Charging *ATM* Services: Guaranteed Services

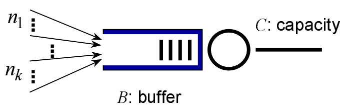
by
Costas Courcoubetis,
Vasilios Siris,
George D. Stamoulis

Winter 1998

Contents

Effective Bandwidths Reminder

- *k* traffic classes
- class *i* contributes n_i sources



- Which $(n_1,\dots,n_k)$ do not violate QoS constraint (e.g. CLP ≤ e$^{-ä}$) ?

➡ **CAC** can use *Effective Bandwidths:*

$$n_1 \cdot \alpha_1 + \dots n_k \cdot \alpha_k \le C^* = C + \frac{1}{t}(B - \frac{\gamma}{s})$$

Effective Bandwidth Formula

- Effective bandwidth of a source of type j

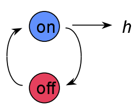
$$\alpha_j(s,t) = \frac{1}{st}\log E\left[e^{sX_j[0,t]}\right]$$

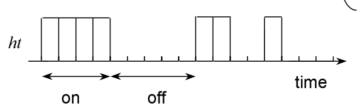
$X_j[0,t]$: load produced by source of type j in window t

- The effective bandwidth $\alpha_j(s,t)$ quantifies resource usage for a *particular operating point* (s,t)

ON/OFF Source Model

- At each interval of duration *t*:
  - either OFF: $X=0$ (with probability 1-p)
  - or ON: $X=ht$ (with probability p)





- mean rate $m = p^*h$
- effective bandwidth $\alpha_{on/off}(s,t) = \frac{1}{st}\log\left[1 + \frac{m}{h}\left(e^{sth} - 1\right)\right]$

➡ ON/OFF has the *largest* possible effective bandwidth (over *all* traffic models) for *given* m and h and operating point

---

Charging VBR Services - Idea 1

- Apply CAC with effective bandwidths

and

- Charge (per unit time) each VBR call proportionally to effective bandwidth

▲ Pro: Theoretically justifiable and fair

▼ Con: Requires accurate *a priori* knowledge of complex traffic statistics

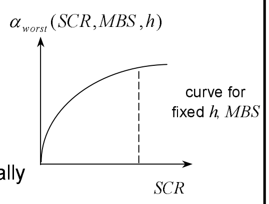
- If *empirical* estimate from sampling previous calls is used, users have the *wrong* incentive to reach this value
- user will tend to *overload* the network => similar to overeating in all-you-can-eat restaurants

Charging VBR Services - Idea 2

- Use traffic contracts specifying:
  - peak rate
  - sustainable cell rate, max burst size
- Apply CAC with the *worst*-case effective bandwidths for contracts

and

- Charge each VBR call proportionally to worst-case effective bandwidth



▲ Pro: No complicated traffic statistics

▼ Cons: *Unfair* to users with mean rate < SCR, or peak rate < *h*
Provides the *wrong* incentive to increase traffic

Why Failed So Far ?

- Previous approaches only based on *static* a priori variables, determined by traffic contract
  - Provided *wrong* incentives to *exhaust* the resource usage permissible by contract
- Charging should also employ *dynamic* a posteriori variables, *measured* during call

Charging VBR Services - Idea 3

- *Measure* the effective bandwidth during each VBR call, and charge proportionally to it

▲ Pro: Provides the *right* incentives

▼ Con: Incompatible with *static CAC*
users may possibly *not* pay for all the resources *reserved* by CAC
e.g. a user with no traffic, would face 0 charge

What is Needed ?

- Charge according to both:
  - static variables reflecting traffic contract and resources reserved by CAC
  - dynamic variables reflecting actual usage
- Final charge should be close to *actual* effective bandwidth
- Allow the user to select a tariff:
  - selection should reveal some important additional *information* to the network

Simple Charging Scheme for VBR Services

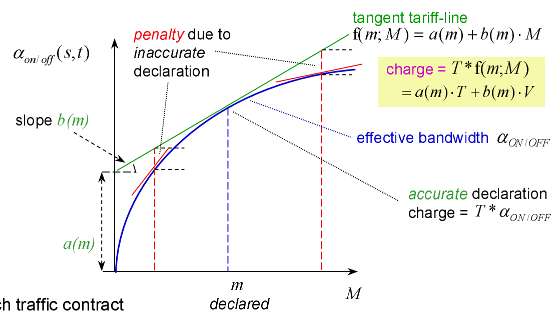
- Assume per VBR call: only *peak* rate is policed
  - *known* peak rate h, *unknown* mean rate M
- ➡ For each VBR call, the user *selects* a tariff, by declaring his *prediction* m for the *mean* rate;
  - e.g., by sampling the volume and duration of previous calls
- ➡ Charge per unit time: $f(m; M) = a(m) + b(m) \cdot M$
  - for good choice of tariff, charge is close to the worst-case (ON/OFF) effective bandwidth permissible by contract

→ *actual* mean rate, measured in call as $M = V/T =$ *Volume / Time*

→ user *prediction* of mean rate - defines tariff

- F.P.Kelly, "Tariffs and Effective Bandwidths in Multiservice Networks", ITC 94

Simple Charging Scheme



- Each traffic contract defines a family of tariff lines

Properties of Simple Charging Scheme

- Total charge = $T \cdot f(m; M) = a(m) \cdot T + b(m) \cdot V$
- Fair: Charges both for
  - resource *usage* => volume-component
  - resource *reservation* => time-component
- Simple Accounting
  - Requires only *simple* measurements: *T* and *V*
- Choice of tariff *reveals* important user-information

Incentives to the User

- Provides the user with the *right incentives*. In particular:
  - Incentive to accurately declare the mean rate M, if known a priori
  - For random mean rate M: *Expected* charge is *minimised* for

    $m=E[M]$
    - user has the incentive to *estimate* this (from empirical information), and declare it to the network
  - Incentive to *shape* traffic, thus reducing peak rate h and the charge

Incentive Compatibility

- User's optimal declaration of m is *informative* to the network provider
- Can be used by network provider in more efficient allocation of resources, thus improving operation of the network
- User's incentive to shape traffic reduces burstiness, thus also leading to more efficient operation of the network

Computation of $a(m)$ and $b(m)$

- Both $a(m)$ and $b(m)$ can be expressed in *closed* form
- Appropriate values of s,t can be derived /numerically

Examples of Tariffs

$h=3$ Mbps	$st=1$ sec/Mbit	
M	$a(m)$	$b(m)$
0.20 Mbps	0.26	2.80
0.75 Mbps	0.93	1.10
1.50 Mbps	1.46	0.60
2.25 Mbps	1.81	0.41
2.80 Mbps	1.98	0.34

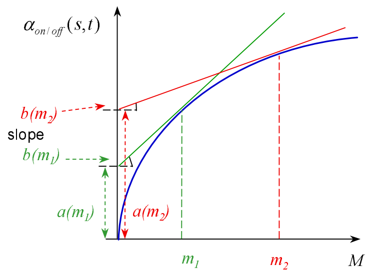
$h=1.5$ Mbps	$st=1$ sec/Mbit	
M	$a(m)$	$b(m)$
0.20 Mbps	0.06	1.59
0.75 Mbps	0.37	0.85
1.50 Mbps	0.72	0.52

$h=3$ Mbps	$st=2$ sec/Mbit	
M	$a(m)$	$b(m)$
0.20 Mbps	1.18	2.41
0.75 Mbps	1.82	0.66
1.50 Mbps	2.16	0.33
2.25 Mbps	2.36	0.22
2.80 Mbps	2.46	0.18

$a(m)$ => $/sec
$b(m)$ => $/Mbit

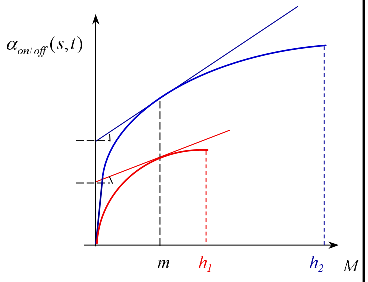
How $\mathbf{a}(m)$ and $b(m)$ Vary

- For fixed h, s, t, as m increases:
  - $a(m)$ *increases*
  - $b(m)$ *decreases*
  - charge for time *increases*, while charge for volume *decreases*, because the ability for multiplexing diminishes



How a(m) and b(m) Vary (continued)

- For fixed m, s, t, as h increases :
  - both $a(m)$ and $b(m)$ increase
  - ➜ the source is *more bursty*, thus reserving and using more resources



Simple Charging under Multiple Leaky Buckets

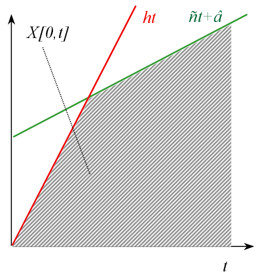
- ON/OFF bound corresponds to a *single* leaky bucket, constraining the peak rate h

$$\alpha_{on/off} = \frac{1}{st}\log\left[1+\frac{m}{h}\left(e^{sth}-1\right)\right]$$

- Under *multiple* leaky buckets, we can use the *tighter* bound

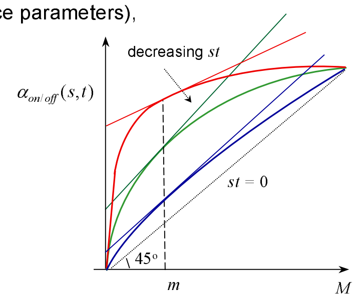
$$\alpha_{leaky\_bucket} = \frac{1}{st}\log\left[1+\frac{tm}{H(t)}\left(e^{sH(t)}-1\right)\right]$$

where $H(t) := \min_{k\in K}\{\rho_k t+\beta_k\}$



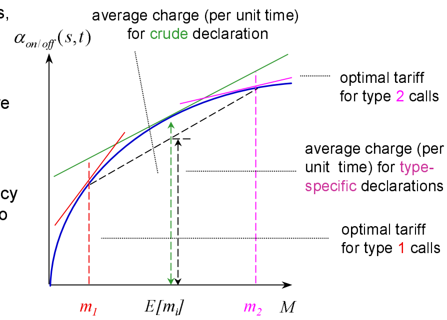
How a(m) and b(m) Vary (continued)

- For fixed m, h (i.e. source parameters), when st decreases :
  - $a(m)$ decreases
  - $b(m)$ increases and then *decreases* to 1
  - ➜ for small st time-charge *decreases*, and volume-charge *increases*, because the ability for multiplexing increases

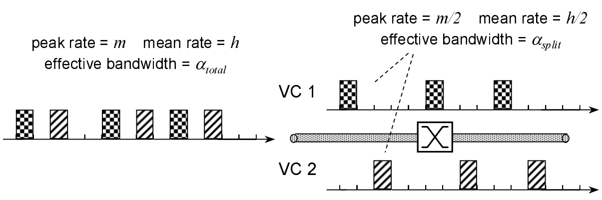


More on Incentives

- Two *types* of calls, with mean rates m_1 and m_2
- User has incentive to *differentiate* declarations
- Additional accuracy also *informative* to network



Splitting of Traffic



- Splitting leads to *benefit* for the user and *loss* for the network
  - user pays *less* total charge, because

$$2\alpha_{split} < \alpha_{total}$$

  - correlated traffic streams are erroneously charged as independent ones

Discouraging Splitting of Traffic - Fixed Charge

- Traffic splitting is undesirable to provider, because:
  - may lead to reduced revenue
  - set of available VPI/VCI may be exhausted
  - increased signaling overhead for setting more VCs
- ➜ Splitting should be discouraged => add a *fixed* charge per VC
- ➜ Total Charge = $a(m)\cdot T + b(m)\cdot V + c(m)$
- However, traffic splitting could be *beneficial* to provider, if sub-streams can only be accommodated through *different routes*

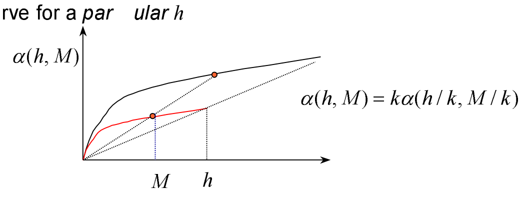
Examples of a, b, c Tariffs

$h = 3$ Mbps st = 1 sec/Mbit			
M	$a(m)$	$b(m)$	$c(m)$
0.20 Mbps	0.26	2.80	1.26
0.75 Mbps	0.93	1.10	5.65
1.50 Mbps	1.46	0.60	8.30
2.25 Mbps	1.81	0.41	10.05
2.80 Mbps	1.98	0.34	10.90

- Fixed charge $c(m)$
  - is expressed in $
  - was taken (in the examples) as $a(m)$*5sec+1

Alternative Discouraging of Traffic Splitting

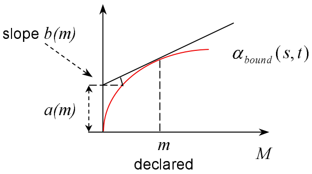
- Use *homothetic* tariffs, starting from the effective bandwidth curve for a *par ular h*



- Pros: *convexity* makes users reveal their mean rates, no incentive to split
- Cons: charge *not* proportional to effective bandwidth (but close!)

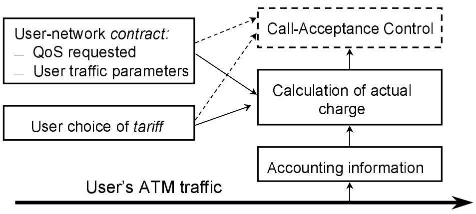
Extensions to the Simple Charging Scheme

- Simple charging scheme bounds the effective bandwidth according to the ON/OFF model
- *Other concave* bounds, parameterised with mean rate, can also be used
  - *Same* approach: charge per unit time derived according to the *tangent* selected by the user



Outline of the VBR-Charging Approach

- Total charge for a call is $a(...)T+b(...)V+c(...)$ where
  - T = duration of call (e.g. seconds)
  - V = volume of call (e.g. Mbits or Mcells)
  - $a(...)$, $b(...)$, $c(...)$ capture QoS, traffic contract parameters, and user's choice of tariff



Simpler Charging: Dispensing with Duration

- The time-component of charge can be eliminated
  - total charge = $b(m)\cdot V + c(m)$
  - tariff will be simpler
  - dependence of usage-charge on QoS will be clearer
- Motivation:
  - $c(m)$ can be set to account for *typical time-charge* too
  - For reasonable user declarations: $T \approx V/m$ and

$$a(m)\cdot T + b(m)\cdot V + c(m) \approx a(m)\cdot(V/m) + b(m)\cdot V + c(m) = b'(m)\cdot V + c(m)$$

- However, users will have *no* incentive to close connections
  - set of available VPI/VCI may be exhausted
  - provider can limit the maximum number of VPI/VCIs permissible per user

Charging CBR Services

- Simple charging scheme can also be applied to CBR services
  - users should declare $m = h$
  - Total Charge = $a(h)\cdot T + b(h)\cdot V + c(h)$
  - Volume-charge does *not* vanish, because $b(h) \neq 0$
    - makes sense because unused bandwidth is taken by *ABR*
- CBR services should be charged *only* on the basis of time, if their peak rate is really *reserved*, and CBR is *not* multiplexed statistically
  - simpler scheme
  - already adopted in practice

Charging PVCs

- So far have only dealt with *Switched* VCs for VBR services (SVCs)
- Simple charging scheme can also be applied to *Permanent* VCs (PVCs) for VBR services
- However, PVCs can also be charged *only* on the basis of time, if they are *not* multiplexed statistically, due to their long duration
  - simpler scheme
  - already adopted in practice