Multiplexing example

Offered load: 3 sources with total load of 8 units/second

- 4/sec
- 3/sec
- 1/sec

Separate circuits: 4 units per second each (capacity of 12 units/sec)

- 4/sec
- 3/sec
- 1/sec

Multiplexed on one channel (capacity of 8 units/sec)

When arrivals are deterministic, we can time-division multiplex in a pre-determined (circuit-switched) manner. But arrivals are likely to be bursty.

Statistical multiplexing

- Bursty network traffic ⇒
  - Dedicated channel may often be wasted
  - Circuit switching may be inefficient
    - (may not be important for optical links)
  - Non-dedicated channel may occasionally be overloaded
  - Packet-switches need buffering and may drop packets due to congestion.

Statistical Multiplexing

Observations
1. The bigger the buffer, the lower the packet loss.
2. If the buffer never goes empty, the outgoing line is busy 100% of the time.

Duration/volume of traffic flow

- Transactions tend to be short: request/response
- Interactions tend to be relatively long
- Duration of transfer determines significance of signaling overheads
  - e.g. 80ms round-trip time (RTT) insignificant compared to voice call of minutes, but appreciable compared to transaction of single packet
- Often can’t predict duration/volume of transfer before it happens
  - Important exception: file transfers ⇒ scheduling
**Outline**

Initiation

**Transfer initiation:**

*Pull vs push access*

- **Pull:** The receiving end initiates the transfer  
  e.g. Web browsing: small/minor transfer to server (URL)  
  large/important transfer to client (page+images)

- **Push:** The source end initiates the transfer  
  e.g. email

*Alternative: initiation by a third party*

Has implications for:

- Signalling between switches:
  - Pushing reduces number of passes of signalling messages,
  - But is difficult when multicasting.

- Caching (and hence link dimensioning): Caches are useful only when pulling.

**Transmission unit length**

No universal optimal length:

- **Transmission overhead:** Large is good: Packet overheads (address, interrupts) are rarer

- **Retransmission overhead:** Small is good: Small payload requires much padding for large fixed-length packet

- **Small can reduce delays:**
  - Serialisation delay: Time to fill a packet from a serial source (e.g. PCM voice)
  - Multiplexing delay: Time to wait for another access to end before gaining access.
  - Store-and-forward delay: Time to receive a whole packet before being able to start forwarding it.
  - Playout delay: Depends on jitter, which depends on multiplexing delay.

**Fixed or variable length transmission units?**

Benefits of fixed length units (segments, cells, slots):

- **Simpler,** e.g. buffer management
- **Predictable,** e.g. time when next packet will arrive
- **More secure:** packet length doesn’t leak information
- **Fixed processing time helps concurrency,** pipelining

Internal switch fabric can use fixed length units

While external line interfaces use variable-length units

With port processors converting between the two through segmentation & reassembly

To Be Continued in discussion about ATM...
Outline

“Quality of Service” (QoS)

Anything that is desirable contributes to quality.

Service Level Agreements (SLAs) generally encompass measures such as timeliness, reliability, availability (discussed shortly).

Users often want “guarantees” of service ⇒ "QoS" often discussed in contrast to "best effort" alternative.

"best effort" service = "I’ll do my best but make no guarantees"" best effort" service better than other service.

QoS in the research community vs deployment:

• Work has continued for decades – Practical problem: all systems on the end-to-end path must support it; any intermediary can disrupt the QOS

• ISPs don’t like QoS – it costs them money to install the equipment, and they can’t get the returns – goes to content providers (ISPs are commodity dealers)

QoS mechanisms only matter when resources are scarce. Can often avoid fancy mechanisms by over-dimensioning system.

Timeliness requirements: 1st pass

Superficial dichotomy:

• “real-time”: Interactive systems
  • Continuous-time signals, e.g. voice and video (though only if ends interact “live” e.g. videoconference; tolerate delays for movie playback)
  • Other interactive systems: games, stock ticker, process control

• “data”: file transfers, email, web browsing
  • The protocols used here have some interactivity – e.g. timeout after a certain period

“Real-time” requires QoS, whereas “best effort” is sufficient for data

Timeliness requirements in depth

Important aspects of delay:

• Mean - low for interaction

• Range/variation (CDV) - low for streaming media

† Cell Delay Variation – cell = short packet of fixed length
Reliability

6 aspects of reliability defined:

• **Integrity**: What is received is what was transmitted.
• **Completeness**: Everything transmitted is received.
• **Uniqueness**: Information is only received once.
• **Sequence**: Information is received in the correct order.
• **Relevance**: Information from extraneous sources is not inserted in the midst.
• **Delivery**: Source receives acknowledgement indicating that destination has received information.

Ties between switching and reliability

**Integrity**: Affects switch forwarding modes (prefer store-and-forward)

**Completeness**:
- Packet discard strategies to avoid unnecessary retransmissions.
- Loss priorities for layered video (baseband + enhancement)

**Uniqueness**: Negligible impact on switches.

**Sequence**:
- Switches may misorder when rearranging or recirculating traffic
- Deflection networks may misorder
- Preservation of sequence is a benefit of connection-oriented switching

**Relevance**: Protect addressing information in headers with CRCs

**Delivery**: Middleboxes may break end-to-end semantics (e.g. wireless Access Point may send TCP ack before segment received by end-system)

CRC = Cyclic Redundancy Code

Loss vs delay variation

[Diagram showing the relationship between loss and delay variation]

Fig. 5.2 from McDysan based on Fig. 3 of G. Woodruff and R. Kosupahom: 'Multimedia traffic management principles for guaranteed ATM network performance', *IEEE J. Sel. Areas in Comm.* 11(7):437-46, Apr. 1990

References in Keshav

Layering: Chapter 5
Addressing: Chapter 10