Wahlpflichtfacht Multimediasysteme

Kapitel 4: Dienstleistungsgüte

Simon Schwantzer

schwantzer@icsy.de

University of Kaiserslautern
Integrated Communication Systems

http://www.icsy.de

Wahlpflichfach
Multimediasysteme
TU Kaiserslautern

Oktober 2009

Content

- Quality of Service Overview
- Scaling and Adaptation
- Resource Reservation
- Throughput
- Delay
- Round-Trip-Delay (RTT)
- Delay Variation
- Delay vs. Loss
- Reliability
- Error Detection and Error Correction

Quality of Service Overview

- Challenge: Media processing is (often) related to "real-time".
- Problem:
  - classical computer architectures, system software, and data networks were designed for batch processing
  - Enabling appropriate media processing requires that sufficient resources are available at specific points in time within all steps of media processing

Definition:
Quality of Service (QoS) denotes the well-defined and managable behaviour of a system according to measurable parameters.

- The implementation of QoS requires the availability of resource management!
- There are two major approaches for enabling QoS:
  - Scaling and Adaptation of media
  - Resource Reservation
- Note: several other definitions of QoS exist

Scaling and Adaptation

- Requires determination of available resources
  - automatic or manual
    - a priori measurements
    - monitor error rates during processing (within control loops)
    - manual selection of media quality by user or configuration
  - abilitiy to adapt to changing amount of resources
- Ability of Adaptation
  - range (high to medium or very high to very low quality?)
  - granularity (few classes of quality or fine grained adjustment?)
  - static (a priori) or dynamic (at runtime)
    in case of dynamic adaptation: spontaneous or negotiation required
  - type of resources considered by adaptation

Resource Reservation

- End-to-End reservation of resources:
  - amount of resources must be known for all types of critical resources
    - a priori reservation
    - adaptation of reserved resources
  - resource reservation steps
    - describe amount of required resources
    - signal/negotiate resource reservation
      - policy control (who is allowed?)
      - admission control (how much is available?)
      - acknowledge availability of resources
    - perform reservation and adjust/initialise "resource scheduler"
    - during resource usage
      - user: monitor provision of resources
      - provider: monitor / adjust resource usage

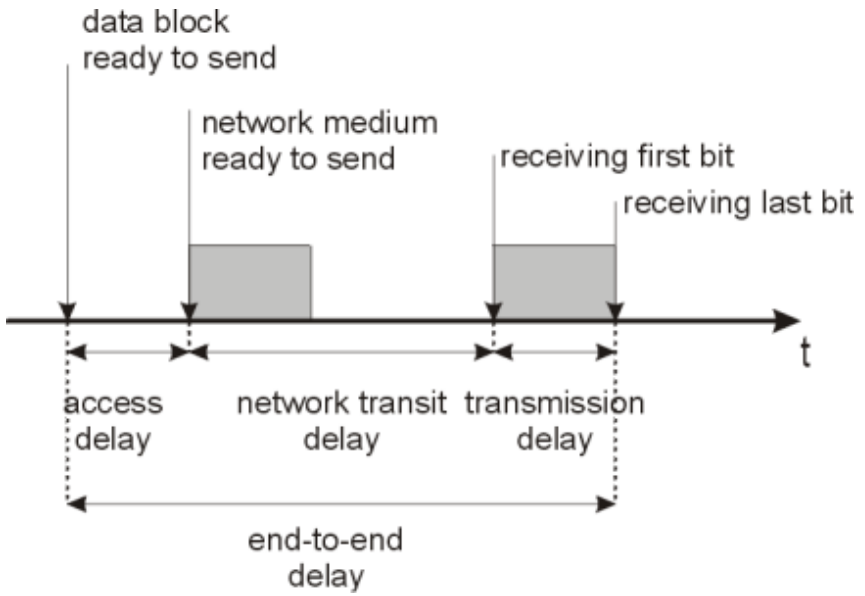
Throughput

The bit rate between two communication endpoints is the number of binary digits that the network is capable of delivering and accepting per time unit

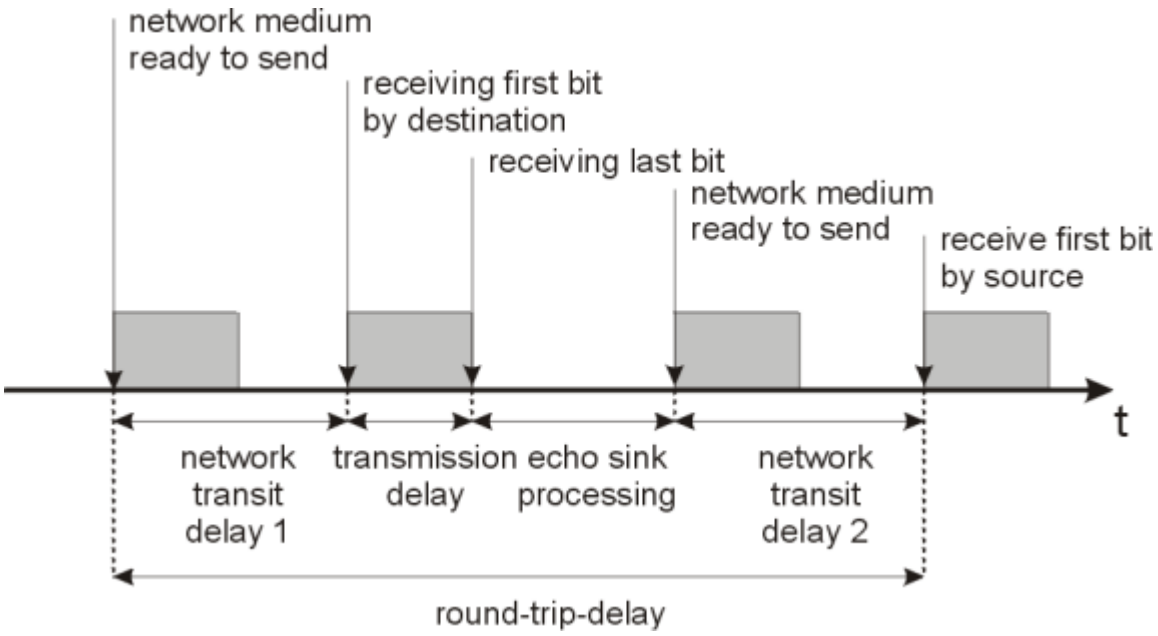
- The commonly used time unit is one second, but often the actually observed time unit is another one
  - ISDN offers a throughput of 8 bit per 125 µs
- distinguish individual and aggregate bit rates
- access speed versus bit rate
  - access speed refers to the frequency at which bits may be sent or received
- bit rate could be constant or variable
- burstiness
  - peak bit rate (PBR): maximum bit rate during a short time interval
  - mean bit rate (MBR): averaged bit rate over a longer time interval
  - burstiness is the relation between PBR and MBR

Delay

- Delay of a service object is the sum of the following delays:
  - **access delay**
    the time between the request to send a data block and the emission of the first bit
  - **(network) transit delay**
    the time between emission of the first bit of a data block and its reception at the destination system
    (network transit delay is also called **latency**)
  - **transmission delay**
    the time between emission of the first and the last bit of a data block



Round-Trip-Delay (RTT)



- The RTT is also called response time
- RTT is a good metric for interactive applications

Delay Variation

- **The Delay variation relates to variation of end-to-end delay**
  - also called "jitter"
  - distinguish average and maximum delay variation
- **Originator of delay variation**
  - the physical environment causes jitter in magnitude of nano seconds
  - intermediate switches/routers may cause jitter by
    - variation of medium access time, e.g. CSMA/CD (usually micro-seconds)
    - store and forward switching delays, e.g. by internal node congestion (micro-seconds)
    - waiting time caused by flow control mechanisms (micro- or even milli-seconds)
- In ISDN there is only a physical jitter, whereby in IP networks all types of jitter occur.

Delay vs. Loss

Reliability

- Mechanisms
  - Error detection
  - Error handling
    - requires error detection
    - posteriori, e.g. retransmission
    - a priori, e.g. forward error correction
- Error types
  - Data loss
  - Data alternation
  - Data duplication, miss insertion or wrong delivery
  - Failure of components usually not considered for multimedia systems

Error Detection and Error Correction

- Motivation:

  Many applications require that data is transmitted correctly, i.e. that data reaching its destination is the same than the data transmitted.

- Problem:
  - Channels are not an ideal medium
    - noise/interferences on the channel
    - losses on purpose (e.g. dropping packets in case of congestion)

Error Detection and Error Correction

- Error detection:
  - Means:
    - checksums (e.g. parity bits, CRC)
  - Not all conceivable errors can be detected!
- Error correction (if required):
  - Means:
    - retransmission of data (e.g. in TCP)
    - employing coding theory
  - Not all conceivable errors can be corrected!
- Terms:
  - EDC: Error Detection Code
  - ECC: Error Correction Code
  - FEC: Forward Error Correction

Questions?

Contact: Simon Schwantzer, schwantzer@informatik.uni-kl.de, http://www.icsy.de