES in LAPF core	Optional in LAP-F control (at the edge)	Through VCs One priority level per VC
ATM (core)+ AAL (edge)	Through physical layer	Through VC	YES in AAL (edge)	NO	Through VCs One priority level per VC
Ethernet MAC	Silence	YES	YES	NO	Priority in VLAN

Internet protocol architecture

212

Internet protocol suite



- *Internet in practice: ''Network Measurement Laboratory''*

213

IP: Internet Protocol

- Layer 3 protocol
- Defines
  - Packet format
  - Address format
  - Data (named datagram) forwarding procedures
- Best-effort service
  - connectionless
  - unrealiable
  - With no QoS guarantess
- Specified in RFC 791 (november 1981)

214

IP protocol

- Connectionless delivery
  - Stateless approach
    - No state information on datagram kept in routers
    - No connection concept at IP layer
  - Each datagram routed independently
    - Two packets with the same source and destination can follow two different paths
    - In practice, most packets follow a fixed route, unless
      - Link failure
      - Parallel links among routers
- No QoS guarantees
  - All packets treated fairly
  - Extensions to the traditional IP QoS model

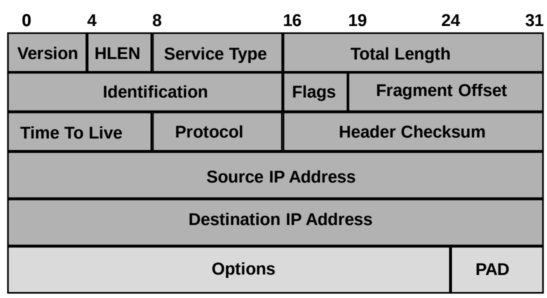
215

IP protocol: unreliable delivery *

- In case of
  - Failure (ex. out of service router, link failure)
    - Datagram dropped end error message sent to the source
  - Buffer shortage
    - Datagram dropped (no error message sent, since the datagram cannot be stored)
  - Checksum error (error control only over the header!)
    - Datagram dropped
    - No error message sent, since address may be wrong

216

IP packet header



Standard size: 20 byte

217

IP header fields

- VER: IP protocol version (currently used: 4, most recently defined: 6)
- HLEN: header length measured in 32 bit (equal to 5, if no options are used)
- Type of service (TOS): type of service required by the datagram (minimize delay, maximize throughput, maximize reliability, minimize cost). Traditionally ignored by routers. RFC 1349
- Total Length: datagram length in byte (header included).
  - Maximum size of IP datagram: 65535 byte

Fragmentation

- MTU (Maximum Transfer Unit): maximum size of an IP datagram, including header
  - Derived from layer 2 size constraints
  - Ethernet: 1500 B
- Minimum default MTU: 576 B
- When the link layer has a smaller MTU, IP datagram must be fragmented
- Fragments
  - Are independent datagrams, with almost the same hader as the original datagram (different fields: fragmentation fields (identification, flags, offset), length, CRC)
  - Reassemled only at the destination! (router never reassemble datagram, unless they are the final destination)
- Fragmentation process transparent to layer 4
- Can be applied recursively
- Specified in RFC 791, RFC 815
- It exist a path MTU Discovery (RFC 1191) algorithm to determine the "optimal" datagram size

Fragmentation *

- Fragmentation is harmful
  - More header overhead, duplicated over each fragment
  - Loss of a single fragment implies that the full datagram is lost; increses the loss probability
  - Creates "useless" traffic
    - fragments belonging to a datagram for which at least a fragment was lost are transported with no use
  - Reassemlby timers are needed at the receiver
- Reassembly normally done at network edge (hosts, not routers) to keep router complexity low
- IP over ATM needs AAL to avoid IP fragmentation on ATM celles (20B of IP header in each 48B ATM cell)

IP header fields

- Identification, Flags, Fragment offset: to control fragmentation operation
  - Identification:
    - Unique code for each datagram, generated at the source
    - Fragments originated by the same datagram have the same identification field
  - Fragment offset:
    - Specifies the position of fragment data with respect to the original datagram, as a multiple of 8 byte (first fragment has offset 0, last segment has offset = datagram size less last fragment size)
  - Flags (3 bit): don't fragment e more fragments (to identify the last fragment)

IP header fields

- TTL (time to live):
  - Datagram lifetime (in hops)
  - Initial value freely chosen at the source (typical values 64, 128, 256)
  - Each router decrements the TTL value by 1
  - If TTL=0, the router discards the datagram and send an ICMP error message to the source (can be disabled)
- Protocol: higher layer protocol code. RFC 1700 lists the protocol codes

Protocol	Name
1	ICMP
4	IP in IP
6	TCP
17	UDP
89	OSPF

IP header fields

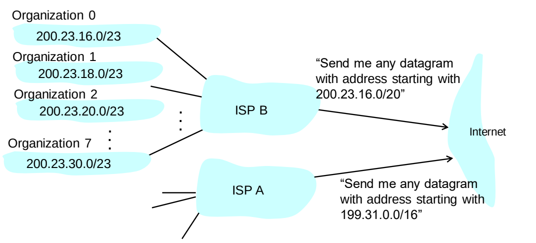
- Header Checksum: error control only over the header, non over user data.
  - Specified in RFC 1071,1141,1624,1936. Complement to 1 sum, aligning the header over16 bits
  - The header checksum can be computed incrementally (useful since each router decrements the TTL field and must re-compute the header).
- Source e Destination Address (32 bit): source and destination address of the hosts (may be routers) exchanging the datagram
  - Composed by a net_id and host_id
  - Masks to overcome the lack of available addresses

Hierarchical routing

- Ideal (conceptually simpler) case
  - All routers are identical
  - Flat network, no hierarchy
- Not useable in practice
  - Scalability: with 100 million of destination :
    - All destinations in a single routing table?
    - Routing info exchange would require too much bandwidth
  - Administrative autonomy
    - Internet = network of networks
    - Each network administrator is willing to control routing functions within its domain

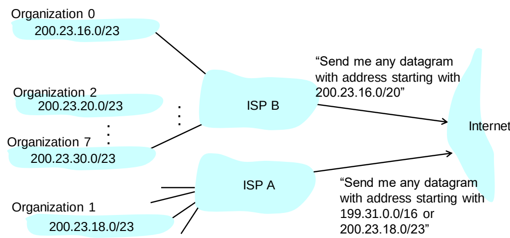
Hierarchical routing: route aggregation

- Hierarchical addressing permits more efficient announcements of routing infos



Hierarchical routing: route aggregation

- If ISP A has a more specific path to Organization 1



Hierarchical routing

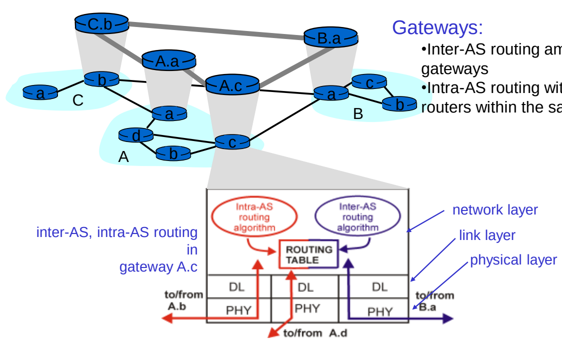
- Router aggregated in Autonomous System (AS)
  - Networks with complex structure (many subnets and routers) but with the same administrative authority
  - Router within the same AS use the same routing protocol
  - Intra-AS routing protocols: (IGP: Interior Gateway Protocol)
    - Routers belonging to different AS may use different IGP protocols

Hierarchical routing

- In each AS there exist "gateway" routers
  - Responsible to route to destinations external to the AS
  - Run intra-AS routing protocols with all other AS routers
  - Run also inter-AS routing protocols (EGP: Exterior Gateway Protocol)
- We can identify an internal routing (IGP) and an external routing (EGP)

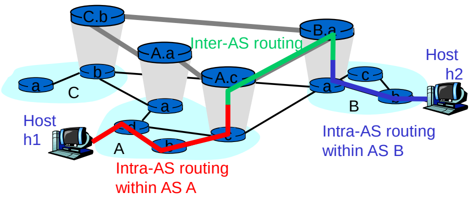
Routing Intra-AS and Inter-AS

Gateways:
- Inter-AS routing among gateways
- Intra-AS routing with routers within the same AS



Computer Networks Design and Control – Class intro and review

Routing Intra-AS and Inter-AS



Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 230

230

Network node comparison *

Device	Switching type	Address used for forwarding	Routing	If packet address not found	Routing table content	Forwarding table content
Ethernet or Layer 2 switch	Packet switching with datagram	Dest MAC address	Local via backward learning (look at source MAC)	Forward on all ports	None	MAC address Out port
IP or layer 3 router	Packet switching with datagram	Dest IP address	Distributed link state dist. vector Policy based	Default route or drop	IP address Mask Output port	None
Frame relay or ATM or any VC-based technology	Packet switching with virtual circuit	Label	Distributed	Drop	User unique address Output port	Input label Input port Output label Output port
SDN switch	Will see	Will see	Will see	Will see	Will see	Will see

Device architecture: "Cloud computing and data center design lab"

Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 231

231

Internet transport layer

- Two alternative protocols: TCP e UDP
- Different service models:
  - TCP is connection oriented, reliable, it provides flow and congestion control, it is stateful, it supports only unicast traffic
  - UDP is connectionless, unrealiable, stateless, it supports multicast traffic
- Common characteristics:
  - Multiplexing and demultiplazione of application processes through the port mechanism
  - Error detection over header and data (optional in UDP)

Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 232

232

Mux/demux: ports

- Final destination of data is not the host but an application process running in the host
- Interface between application processes and the network architecture is named port
  - Integer number over 16 bit
  - There is an association between ports and processes
    - Public server process are statically associated to well-know ports, with an identifier smaller than 1024 (e.g.: 80 for WWW, 25 for email)
    - Client processes use ports dynamically assigned by the operating system, with an identifier larger than 1024
      - Each client process on a given host has a unique port number within that host

Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 233

233

UDP: User Datagram Protocol

- Connectionless transport protocol
- No delivery guarantee
- Two main functions:
  - Application process multiplexing through port abstraction
  - checksum (optional) to verify data integrity
- Applications using UDP should solve (if interested)
  - Reliability issues
    - Data loss, data duplication
  - Sequence control
  - Flow and congestion control
- Standardized in RFC 768

Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 234

234

UDP: packet format

0 4 8 16 19 24 31

UDP Source Port	**UDP Destination Port**
UDP Message Length	**UDP Checksum**
DATA	

Andrea Bianco – TNG group - Politecnico di Torino

Computer Networks Design and Control- 235

235

UDP: applicability

- UDP is useful when:
  - Operating in local area, a reliable network (NFS)
  - All application data are contained in a single packet, so that it would be useless to open a (TCP) connection for a single packet (DNS)
  - Full reliability is not fundamental (some interactive video/audio service)
  - A fast protocol is needed
    - Connection opening overhead avoided
    - Retransmission mechanisms to ensure reliability cannot be used due to strict timing constraints
  - Application manages retransission mechanisms (DNS)
  - Need to send data at constant rate or at a rate independent from the network (some interactive video-audio services)

TCP protocol

- TCP (Transmission Control Protocol) is
  - Connection oriented
  - Reliable (through retransmission)
    - Correct and in-order delivery of stream of data
  - Flow control
  - Congestion control
- Used by applications requiring reliability
  - telnet (remote terminal)
  - ftp (file transfer protocol)
  - smtp (simple mail transfer protocol)
  - http (hypertext transfer protocol)

TCP

- Multiplexing/demultiplexing through ports
- Connection opened between two TCP entities (service similar to a virtual circuit)
  - bidirectional (full duplex)
  - With error and sequence control
- It is more complex than UDP, it requires more CPU and memory, state information (port numbers, window size, Packet and ACK numbers, timeout, etc) must be kept in each host for each TCP connection

TCP

- TCP freely segments and reassembles data:
  - Manages byte stream generated by application protocols; unstructured data at TCP level
  - A FIFO buffer byte oriented is the interface between TCP and application processes
- Window protocol to ensure reliability
- Flow control and congestion control operates on the transmitter window size
  - Flow control executed by the TCP receiver exploiting the window field in the TCP header
  - Congestion control autonomously executed by the TCP transmitter

TCP: connection identification

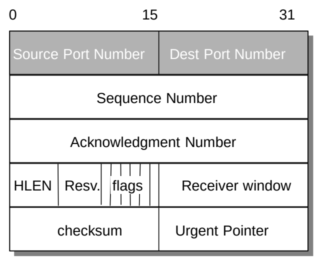
- A TCP connection is identified uniquely by:
  - Source and destination IP addresses (layering principle violation)
  - Source and destination port numbers
  - Example: TCP connection identifed by porta 15320 on host with IP address 130.192.24.5 and port 80 on host with IP address 193.45.3.10
- Note that TCP and UDP use port numbers are independent

TCP: header

0 4 8 16 19 24 32 bit

Source Port		Destination Port	
Sequence Number			
Acknowledgment Number			
HLEN	Resv	Control flag	Window
Checksum		Urgent Pointer	
Options			Padding

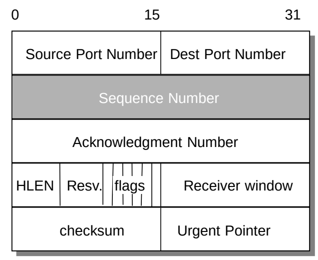
TCP header



Identify the application processes sending and receiving data

242

TCP header



- Identifies, in the data stream, the position of the first byte in the data carried in the segment
- Each side of the connection uses different and independent sequence numbers

243

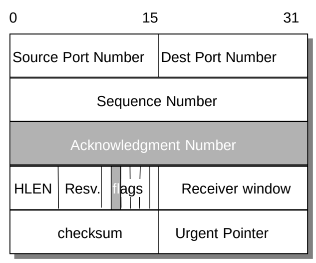
Sequence number

- The sequence number field is 32 bit long
- Depending on the available bit rate, there are different Wrap Around times (the same sequence number is seen again)

Capacity		First wrap around time	
T1	(1.5Mbps)	6.4	hours
Ethernet	(10Mbps)	57	minutes
T3	(45Mbps)	13	minutes
FDDI	(100Mbps)	6	minutes
STS-3	(155Mbps)	4	minutes
STS-12	(622Mbps)	55	seconds
STS-48	(2.5Gbps)	14	seconds

244

TCP header

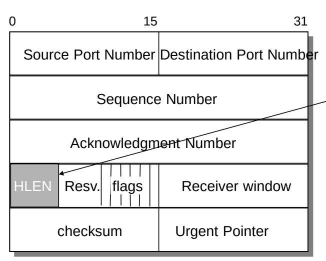


Sequence number +1 of the last byte correctly received

It is meaningful only if the ACK flag is set (almost always, unless at conection startup)

245

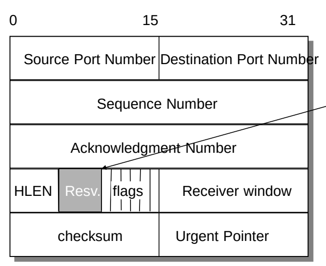
TCP header



Header lenght in 32 bit words (default value 5)

246

TCP header

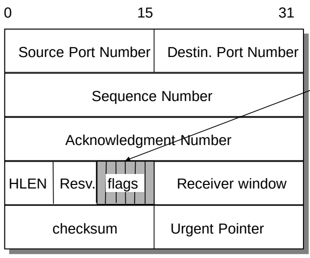


Reserved for future use (ECN)

247

TCP header

- Connection management

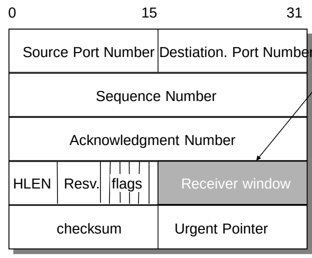


- Six flags, one or more can be set at the same time:
  - URG: urgent pointer valid
  - ACK: ack number valid
  - PSH: pass immediately data to the application
  - RST: connection ReSeT
  - SYN: SYNchronize seq. No. Connection opening
  - FIN: connection closing

Andrea Bianco – TNG group - Politecnico di Torino   **Computer Networks Design and Control- 248**

248

TCP header



Amount of data (in bytes) the receiver is willing to store (flow control) Maximum value 65535 byte, unless the *window scaling option* is used

Andrea Bianco – TNG group - Politecnico di Torino   **Computer Networks Design and Control- 249**

249

Window needed to fully exploit available bit rate

- Maximum amount of data flowing per RTT:
  - 16-bit rwnd = 64kB max
- Bit rate x delay product for RTT=100ms

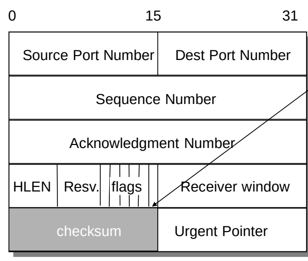
Bit rate		Possible Bit rate x delay
T1	(1.5Mbps)	18KB
Ethernet	(10Mbps)	122KB
T3	(45Mbps)	549KB
FDDI	(100Mbps)	1.2MB
STS-3	(155Mbps)	1.8MB
STS-12	(622Mbps)	7.4MB
STS-48	(2.5Gbps)	29.6MB

- Can be overcome with the window scale option

Andrea Bianco – TNG group - Politecnico di Torino   **Computer Networks Design and Control- 250**

250

TCP header

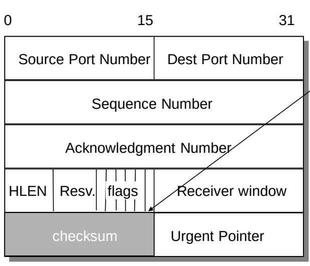


Checksum over header and data, plus a pseudo-header including IP addresses and protocol type (it violates the layering principle)

Andrea Bianco – TNG group - Politecnico di Torino   **Computer Networks Design and Control- 251**

251

TCP header



- Checksum algorithm
  - Align to 16 bit TCP header, data and pseudo-header
  - One complement sum
  - The 32-bit results is divided into two 16-bit parts
  - One complement sum of the two parts is inserted in the checksum field

Andrea Bianco – TNG group - Politecnico di Torino   **Computer Networks Design and Control- 252**

252