An implementation of a service class providing assured TCP rates within the AQUILA framework

Christof Brandauer

Peter Dorfinger

Salzburg Research

Salzburg, Austria

http://www.salzburgresearch.at/

AQUILA PMM class

- **for ...**
  - greedy, TCP-based applications
  - minimum rate requirement
  - no delay requirement
- **QoS objective**
  - P (achieved rate > R) very high,   R = requested rate
- **sample applications**
  - (large) file transfer within guaranteed time
    - x-ray images, content distribution, ...
  - streaming media (TCP)

TCP rate control

- **"invasive"**
  - on-the-fly modification of TCP header (receiver window, ACK number)
  - accelerate / delay ACKs
  - e.g. Packeteer
- **packet marking / preferential dropping**
  - mark packets at domain ingress
  - preferential dropping in core routers
- **other**
  - receiver window
    - effectively disable congestion avoidance/control mechanisms
    - control via advertised receiver window
  - rate shaper
    - exclusiv control over p, RTT [PROMS 02]

Token bucket marking / preferential dropping

- investigated feasibility for PMM implementation
- existing model of TCP + token bucket marking (TBM)
- fits well into AQUILA architecture
- can be implemented with today's equipment (trial)

TCP + TBM

analytical model

- S. Sahu, P. Nain, D. Towsley, C. Diot, V. Firoiu, 2000
  "On Achievable Service Differentiation with Token Bucket Marking for TCP"
- derived from accurate TCP model

"under-subscribed" scenario

- def: $p_1 = 0$, $p_2 > 0$
- requested rate can be achieved
- formula for required marking profile

$$A = \begin{cases} R - \frac{3}{2Rp_2T^2} & R \leq \frac{3W}{2T} \\ \frac{4}{3}\left(R - \frac{3}{2T\sqrt{2}}\sqrt{Z + \frac{1}{p_2}}\right. & R > \frac{3W}{2T} \end{cases}$$

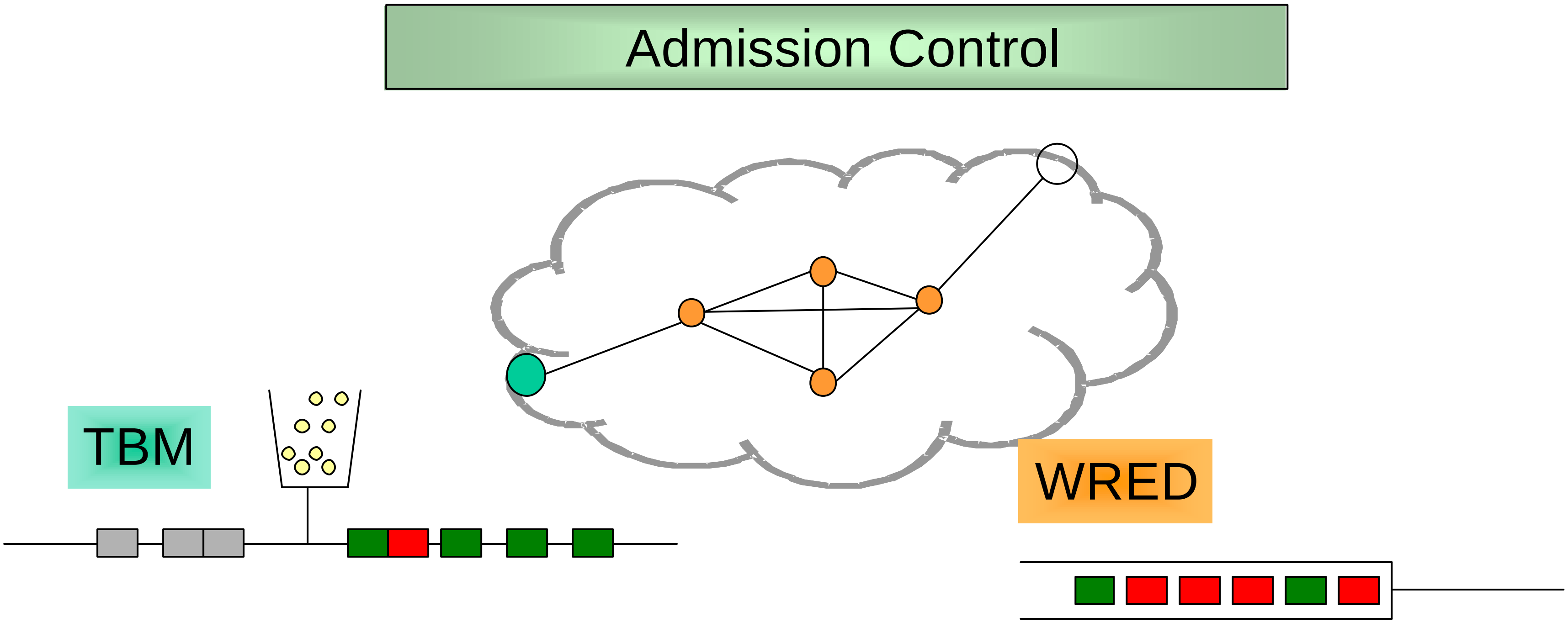
$$W = \sqrt{2(Z + 1/p_2)} + 2\sqrt{2Z}$$

A	token bucket rate
Z	token bucket size
R	requested rate
p_1	pdrop IN-profile
p_2	pdrop OUT-profile
T	round-trip-time

TCP + TBM cont.

- **promising model**
  - closed loop formulae for computation of marking profile
  - simple computation
  - model is shown to be accurate
    - wide range of p_2, T, R
- **PMM implementation**
  - on the basis of the TBM model
  - enforce under-subscribed scenario
    - admission control
    - queue management

PMM architecture



S. Sahu, P. Nain, D. Towsley, C. Diot, V. Firoiu "*On Achievable Service Differentiation with Token Bucket Marking for TCP*", 2000

WRED model, enforces conditions required by TBM model

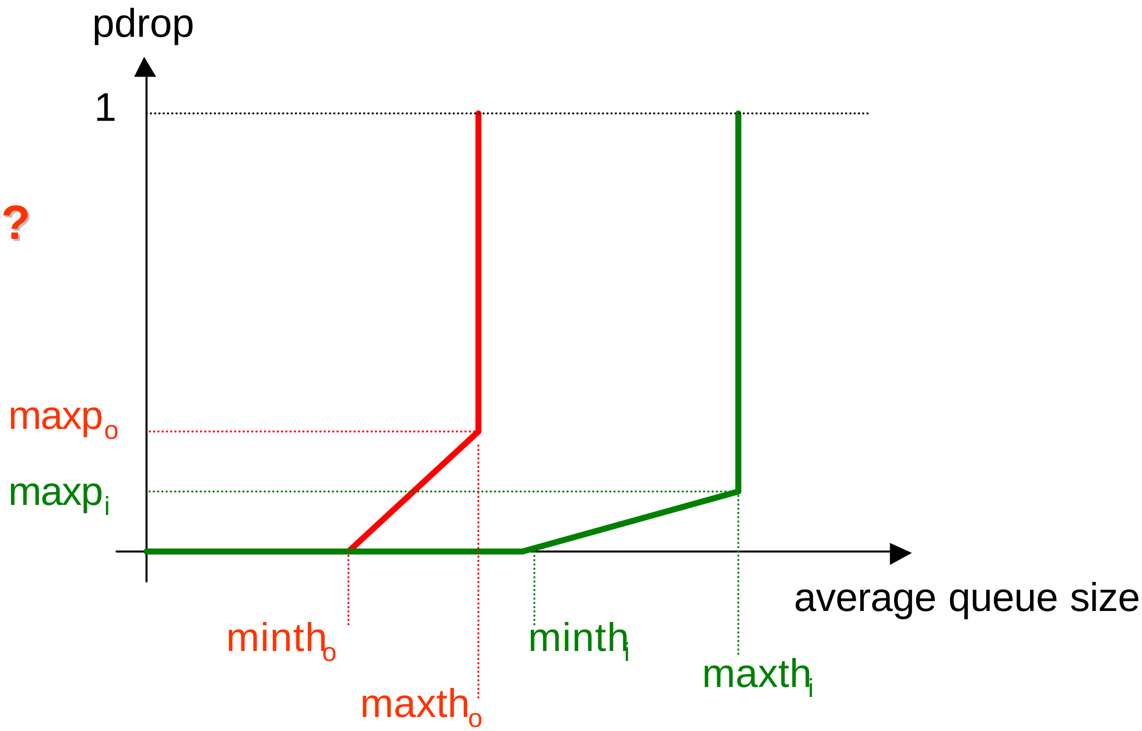
Admission Control

- **Declaration Based**
  - input: requested rate R
- **AC computation**
  - compute token bucket rate A
  - if A > R: require A resources
  - if A ≤ R: require R resources
  - -> require max(A,R)
- **AC rules**
  - $\sum_{i=1}^{N} \max(A_i, R_i) \leq \rho C$
  - R ≥ R_0 , $R_0 = \frac{1}{T}\sqrt{\frac{3}{2p_2}}$

C ... PMM capacity
ρ ... over-provisioning factor

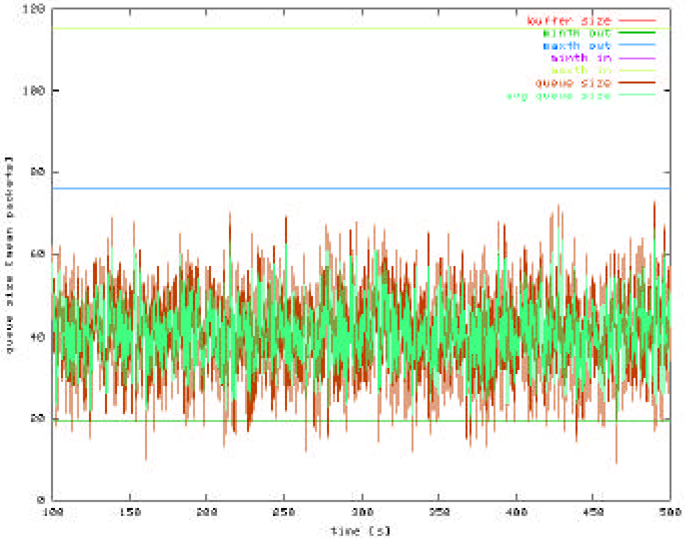
Queue Management

- **under-subscribed scenario**
  - drop no IN-packet, drop some OUT-packets
  - color-aware
- **two-color extension of RED**
- **how to configure ??**



WRED dynamics

- **wish list**
  - average queue size converges between $minth_o$ and $maxth_o$
  - bounded oscillation
  - 0 < instantaneous queue < buffer size



- **benefits**
  - $p_1 = 0$, $p_2 > 0$
  - full utilization of available capacity
  - rather constant pdrop for TCP flows
  - no timeouts, smooth sending behavior

RED model

- **quantitative RED model (IWQoS 01)**
  - developed for long-lived TCP flows
  - input: bottleneck bandwidth, RTT, #flows
  - output: RED parameters { minth, maxth, maxp, wq }

- **model composition**
  - analytical model of TCP sending behavior [PFTK]
  - analytical model for wq [Firoiu, Borden]
  - empirical model for maxth-minth

Idea of RED model

$$S_i \sim \frac{1}{T_i \sqrt{p}}$$

here: simple TCP model

$$p = \frac{\max p}{2}$$

$$\sum_{i=1}^{N} S_i = C$$

$$\Rightarrow C = \sum_{i=1}^{N} \frac{1}{T_i \sqrt{(\max p / 2)}}$$

$$\max th - \min th = f(C, T, N)$$

$$wq = g(p, T, T_0, RTO)$$

S	TCP sending rate
p	packet drop probability
T	average RTT
T_0	minimum RTT
C	link capacity
RTO	basic retransmission timeout

WRED model

- **extension of the RED model**
- **handling of OUT packets**
  - can only drop OUT packets
  - increase maximum drop probability for OUT packets to get same overall pdrop
  - need estimate (*adapt*) of out-share for correct adaptation
- **Handling of IN packets**
  - IN packets not needed for congestion control / avoidance (have AC)
  - thresholds for IN packets are not critical
  - minimize prob. of IN drop
    - set min / max thresholds to buffer size
    - no randomness

http://www.salzburgresearch.at/~cbrand/WREDmodel/

Service class parameterization

WRED

- estimate of out-share: *adapt*
  - varies over time
  - portion of “sold“ capacity
  - size of requested rate
- # flows: N
  - inherited from RED
  - $N_{min} = 0$; $N_{max} = C/R_{min}$

TBM

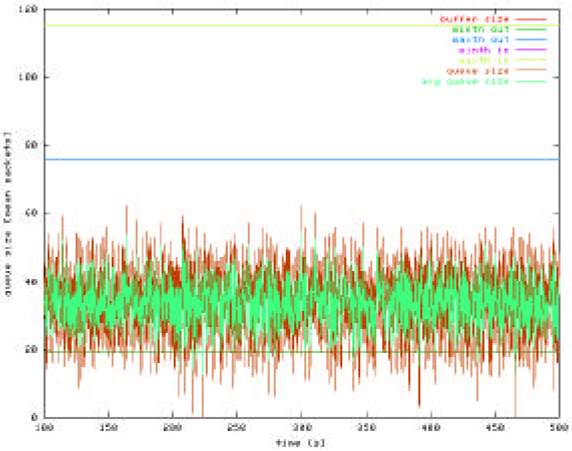
- p_2 = f (level of congestion, out-share)
- worst-case estimation not possible

RTT

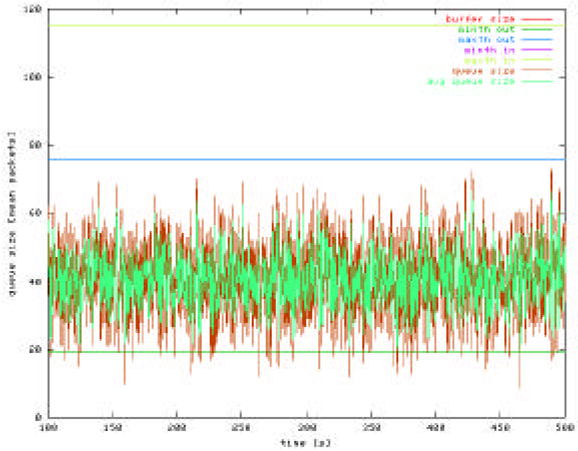
- propagation + transmission + WRED queueing delay
- ? number of bottlenecks
- ? different RTTs

WRED behavior

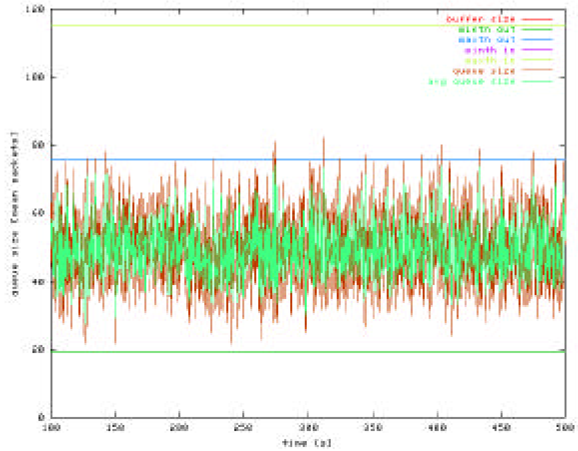
- desired convergence behavior
- no force OUT drops; no IN drops
- little sensitivity to correct estimation of out-share



90% OUT



50% OUT



10% OUT

PMM operation

PMM model

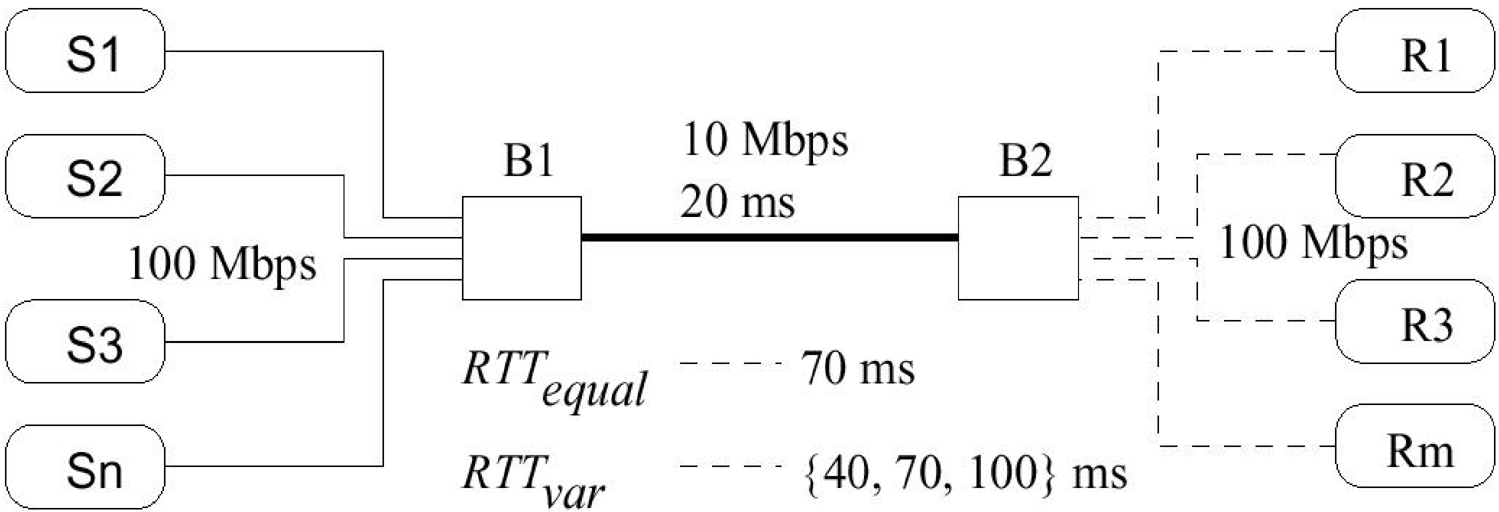
- ? behavior of AC + TBM + QM
- no analytical model that captures composite behavior

PMM configuration

- feasibility in general
- good / bad parameter sets
- dependencies between parameters
- influence of wrong estimation on usability of mechanisms
- support for testbed trials

Simulation study

- **explore parameter space**
  - many simulations
- **topology**



Variation of input parameters

- **r of AC**
  - ∈ {0.7, 0.8, 0.9}
- **# flows N (WRED)**
  - high: $N = rC / R_{\min}$
  - low: $N = rC / RR_{avg}$
- **RTT of flows**
  - ∈ {equal, unequal}
- **error in RTT estimation**
  - T_{dev} ∈ {0%, 25%, 50%, 75%, 100%}
- **requested rate factor rF**
  - rF ∈ {3, 5, 8}
- **requested rate distance rD**
  - rD ∈ {100, 200, 300}

full combination ⇨ 540 scenarios

Requestable rates [kbps]

- **rF = 3**
  - rD = 100: 200 | 300 | 400 | 500 | 600
  - rD = 200: 200 | 400 | 600
  - rD = 300: 200 | 500
- **rF = 5**
  - rD = 100: 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000
  - rD = 200: 200 | 400 | 600 | 800 | 1000
  - rD = 300: 200 | 500 | 800 | 1100
- **rF = 8**
  - rD = 200: 200 | 400 | 600 | 800 | 1000 | 1200 | 1400 | 1600
  - rD = 300: 200 | 500 | 800 | 1100 | 1400
  - rD = 400: 200 | 600 | 1000 | 1400

Simulation setup

- **TBM config**
  - p_2: 0.1
  - T = propagation delay + ($minth_o$ + $maxth_o$)/2
  - $T = T * (1 + T_{dev})$
- **WRED config**
  - number of flows
    - overestimated (N==high)
    - underestimated (N==low)
  - rho = <rho>
  - adapt = 0.5, packet size = 1500, delayed ACKs = Yes

Simulation setup

Traffic

- only PMM traffic
- simulated as bulk-data transfer (FTP data)
- flow duration: uniform(50-150s)
- request interarrival time: exp(2s)
- for each combination of rate and RTT: 1 traffic template
- for each template measure the “success“ ratio s_t $$s_t := \frac{\#\ flows : gput \geq R}{\#\ flows}$$
- success of a simulation S $$S = \min(s_t)$$

Runtime

- 50000 s
- bias against R_{max}

Simulation results

- 0 £ S £ 1
- configuration that achieves S > 0.99

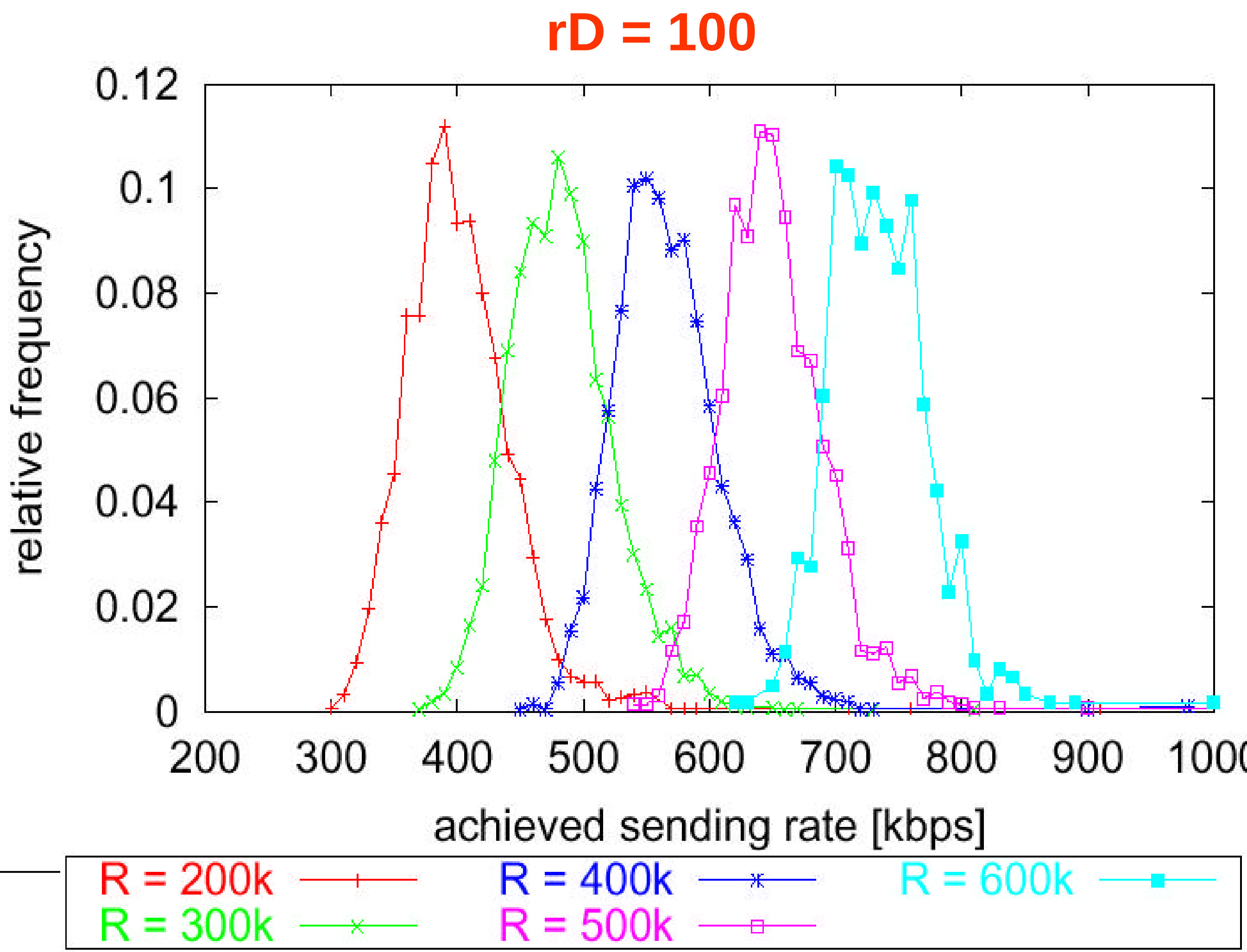
ρ of admission control	$\in \{0.7\}$
number of flows for the WRED model	$\in \{N_{high}\}$
link delays in topology	$\in \{RTT_{equal}, RTT_{var}\}$
error in RTT estimation T_{dev}	$\in \{0\%, 25\%, 50\%, 75\%, 100\%\}$
requested rate factor rF	$\in \{3\}$
requested rate distance rD	$\in \{100, 200, 300\}$

Good configuration

- r of AC = 0.7
  - enough safety margin to compensate imperfections
- # flows N = high
  - convergence behavior according to objective of the model
  - justifies correctness and usability
- varying RTTs
  - no major problem under “good“ configuration
  - some unfairness remains
- rF = 3
  - broad range of rate offering not possible
- choice of rD
  - no impact on success
  - impacts service differentiation witin PMM

Exemplary results

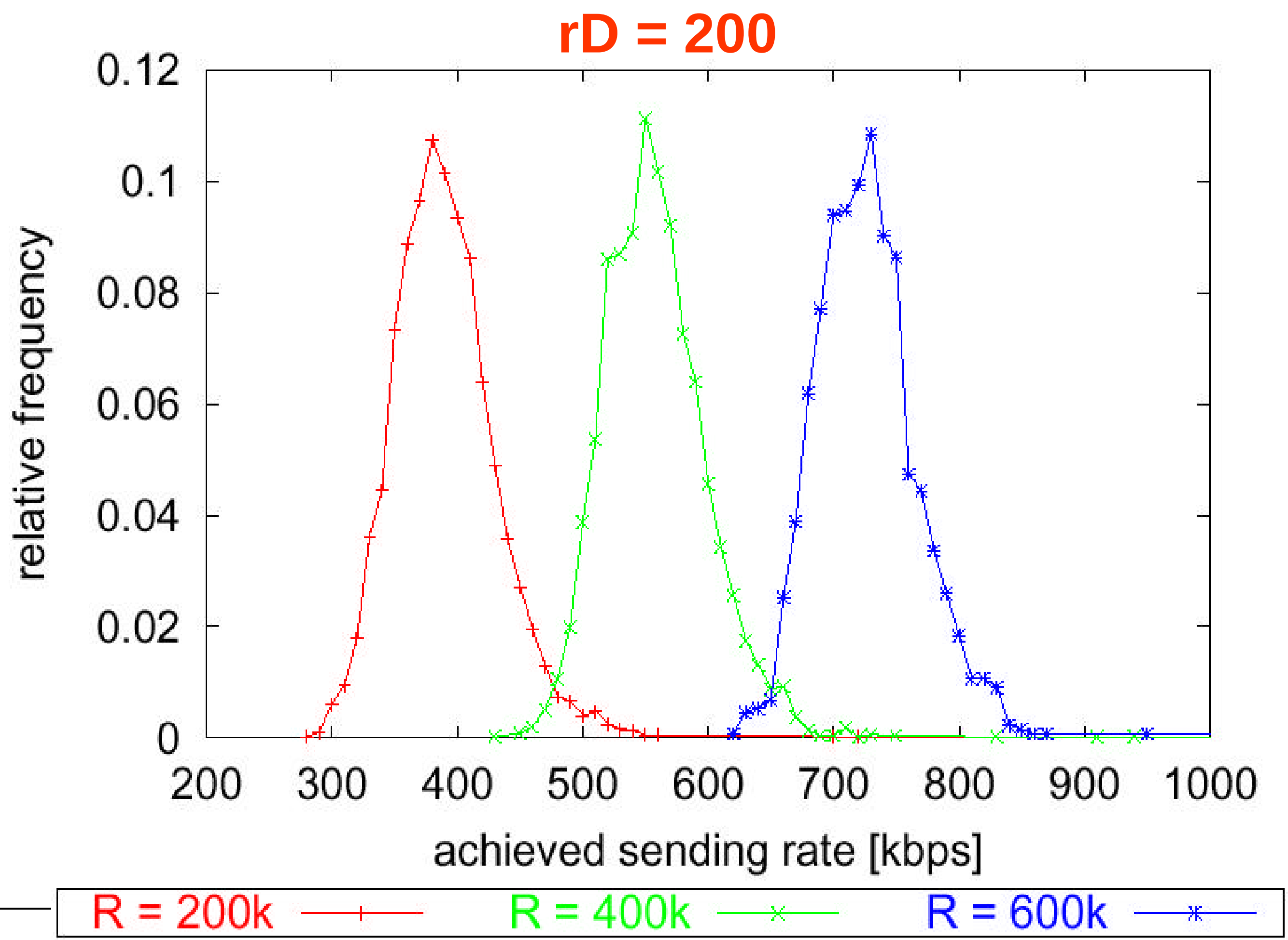
- r = 0.7
- N_{high}
- rF = 3
- T_{dev} = 50%
- RTT_{equal}



AQUILA

Exemplary results

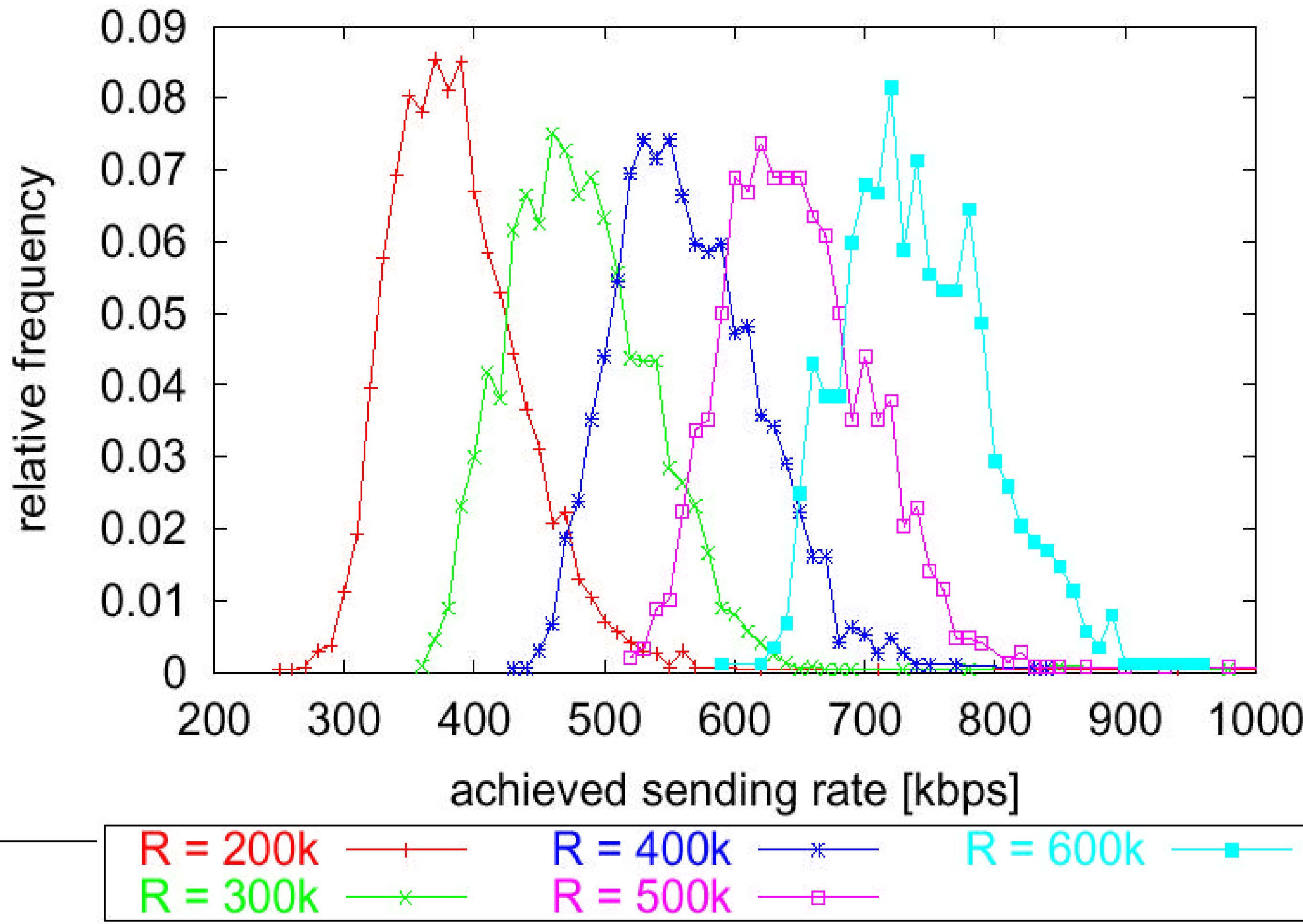
- r = 0.7
- N_{high}
- rF = 3
- T_{dev} = 50%
- RTT_{equal}



Exemplary results

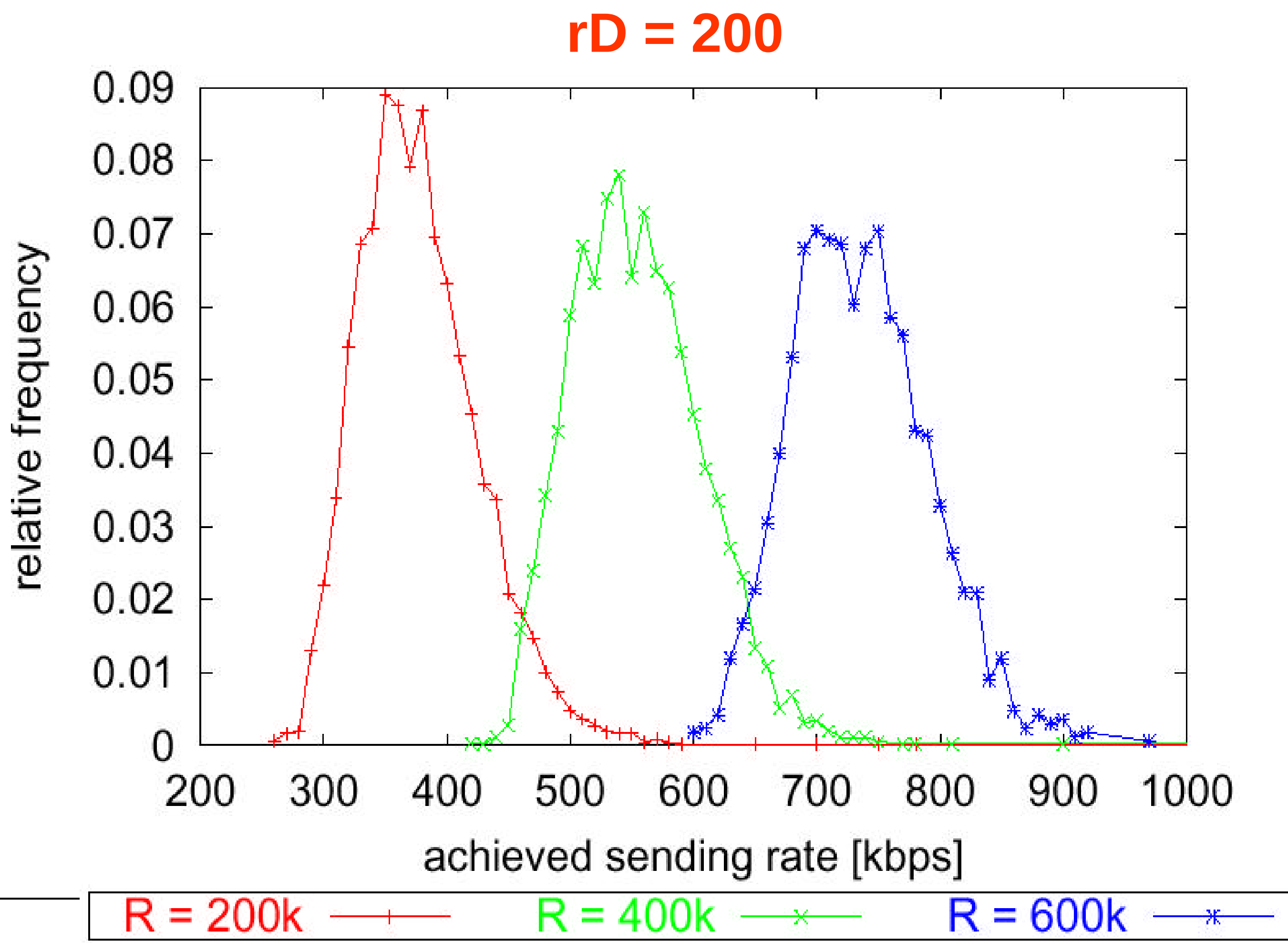
- r = 0.7
- N_{high}
- rF = 3
- T_{dev} = 50%
- RTT_{var}

rD = 100



Exemplary results

- r = 0.7
- N_{high}
- rF = 3
- T_{dev} = 50%
- RTT_{var}



Conclusions

- **PMM service feasible**
  - with AC / TBM / QM approach
  - despite imperfections
  - even under unrealistically high call blocking probabilities
- **Parameter setting**
  - insight on how to configure PMM
  - WRED model works nicely despite varying load / error in input params
- **Testbed trials**
  - significant router restrictions
    - only 64 packets buffer space for ALL service classes
    - cannot employ WRED configuration
  - throughput degradations
  - unfairness

Thank you!

Conclusions

- **Real-world experiments**
  - modeling of TCP sending behavior feasible ?
  - many different TCP implementations
  - applies to all approaches based on TCP models