Class goals

• Describe, mainly in a qualitative way, techniques and algorithms to offer quality of service to users and to ease network management in telecommunication networks
  – Algorithms
  – Standardization
    • Telephone network
    • Internet
    • Frame-relay network (ISDN)
    • ATM network (B-ISDN)
    • Ethernet
Course syllabus

- Technology: review
  - Internet, Frame-Relay, ATM, Ethernet
- Introduction to quality of service
- Quality of service standardization efforts
  - Frame Relay
  - ATM
  - Internet
    - Intserv
    - Diffserv
  - Ethernet

- Algorithms
  - Policing / shaping
  - CAC: Connection Admission Control
  - Scheduling and buffer management
  - Congestion control
  - Network protection and restoration
- SNMP and network management
- SDH, WDM and capacity planning
- Lab classes (deal with scheduling and shaping algorithms using Click modular router)
Other info

- Class web site
  - http://www.telematica.polito.it/Network_Management_and_QoS_Provisioning/
  - Linked from the Politecnico portal http://didattica.polito.it/
- Use of mail addresses for announcements (delayed or cancelled lessons, …)
- Teaching material
  - 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 date and a phone number

Review

- Multiplexing and multiple access
- Switching techniques
- ISDN
  - X.25
  - Frame Relay
- B-ISDN
  - ATM
- Ethernet
- Internet (TCP/IP)
- “Low” layers in ISDN, B-ISDN and Ethernet, “high” layers in Internet
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 cellular network, PC accessing via a Wi-Fi hot-spot

Channel sharing techniques

- Frequency (FDM - FDMA)
- Time (TDM - TDMA)
- Code (CDM - CDMA)
- Space
Frequency division

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

Time division (TDM – TDMA)

- Each flow exploits different time intervals (slots)
- Define frame over which slot allocations are repeated
- Need for time guard
Code division
(CDM – CDMA)

• Each flow exploits a different code (waveform with higher frequency than the bit tx rate)
• Need for orthogonal codes
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

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

- Over the channel, transmitted signals sum up (need to equalize power)
  - Transmissions of 1+2+3: +3 +1 +1 -1
  - Transmissions of 2+3: +2 0 +2 0
### Code multiplexing

- **Un esempio (segue):**
  - 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

---

### 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 an example of space reuse
**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)

---

**FDM and TDM**

**FDM**
- Frequency
- Time

**Example:**
- 4 users

**TDM**
- Frequency
- Time
- 125 µs
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

Statistical Multiplexing

- Sequence of A & B packets does not have fixed pattern, bandwidth shared on demand
- Dynamic TDM scheme
Circuit switching

- Opening
- Data transfer
- Closing

Space vs time switching

125 µs

125 µs
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)

Packet switching

[Diagram showing 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

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
Grouping virtual circuits

- A virtual circuit is logically identified by a label
- Label = often a pair of identifiers (VCI-VPI in ATM)
  - Virtual channel (VC): identifies a single connection
  - Virtual path (VP): identifies a group of virtual channels
Grouping virtual circuits

- The grouping allows flow aggregation
  - Eases network management
  - Increases scalability
- Possible use
  - LAN inter-connection to create a VPN (Virtual Private Network)
  - Multimedia flows (video, audio, data)

Virtual circuits and paths (ATM)
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

X.25 – ISDN B channel

Gruppo Reti TLC
nome.cognome@polito.it
http://www.tlc-networks.polito.it/
ISDN

- ISDN: Integrated Services Digital Network
- Integrated Services: different services are provided to users using the same network resources (not a dedicated network, rather an integrated network)
- Digital: data are transferred in digital format (bits or symbols), independently of their original nature, up to the user terminal

Integrated vs dedicated networks

- Telecommunications networks were traditionally defined to provide a specific service
  - one service one network paradigm
    - Telephone network for the interactive human voice transportation service
    - Internet for data exchange among computers
    - TV or radio distribution for the TV or radio system
- Integrated networks
  - one network for any service
    - narrowband ISDN or N-ISDN
    - broadband ISDN or 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 performance and increased control complexity

ISDN: Main characteristics

- Connection oriented
  - Time-based billing
- Public and private
- Digital end-to-end
- Plesiochronous (TDM frame) at the physical layer
- Offers both circuit and packet services (phone calls, fax, data transmission) but on a circuit-switched based network
- POTS supported through D/A conversion at user premises
ISDN

- ISDN main goals
  - Extend telecommunication services of traditional POTS architectures
  - Uniform and standardized network
  - Provide a fully-digital user interface for all services
- ISDN
  - Standardized by CCITT (now ITU-T) from 1980 to 1990
  - Commercial services available to users starting from late 80s

ISDN: Transmission structure

- Based on two (TDM separated) flows:
  - B (Bearer) channel - 64 kb/s
    - voice, data, fax, low-resolution video
  - D (Data) channel - 16 kb/s (or 64 kb/s)
    - Signaling, data, remote-control
- An ISDN access could freely combine B and D channels
  - nB + mD (n and m can take arbitrary values)
- Classical commercial offer permit only few combinations of m and n.
ISDN: Transmission structure

- Standard interfaces:
  - BRI - Basic Rate Interface –
    - 2B + D (128kb/s)
  - PRI - Primary Rate Interface –
    - 30B + D (EU)
    - 23B + D (USA)

ITU-T (CCITT) X.25

- Recommendation that describe the first three (lower) layers in data public networks
- Packet transfer, connection oriented, low speed (75 bit/s up to 192 Kbit/s)
- Packet switched network with virtual circuit service
- Included in the ISDN (Integrated Services Digital Network) specification
**X.25**

- DTE: Data Terminal Equipment
- DCE: Data Circuit-terminating Equipment

**X.25 network**

- X.25 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
X.25

- PSE: Packet Switching Exchange

X.25 network architecture

Application protocol
Presentation protocol
Session protocol
Transport protocol

Unspecified internal protocols
Layered architecture: X.25 and ISDN B channel

- **Physical layer**
- **Data link layer:** LAPB (derived from HDLC)
  - Packet delineation
  - Addressing
  - Flow and sequence control, error recovery
- **Packet layer:**
  - Defines
    - the use of virtual circuits
    - data unit format
  - Flow and sequence control (per virtual circuit)
X.25 packet layer functions

- Virtual circuit opening and closing
- Data transfer over virtual circuits
- Error recovery per virtual circuit
- Flow control per virtual circuit
- Sequence control per virtual circuit
- Virtual circuit multiplexing
- Routing functions are missing
  - “Interface” standard

Virtual circuit identifiers

- To each SVC (switched circuit) and PVC (permanent virtual circuit) are assigned
  - Logical channel group identifier (< 16)
  - Logical channel number (< 256).
- To avoid conflicts, when opening a virtual circuit, the DTE uses first high numbers, DTE start assigning ids from low numbers.
- Small numbers are reserved to PVCs
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

X.25 and 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 in X.25 is a variant of the ISO HDLC (High-Level Data Link Control) named LAPB (Link Access Procedure Balanced)
X.25 encapsulation

Layer 3 packet

- Upper layer data and/or extended layer 3 header

Data format (of many layer 2 protocols)

- PDU format:

<table>
<thead>
<tr>
<th>Format</th>
<th>address</th>
<th>control</th>
<th>data</th>
<th>CRC</th>
<th>01111110</th>
</tr>
</thead>
<tbody>
<tr>
<td>01111110</td>
<td>8</td>
<td>8/16</td>
<td>&gt;=0</td>
<td>16</td>
<td>8</td>
</tr>
</tbody>
</table>

- Bit oriented protocol, with bit-stuffing to ensure data transparency (the flag fields 01111110 must not appear in other fields)
- Address field is used in multi-point configuration (master/slave)
- Control field differentiates the PDUs
HDLC: Master/Slave Configuration

- Protocol used to manage the communication among master and slaves
- PDUs sent from the master are named command, PDUs sent from slaves are named response

HDLC: operational modes

- Normal Response Mode (NRM)
  - Suited for point-to-point of unbalanced multi-point.
  - One primary station (Master) and several slaves
  - The Master sets the P/F bit to 1 to enable slave transmission
  - The slave sets the P/F bit to 1 in the last PDU
**HDLC: operational modes**

- **Asynchronous Response Mode (ARM)**
  - Unbalanced configuration. The slave can send data without waiting to be polled by the master.

- **Asynchronous Balanced Mode (ABM)**
  - Balanced point-to-point configurations. The P/F bit requires immediate response.

---

**HDLC: three types of PDUs**

- **Control field**
  - Differentiates three types of PDUs

<table>
<thead>
<tr>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supervision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unnumbered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

- **P/F = poll/final bit**
**HDLC: three types of PDUs**

- **Information**
  - Data sent after connection opening
- **Supervision**
  - ACKs (positive and negative)
- **Unnumbered**
  - Link management
  - Data sent in connectionless mode

**HDLC: two numbering schemes**

- Normal numbering (modulo 8) and extended numbering (modulo 128)
- Control field equal to either 1 or 2 byte

<table>
<thead>
<tr>
<th>Information</th>
<th>0</th>
<th>P/F</th>
<th>N(S)</th>
<th>N(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervision</td>
<td>1 0 S S x x x</td>
<td>P/F</td>
<td>N(R)</td>
<td></td>
</tr>
<tr>
<td>Unnumbered</td>
<td>1 1 M M P/F M M M</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
HDLC vs LAP-B

- LAP-B (ISDN B channel) uses only a subset of PDUs defined by HDLC
- We briefly recall the PDUs used by LAP-B

LAP-B: information PDUs v(I)

- Data transfer
- N(S) and N(R) fields allow the error and sequence control to be obtained through a window protocol
  - N(S) = transmitted PDU sequence number
  - N(R) = Acknowledge number, refers to the expected PDU at the receiver
**LAP-B: supervision PDUs (S)**

- **ACK transfer**

- **RR** (Receiver Ready - C/R)
  - Positive ACK

- **RNR** (Receiver Not Ready - C/R)
  - Positive ACK and flow control signal sent from the receiver which is unavailable (ON-OFF flow control)

- **REJ** (Reject - C/R)
  - Request for retransmission of all PDU starting from N(R)

---

**LAP-B: unnumbered PDUs (U)**

- Mainly PDUs used to control the connection management

- 5 M bits, permit to define up to PDU. Only a limited number is used by LAP-B

- **Command PDUs:**
  - **SABM**(E) (Set Asynchronous Balanced Mode), used to (re)open the connection
    - E = Extended numbering scheme
  - **DISC** (Disconnect): the connection is aborted
LAP-B: unnumbered PDUs (U)

- **Response PDUs**
  - **UA** (Unnumbered Acknowledgment):
    - ACK for initializing PDUs or to answer to DISC commands
  - **DM** (Disconnect Mode)
    - Connection was not set up correctly
  - **FRMR** (FRaMe Reject)
    - Answer to the reception of a correct but unknown PDU
    - 24 additional bits to explain the reason why the PDU was rejected

**LAPB: command and response PDUs**

<table>
<thead>
<tr>
<th>format</th>
<th>command</th>
<th>response</th>
<th>code in control field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>I (Information)</td>
<td></td>
<td>0 N(S) P N(R)</td>
</tr>
<tr>
<td>Supervision</td>
<td>RR (Receiver Ready)</td>
<td>RR (Receiver Ready)</td>
<td>1 0 0 0 P/F N(R)</td>
</tr>
<tr>
<td></td>
<td>RNR (Rec. Not Ready)</td>
<td>RNR (Rec. Not Ready)</td>
<td>1 0 1 0 P/F N(R)</td>
</tr>
<tr>
<td></td>
<td>REJ (Reject)</td>
<td>REJ (Reject)</td>
<td>1 0 0 1 P/F N(R)</td>
</tr>
<tr>
<td>Unnumbered</td>
<td>SABM (Set Asynchr. Balanced Mode)</td>
<td></td>
<td>1 1 1 1 P 1 0 0</td>
</tr>
<tr>
<td></td>
<td>DISC (Disconnect)</td>
<td></td>
<td>1 1 0 0 P 0 1 0</td>
</tr>
<tr>
<td></td>
<td>DM (Disconnect Mode)</td>
<td></td>
<td>1 1 1 1 F 0 0 0</td>
</tr>
<tr>
<td></td>
<td>UA (Unnumbered Acknowledgement)</td>
<td></td>
<td>1 1 0 0 F 1 1 0</td>
</tr>
<tr>
<td></td>
<td>FRMR (Frame Reject)</td>
<td></td>
<td>1 1 1 0 F 0 0 1</td>
</tr>
</tbody>
</table>
LAPB: Poll/Final Bit

- In *command* PDUs
  - the P/F bit is used to poll stations (i.e. to require an answer) when set to 1
- In *response* PDUs
  - the P/F bit is used to answer (final) to command PDUs with a P/F bit set to 1
- Poll bit set to 1 by DTE (or DCE) is a request to answer for DCE (or DTE).
- Final bit set to 1 specifies the answer to the poll request
- DTE and DCE cannot send a command PDU with the P bit set to 1 unless a response PDU with the F bit set to 1 was received

LAPB addresses

- DTE has address 00000011 (3)
- DCE has address 00000001 (1)
- 3 is the value of the address field in
  - *command* PDUs: DCE ⇒ DTE and in
  - *response* PDUs: DTE ⇒ DCE
- 1 is the value of the address field in
  - *command* PDUs: DTE ⇒ DCE and in
  - *response* PDUs: DCE ⇒ DTE
- The address field permits to distinguish *command* PDUs from *response* PDUs and to understand whether the P/F bit is a *poll* or a *final*
Frame relay

Telecommunication Network Group
firstname.lastname@polito.it
http://www.telematica.polito.it/

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
- Physical speeds 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
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

Logical topology design

• 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
Physical topology

Logical topology
Frame Relay

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

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
Core and edge approach

L >= 3
L2
L1

Higher layer protocols
DL-CONTROL

DL-CORE
Physical

Frame Relay switching node

Higher layer protocols
DL-CONTROL

DL-CORE
Physical

User terminal

User terminal

Error control

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)
LAPF packet

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

**Network Management and QoS Provisioning - 79**

Andrea Bianco – TNG group - Politecnico di Torino

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

**Network Management and QoS Provisioning - 80**

Andrea Bianco – TNG group - Politecnico di Torino
B-ISDN

- Private and public networks
- Integrated network
  - Support all type of services, with different transmission speeds and quality of service requirements over the same network infrastructure
- Standardized by ITU-T and ATM Forum
- ISDN (re)evolution
- Exploit ATM as a transport, multiplexing and switching technique
### ITU-T definition

- 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.
- A B-ISDN will contain intelligent capabilities for the purpose of providing advanced service characteristics, supporting powerful operation and maintenance tools, network control and management.

### B-ISDN requirements

- **Broadband network**
  - Large area coverage
  - Large number of users
  - High speed
- **Integrated network**
  - Heterogeneous traffic over a single infrastructure
- **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
ATM: packet switching with virtual circuit service

- Three phases:
  - Virtual circuit opening
  - Data transfer
  - Virtual circuit closure

- Data sent over a connection from the same source to the same destination follow the same path ⇒
  - Routing (which is the best path?) performed only when opening the circuit
  - Data sequence guaranteed by the network
  - Smaller variability in delays
ATM: switching

- Switching implies network resource allocation to virtual circuits
  - At least labels are assigned
- ATM switching
  - Spatial switching
  - Label swapping (virtual circuit identifiers have local meaning)
  - Time switching
**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

**STM transport**

- Slot
- Frame
- Time
- Free slot
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 explicitly identify the connection ⇒ label switching
  - Statistical multiplexing
  - Suited to both circuit and packet services
B-ISDN: Protocol architecture

Telecommunication Network Group
firstname.lastname@polito.it
http://www.telematica.polito.it/

B-ISDN: reference model

Management plane

Control plane

User plane

Higher layers

Higher layer

AAL (ATM Adaptation Layer)

ATM layer

Physical layer
Opening a SVC

PM = plane management
LM = layer management

Andrea Bianco – TNG group - Politecnico di Torino
Opening a SVC

PM = plane management
LM = layer management
Opening a SVC

PM = plane management
LM = layer management

Andrea Bianco – TNG group - Politecnico di Torino
Network Management and QoS Provisioning - 99

Opening a SVC

PM = plane management
LM = layer management

Andrea Bianco – TNG group - Politecnico di Torino
Network Management and QoS Provisioning - 100
Opening a SVC

PM = plane management
LM = layer management
Core and edge approach in the user plane

Higher layers protocols

\[ \begin{array}{c}
L_1 \\
\text{ATM} \\
\text{Physical}
\end{array} \]

Error detection only on-demand for some AAL

Higher layers protocols

\[ \begin{array}{c}
L_2 \\
\text{ATM} \\
\text{Physical}
\end{array} \]

\[ \begin{array}{c}
L_3 \\
\text{AAL}
\end{array} \]

User terminal

ATM switching node

User terminal

B-ISDN: reference model

Management plane

Control plane

User plane

Higher layers

Higher layer

AAL (ATM Adaptation Layer)

ATM layer

Physical layer
ATM protocol layer

- Main functions:
  - Switching
  - Cell multiplexing
- Rate adaptation between physical layer and AAL layer
- Connection management through OAM and RM cells

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
- Slightly different format at network edge and core
B-ISDN interfaces

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

ATM cell format

DATA

UNI CELL

DATA

NNI CELL
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 format

- 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.
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

<table>
<thead>
<tr>
<th>PT</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>User cell EFCI No congestion AAL 5 indication=0</td>
</tr>
<tr>
<td>0 0 1</td>
<td>User cell EFCI No congestion AAL 5 indication=1</td>
</tr>
<tr>
<td>0 1 0</td>
<td>User cell EFCI Congestion AAL 5 indication=0</td>
</tr>
<tr>
<td>0 1 1</td>
<td>User cell EFCI Congestion AAL 5 indication=1</td>
</tr>
</tbody>
</table>
Campos PT (Payload Type)

<table>
<thead>
<tr>
<th>PT</th>
<th>SIGNIFICATO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0 0</td>
<td>OAM cell (Operation and Maintenance)</td>
</tr>
<tr>
<td>1 1 0</td>
<td>RM cell (Resource Management)</td>
</tr>
<tr>
<td>1 1 1</td>
<td>Not used Reserved for future use</td>
</tr>
</tbody>
</table>

**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
ATM cell format

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

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
ATM layer function

- Performance monitoring
  - Delay management
  - CLP bit management
    - Selective discarding
  - User parameter control
  - ECN (Explicit Congestion Notification)
  - Cell type discrimination
    - User, OAM, Control

B-ISDN: reference model

- Management plane
- Control plane
- User plane
- Higher layers
  - AAL (ATM Adaptation Layer)
- Higher layer
- ATM layer
- Physical layer
AAL: ATM Adaptation Layer

- Integrates ATM transport to offer service to users
- Service dependent layer
- Examples of AAL functions:
  - Transmission errors detection and management
  - Segmentation and reassembly
  - Cell loss management
  - Flow control
  - Synchronization

AAL: ATM Adaptation Layer

- It defines four classes of service (service classes)
  - Through three main parameters:
    - Source transmission speed
    - Type of connection (connection oriented/connectionless)
    - Temporal relation between end user
AAL: 4 service classes

- A: CBR traffic, constant but rate, connection oriented, synchronism required ⇒ AAL 1
- B: VBR traffic, connection oriented, synchronism required ⇒ AAL 2
- C: VBR traffic, connection oriented, synchronism not required ⇒ AAL 3/4
- D: VBR traffic, connectionless, synchronism not required ⇒ AAL 5

### AAL service classes

<table>
<thead>
<tr>
<th></th>
<th>Class A</th>
<th>Class B</th>
<th>Class C</th>
<th>Class D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchronism</td>
<td>required</td>
<td>not required</td>
<td></td>
<td></td>
</tr>
<tr>
<td>required between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>source and dest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>constant (CBR)</td>
<td>variable (VBR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connection type</td>
<td>Connection oriented</td>
<td></td>
<td>connection less</td>
<td></td>
</tr>
<tr>
<td>AAL type</td>
<td>AAL 1</td>
<td>AAL 2</td>
<td>AAL 3/4 - 5</td>
<td></td>
</tr>
<tr>
<td>Possible applications</td>
<td>voice 64kbit/s video CBR</td>
<td>video/audio VBR</td>
<td>data</td>
<td>data</td>
</tr>
</tbody>
</table>
AAL layer: architecture

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

AAL architecture

- CS convergence sublayer
- SAR segmentation and reassembly
- SSCS service specific CS
- CPCS common part CS
- Some sub-layers can be empty
### Formato dati AAL

<table>
<thead>
<tr>
<th>AAL 1</th>
<th>ATM Cell Header</th>
<th>SN</th>
<th>SNP</th>
<th>SAR - SDU</th>
<th>47 byte</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAL 2</td>
<td>ATM Cell Header</td>
<td>SN</td>
<td>IT</td>
<td>SAR - SDU</td>
<td>LI</td>
</tr>
<tr>
<td>AAL 3</td>
<td>ATM Cell Header</td>
<td>ST</td>
<td>SN</td>
<td>RES</td>
<td>SAR - SDU</td>
</tr>
<tr>
<td>AAL 4</td>
<td>ATM Cell Header</td>
<td>ST</td>
<td>SN</td>
<td>MID</td>
<td>SAR - SDU</td>
</tr>
<tr>
<td>AAL 5</td>
<td>ATM Cell Header</td>
<td>ST</td>
<td>SN</td>
<td>MID</td>
<td>SAR - SDU</td>
</tr>
</tbody>
</table>

#### AAL 1

- Convergence Sublayer
  - Packetization
  - Adaptive source clock recovery
  - Timing information transfer
- SAR sublayer
  - Sequence counter (modulo 8)
  - Counter error recovery
  - Cell loss notification
### CS AAL 3/4 header

- **CPI (Common Part Indicator):** unit of measure for Length e BA size (up to now, only bytes admitted)
- **BTag e ETag:** CS PDU delimitator
  - Assume the same value (BTag=ETag)
- **BA (Buffer Allocation) size:** buffer to be allocated at the receiver
- **PAD:** padding field, to align the PDU size to a multiple of 4 byte
- **AL:** alignment byte
- **Length:** PDU length measured according to the CPI field
AAL 3/4 SAR header

- **ST (Segment Type):**
  - BOM (Begin of Message), COM (Continuation), EOM (End), SSM (Single Segment)
- **SN (Sequence Number):** increasing number
- **LI (Length Indicator):** PDU length (in byte)
  - Equal to 44 for BOM, SSM and COM cells
- **MID (Multiplexing Identifier):** multiplexing management
- **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 3/4 CS function

- Mapping (between VC and AAL-SAP)
- AAL SDU Blocking / deblocking or segmentation/reassembly
- Error control over CS-PDU, with retransmission in class C
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

- 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
LANs (Ethernet) : Protocol architecture

Telecommunication Network Group
firstname.lastname@polito.it
http://www.telematica.polito.it/

LANs
• Small geographical extension
• Shared transmission medium (originally) ⇒ 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
    • Need to use address to identify node for unicast traffic
• Many topologies
  – bus, ring, star
### Possible solutions for medium access

- **Static channel division**
  - Fixed assignment of portion of channels
    - Time Division
    - Frequency Division
    - Code Division

- **Not suited to bursty traffic**
  - (N queues and servers at speed C are worse than 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**

---

### 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
Random access protocols

• Free access
  – Each node send at the channel speed $R$
  – No coordination among nodes
• If two concurrent transmissions $\Rightarrow$ 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

CSMA: Carrier Sense Multiple Access

• Sense the channel before transmission
  – If the channel is $\textbf{sensed}$ 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
CSMA: collisions?

- Collisions 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

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 disabled)
CSMA/CD: performance

- Throughput performance strongly depend on the 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)

Random access protocols performance
Network Management and QoS provisioning—Class intro and review

## Standard IEEE 802

<table>
<thead>
<tr>
<th>802.1 ARCHITECTURE</th>
<th>802.2 LOGICAL LINK CONTROL</th>
<th>802.3 MEDIUM ACCESS</th>
<th>802.4 MEDIUM ACCESS</th>
<th>802.5 MEDIUM ACCESS</th>
<th>802.6 MEDIUM ACCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.1 INTERNETWORKING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>802.2 LOGICAL LINK CONTROL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other committees:
- 802.7: Broadband Technical Advisory Group
- 802.8: Fiber-Optic Technical Advisory Group
- 802.9: Integrated Data and Voice Networks
- 802.10: Network Security
- 802.11: Wireless Networks
- 802.12: 100 base VG
- 802.13: 100 base X
- 802.15: Bluetooth
- 802.17: Resilient Packet Ring
**LLC Addresses**

- Enable higher layer protocol multiplexing

![Diagram showing LLC Addresses]

- Identify each NIC (Network Interface Card) on a local area network

![Diagram showing MAC Address]
MAC Address

• 6 byte
• Available on ROM in the card
  – Originally, established by the card producer
  – Today, partly configurable
• Two parts:
  – Most significant 3 bytes: assigned to each NIC producer (Organization Unique Id)
  – Less significant 3 bytes: progressive card number
  – E.g.: 02-60-8C-07-9A-4D is a 3com NIC

MAC Address

• 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
• Two types of multicast:
  – Solicitation: request a service to a multicast group
  – Advertisement: periodic diffusion of info related to membership to a multicast group
Indirizzi MAC

- 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
- Promiscuous mode bypass any control

Ethernet and IEEE 802.3

- 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
  - Ethenernet and 802.3 NICs co-exist in the same LAN
- Protocol
  - CSMA/CD 1 persistent
  - No ACK is sent to confirm packet reception
**Ethernet: packet format**

<table>
<thead>
<tr>
<th>Component</th>
<th>Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble = 101010…</td>
<td>7</td>
</tr>
<tr>
<td>SFD = 10101011</td>
<td>1</td>
</tr>
<tr>
<td>MAC Destination Address</td>
<td>6</td>
</tr>
<tr>
<td>MAC Source Address</td>
<td>6</td>
</tr>
<tr>
<td>Protocol type &gt; 1500</td>
<td>2</td>
</tr>
<tr>
<td>DATA</td>
<td>46 - 1500</td>
</tr>
<tr>
<td>FCS</td>
<td>4</td>
</tr>
<tr>
<td>Inter Packet GAP (silence)</td>
<td>12 bytes time</td>
</tr>
</tbody>
</table>

**IEEE 802.3: packet format**

<table>
<thead>
<tr>
<th>Component</th>
<th>Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preamble = 101010…</td>
<td>7</td>
</tr>
<tr>
<td>SFD = 10101011</td>
<td>1</td>
</tr>
<tr>
<td>MAC Destination address</td>
<td>6</td>
</tr>
<tr>
<td>MAC Source address</td>
<td>6</td>
</tr>
<tr>
<td>Length (&lt;1500)</td>
<td>2</td>
</tr>
<tr>
<td>DATA</td>
<td>0 - 1500</td>
</tr>
<tr>
<td>Padding</td>
<td>0-46</td>
</tr>
<tr>
<td>FCS</td>
<td>4</td>
</tr>
<tr>
<td>Inter Packet GAP (silence)</td>
<td>12 bytes time</td>
</tr>
</tbody>
</table>
LAN Interconnection

• Needed to
  – Extend LAN physical size
  – Increase the number of access nodes
  – No need to modify protocol architecture

• May increase LAN throughput performance
  – More space diversity
  – Exploits traffic locality

Interconnecting devices

• Repeater or Hub (layer 1)
  – Not an interconnecting device
  – Permit to extend cable lengths

• 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
Repeater or Hub

- Multi-port device
- Operates at the bit level
- Extend the cable length
  - No increase in network capacity
- Regenerates strings of bits and forwards them on all the ports
- Shared bandwidth on all ports
- 3R: re-generation, re-shaping, re-timing
  - May introduce delay
- Repeaters
  - On coaxial cable
  - Tree-like topology
- Hubs
  - On twisted-pair
  - Star-based topology
Bridge or Switch

- Store and forward devices
- Dedicated bandwidth per port
- Transparent to users
- Do not modify packet content
- Limited routing capability
- Bridge
  - Operates on coaxial cable
  - Interconnect LANs
- Switches
  - Operates on twisted pair
  - Interconnect LANs or single users
  - Support VLANs
Bridge/Switch

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

![Bridge/Switch Diagram]

Bridge/switch operations

- Each bridge/switch has a unique ID
- Each bridge/switch port has a unique id
- Three fundamentals 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
**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

---

**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
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

Backward learning over a loop
**Backward learning over a loop**

- Q transmits to X
  - B1 and B2 receive the packet and assume that Q can be reached using port B
- If B1 and B2 have the MAC address of X in the forwarding table
  - B1 sends the packet on port A
    - B2 assumes that Q can be reached using port A (true, but via a loop)
  - B2 sends the packet on port A
    - B1 assumes that Q can be reached using port A
- Thus
  - X receives two copies of the packet
  - B1 and B2 are unable to reach Q

---

**Backward learning over a loop**

- Q sends to X
  - B1 and B2 receive the packet and assume that Q can be reached using port B
- If the MAC address of X is NOT found in the forwarding tables
  - B1 sends the packet on port A
    - B2 assumes that Q can be reached using port A (true, but via a loop)
  - B2 sends the packet on port A
    - B2 assumes that Q can be reached using port A (true, but via a loop)
- B1 and B2 keep sending packets forever
Bridge/Switch properties

- From a multiple-access network to a multiplexed network
  - Reduce collision probability by partitioning the network in independent segments
- Throughput performance may increase
  - Space diversity
  - Exploits traffic locality
- Security mechanisms
  - Traffic separation
- Introduce store and forward (and queueing) delays
  - Worse delays than hubs
- Potential packet losses when queues are filled-up
- Unfairness in resource access

VLAN (Virtual LAN)

- Hosts 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

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Packet delimitation</th>
<th>Layer 3 protocol multiplexing</th>
<th>Error detection</th>
<th>Error correction (window protocol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAPB + Layer 3</td>
<td>Flag</td>
<td>Through VC at layer 3</td>
<td>YES in both layers</td>
<td>Yes in both layers</td>
</tr>
<tr>
<td>LAPF core + LAPF control</td>
<td>Flag</td>
<td>Through VC</td>
<td>YES in LAPF core</td>
<td>Optional in LAP-F control (edge)</td>
</tr>
<tr>
<td>ATM (core)+ AAL (edge)</td>
<td>Through physical layer</td>
<td>Through VC (edge)</td>
<td>YES in AAL (edge)</td>
<td>NO</td>
</tr>
<tr>
<td>Ethernet MAC</td>
<td>Silence</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
</tbody>
</table>
Internet

Gruppo Reti TLC
nome.cognome@polito.it
http://www.tlc-networks.polito.it/

Internet protocol suite

<table>
<thead>
<tr>
<th>OSI</th>
<th>Internet Protocol Suite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Telnet</td>
</tr>
<tr>
<td></td>
<td>FTP</td>
</tr>
<tr>
<td></td>
<td>SMTP</td>
</tr>
<tr>
<td></td>
<td>SNMP</td>
</tr>
<tr>
<td>Presentation</td>
<td>TCP e/o UDP</td>
</tr>
<tr>
<td></td>
<td>ICMP</td>
</tr>
<tr>
<td></td>
<td>IP</td>
</tr>
<tr>
<td>Session</td>
<td>Protocolli di routing</td>
</tr>
<tr>
<td>Transport</td>
<td>ARP</td>
</tr>
<tr>
<td>Network</td>
<td>Non Specificati</td>
</tr>
<tr>
<td>Data Link</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td></td>
</tr>
</tbody>
</table>

Andrea Bianco – TNG group - Politecnico di Torino
Network Management and QoS Provisioning - 179

Pag. 90
IP: Internet Protocol

• Layer 3 protocol
• Defines
  – Packet format
  – Address format
  – Data (named datagram) forwarding procedures
• Best-effort service
  – connectionless
  – unreliable
  – With no QoS guarantees
• Specified in RFC 791 (November 1981)

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
**IP protocol: unreliable delivery**

- In case of:
  - Failure (ex. out of service router, link failure)
    - Datagram dropped and 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

--

**IP packet header**

<table>
<thead>
<tr>
<th>0</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>19</th>
<th>24</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>HLEN</td>
<td>Service Type</td>
<td>Total Length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td>Flags</td>
<td>Fragment Offset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time To Live</td>
<td>Protocol</td>
<td>Header Checksum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source IP Address</td>
<td>Destination IP Address</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Options</td>
<td>PAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard size: 20 byte
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 header as the original datagram (different fields: fragmentation fields (identification, flags, offset), length, CRC)
  - Reassembled 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; increases the loss probability
  - Creates “useless” traffic
    - fragments belonging to a datagram for which at least a fragment was lost are transported with no use
  - Reassembly timers are needed at the receiver
- Reassembly normally done at network edge (hosts, not routers)

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 sends an ICMP error message to the source (can be disabled)

- **Protocol**: higher layer protocol code. RFC 1700 lists the protocol codes

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ICMP</td>
</tr>
<tr>
<td>4</td>
<td>IP in IP</td>
</tr>
<tr>
<td>6</td>
<td>TCP</td>
</tr>
<tr>
<td>17</td>
<td>UDP</td>
</tr>
<tr>
<td>89</td>
<td>OSPF</td>
</tr>
</tbody>
</table>

- **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 over 16 bits
  - The header checksum can be computed incrementally (useful since each router decrements the TTL field and must re-compute the header).

- **Source and 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
**IP header fields: options**

- **Options format:**
  - option code (option number, option class, copy flag for fragmentation) + option length + data
- **Options**
  - record route: datagram path recorded
  - source route (loose and strict): source specifies datagram path
  - timestamp: 32-bit timestamp of host and routers dealing with the datagram

**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

ISP A

"Send me any datagram with address starting with 199.31.0.0/16 or 200.23.18.0/23"

ISP B

"Send me any datagram with address starting with 200.23.16.0/20"

ISP A

"Send me any datagram with address starting with 200.23.18.0/23"

ISP B

"Send me any datagram with address starting with 200.23.16.0/20"

ISP A

"Send me any datagram with address starting with 199.31.0.0/16 or 200.23.18.0/23"

ISP A

"Send me any datagram with address starting with 200.23.18.0/23"

ISP B

"Send me any datagram with address starting with 200.23.16.0/20"

ISP A

"Send me any datagram with address starting with 199.31.0.0/16 or 200.23.18.0/23"
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

- 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
Internet transport layer

- Two alternative protocols: TCP and 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, unreliable, stateless, it supports multicast traffic
- Common characteristics:
  - Multiplexing and demultiplexing of application processes through the port mechanism
  - Error detection over header and data (optional in UDP)

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 processes are statically associated to well-known 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
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

UDP: packet format

<table>
<thead>
<tr>
<th>0</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>19</th>
<th>24</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP Source Port</td>
<td>UDP Destination Port</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UDP Message Length</td>
<td>UDP Checksum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DATA
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 there is no need to open a 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 retransmission 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, 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 identified by port 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

<table>
<thead>
<tr>
<th>Bit</th>
<th>Source Port</th>
<th>Destination Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-32</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequence Number</td>
<td>Acknowledgment Number</td>
</tr>
<tr>
<td></td>
<td>HLEN</td>
<td>Resv</td>
</tr>
<tr>
<td></td>
<td>Checksum</td>
<td>Urgent Pointer</td>
</tr>
<tr>
<td></td>
<td>Options</td>
<td>Padding</td>
</tr>
</tbody>
</table>
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

Identify the application processes sending and receiving data
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)

<table>
<thead>
<tr>
<th>Capacity</th>
<th>First wrap around time</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 (1.5Mbps)</td>
<td>6.4 hours</td>
</tr>
<tr>
<td>Ethernet (10Mbps)</td>
<td>57 minutes</td>
</tr>
<tr>
<td>T3 (45Mbps)</td>
<td>13 minutes</td>
</tr>
<tr>
<td>FDDI (100Mbps)</td>
<td>6 minutes</td>
</tr>
<tr>
<td>STS-3 (155Mbps)</td>
<td>4 minutes</td>
</tr>
<tr>
<td>STS-12 (622Mbps)</td>
<td>55 seconds</td>
</tr>
<tr>
<td>STS-48 (2.5Gbps)</td>
<td>14 seconds</td>
</tr>
</tbody>
</table>

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 connection startup)
TCP header

0  15  31
Source Port Number  Destination Port Number

Sequence Number

Acknowledgment Number

Resv  flags  Receiver window

check sum  Urgent Pointer

Header length in 32 bit words (default value 5)

Reserved for future use (ECN)
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

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Port Number</td>
<td>Destination. Port Number</td>
</tr>
<tr>
<td>Sequence Number</td>
<td>Acknowledgment Number</td>
</tr>
<tr>
<td>HLEN</td>
<td>Resv.</td>
</tr>
<tr>
<td>Receiver window</td>
<td>checksum</td>
</tr>
<tr>
<td>Urgent Pointer</td>
<td>Amount of data (in bytes) the receiver is willing to store (flow control) Maximum value 65535 byte, unless the window scaling option is used</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Bit rate</th>
<th>Bit rate x delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 (1.5Mbps)</td>
<td>18KB</td>
</tr>
<tr>
<td>Ethernet (10Mbps)</td>
<td>122KB</td>
</tr>
<tr>
<td>T3 (45Mbps)</td>
<td>549KB</td>
</tr>
<tr>
<td>FDDI (100Mbps)</td>
<td>1.2MB</td>
</tr>
<tr>
<td>STS-3 (155Mbps)</td>
<td>1.8MB</td>
</tr>
<tr>
<td>STS-12 (622Mbps)</td>
<td>7.4MB</td>
</tr>
<tr>
<td>STS-48 (2.5Gbps)</td>
<td>29.6MB</td>
</tr>
</tbody>
</table>

- Can be overcome with the window scale option

TCP header

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