End-to-end IP Service Quality and Mobility
- Lecture #2 -

Special Course in Networking Technology

S-38.215

vilho.raisanen@nokia.com

Agenda

- Agenda
- Goal of this lecture
- Rôles
  - Endpoint
  - Access
- Characteristics of applications
- Mobile applications

Goal of this lecture

- Definition of the characteristics of applications to be used in mobile environment.
  - “TSpec” this lecture, “RSpec” the next one.
- Impact of multi-access mobile environment
  - Endpoint
  - Access technology
- Analysis of services:
  - Service events that make up the service instance.
  - Service event characteristics.
  - How service is instantiated.

Vilho Räisänen

Rôle of the endpoint

- In multi-access network, the endpoint capabilities vary.
- PC-type endpoint:
  - Lots of processing power, large memory.
  - Large display, stereo speakers.
  - Keyboard & mouse available.
- PDA/cellular phone –type endpoint:
  - More limited processing power & memory.
  - Display smaller, sound capabilities limited.
  - Input devices more constrained.
- **Conclusion**: preferably service instantiation dependent on the type of endpoint.

Vilho Räisänen

Rôle of access technology

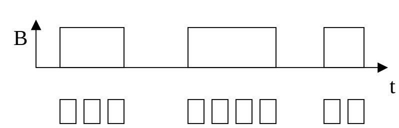
- Access technology has an effect on services:
  - Available bandwidth.
  - RTT, delay variation, packet loss.
  - Service availability.
  - Service instantiation time varies.
  - Service quality support mechanisms vary per access technology.
    - Service quality support mechanisms available.
      - Access network may not be able to provide support for real-time services.
      - There may be no guarantees.
  - Mobility support varies.

Case VoIP media stream

- VoIP with 8 kHz PCM coding, 8 bit sample resolution
  - Bi-directional
  - “On/off” traffic pattern, talkspurts & silence periods
  - 20 ms frame size => 160 bytes / sample
  - 1 sample / packet

  ⇒During talkspurts, a 200-byte packet is transmitted every 20 ms (IPv4); during silence periods no packets (VAD).

  ⇒Requires 80 kbit/s IP layer throughput



Case HTTP browsing

- HTTP browsing without pipelining.
  - Request / response pattern.
  - HTTP GET requests typically small, HTML page sizes vary.
  - Large non-textual objects such as can be embedded into HTML pages.
  - User wants interactivity: something happens "soon" when a hyperlink is clicked.
  - Non-first time users understand that downloading large content takes time.
    - It is the total downloading time that counts

  ⇒Small messages best given some statistical capacity in "uplink" direction, downlink can vary more.

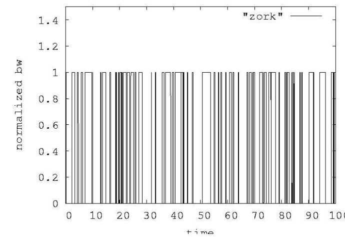
Token bucket parameters

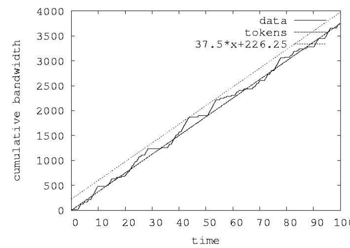
- Bursty VoIP streams can be shaped
  - Token bucket parameters (r, B, M, m) need to match the nominal bit rate and media stream characteristics of the codec (see **example**).
  - Buffering delay must be minimized.
- Bursty browsing stream (downlink) has no intrinsic bandwidth requirement.
  - End-to-end delay can be larger => longer shaping buffering possible.
  - HTTP is run on top of a reliable transport protocol

  =>throughput is a design parameter.
- TB parameters establish a bandwidth envelope: $\int_0^T d(t)dt \leq rT + B$

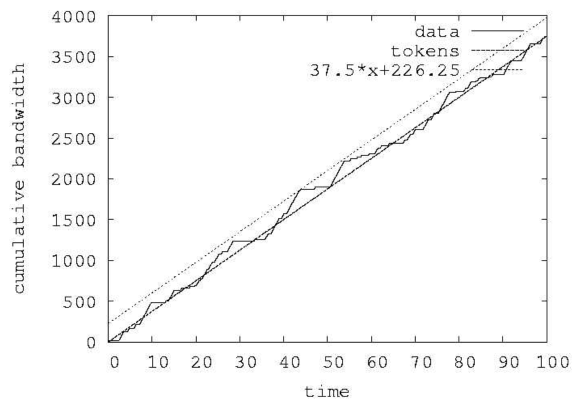
Conditioning example

- VoIP media stream: Poisson modulated ON/OFF process.
- Let us assume
  - Average talkspurt length of 0.6 seconds
  - Average silence period length of 1 seconds
  - Imaginary codec with 100 units/s output rate
- Example trace:
  - R = 37.5 units/s
  - B = 226.25 units



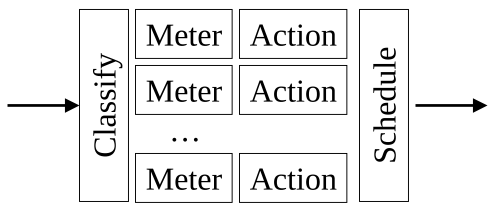


Conditioning example / 2



Edge conditioning

- Edge conditioning consists of three parts:
  - Classification
  - Metering
  - Action
- Metering is based on token bucket parameters
  - Per flow
  - Per traffic aggregate
  - Based on SLA
- Action can be
  - Shaping
  - (Re)marking
  - Dropping
  - Null action



Bandwidth utility

- *Bandwidth utility U(b)* tells how much an application can benefit from bandwidth.
- No single measure exists.
- Phenomenological differences between applications:
  - CBR VoIP codec: U(b) almost a step function.
  - G.723.1: two-step function.
  - AMR codec: more steps.
  - Downloading of a large HTTP page: “continuous” function.
- If utility function is known, bandwidths can be allocated by maximizing the total utility of allocated bandwidths.
- In real systems, utility can only be approximated.

$$U ! = \max_{\{b_i\}} \sum_i U(b_i)$$

[F.Kelly, European Trans. Telecom. Vol. 8, no.1, 1997, p.33 ff.]

Bandwidth utility examples

- Simple utility for elastic applications: linear in bandwidth.
- Can be made more elaborate by taking into account application type with a per-app linear coefficient of per-app B parameter.
- Simple-minded application of super-linear utility functions dangerous: utility maximized by non-homogeneous resource allocation.
  - In example, criterion M>C/B.

=> **Fairness important.**

- Max-min utility schemes.

$$U(b) = \begin{cases} 1, b \geq B \\ \frac{b}{B}, b < B \end{cases}$$

$$U_i(b) = C_i \begin{cases} 1, b \geq B \\ \left(\frac{b}{B}\right)^{\alpha}, b < B \end{cases}$$

[Kao, Zhegura, Utility max-min: an application-oriented…, Infocom'99 p.793 ff.]

Vilho Räisänen

Average throughput

- Some applications are inherently bursty, yet require near-constant throughput.
  - VoIP
- Another example: H.261 video streaming codec
  - Full picture sent periodically => bandwidth peak.
  - Intermediate frame size depends on motion.
- Token bucket regulator must accommodate momentary bandwidth peaks, but typically end-to-end delay allows for shaping to smoothen the bandwidth.
- Contrast with browsing: TB regulator does not need to accommodate all bandwidth peaks; longer delay allowable.

[A. Adas, Traffic models in broadband networks, IEEE Comm. Mag. July 1997, p.82 ff]

Vilho Räisänen

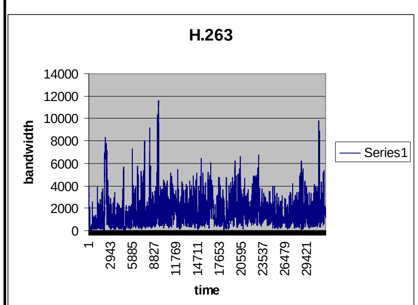
Estimation of token bucket parameters

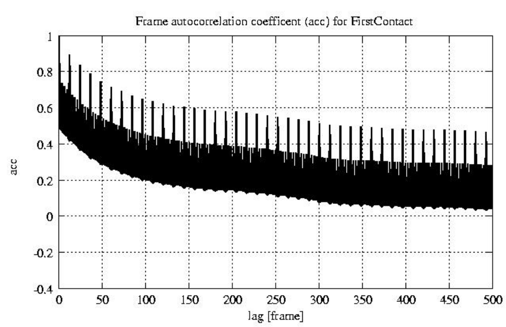
- Straightforward method:
  - Compute token rate
  - Get bucket depth from arrivals process
- Need to have representative sampling.
- In IP affected by
  - header overhead
  - MTU size.

  => Real token bucket parameters used by the network cannot always be fully determined on application layer without referring to link layer.

Vilho Räisänen

Example: H.263 trace (UBR)





[http://www-tkn.ee.tu-berlin.de/research/trace/trace.html]

Vilho Räisänen

Application characteristics

- Types of service events involved.
  - Characteristics may be different per component.
- Unidirectional / bi-directional.
  - Asymmetry of characteristics
- Type of bandwidth utility: smooth vs. step-like.

  => Designed bandwidth possible or not.
- Burstiness characteristic: token bucket parameters

Vilho Räisänen

Useful mobile applications

- Mobile environment creates new needs for applications.
- Mobile application types:
  - Telephony / conferencing
  - Instant / near-instant messaging
  - M-mail
  - Fetching of information
  - Killing time

Vilho Räisänen

Mobile applications

- Off-line data transfer (e-mail)
- On-line data transfer
- Browsing
- Messaging
- Presence
- Chat
- Games
- Streaming
- Multimedia-over-IP conferencing

Data transfer

- Examples: e-mail, downloading of a picture, a piece of music or a game.
- Characteristics:
  - Content size can be large.
  - E-mail: more or less symmetric, downloading: unidirectional/asymmetric.
  - Irregular arrival process for service events.
  - Traffic can be bursty.
  - Applications typically very elastic with respect to utility.
  - TCP-based.
- Conclusion:
  - Bandwidth designed

Interactive applications

- Examples: browsing, M-commerce.
- Characteristics:
  - Content size varies.
    - Symmetric to asymmetric bandwidth characteristics.
  - Some services have real service instances.
    - Arrival process for service instances random.
    - Arrival process for service events also random, but may be frequent.
    - Request and response temporally close to each other.
  - TCP-based.
- Conclusion: sufficient r & B for most HTTP/WAP requests, downlink throughput can be designed.

Simple SIP services

- Examples: (instant) messaging, chat/presence.
- Characteristics:
  - Content size varies.
  - Can be unidirectional, symmetric, or asymmetric.
  - Arrival process random for service instances and service events.
    - Some degree of interactivity expected of IM, chat.
  - TCP or UDP based.
- Conclusion:
  - Sufficient throughput for small messages, throughput for large messages may be designed.
    - UDP: shaping rather than dropping.

Streaming

- Examples: audio or video streaming.
- Characteristics:
  - RTSP signalling:
    - Interactivity.
    - TCP or UDP.
  - Media flow:
    - Periodic PDU stream.
    - Can be bursty or already shaped.
    - UDP, for firewall traversal reasons also TCP.
- Conclusion:
  - Need to provide enough bandwidth, but can apply shaping to media flow.

Multimedia conferencing

- Examples: audio or video conferencing.
- Characteristics:
  - SIP signalling:
    - Call set-up time requirements.
    - TCP or UDP.
  - Media flow:
    - Periodic PDU stream.
    - Can be bursty (video) or already shaped.
    - UDP, for firewall traversal reasons also TCP.
- Conclusion:
  - Need to provide enough bandwidth for media stream & signalling.

Summary

- The following characteristics of applications need to be known:
  - Token bucket parameters.
    - Not necessary same on application level and network level.
  - Bandwidth utility.
    - May be application type dependent.
    - Fairness considerations.
- Mobile application categories:
  - Data transfer
  - Interactive applications
  - Simple SIP services
  - Streaming
  - Multimedia conferencing