ΗΥ537: Έλεγχος Πόρων και Επίδοση σε Ευρυζωνικά Δίκτυα

Βασίλειος Σύρης

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης
Εαρινό εξάμηνο 2008

Network services
Contracts – Service Level Agreements (SLAs)

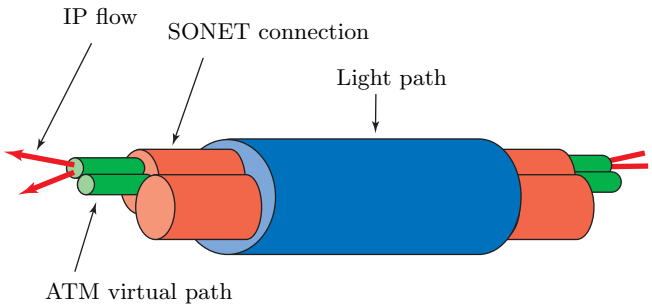
Service layering

- A service at one layer utilizes lower layer services, and adds to their functionality
- Higher layer services: interface directly with user, e.g., web browsing
  - ASP: Application Service Providers
- Lower layer services: users use indirectly, e.g. Internet transport
- Quality of higher layer services depends on quality of lower layer services
  - Higher layer: response time, transaction rate, etc
  - Lower layer: throughput, delay, jitter, etc
- Access services: connects user equipment to network

Transport technologies

- IP
- Ethernet
- ATM (Asynchronous Transfer Mode)
- MPLS (MutliProtocol Label Switching)
- SDH (Synchronous Digital Hierarchy), SONET (Synchronous Optical Network)
- DWDM (Dense Wave Division Mux)

Transport service layering



Transport service classification

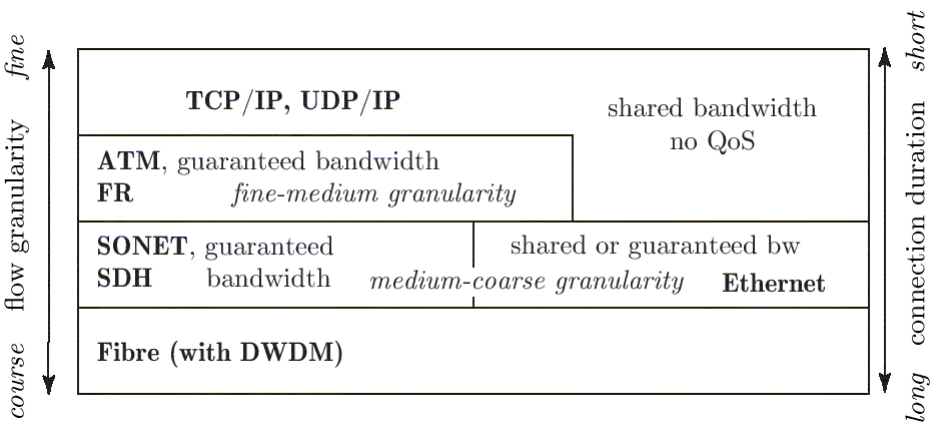
- Connection-oriented
- Connectionless

- Synchronous, e.g. ISDN, SDH/SONET
  - data transport at regular intervals
- Asynchronous, e.g. ATM (cell switching), IP (packet switching)
  - take advantage of statistical multiplexing

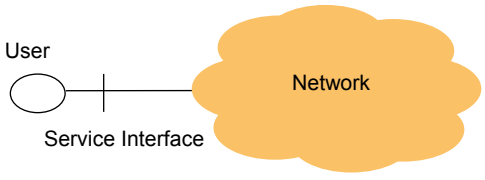
- Best-effort
- Guaranteed

- Flow isolation: used to guarantee performance
- Flow granularity

Flow granularity



Service contracts



- Service contract: network commits to deliver a service with specific quality and performance characteristics and user commits to interact with network at given way

Service contracts (cont.)

- Service contracts also known as Service Level Agreements (SLA)
- Service contract can include:
  - Billing/tariff issues
  - Geographic coverage
  - Reliability/uptime
  - Response to failures
  - Help desk/call center support
  - QoS parameters
  - Traffic description
- Traffic contract part of service contract that is concerned with characteristics of transport service

ATM Quality of Service (QoS) Parameters

ATM QoS parameters characterise performance of an ATM layer connection

- **Max Cell Transfer Delay (CTD):** Delay experienced by a cell between network entrance and exit points
- **peak-to-peak Cell Delay Variation (CDV):** max - min cell transfer delay
- **Cell Loss Ratio (CLR):** Percentage of cells that are lost
- **Minimum Cell Rate (MCR):** Minimum cell rate user is allowed to always send
- The above are user negotiable
- Also, a number of additional, *non-negotiable* QoS parameters have been defined

Quality specification

- Deterministic, e.g. 0% loss
- Statistical, e.g. 10^{-5} loss percentage
- Relative, e.g. class A received better quality than class B

Traffic Parameters

Traffic parameters describe inherent characteristics of a traffic source

- **Peak Cell Rate (PCR):** Maximum instantaneous rate
- **Sustained Cell Rate (SCR):** Average cell rate measured over some long interval
- **Maximum Burst Size (MBS):** Maximum burst size (# of cells) that can be sent at the peak rate

Leaky bucket traffic descriptor

- Used for traffic characterization
- Defined by leak rate r and bucket size b
- Conforming traffic stream must satisfy

$$X[t,t'] \le r(t'-t) + b \quad \text{for all} \quad t < t'$$

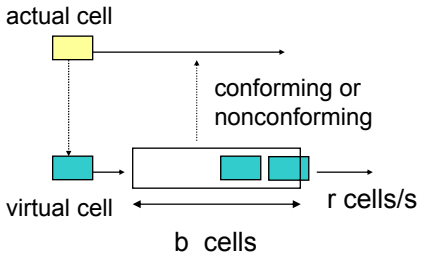
- Equivalently

$$\frac{X[t,t']}{t'-t} \le r + \frac{b}{t'-t} \quad \text{for all} \quad t < t'$$

- Larger width [t,t') => rate bounded by leak rate r
- In window [t,t') source allowed to produce burst b above r(t'-t)
- Leaky bucket permits arbitrary large peak rate

Leaky Bucket Algorithm

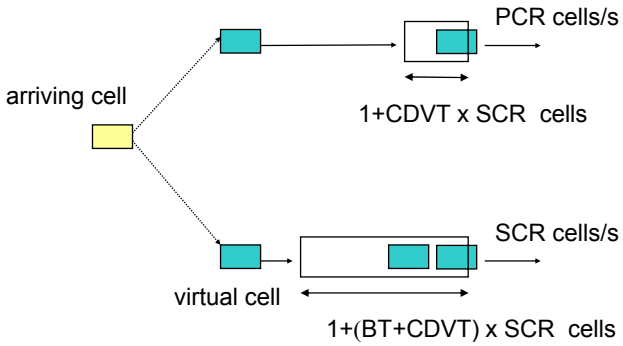
- Bucket size b
- Leak rate r
- Bucket contents by 1 or pkt size for each conforming cell/pkt



B: bucket contents

If B+1> b
  cell nonconforming
else
  cell conforming
  B=B+1

Multiple leaky buckets



Traffic control using leaky buckets

- Assume leaky bucket (r,b) and access rate h
- Maximum time t user can transmit at rate h

$$ht \le rt + b \Rightarrow t = \frac{b}{h - r}$$

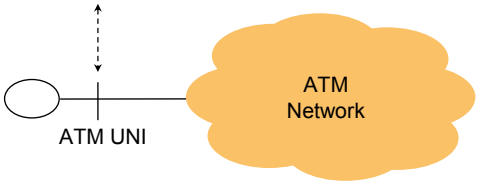
Traffic Contract Negotiation

Traffic Contract:

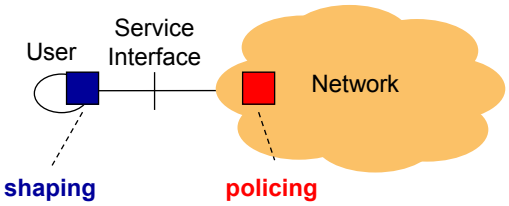
Traffic parameters: peak rate, sustainable cell rate, burst size, minimum cell rate

QoS parameters: cell delay, cell delay variation, cell loss ratio

Conformance definition: Generic Cell Rate Algorithm (GCRA)



Policing and shaping



- Policing: network verifies that user's traffic is within traffic profile specified in contract
- Shaping: buffer (delay) packets to conform with profile
- Charging: loose form of policing

Contract parameters

- Static parameters: set when contract is established, and remain constant throughout its life
- Dynamic parameters: updated during contract's lifetime

Contract A	Contract B
No loss provided rate < 1 Mbps	No loss provided rate < h Mbps, where network can vary h between 1<h<2 Mbps

- Both A and B: static parameters zero loss, 1 Mbps
- Contract B: dynamic parameter extra rate 0<x<1 Mbps

Contract parameters (cont.)

Contract A	Contract B	Contract C
No loss provided rate < 1 Mbps	No loss provided rate < h, where *network* can vary h between 1<h<2 Mbps	Loss < 10^{-5} provided rate < h, where *user* can vary h between 1<h<2 Mbps

- Contract B: network is varying rate h, depending on spare capacity; appropriate for elastic applications
- Contract C: user varies h; network must always be able to provide 2 Mbps, with loss < 10^{-5}; network takes advantage of statistical mux
- Contract C: good idea to charge user for h-1

Contract parameters (cont.)

- Static parameters
  - network must reserve resources for these
- Dynamic parameters
  - Network dynamically reserves resources for these
  - Must specify who is responsible for changing dynamic parameters
  - With pricing, network can influence value of dynamic parameters, even if he is not responsible for choosing them
- Example of service with purely dynamic parameters: TCP
  - TCP software dynamically controls maximum rate h
  - Implicit guarantee of small pkt loss
- Measured parameters: known after connection starts, e.g. duration, transferred volume