Class-of-Service in IP Backbones: Informing the Network Neutrality Debate

Murat Yuksel (University of Nevada, Reno, yuksem@cse.unr.edu), **K. K. Ramakrishnan** (AT&T Labs-Research, kkrama@research.att.com),
Shiv Kalyanaraman (shivkuma@ecse.rpi.edu, Rensselaer Polytechnic Inst.), **Joseph D. Houle** (jdhoule@att.com, AT&T), **Rita Sadhvani** (rsadhvani@att.com, AT&T)

ACM SIGMETRICS, Annapolis, MD, June 2008

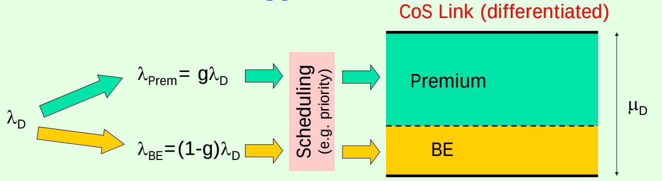
Goal

Quality of Service has been studied for many years. For an IP network, how much capacity do we need to over-provision with a "best-effort" service compared to one that provides class-of-service differentiation? This issue has been raised in the context of recent network neutrality debate. We quantify this required extra capacity.

Motivation

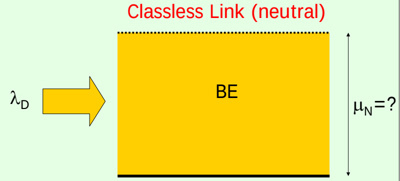
- Media-rich applications require performance guarantees:
  - e.g.: VoIP requires <300ms round-trip delay, <1% loss
- How to respond to these application needs?
  - **CoS approach:** provide priority (i.e. higher class) to premium traffic
  - **Classless (best-effort) service approach**: over-provision the capacity

Two Service Types: CoS vs. Classless



GIVEN: μ_D, λ_D and a **performance target** (i.e. t_{target} or p_{target})

FIND: What is **the minimum** μ_N that gives the same performance as in the premium class of the CoS case?



REC: Required Extra Capacity

REC = <required classless link capacity> - <CoS link capacity>

= $\mu_N - \mu_D$ (rate)

= $100(\mu_N/\mu_D - 1)$ (%)

How to quantify REC for a realistic network?

- First quantify REC for a link, i.e., "link model".
- Then, extend the quantification to a network, i.e., "network model".

Poisson Traffic:

M/M/1 and M/M/1/K Link Models

$$t_{target} = \frac{1+(1-g)\rho}{1-g\rho}$$

$$p_{target} = \frac{1-g\rho}{1-(g\rho)^{K+1}}(g\rho)^K$$

$$REC_{delay} = 100\left[\frac{(1-g\rho)}{1+(1-g)\rho}+\rho-1\right]$$

$$REC_{loss} = 100\left[\frac{\rho}{\sqrt{g\rho}}\sqrt[2K]{\frac{1-(g\rho)^{K+1}}{1-g\rho}}-1\right]$$

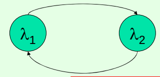
where $\rho = \lambda_D/\mu_D$ is the aggregate traffic load at the CoS link, and K is the buffer size available to each of the two CoS classes

Both the performance target and the REC can be expressed in terms of two key parameters: (i) ρ – utilization, (ii) g – proportion of premium traffic.

Higher utilization causes higher REC.

More Bursty, Short-Range Dependent: MMPP Traffic

Two-state MMPP traffic

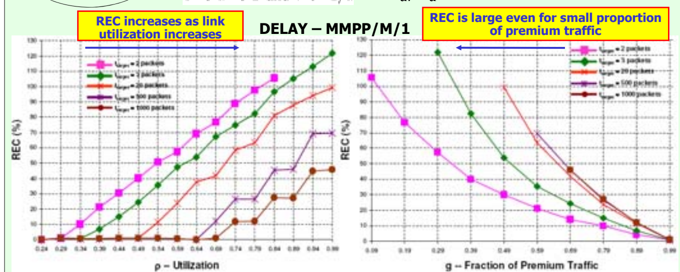


$$\lambda_1 = a\lambda_t$$
$$\lambda_2 = ar\lambda_t$$
$$0 < a < 1 \text{ and } r > 1/a$$

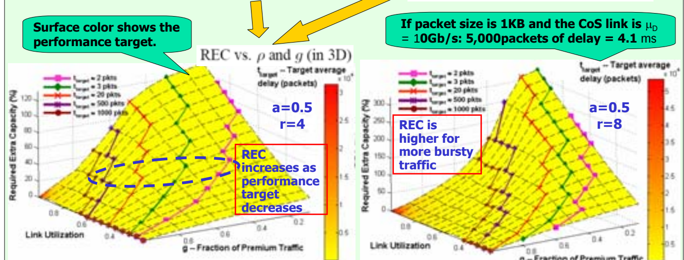
$$\pi_1 = \frac{ar-1}{ar-a} \qquad \pi_2 = \frac{1-a}{ar-a}$$

- The target average traffic rate is in between λ_1 and λ_2.
- Larger r means more bursty traffic.

DELAY – MMPP/M/1

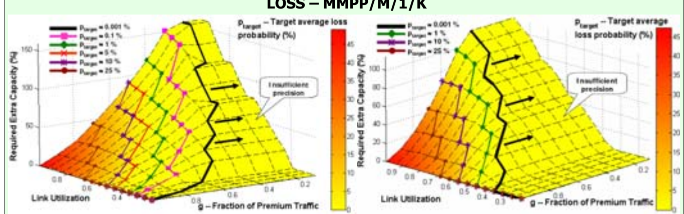


REC vs. ρ for fixed t_{target} (in 2D) REC vs. g for fixed t_{target} (in 2D)



REC vs. ρ and g (in 3D)

LOSS – MMPP/M/1/K

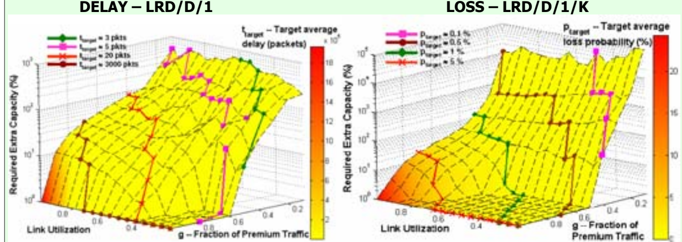


(a) $a = 0.5, r = 4, K = 25ms$ (b) $a = 0.5, r = 4, K = 100ms$

REC for the same performance target decreases as buffer size increases

Long-Range Dependent (LRD) Traffic

DELAY – LRD/D/1 **LOSS – LRD/D/1/K**



(a) DELAY – LRD/D/1: Hurst=0.75 (b) LOSS – LRD/D/1/K: K=100ms, Hurst=0.75

- Internet traffic is known to be LRD with Hurst parameter value ranging between 0.7 and 0.9.
- REC for Hurst=0.75 is significantly higher than our 2-state MMPP model results.
- We also observed that REC increases as Hurst value increases towards 0.9.

Network REC (NREC) for Edge-to-Edge (g2g) Performance Targets

- We build a network model to calculate "Network REC", i.e., the required extra capacity to provide the same g2g performance as the premium class performance of a CoS network.
  - We used Rocketfuel topologies, and gravity-based traffic demand and BFS-based link capacity models.

Input to the network model

- *The traffic matrix:* $T_{N\times N}$
- *Topology information:* Adjacency matrix $A_{N\times N}$, link weight matrix $W_{N\times N}$, link propagation delay matrix $S_{N\times N}$, link capacity matrix $C_{N\times N}$
- *The link model:* The link model formulates the REC ($\mu_N - \mu_D$) for a given traffic load (i.e., $\lambda_{F\times 1}$) and performance goal.
- *Premium class performance:* t_{target} or p_{target}, the performance goal to be achieved.

Steps of NREC calculation

- *Step 1:* Construct the *routing matrix* $R_{F\times L}$ based on shortest path first (Djiskstra's) algorithm using the topology information $A_{N\times N}$ and $W_{N\times N}$.
- *Step 2:* Form the *traffic vector* $\lambda_{F\times 1}$ from $T_{N\times N}$.
- *Step 3:* Calculate the *traffic load on each link* by performing the matrix operation $Q = R^T\lambda$, where $Q_{L\times 1}$ is the link load vector (in Mb/s).
- *Step 4:* Check the *feasibility* of the traffic load and routing. If any link's capacity is less than the load on to that link, then we fix the infeasibility by increasing the capacity of that link.
- *Step 5:* Calculate the *per-link* REC by using Q_i as the total traffic rate for the ith link and the performance goal t_{target} or p_{target} for that link i.
- *Step 6:* Calculate the *network* REC (NREC) by averaging the per-link RECs from Step 5.

NREC

$$NREC_A = \frac{\sum_{l\in L} c_l REC_l}{\sum_{l\in L} c_l}$$

total extra capacity needed on the whole network

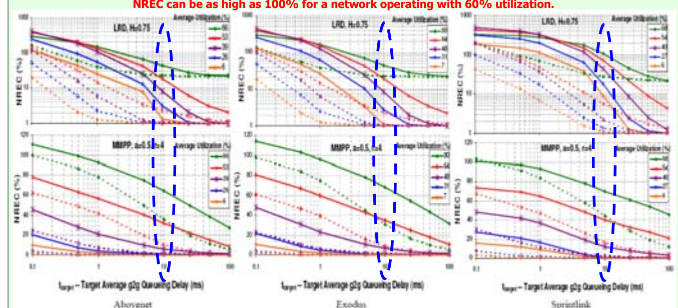
$$NREC_I = \frac{\sum_{l\in L} REC_l}{|L|}$$

average extra capacity needed on each link

NREC for Abovenet, Exodus, Sprintlink

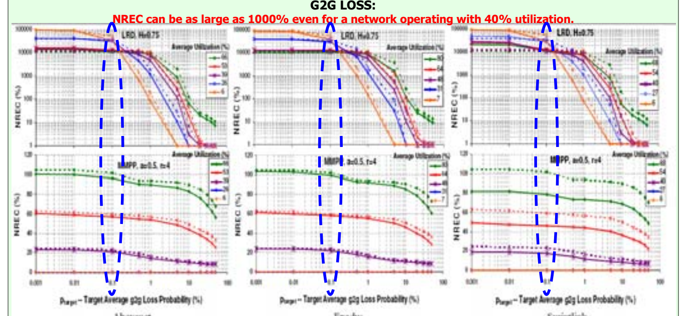
G2G QUEUEING DELAY:

NREC can be as high as 100% for a network operating with 60% utilization.



G2G LOSS:

NREC can be as large as 1000% even for a network operating with 40% utilization.



Conclusions

- REC can be large as the amount of premium traffic is small – a situation likely initially
- REC increases
  - as the burstiness of traffic increases.
  - as the network utilization increases.
- REC is of particular concern when the proportion of premium class traffic requiring delay or loss assurances is small.
- With CoS in IP backbones, the *capacity needed* can be significantly smaller than the classless (i.e., over-provisioning) approach.
- However, further research is necessary to estimate the actual costs of the two approaches, as scheduling and management complexity need to be considered.