

Network control and effective bandwidths

Vasilios A. Siris

Computer Science Department
University of Crete
HY536 - Network Technology Lab II
2000 - 2001

Network control for various service types

- ATM and IP supports statistical multiplexing
- Network control different for guaranteed and elastic services
- **Guaranteed services**
 - User-network contract
 - Call Admission Control - CAC
 - Open loop control
- **Elastic services (ABR):**
 - no CAC, except for MCR (Minimum Cell Rate)
 - Closed loop control

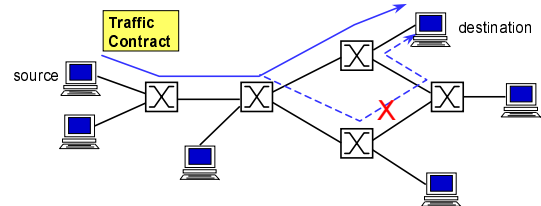
Network control & effective bandwidths - 2

Time scales of network control

Traffic & Congestion Control Functions	Response time
Selective cell discard, frame discard, priority control and scheduling, Usage Parameter Control (UPC), traffic shaping	Cell time
Feedback controls	Round-trip propagation time
Routing, Call Admission Control (CAC)	Connection interarrival time
Network management control	order of minutes
Pricing	months, years

Network control & effective bandwidths - 3

Routing and CAC



- **Routing:** find path from source to destination that fulfils user requirements (bandwidth, QoS)
- **Call Admission Control (CAC):** performed at every switch, determines whether there are enough resources to accept a call

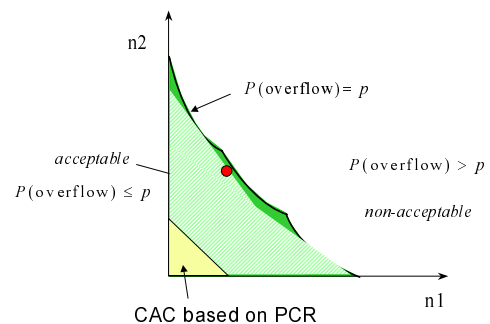
Network control & effective bandwidths - 4

Call Admission Control (CAC)

- k traffic classes (actual or contract types)
 - n_1
 - n_2
 - n_k
- class i contributes n_i sources
- QoS constraint (contract obligation): $CLP \leq p$ (e.g. $p=10^{-8}$)
- What (n_1, \dots, n_k) do not violate QoS constraints?
- Approaches to CAC:
 - Non-dynamic: based only on traffic contract parameters
 - Dynamic: includes on-line measurements and contract parameters

Network control & effective bandwidths - 5

Acceptance region



Network control & effective bandwidths - 6

Simplifying the problem of CAC

Use Effective Bandwidths:

acceptance condition: $n_1 \cdot \alpha_1 + \dots + n_k \cdot \alpha_k \leq C^*$

- Can we define $\alpha_1, \dots, \alpha_k, C^*$ such that
- α_i depends on **source traffic statistics**, as well as **traffic mix, capacity, buffer, QoS**
- * depends on **traffic mix, capacity, buffer, and QoS**
- Calculation of α_i can be done off-line
- The **true** acceptance region is well approximated
- YES!**

Network control & effective bandwidths - 7

Statistical multiplexing gain

Important consideration:

- Statistical multiplexing gain:** increased utilization when allowing some loss rather than zero loss

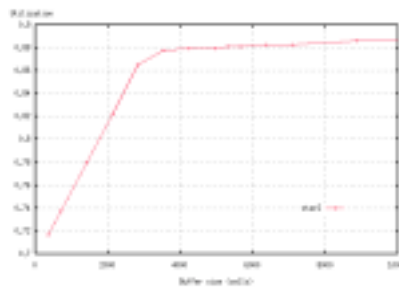
$$SMG = \frac{N_{stat.}}{N_{det.}}$$

- Why: 99.999% is much cheaper than 100%
- Enabled by burstiness of traffic sources and large capacity of transmission links

Network control & effective bandwidths - 8

Statistical multiplexing gain (cont.)

- Example of statistical multiplexing gain

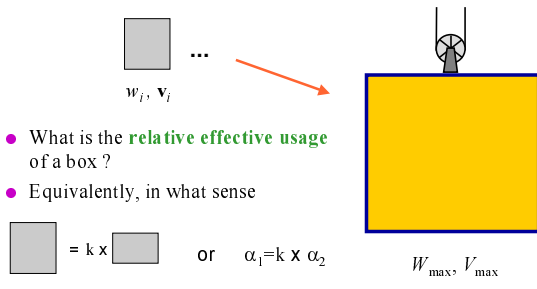


$C=155$ Mbps, $CLP \leq 10^{-7}$, *Star Wars* traffic
mean=0.26 Mbps, peak=3.46 Mbps

- Peak rate allocation achieves utilization 7.5%!

Network control & effective bandwidths - 9

Loading an elevator with boxes



- What is the **relative effective usage** of a box?
- Equivalently, in what sense

$$\text{Grey box} = k \times \text{Yellow box} \quad \text{or} \quad \alpha_1 = k \times \alpha_2$$

Key notion: substitution

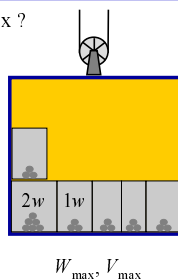
Network control & effective bandwidths - 10

Loading an elevator (cont.)

- What is the relative effective usage of a box?
 - Depends on which constraint is active: **max. weight** or **max. volume**
 - Determined by operating point
- If **max. weight** is active, then effective usage equals box's **weight**

$$\sum_i w_i = W_{max}$$

$$\sum_i v_i < V_{max}$$



- Effective bandwidth = weight**

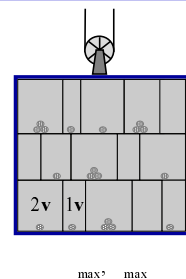
Network control & effective bandwidths - 11

Loading an elevator (cont.)

- If **max. volume** is active, then effective usage equals box's **volume**

$$\sum_i v_i = V_{max}$$

$$\sum_i w_i < W_{max}$$

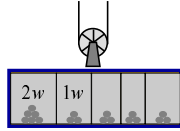


- Effective bandwidth = volume**

Network control & effective bandwidths - 12

Loading an elevator (cont.)

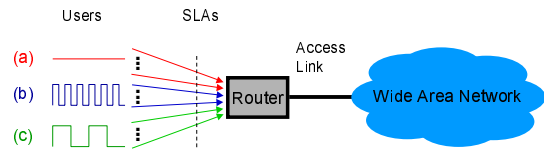
- What is the relative effective usage of a box ?
 - Depends on which constraint is active: **ma weight or ma volume**
 - Determined by operating point



- **Effective bandwidth = ~~weight~~ = volume**

Network control & effective bandwidths - 13

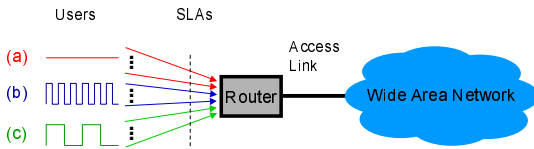
Example: resource usage



- All user connections have the same duration and volume
- Some QoS is supported at the access link
- What is the resource usage of each user?

Network control & effective bandwidths - 14

Example: resource usage (cont.)

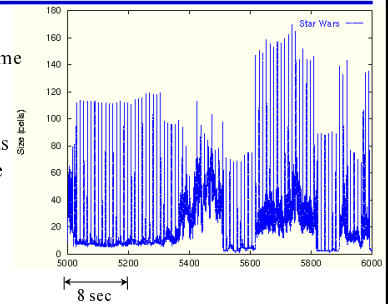


- All user connections have the same duration and volume
- Some QoS is supported at the access link
- What is the resource usage of each user?
- Answer: **depends!**
 - “small” capacity: $a < b, c$
 - “medium” capacity: $a = b < c$
 - “large” capacity: $a = b = c$

Network control & effective bandwidths - 15

Effective bandwidth of traffic streams

- Broadband traffic has burstiness in different time scales
- Effective bandwidth (resource usage) depends on time scales which are important for buffer overflow
- How can we identify which time scales are important for overflow?
- Dependence on context



Network control & effective bandwidths - 16

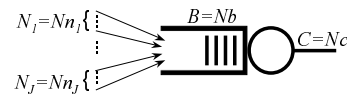
An effective bandwidth formula

- Effective bandwidth of a source of type j

$$\alpha_j(s, t) = \frac{1}{st} \log E \left[e^{sX_j[0, t]} \right]$$
- $X_j[0, t]$: load produced by source of type j in window t
- (s, t) = **op ting point of th link**
 - depends on the link param. (C, B) , traffic mix, and CLP ($= e^{-\gamma}$)
 - t : **time** parameter, related to time for buffer overflow
 - s : **space** parameter, depends on link's multiplexing capability, exponential tilt parameter of distributions
 - $s = \frac{\partial \gamma}{\partial B}$, $st = \frac{\partial \gamma}{\partial C}$ where $\gamma = -\log \text{CLP}$

Network control & effective bandwidths - 17

Mathematical justification



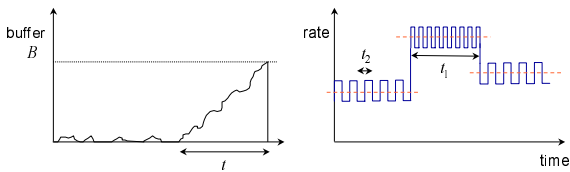
- **Many sources asymptotic** $P(\text{overflow}) = e^{-Nl + o(N)}$
- $\lim_{N \rightarrow \infty} \frac{1}{N} \log P(\text{overflow}) = \sup_t \inf_s \left[\sum_{j=1}^J n_j \log E e^{sX_j[0, t]} - s(b + ct) \right] = -l$
- If $P(\text{overflow}) = e^{-\gamma}$ then we must have

$$\sum_{j=1}^J N_j \alpha_j(s, t) \leq C + \frac{1}{t} \left(B - \frac{\gamma}{s} \right) \text{ where } \alpha_j(s, t) = \frac{1}{st} \log E \left[e^{sX_j[0, t]} \right]$$

$$= C_{\text{eff}}$$

Network control & effective bandwidths - 18

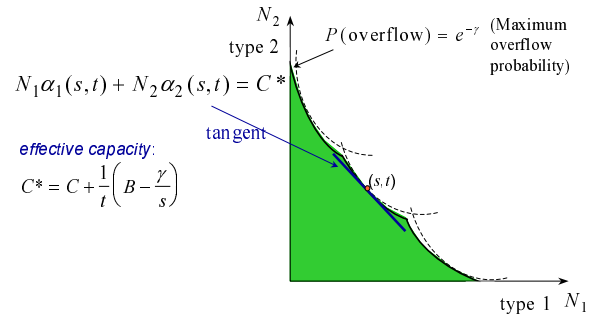
Operating point parameters s, t



- During the overflow, the inputs have a different distribution with higher means: exponentially tilted distribution with parameter s (= distribution of most probable behaviour)
- Overflow period has duration $t \Rightarrow$ we care for contribution of input sources in window t
 - time scale of relevant burstiness = t_1 , not t_2

Network control & effective bandwidths - 19

Acceptance region for two source types

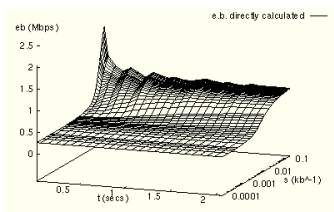


$$N_1 \alpha_1(s, t) + N_2 \alpha_2(s, t) = C^*$$

$$C^* = C + \frac{1}{t} \left(B - \frac{\gamma}{s} \right)$$

Network control & effective bandwidths - 20

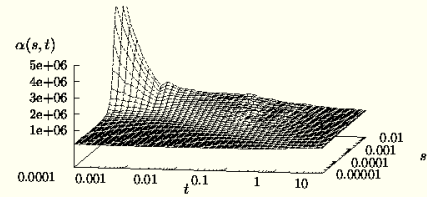
Effective bandwidth for MPEG-1 traffic



- Star Wars MPEG-1 trace

Network control & effective bandwidths - 21

Effective bandwidth for Ethernet traffic



- Bellcore Ethernet trace

Network control & effective bandwidths - 22

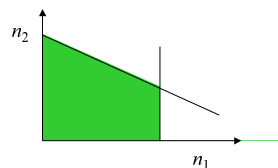
Multiple QoS constraints

- Acceptance region described by multiple constraints
- Example: Priority queuing
 - two classes: $J_1 > J_2$
 - for J_1 : $P(\text{delay} > B_1 / C) \leq e^{-\gamma_1}$
 - for $J_1 \cup J_2$: $P(\text{buffer overflow}) \leq e^{-\gamma_2}$

- Two constraints:

$$n_1 \alpha_1(s_1, t_1) \leq K_1$$

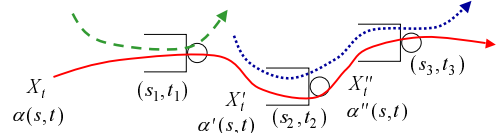
$$n_1 \alpha_1(s_2, t_2) + n_2 \alpha_2(s_2, t_2) \leq K_2$$



Network control & effective bandwidths - 23

Decoupling bandwidths

Is the effective bandwidth of a stream constant in a network?



- Effective bandwidth of a traffic stream defined for single link
- In general $\alpha(s, t) \neq \alpha'(s, t)$
- Under realistic multiplexing conditions (conditions for "decoupling") $\alpha(s, t) = \alpha'(s, t)$
- Economic arguments suggest that (s, t) does not vary much

$$s = \frac{\partial \gamma}{\partial B}, \quad st = \frac{\partial \gamma}{\partial C} \quad t = \frac{MC \text{ per unit } c}{MC \text{ per unit } b}$$

Network control & effective bandwidths - 24

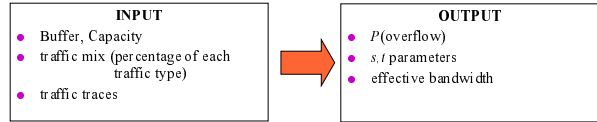
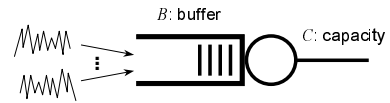
The msa and lb tools

- The tools work on files with measurements of load in consecutive fixed length intervals
- The input has the following format:

```
# epoch_in_msecs = 40
# bits_per_info_unit = 424
65
4
5
8
...
```

Network control & effective bandwidths - 25

Traffic analysis tools



Traffic trace:
e.g., 1 epoch=10 msec

Epoch	Cell or packet count
0	111
1	24
2	20
...	...

Network control & effective bandwidths - 26

The msa tool

Functions:

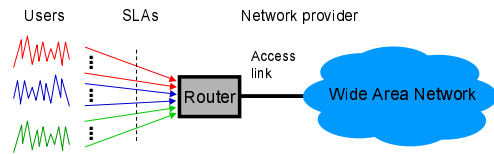
- Trace file, link operating point => resource usage
- Link resources (capacity, buffer), traffic mix, load => QoS
- Link resources (capacity, buffer), traffic mix, QoS => load
- Capacity (buffer), traffic mix, load, QoS => buffer (capacity)

With the use of scripts:

- QoS as a function of buffer size
- Load as a function of buffer size
- Acceptance region for two traffic types

Network control & effective bandwidths - 27

Some management questions

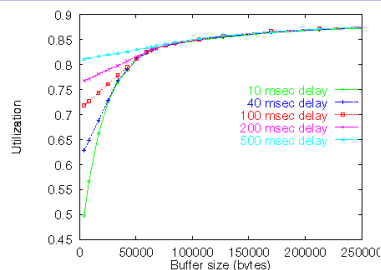


- What combinations of user types can the provider accept?
- What is the effect of increasing the link resources or of traffic shaping on the multiplexing capability of the link?
- What traffic parameters should a user choose?

Network control & effective bandwidths - 28

Resource dimensioning and traffic shaping

Internet WAN
 $p=2.3$ Mbps, $m=6.4$ Kbps
 $C=34$ Mbps
 $P(\text{overflow})=10^{-6}$



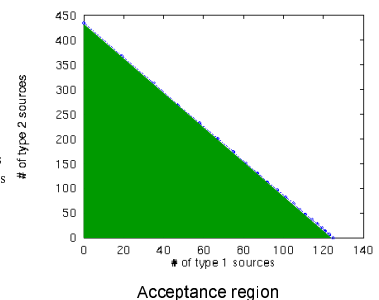
- ➔ Effects of traffic shaping depend on buffer size

Network control & effective bandwidths - 29

Acceptance region

$C=155$ Mbps, $B=750$ Kbytes
 $P(\text{overflow})=10^{-6}$

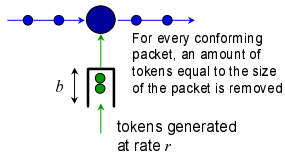
Internet WAN traffic:
 Type 1: $m=1$ Mbps, $p=10.2$ Mbps
 Type 2: $m=0.3$ Mbps, $p=8.9$ Mbps



Network control & effective bandwidths - 30

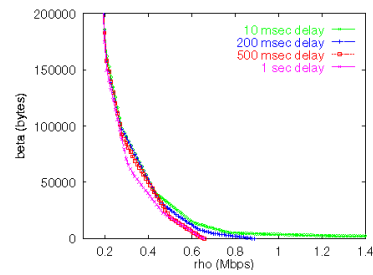
Token bucket parameter selection

- Token (or leaky) bucket is a widely used traffic descriptor:
 - ATM (GCRA)
 - Internet (Integrated & Differentiated Services, MPLS)
 - Policy-based management
- Token or leaky bucket (r, b) : r is token rate, b is bucket size



Network control & effective bandwidths - 31

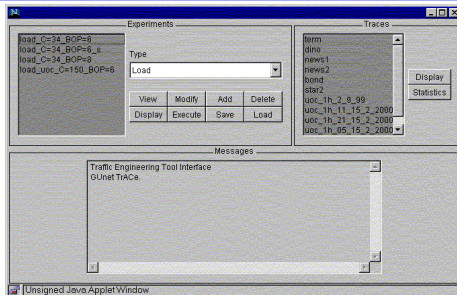
Indifference curve



- Traffic shaping affects the lower right portion of the indifference curve

Network control & effective bandwidths - 32

Web-based interface for traffic engineering tools



- Interface: <http://trace.ucnet.uoc.gr/interface/src/index.html>
- Tools & manuals: <http://www.ics.forth.gr/netgroup/msa/>

Network control & effective bandwidths - 33

Conclusions

- Acceptance region defines the “**technology set**” of a network for guaranteed services
- Effective bandwidths provide a mathematically rigorous approximation of the acceptance region
 - can be approximated by a set of **linear constraints**
- The effective bandwidth of a stream is a function of the **operating point** of the link
 - defined by network resources and consistency of traffic mix
- The effective bandwidth is a **good “proxy”** for the relative resource usage among traffic streams

Network control & effective bandwidths - 34