University of Glasgow

Real-time and Interactive Applications

Networked Systems (H)

Lecture 7

Lecture Outline

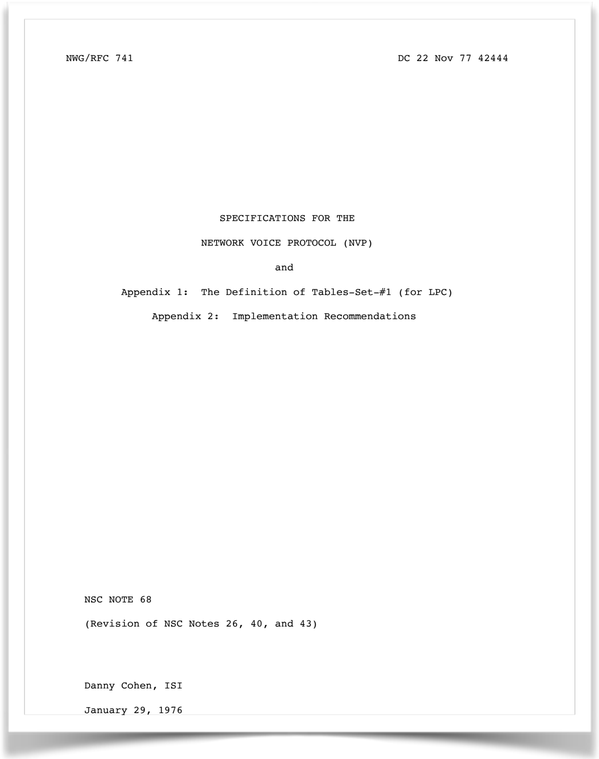
- Real-time applications and the Internet
  - What is real-time traffic?
  - Requirements and constraints
  - Quality-of-service and quality-of-experience
- Interactive applications
  - Conferencing application architecture
  - Signalling, session description, and real-time media traffic
- Streaming applications
  - HTTP adaptive streaming for video on demand

Real-time Media Over The Internet

- What is real-time traffic?
- Requirements and constraints
- Quality of service vs. user experience

Real-time Traffic on the Internet

- Long history of real-time traffic over the Internet:
  - Telephony and voice-over-IP (VoIP), Internet radio, video conferencing, streaming video and TV, gaming, sensors, industrial control
  - Initial packet voice experiments → 1970s
    - Network Voice Protocol
  - Core standards defined → mid-1990s
    - SIP, SDP, RTP, RTSP, H.323
  - HTTP adaptive streaming → mid-2010s
    - MPEG DASH
- The design of the network and its transport protocols has evolved to support real-time media



NWG/RFC 741 DC 22 Nov 77 42444

SPECIFICATIONS FOR THE

NETWORK VOICE PROTOCOL (NVP)

and

Appendix 1: The Definition of Tables-Set-#1 (for LPC)

Appendix 2: Implementation Recommendations

NSC NOTE 68

(Revision of NSC Notes 26, 40, and 43)

Danny Cohen, ISI

January 29, 1976

https://datatracker.ietf.org/doc/rfc741/

What is Real-time Traffic?

- Defining characteristic: **deadlines**
  - The system fails if data is not delivered by a certain time
    - Railway signalling must deliver data to change a signal before train arrives – **hard real-time**
    - Streaming video should deliver a new frame every 60th of a second, else the movie playback stutters – **soft real-time**
    - What is the probability a deadline is missed? What are the consequences? Engineer accordingly – **no system is 100% reliable**
  - Deadlines can be absolute or relative
    - Data must be delivered before a certain time – **absolute deadline**
    - Data must be delivered periodically, within some time after the previous – **relative deadline**
- Real-time is not necessarily high performance
  - It requires **predictable timing**, not necessarily high bandwidth or low latency

Requirements for Streaming Applications

- Video on demand has no absolute deadline, but needs regular data to prevent stalls once started
  - Acceptable if a takes a few seconds to start the video
  - Playback should be smooth once it starts
- Live video may have absolute deadlines
  - e.g., live sports playout

NETFLIX

BBC iPlayer

YouTube

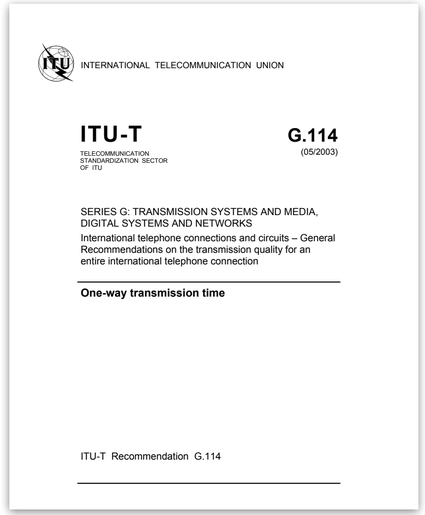
- Bit rate depends on desired quality
  - Higher quality is better, up to limits of display
  - Predictable, but lower, quality often preferred to more variable, but on average higher, quality
- For a given bit rate, trade-off between frame rate and frame quality
  - Smoother motion or higher resolution?

Requirements for Interactivity

- Requirements for interactive applications determined by task and human perception
  - Phone call or video conference
    - One-way mouth-to-ear delay ~150ms maximum for telephony
    - Video conferences want to lip-sync audio and video
      - Audio should be no more than 15ms ahead, or 45ms behind, video
  - Lecture style
    - Mostly unidirectional with occasional questions → can tolerate much higher latency
  - Distributed music performance
    - One-way latency ≪50ms desirable
    - Speed of sound: ~15ms to go from one side of a large orchestra to the other

ITU Recommendation G.114

ATSC IS-191 “Relative Timing of Sound and Vision for Broadcast Operations”, 2003


INTERNATIONAL TELECOMMUNICATION UNION

ITU-T
TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU

G.114
(05/2003)

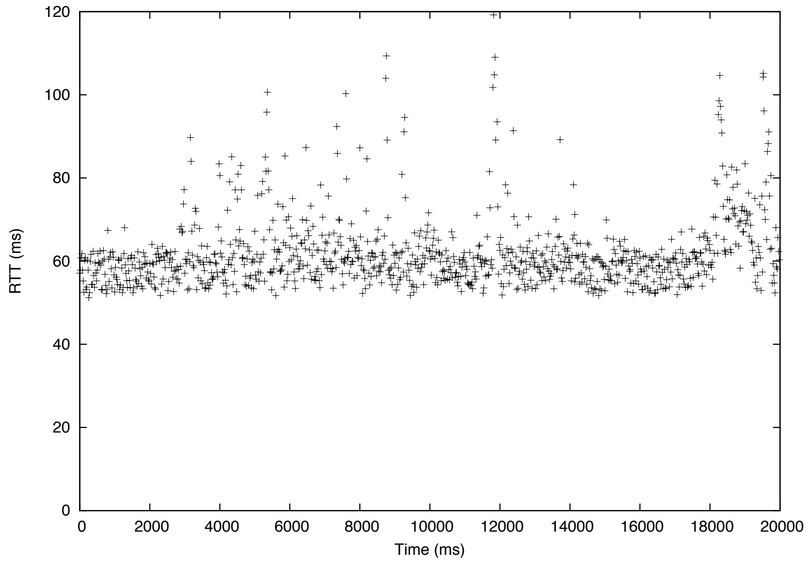
SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

International telephone connections and circuits – General Recommendations on the transmission quality for an entire international telephone connection

One-way transmission time

ITU-T Recommendation G.114

Real-time Traffic Must Be Loss Tolerant

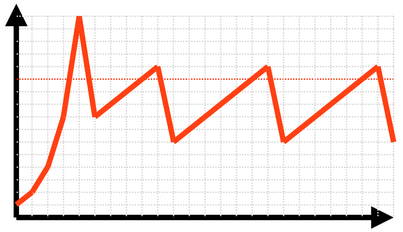


- The Internet is a best effort packet network
  - Timing is not guaranteed – depends on the amount of queueing on the path and on the path taken
  - Packets can be lost – retransmissions take time, and may not arrive before the deadline
  - In some networks these effects can be significant – with careful engineering they can be insignificant
- Real-time applications need to be loss tolerant

Data transport can be timely or reliable:
not possible to be both simultaneously

Real-time Traffic Has Limited Elasticity

- TCP congestion control adapts transmission speed to match available capacity
- Many transfers are **elastic** – faster is better, but it doesn't matter what rate the congestion control selects
  - The application can adapt to any chosen rate
- Real-time traffic is **inelastic**
  - Media has a minimum rate, below which it cannot be used
    - e.g., a minimum quality level below which the speech is unintelligible, that needs a certain bit-rate to transmit
    - If the congestion control algorithm cannot sustain this rate – if the network has insufficient capacity – real-time traffic cannot be used
  - Media has a maximum rate → cannot consume more
    - Video limited by capture frame rate and resolution
    - Audio limited by capture sampling rate



Quality of Service (QoS) Guarantees

- It may be possible to reserve network capacity for real-time traffic
  - Using RSVP, MPLS, 5G network slicing, …
  - Requires flow setup signalling, authorisation, and accounting
    - Need to tell the network what resources the real-time traffic requires
    - Need to demonstrate that sender is allowed to reserve these resources and can pay for the reserved capacity – reservations prevent other users from accessing the network, so have some cost
  - If the network has capacity for the traffic, reservations don't help
  - If the network does not have capacity, they allow the operator to discriminate in favour of customers who are willing to pay
    - Some customers care enough to pay for reserved capacity
    - Inter-domain reservations and accounting can be difficult to agree and implement
    - Many operators find it cheaper and easier to just buy more capacity, so no need to reserve

Quality of Experience (QoE)

- What ultimately matters is subjective quality of experience
  - Does the application meet user needs?
  - Does the application allow users to communicate effectively?
  - Does the application provide compelling entertainment?
- QoE is not a one-dimensional metric
  - What aspect of user experience are you evaluating?
    - e.g., “Does it sound good” is not the same as “Can you understand it?”
  - How does your metric relate to the task being performed?
  - Some aspects of user experience can be estimated from technical metrics
    - e.g., the ITU-T E-model for speech quality evaluation based on latency and packet loss
  - Some aspects are subjective and task dependent – need user trials

Real-time Media Over The Internet

- What is real-time traffic?
- Requirements and constraints
- Quality of service vs. user experience