Flow Level Models of DiffServ Packet Level Mechanisms

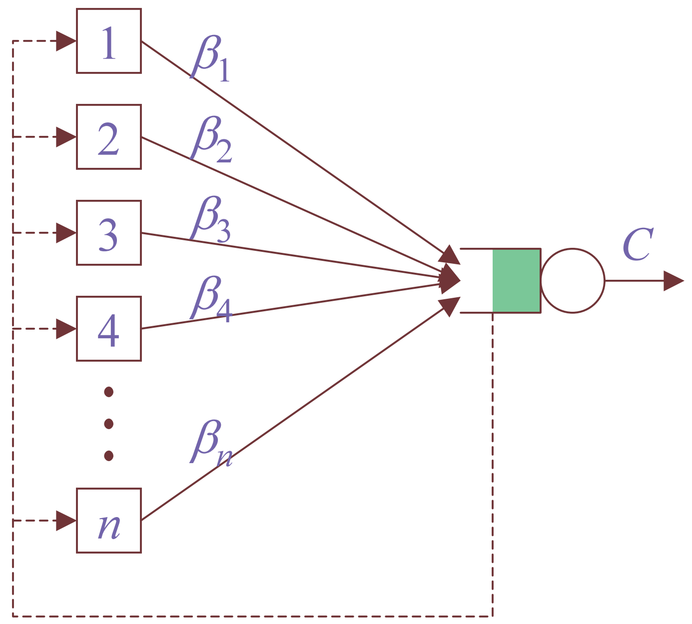
Samuli Aalto & Eeva Nyberg
Networking Laboratory
Helsinki University of Technology

samuli.aalto@hut.fi

Background: Ideal flow level model of Best Effort TCP

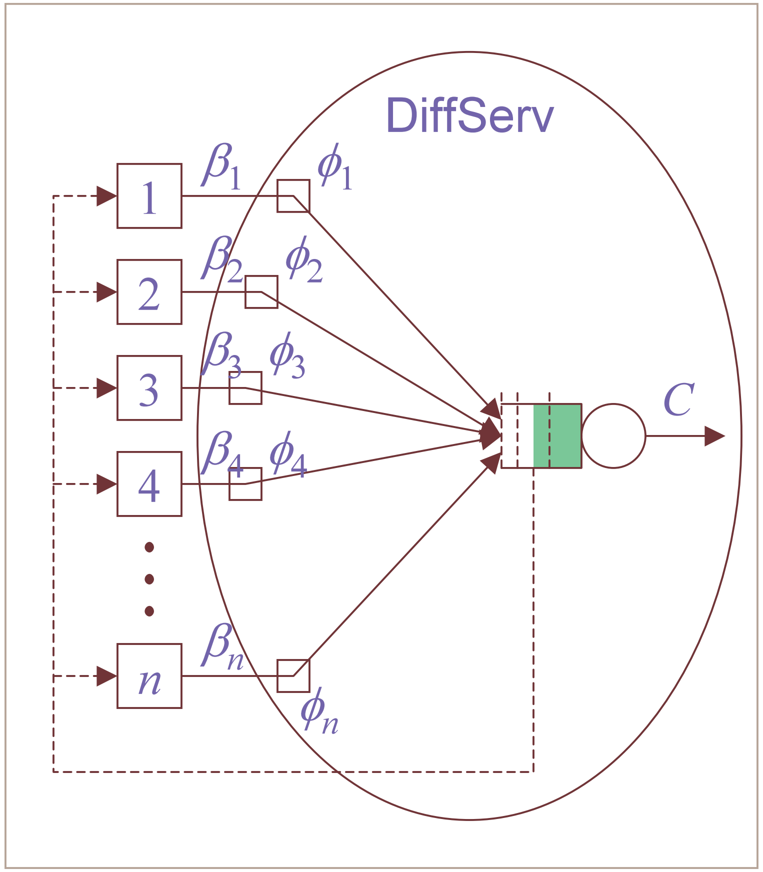
- Setting:
  - single (bottleneck) link with capacity $C = 1$
  - loaded by a **fixed** number, n, of TCP flows with similar RTT
- **Ideal** model for bandwidth sharing:
  - fair sharing = equal bw shares
    - due to elasticity and cooperation of TCP

$$\beta_m \equiv \beta = \frac{1}{n}$$



The General Problem: Bandwidth sharing among TCP flows in a DiffServ cloud

- Setting:
  - single link in a DiffServ cloud
  - loaded by a fixed number, n, of TCP flows with similar RTT and belonging to the same PHB class
- What are the target bandwidth shares in this case?
  - Assured Service approach
  - Relative Service approach
- Is it possible to realize these bw shares by DiffServ mechanisms?
  - Without any per flow scheduling scheme, a non-trivial problem



Target bandwidth shares: Assured Service vs. Relative Service

- Starting point:
  - the traffic profile of a flow is defined by a single parameter, the **reference** (contracted) **rate** ϕ
- **Assured Service** approach:

$$\beta = \phi + \frac{R}{n}$$

  - bw share = reference rate + equally shared leftover capacity
  - problem: without admission control, ϕ cannot be guaranteed
- **Relative Service** approach:

We take this one!

$$\beta \propto \phi$$

  - bw share should be proportional to the reference rate
  - problem: what to do with BE flows with $\phi = 0$

Conditioning of flows: Priorities

- Consider a flow with reference rate ϕ
- Traffic of this flow is conditioned at a boundary DiffServ node
  - sending rate θ measured
  - packets marked based on θ and ϕ
- We assume I different marks corresponding to I **priority levels**
  - marks $1,2,\ldots,I$, mark 1 = lowest priority, mark I = highest priority
  - logaritmic threshold function: priority decreases from i to $i-1$ whenever measured rate θ exceeds threshold $t(\phi,i)$, where

$$t(\phi,i) = \phi \cdot 2^{(I-2i+1)/2}$$

  - Note: AF specification defines three levels of drop precedence indexed reversely to our priorities

Conditioning of flows: Marking principles

- **Per flow marking**:
  - All packets of a flow are marked to the same priority level c, where

$$c = \max\{i = 1,\ldots,I \mid \theta \leq t(\phi,i)\}$$

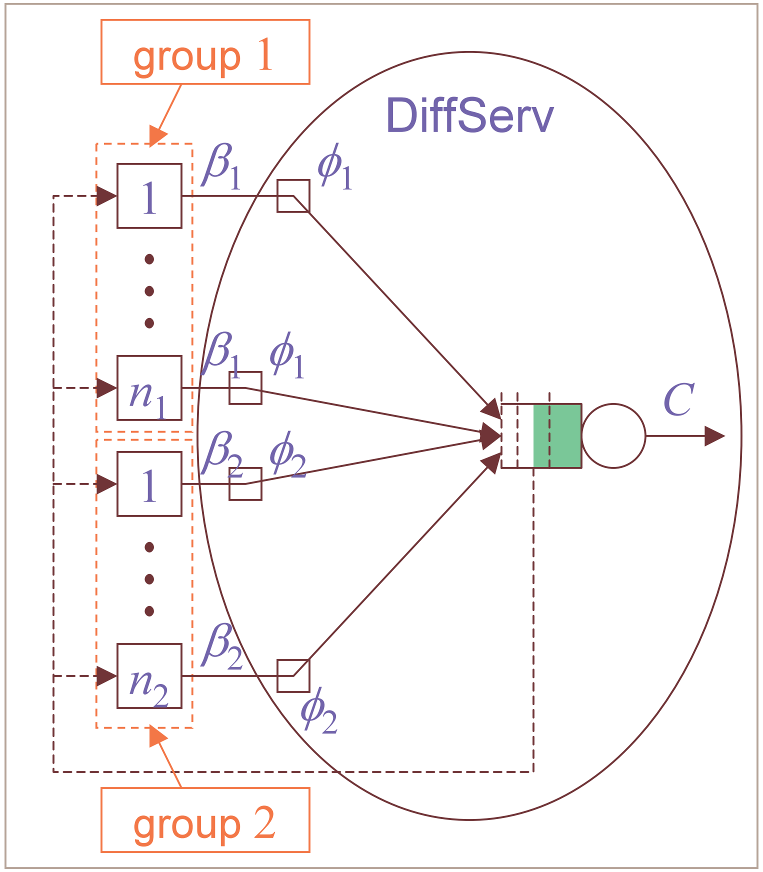
- **Per packet marking**:
  - Packets of a flow are marked to priority levels $i = c,\ c{+}1,\ldots,\ I$ resulting in substreams i with rates

$$\theta(i) = \min\{\theta, t(\phi,i)\} - \min\{\theta, t(\phi,i+1)\}$$

- Hypothesis:
  - **EWMA** (referred to in SIMA proposal) applies per flow marking
  - **LB** (referred to in AF specification) applies per packet marking

The Specific Problem:
Effect of the marking principle on the bandwidth sharing

- Setting:
  - single link in a DiffServ cloud
  - loaded by two groups
    - n_1 TCP flows with reference rate ϕ_1 and
    - n_2 TCP flows with reference rate ϕ_2 (such that $\phi_2 > \phi_1$)
  - but only one PHB class
- For each marking principle separately, we develop a simple flow level model to approximate the bw shares β_l for flows in groups $l = 1,2$



Handling of flow aggregates

- Modelling assumptions:
  - **Strict priority principle**
    - Between priority levels,
      the bandwidth is shared according to strict priorities
  - **Ideal TCP principle**
    - Within each priority level,
      the bandwidth is shared as fairly as possible
- Remark:
  - The strict priority principle (leading typically to starvation problems) is just for our modelling purposes. However, due to elasticity of flows, the starvation problem is avoided here!

Bandwidth shares: SIMA-NRT class

- SIMA-NRT class (applying per flow marking):

$$\beta_1(i) = \min\{\frac{C(i)}{n_1(i)+n_2(i)}, t_1(i)\}$$

$$\beta_2(i) = \min\{\max\{\frac{C(i)}{n_1(i)+n_2(i)}, \frac{C(i)-n_1(i)t_1(i)}{n_2(i)}\}, t_2(i)\}$$

$$C(i) = \max\{C(i+1) - n_1(i+1)t_1(i+1) - n_2(i+1)t_2(i+1), 0\}$$

- Notation:
  - $\beta_l(i)$ = bandwidth share for a flow in group l and at priority level i
  - $t_l(i)$ = threshold rate for flows in group l and at priority level i
  - $n_l(i)$ = number of flows in group l and at priority level i
  - $C(i)$ = remaining capacity at priority level i (with $C(I) = C = 1$)

Bandwidth shares: AF class

- AF class (applying per packet marking):

$$\beta_1(i) = \min\{\beta_1(i+1) + \frac{C(i)}{s_1(i)+s_2(i)}, t_1(i)\}$$

$$\beta_2(i) = \min\{\beta_2(i+1) + \max\{\frac{C(i)}{s_1(i)+s_2(i)}, \frac{C(i)-s_1(i)\delta_1(i)}{s_2(i)}\}, t_2(i)\}$$

$$C(i) = \max\{C(i+1) - s_1(i+1)\delta_1(i+1) - s_2(i+1)\delta_2(i+1), 0\}$$

- Additional notation:
  - $\delta_l(i) = t_l(i) - t_l(i+1)$
  - $s_l(i) = n_l(1) + \ldots + n_l(i)$

Interaction between TCP and DiffServ mechanisms

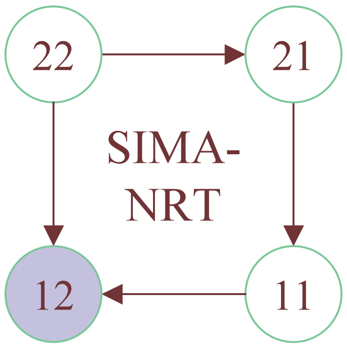
- Modelling assumption:
  - **Individual optimisation principle**
    - Interaction between TCP and DiffServ traffic conditioning makes the flows to maximize their bandwidth share individually
- This assumption leads to
  - a **game between the two groups**
- This assumption is needed to
  - determine the priority levels c_l of the two groups as a function of the number n_l of flows in each group
- Note:
  - Priority levels c_l determine the $n_l(i)$'s for all l and i, from which the bandwidth shares $\beta_l(c_l)$ for each group l can be calculated

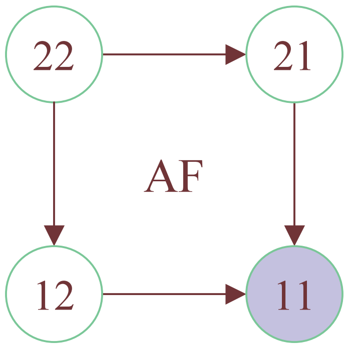
Game between the two groups: Numerical example

$$\phi_1 = 0.040, \phi_2 = 0.080, n_1 = n_2 = 10, I = 2$$

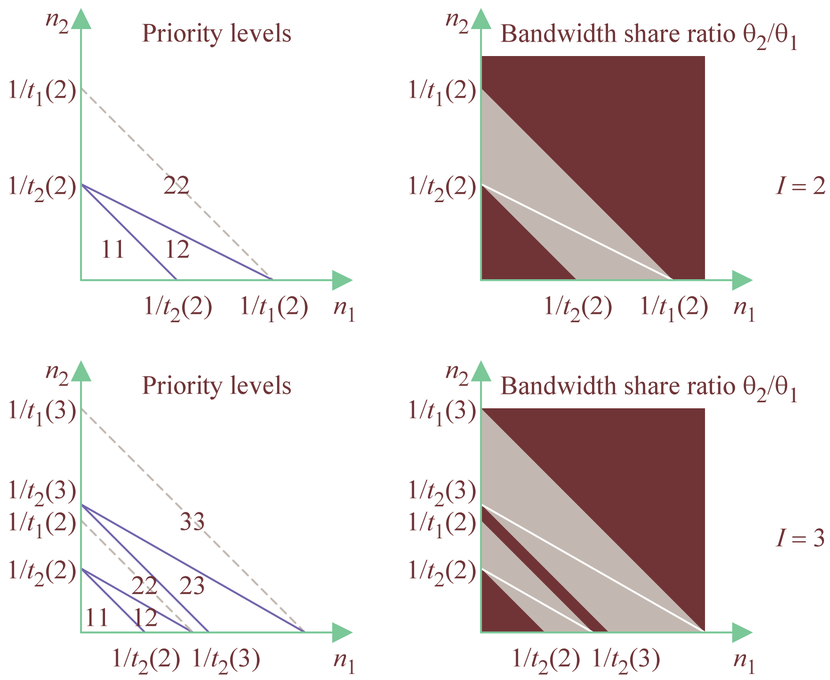
SIMA-NRT		$l = 2$, $c_2 = 2$	$l = 2$, $c_2 = 1$
$l = 1$	$c_1 = 2$	0.028 0.057	0.028 0.072
$l = 1$	$c_1 = 1$	0.043 0.057	0.050 0.050

AF		$l = 2$, $c_2 = 2$	$l = 2$, $c_2 = 1$
$l = 1$	$c_1 = 2$	0.028 0.057	0.028 0.072
$l = 1$	$c_1 = 1$	0.043 0.057	0.036 0.064





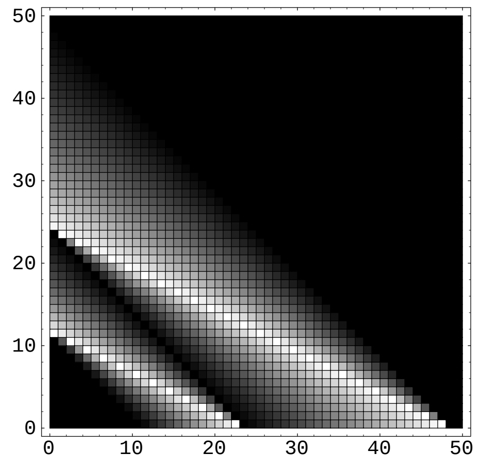
Results:
SIMA-NRT class



White: $\theta_2/\theta_1 = \phi_2/\phi_1$

Gray: $1 < \theta_2/\theta_1 < \phi_2/\phi_1$

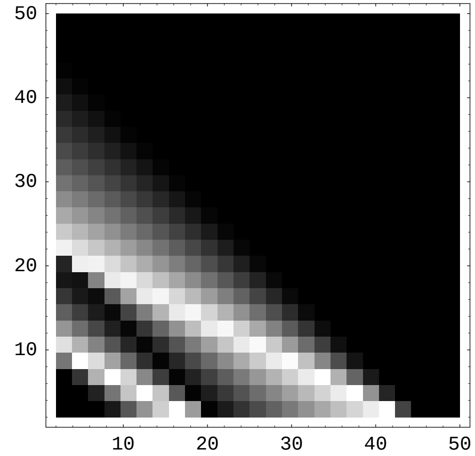
Black: $\theta_2/\theta_1 = 1$



Flow level model

$I = 3$

$\phi_2/\phi_1 = 2$

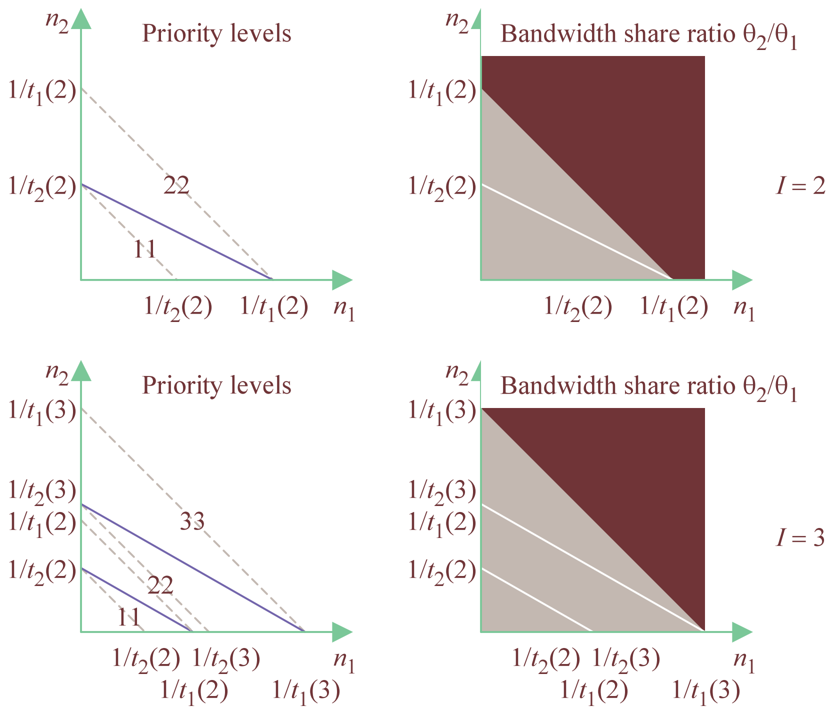


Packet level model

$I = 3$

$\phi_2/\phi_1 = 2$

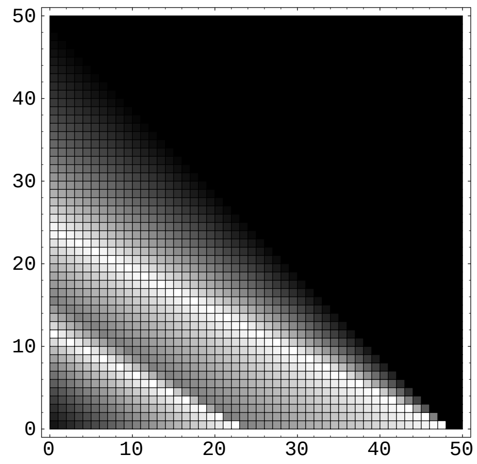
Results:
AF class



White: $\theta_2/\theta_1 = \phi_2/\phi_1$

Gray: $1 < \theta_2/\theta_1 < \phi_2/\phi_1$

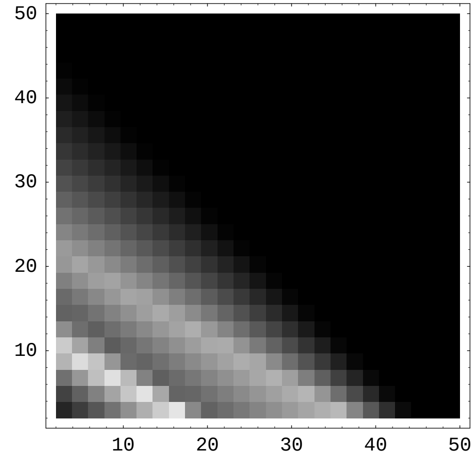
Black: $\theta_2/\theta_1 = 1$



Flow level model

$I = 3$

$\phi_2/\phi_1 = 2$



Packet level model

$I = 3$

$\phi_2/\phi_1 = 2$

Discussion

- Observations:
  - Ideal bandwidth shares (according to the Relative Service approach) are not possible to be achieved comprehensively by the DiffServ packet level mechanisms
  - According to our static flow level model, restricted to a single PHB class, a better approximation of this ideal is achieved by the AF scheme applying per packet marking principle
  - The more priority levels, the better the approximation achieved by the DiffServ mechanisms!
- Future work:
  - multiple parallel PHB classes (with TCP and UDP traffic)
  - more general topologies
  - **dynamic** flow level model where the number of flows varies randomly

THE END