

# Applications and TCP

Markus Peuhkuri

2002-10-17

## Lecture Topics

- What network requirements
- Requirements by application
- Requirements for elastic flows (or look-alike)

## Definition of properties

**Throughput**<sup>1</sup> is the number of binary digits that a system accepts and delivers per unit of time. Traditionally throughput at network level is expressed as bits per second and "kilo" equals 1,000. At application level also bytes or octets per second is used and "kilo" equals *usually* 1024. To avoid vagueness, one should use new SI prefixes of kibi, mebi and gibi when multiplicative prefixes of  $2^N \times 10$  are used.

**Delay**<sup>2</sup> (transit-) is the time elapsing between the emission of the first bit of data block by transmitting system and its reception by the receiving system. Other delays include:

**access delay** is time to wait when data is ready before transmission may start

**transmission delay** or *serialisation* delay is time needed to transmit all of the data block

**Delay variation** expresses how much delay of successive data units vary. Short timescale variation is called jitter<sup>3</sup>.

**Error rate** is count of errors per transmitted data or per time unit. (*PER*: packet error rate) Possible errors are:

**data alteration** usually referred as bit errors. Rare in optical networks but may take place inside networking equipment because of faulty hardware or software. More frequent in wireless networks. (*BER*: bit error rate)

**data loss** is the most common error in packet networks when a packet is discarded because of congestion. It is also possible that intermediate system discards data unit because of data alteration. (*PLR*: packet loss rate)

**data duplication** occurs mostly because of transitory malfunction of network equipment

## Network building blocks

- End systems (hosts)
  - servers
  - PCs, workstations
  - mobile devices
- Routers

---

<sup>1</sup>Läpäisy

<sup>2</sup>viive

<sup>3</sup>Värinä

- access routers: a large number of slow-speed interfaces and few high-speed interfaces. Extensive authentication, accounting and filtering facilities.
- core routers: possibly a large number of high-speed interfaces. Optimised for high-speed routing with large routing tables.
- middleboxes: application gateways / NAT-boxes, basically divide network into two different domains by limiting network visibility end to end.
- Links interconnecting routers and end systems
  - point-to-point: dedicated connection; serial lines, PCM/SDH pipes
  - shared media: fixed total capacity for all systems; "old" Ethernet and wireless LANs
  - overlay network: virtual network hiding physical topology; ATM, MPLS

## End system properties

**Throughput** is determined by speed of

- CPU
- internal busses: peripheral busses (PCI in PCs) and memory bus
- storage devices
- network interface

**Delay** same factors as for throughput, data processing . For example, many (audio) compression methods compress a block of data at time. The delay is length of data in addition to possible processing time. For example, an audio compression method may use 50 ms blocks.

**Delay variation** increases by system load

- interrupt latency
- processor sharing (scheduling)

**Error rate** is usually small

**data alteration** memory or DMA (Direct Memory Access) errors

**data loss** lack of buffer space or too slow processing of data

## Router properties

**Throughput** Modern routers can deliver data at full line speed. Using some advanced features not supported by forwarding hardware such as constrain based routing or accounting may drop performance by a factor.

**Delay** is dependent on traffic load and available buffer space. Small buffers result in small delay but high data loss rate.

**Delay variation** is caused by queue length variations. Short buffers result less delay variation.

**Error rate** is usually small

**data alteration** may happen if there is hardware error

**data loss** is the most common error as queues grow full

**data duplication** happens in routers because of transitory state in packet forwarding. For example a routing table may change just when a packet is being forwarded and the packet is transmitted to both old and new route.

## Link properties

**Throughput** is set by selected technology and depends on

- distance, especially in wireless networks
- number of hosts and their traffic demand, if media is multi-access

**Delay** depends much on distance

**transit delay** is set by distance and signal propagation speed. For optical cables one can use speed of 200,000 km/s (5 ms/1000 km).

**transmission delay** depends on link speed

**access delay** if shared media, ethernet, wireless

**Delay variation** is caused by variation in access delay. In Ethernet-style networks there is no upper bound; in token passing networks upper bound may be set.

Circuit-based networks do not have much of jitter (typically few nanoseconds or microseconds).

If link is overlay network, the delay variation may be much larger because of buffering in intermediate nodes (as in routers).

**Error rate** on links is caused by bit errors

**optical fibres**  $10^{-9}$ – $10^{-12}$

**satellite links**  $10^{-6}$ – $10^{-8}$

Overlay networks may also lose data if network nodes get congested.

## Other network components

- Middleboxes
  - common term for non-router data-processing network components
  - separates two network parts
  - may be in-line or call-out
  - may result loss of data or extensive delay
  - may keep state information
  - may improve network throughput

- Store-and-forward application relay
  - email servers: Message/mail transfer agents

Store-and-forward application relay differs from a router in sense that s-a-f takes responsibility to store the message until it is delivered to next s-a-f or recipient.

- Multicasting component
  - multicast router
  - document replicator
  - mail distributor: for example mailing list software
- Application translator
  - protocol translators: email, wap gateways
  - media transcoders
- Mirror and cache
  - server replicators
    - \* loose: Usenet news

- \* tight: server mirrors (for geographical distribution) For example Akamai <http://www.akamai.com/> provides service to replicate media content (pictures, video, audio) to multiple servers. The user is directed to closest server (think to be).

- document or media caches: for example web cache

The difference between mirror and cache is that mirror is replicated at time intervals (daily or so) while document is stored into cache when it is requested (on demand).

- Network Address Translator (NAT)
- SOCKS
- IP tunnel
- Load balancer
- Firewall (IP-level and application level)
- Network services
  - name service: DNS in Internet
  - directory service such as LDAP

## Application traffic characteristics

**Peak rate** is highest rate source will transmit data

**Burst size** is maximum amount of data source transmits at peak rate

**Average rate** calculated over time interval. Time interval may be order of few seconds or whole lifetime of connection

**Connection hold time** is time application is active

## Application traffic requirements

**Minimum bandwidth** needed to maintain application fidelity

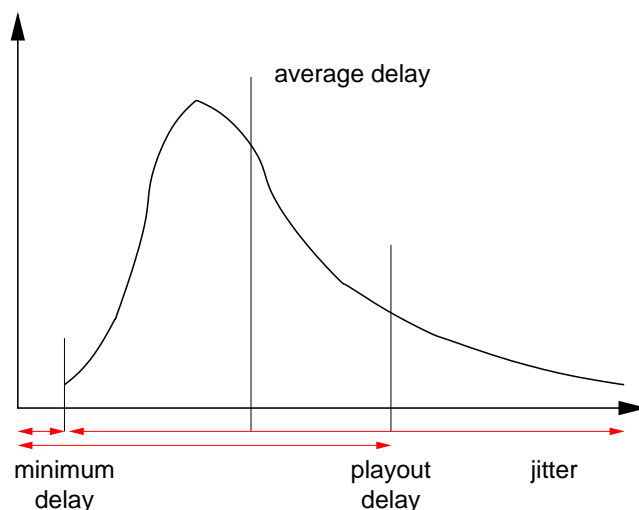
**Delay** application may tolerate. Transmission delay and queueing delay are part of bandwidth requirement.

**Delay variation** is difference between maximum and minimum delay

**Loss rate** allowed without fidelity degradation

## Delay distribution

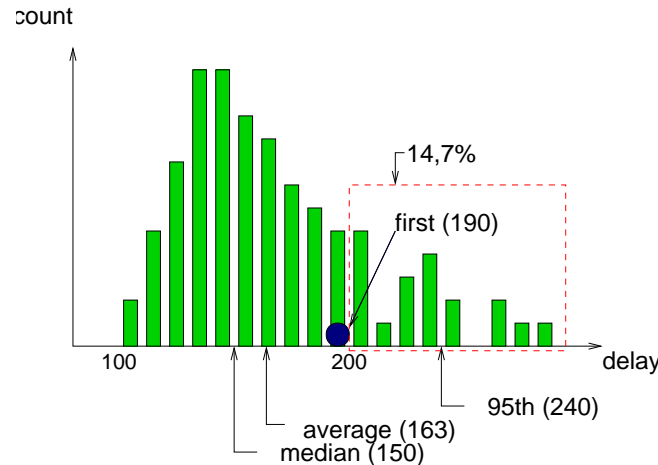
- Packets that arrive before playout delay are ok
- Late packets are “lost”



## Some ways to define delay distribution

These are from [3, Appendix II]

1. For each packet, variation is difference between delay of first packet and current packet delay
2. For each packet, variation is difference between average delay of population and current packet delay
3. Select delay interval in advance and count the proportion of packets that fall outside this interval
4. Distance between two quantiles (like 0.95 and 0.5)



- First packet delay is 190 ms, so variation for each packet  $i$  is  $|d_i - 190|$  ms.
- Average delay is 163 ms, variation is  $|d_i - 163|$
- If acceptable delay interval is 100–200 ms<sup>4</sup> then 14.7 % fall off this region.
- Median delay is 150 ms and 95th percentile is 240 ms so variation is 90 ms.

Different criteria are useful for different applications. For example, the third criteria is suitable for real-time communications if upper bound is set for playout delay. First two are suitable to measure instantaneous delay variation while the fourth could be for generic use.

## Different sources and receivers

- Source

**interacting:** maximum delay may not exceed one second, preferable round-trip delay less than 0.5 s

**live:** greater delay allowed but delay variation should not exceed playout buffer

**stored:** available bandwidth and delay may vary much because data can be sent well in advance. One must take account possible controls and handle those locally.

- Receiver

**human:** sensitive for delay, not much sensitive for transitory errors

- eye integrates information
- less tolerance for audio errors

**recorder:** not sensitive for delay, but prefers error-free transmission

---

<sup>4</sup>The shortest delay is more than 100 ms, so lower bound could be zero as well.

## Audio streams

**Telephone quality** 300–3400 Hz

- uncompressed 64 kbit/s ( $\text{BER} < 10^{-2}$ )
- compressed 4–16 kbit/s

**CD quality** 0–20 kHz

- uncompressed 1.4 Mbit/s
- compressed 128–384 kbit/s ( $\text{BER} < 10^{-4}$ )
- Echo cancelling needed if one-way delay over 24 ms
- Delay for interactive should be less than 100 ms to provide delay offset of 500 ms
- Higher requirements for virtual reality
- Inter-media synchronisation better than 100 ms

## Video streams

**HDTV** (1920x1080x60)

- uncompressed 2 Gbit/s
- compressed 20–34 Mbit/s

**Studio/Broadcast** (625x864x25)

- uncompressed 166 Mbit/s
- compressed 3–6 Mbit/s

**VCR** (625x864x25)

- compressed 1,2 Mbit/s

**Video conference** (352x288x5)

- compressed 112 kbit/s
- Delay requirements set by audio
- $\text{BER } 10^{-4}$ – $10^{-6}$
- Hierarchical coding for multicast delivery
- Compression adds burstiness up to ratio 10:1<sup>5</sup>

## Other application

- "Non-real-time" *do* have requirements to maintain their fidelity
- Web surfing: a page should load in less than 1 second and it must not take more than 10 seconds, otherwise user may get impatient and aborts transfer. [2, p. 135]
- Data transfer applications require minimum bandwidth
  - for example, a server may decide to abort connection if transfer takes too long time

---

<sup>5</sup>This depends on MPEG coding how group of pictures (GOP) defined i.e. how many P and B frames there are for a I frame.

## Interactive applications

- Interactive applications must follow 0.1/1/10 rule [2, p. 135]
  - 0.1 s** immediate feedback to user, to show user that action command (mouse click, key press, voice) was received.
  - 1 s** user's workflow is not interrupted if task completes within one second. If task is not completed, then there should be at least an indicator to show user that task is proceeding with some estimate of time to complete.
  - 10 s** user loses attention to task in question if task does not complete within 10 seconds. User either interrupts task or switches to another task.
- Acceptable performance needs
  - small delay
  - sufficient bandwidth

Transferring 7 KiB document within 1 s using HTTP

|                    | (a)       | (b)       |
|--------------------|-----------|-----------|
| minimum link speed | 64 kbit/s | 10 Mbit/s |
| 1-way delay        | 4,4 ms    | 198 ms    |

TCP connection is established but it is not closed. Maximum segment size of 1500 bytes is assumed and slow start with initial congestion window of two segments. End system processing time for each segment is assumed to be 0.1 ms in average including time to process and deliver HTTP request and response. No segments are lost.

The major difference between those two cases is caused by packet serialisation time which totals to 0.98 s in case (a) and 0.006 s in case (b). Average network utilisation is 61 kbit/s in both cases (as equal segments as transmitted).

- Interactive TCP applications do need some performance assurance
- Available bandwidth is the most important
- Delay should not vary too much or exceed some limit

## Partially interactive applications

- Application is controlled by a human operator, but transfer takes place in background
  - FTP transfers
  - email MUA-MTA (mail user agent – mail transfer agent) communications
- Much longer transfer times are tolerated
- Some upper limit for total transfer time, depends on application: few minutes for email, half an hour for FTP

## Non-interactive communications

- Communication between end hosts
- Each ongoing communication takes some resources from end system
- Application timeouts [1]
  - FTP** server command timeout of 5 minutes
  - SMTP** per command timeout 2–10 minutes (depending on command), per session timeout of an hour
  - DNS** default timeout 5 seconds with exponential back-off (up to one minute)
- There may be “user behind” waiting for email, for example
- Even longer times tolerated

## Can applications adapt (or should)

- Bandwidth less than expected
  - lower media speed using for example hierarchical coding
  - drop some media
- Losses
  - lower media speed to ease congestion
  - prepare for losses
    - \* FEC: Forward error correction
    - \* redundant information send copy of data with lower fidelity in next packet
- Delay variation too large
  - increase playout delay
    - ⇒ total delay increases
  - prepare to “loss”<sup>6</sup> some of packets

## Summary

- Each network component has its properties
  - throughput
  - delay with variation
  - error rate
- Components are
  - end systems
  - routers
  - links
- Every application has requirements

## References

- [1] R. Braden and Ed. Requirements for Internet Hosts - Application and Support. Request for Comments RFC 1123, Internet Engineering Task Force, October 1989. (Internet Standard) (Updates RFC0822) (Updated by RFC2181) (Also STD0003). URL:<http://www.ietf.org/rfc/rfc1123.txt>.
- [2] Jakob Nielsen. *Usability Engineering*. AP Professional, Boston, 1993.
- [3] Network performance objectives for ip-based services. ITU-T Recommendation Y.1541, International Telecommunication Union, 2002. Pre-published.

---

<sup>6</sup>A packet, which is late, can be handled as packet which is lost