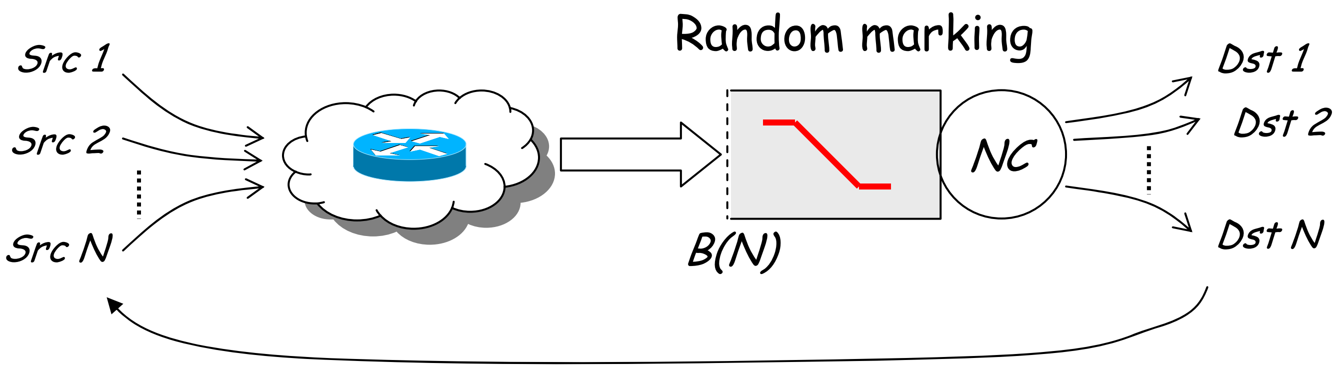
A Doubly-Stochastic Model for a TCP/AQM System under Aggressive Packet Marking

Do Young Eun and Xinbing Wang

Dept. of Electrical and Computer Engineering
North Carolina State University
March 22, 2006

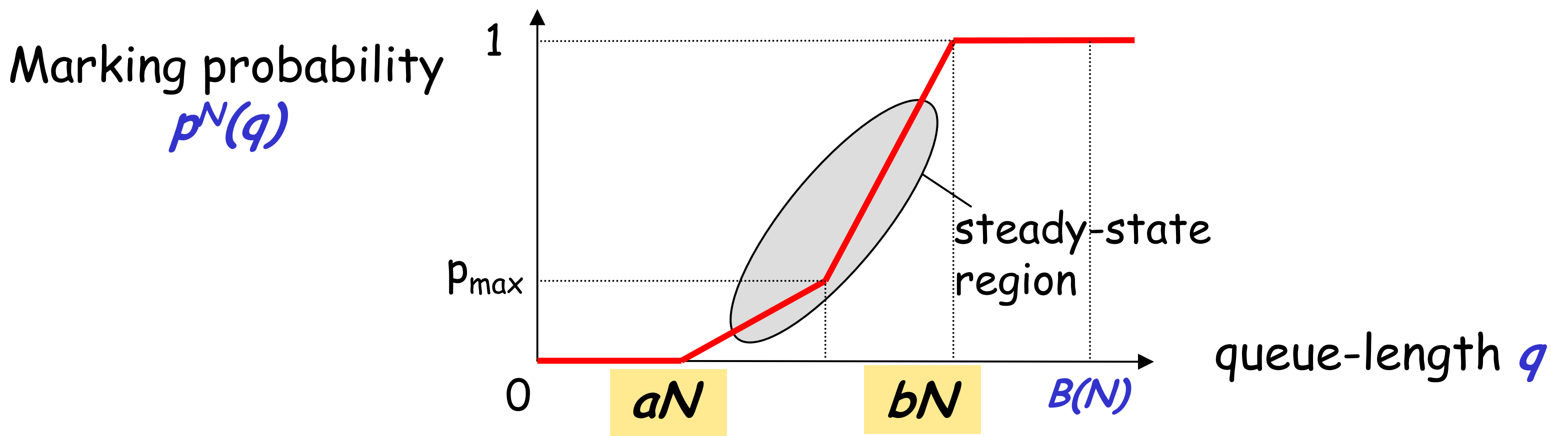
NC STATE UNIVERSITY

TCP/AQM Congestion Control



- More than 90% traffic carried via TCP
- It's a feedback system (equilibrium, stability)
  - TCP AIMD at senders vs. Active Queue Management at Routers
- Fluid approach based on averaged quantities has been the key techniques for design
  - Capacity scaling, Buffer sizing, Choice of TCP/AQM protocols

Existing Scales for Marking Function



- "Structural assumption": $p^N(Nx) = p(x)$
- Most work on TCP/AQM with N flows use this assumption
  - Shakkottai and Srikant 03, 04,   Deb 03
  - Tinnakornsrisuphap and Makowski 03, 04
- Results:
  - Stability criterion in terms of *p(x)* and other normalized network parameters
  - Queue-length increases linearly in N: $$\lim_{N\to\infty} \frac{Q^N(t)}{N} = q(t) > 0$$

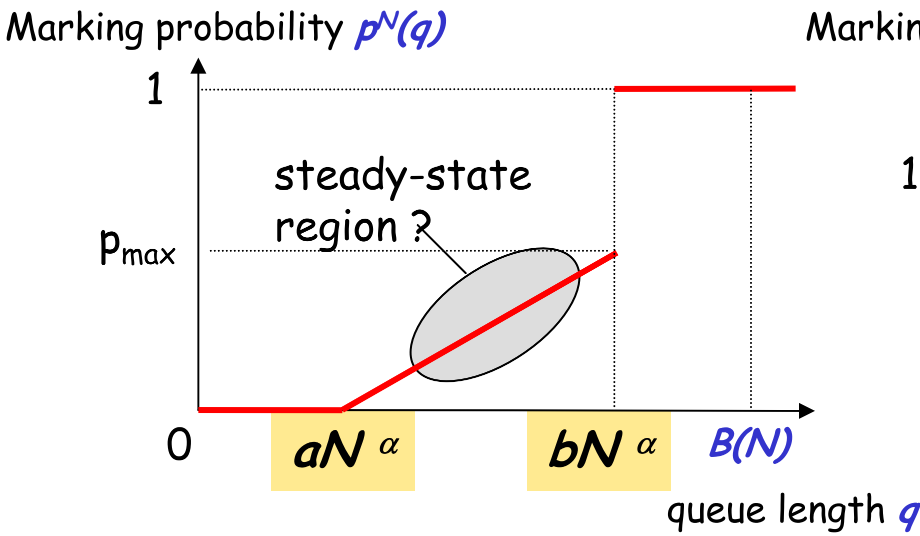
More on scaling

- Rule of Thumb: **buffer size ≈ bandwidth-delay product**
- Linear scaling under N flows and capacity NC
  - Thresholds for packet marking = O(N)
  - buffer size = O(N)
  - To prevent buffer-underflow after all N flows back-off
- For very large N and under drop-tail, B(N) ~ $O(N^{0.5})$ is sufficient to give high utilization [Appenzeller04]
- Why?
  - For large N, N flows becomes independent (no global synchronization !)
  - Average amount of arrivals (NCT) → size of the pipe
  - Buffer will absorb typical fluctuations on the order of $O(N^{0.5})$ by CLT

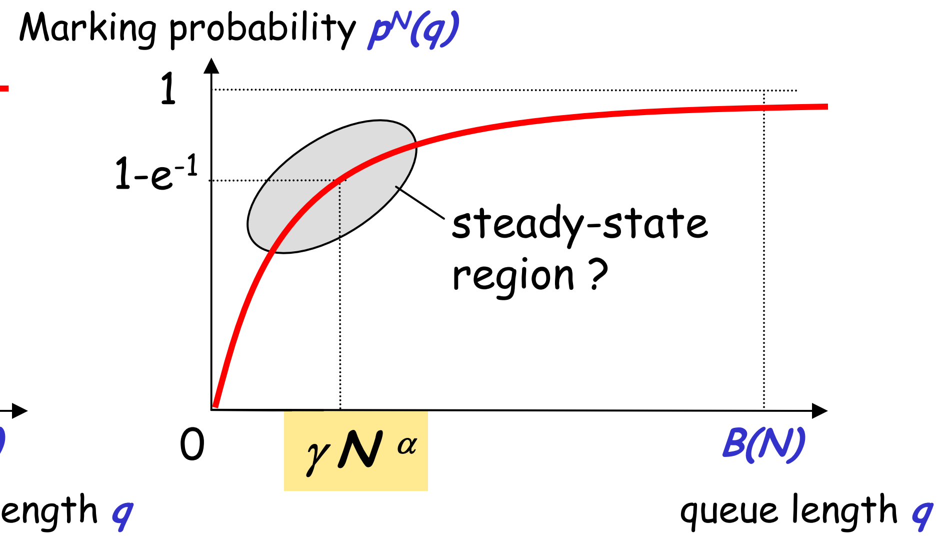
Why Scales? (Why Bothers?)

- Internet is growing…
  - Different scaling → Different growth rate
- **Strategic design of large networks over a long time** → could be more important than short-term optimizations!!
- **Q:** If the number of users (subscriptions) were to double, what would you do for the capacities and buffers at routers (AQM)?
  - Capacity doubles? → investment over long-time scale
  - Buffer doubles? → may not have to…
  - AQM parameters (e.g., packet marking)? → simply reconfigure the router (immediate)

Proposed Scaling: **Aggressive Packet Marking**



RED type



REM type

$$p^N(N^\alpha x) = p(x)$$
$$0 < \alpha < 0.5$$

- Any queue-based AQM
- *p(x)* non-decreasing function with *p(0)=0, p(*∞*)=1*

Proposed Scaling: **Aggressive Packet Marking**

- **<u>Suppose that</u>** queue-length fluctuates as desired, i.e., $Q^N \sim O(N^{\alpha})$, $0 < \alpha < 0.5$
- What do we have then?
  - Packet delay in queue = $Q^N/NC \sim O(N^{\alpha-1}) \rightarrow 0$ as $N \rightarrow \infty$
  - Queue is <u>small enough</u> to have (almost) **zero delay**,
  - Queue is <u>large enough</u> to have **high utilization** (not empty)
  - **Save huge** for buffer cost & **virtually no packet drop** when $B(N) \sim O(N^{\alpha+\varepsilon}) \ll O(N^{0.5})$
- **<u>Is this true?</u>**

Stability Analysis of TCP/AQM

- TCP/AQM is a very complicated **non-linear** system.
- Window sizes & queue-length → <u>deterministic functions (Fluid)</u>
  - Fluid approach or delayed differential equation approach [Misra, Towsley, Srikant, Kunniyur, Shakkottai, etc] :

$$\frac{dx}{dt} = \kappa[\Delta - \beta x(t-T)p(q(t-T))]$$

  - Optimization approach [Kelly, Low, etc] :

$$\max_{x_s \geq 0} \sum_s U_s(x_s), \quad \text{subject to} \quad \sum_{s \in l} x_s \leq C_l$$

- **Linearize** the system around equilibrium point and apply classical stability criterion from control theory
  - E.g., generalized Nyquist criterion
- Global stability → find Lyapunov function for the system.

Stability Analysis of TCP/AQM

- Example: Criterion for linear stability of TCP/RED [Low 03]:

$$\frac{\rho}{2}\cdot\frac{c^3\tau^3}{N^3}(c\tau+N) \;\le\; \frac{\pi(1\text{-}\beta)^2}{\sqrt{4\beta^2+\pi^2(1-\pi)^2}}$$

- Under our setting, this means **$O(N^{1-\alpha}) <$ Const.**
  - Slope for marking function $O(N^{-\alpha})$ gets too steep
- Linear instability can cause [Low 03]
  - Jitter in source rate and delay
  - Subject short-lived flows to unnecessary delay and loss
  - Underutilization of link capacities
- <u>Our scaling always yields **linearly unstable** system for large N</u> → ~~not desirable, forbidden?~~ **NO!**

Performance vs. Stability

- Using *ns*-2 under different scale (α)
- RED with N=1000, RTT ~ [120, 180] ms

	Link utilization	Packet drop ratio	Ave. queueing delay (ms)	Std. of queueing delay (ms)
α = 1	1	0	154	1.94
α = 0.2	0.975	0.02	0.46	0.37

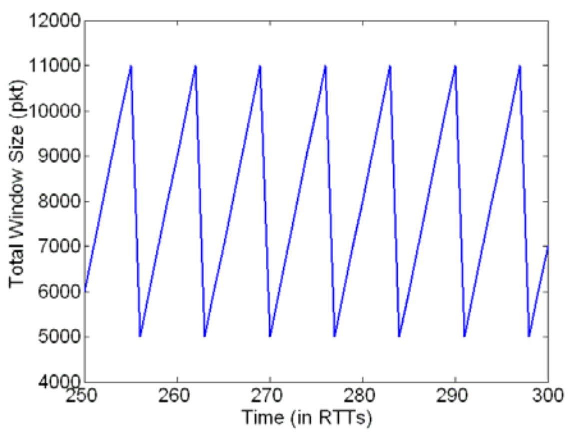
- Advantages of aggressive scale (α < 1)
  - Almost zero queueing delay and much smaller delay jitter!
  - Smaller queue fluctuations!
  - Much smaller buffer size (~10^5 packets → 25 packets !)
- Disadvantages?
  - Utilization: 100% → 97.5%, packet drop ratio: 0 → 2%
  - Difference <u>becomes negligible for larger N</u>

Limitation of RTT-based models with Averaged parameters

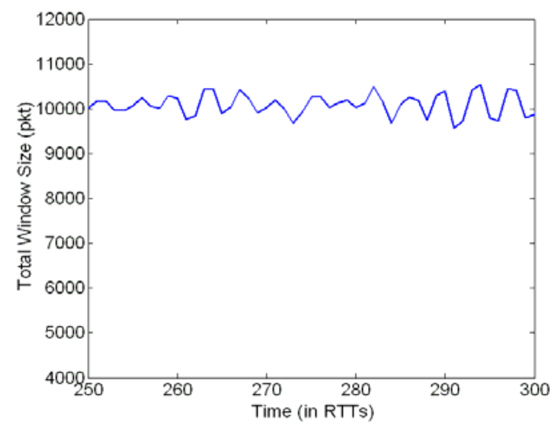
Total window size

$\alpha=0.2$

(a) Lindley recursion



(b) ns-2



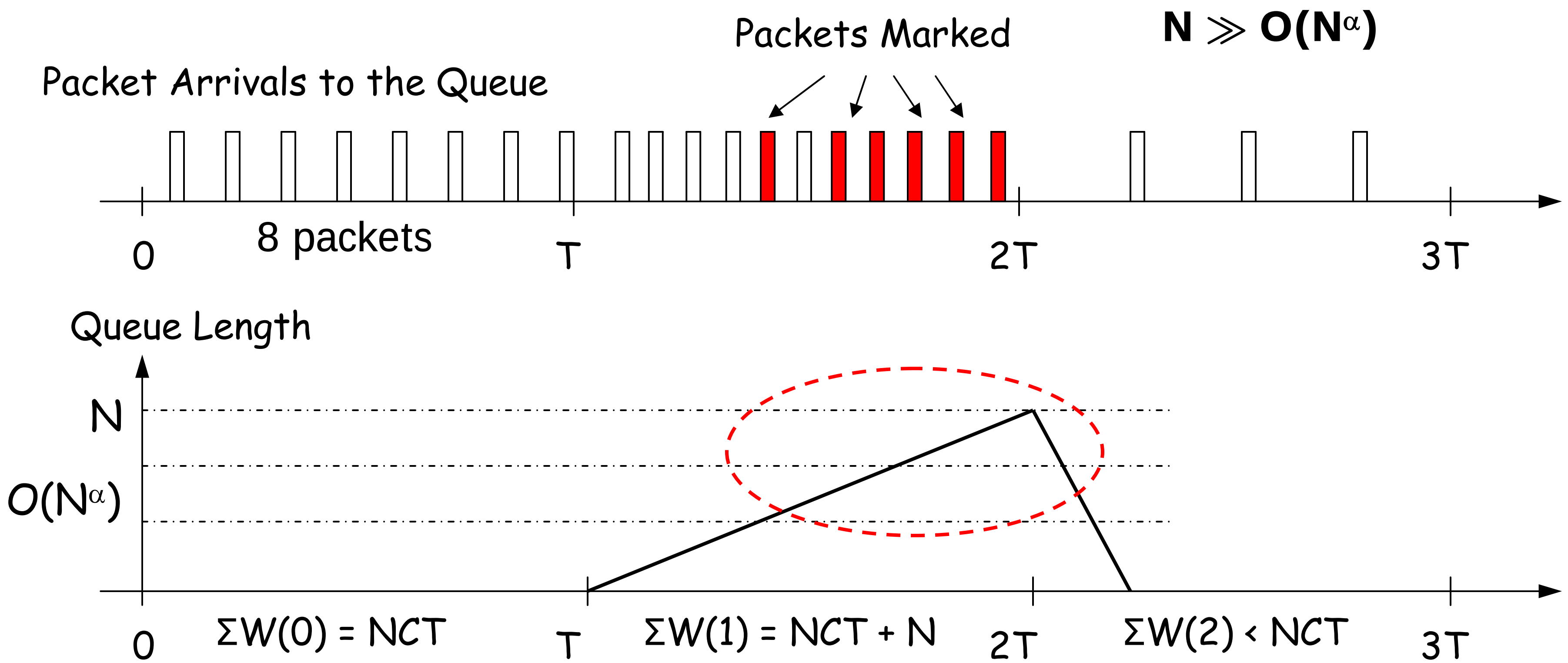
- Lindley recursion with random packet marking vs. ns-2

$$Q^N(k+1) = \min\left\{ \left[Q^N(k) + \sum_{i=1}^{N} W_i^N(k+1) - NC \right]^+ ,\ BN^\alpha \right\}$$

- All packets arrive at the beginning of each RTT

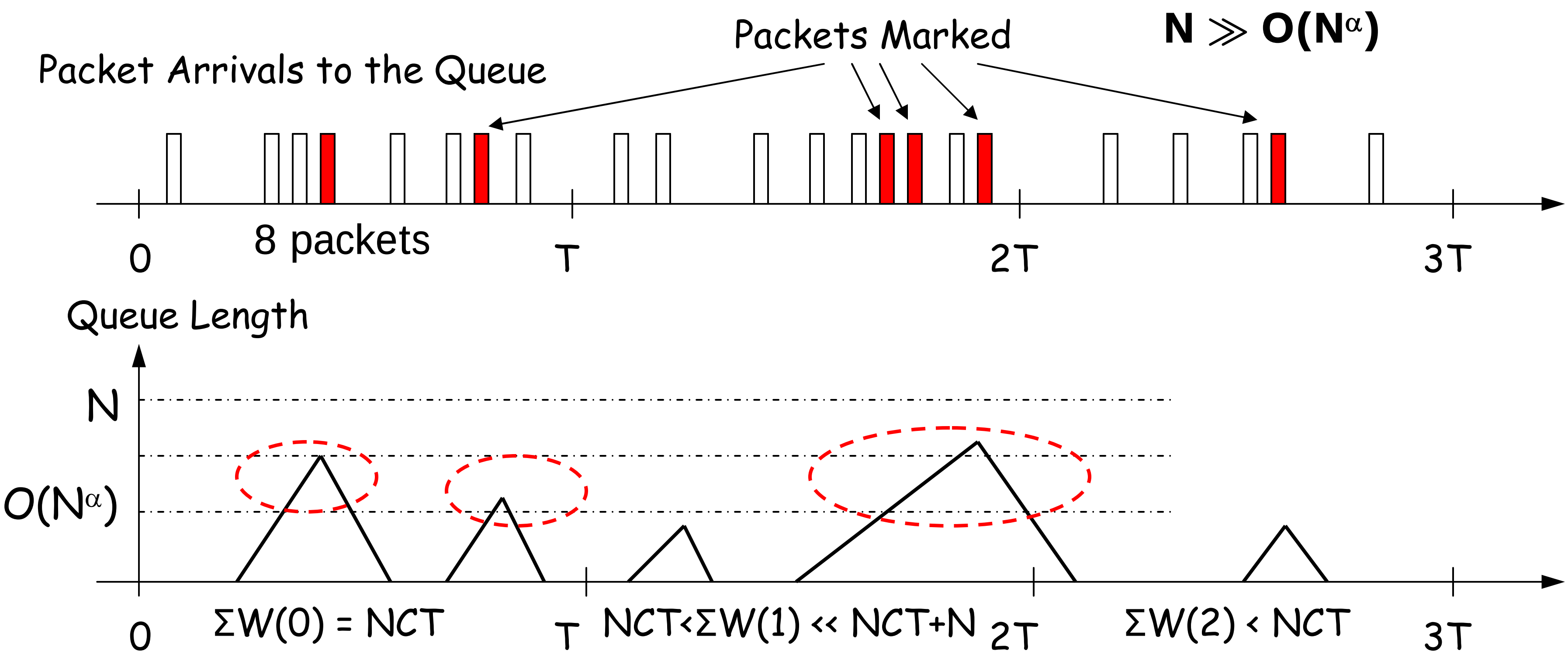
Problem with deterministic arrivals

- **Random** packet marking + **Deterministic** packet arrivals



Random packet arrivals

- **Random** packet marking + **<u>Random</u>** packet arrivals

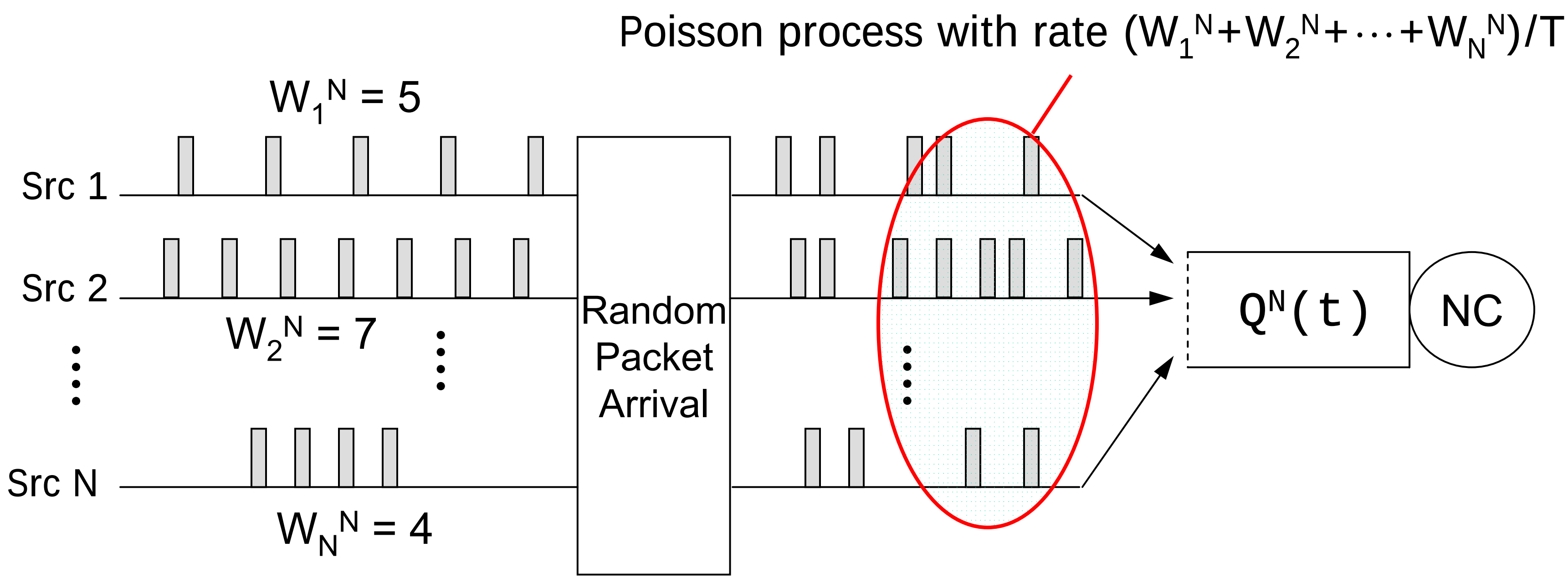


Two sources of randomness

- # of packets is random due to random marking
- *Given the # of packets*, actual arrival instants to the queue are also random.
  - Interaction with other flows at prior links
  - Time-varying queue-lengths
  - Different packet lengths
  - Small difference in RTTs
- Any discrete-time (RTT) based model cannot distinguish the previous two types of arrival patterns.

Doubly-stochastic model for packet arrivals in TCP/AQM

- **<u>Given</u>** all the window sizes ($W_i^N(k) = w_i^N(k)$) of N flow at time t=kT, the arrival to the queue is modeled by a Poisson process with rate $\sum w_i^N(k)/T$



Do Young Eun

NC STATE UNIVERSITY

Model description

- Packet arrivals:
  - Conditional (doubly-stochastic) Poisson process with rate modulated by window sizes
- Window size evolution:

$$W_i^N(k+1) = \begin{cases} \left(W_i^N(k)+1\right) \wedge w_{max} & \text{if no packet marked,} \\ \lfloor W_i^N(k)/2 \rfloor \vee 1 & \text{otherwise.} \end{cases}$$

Model description (2)

- Let $\overline{W^N}(k) := (W_1^N(k), W_2^N(k), \ldots, W_N^N(k))$
- Given $\overline{W^N}(k) = \overline{w^N}(k)$ and if $\sum_{i=1}^N w_i^N(k) < NCT$, then the queue-length ($Q_{\overline{w^N}(k)}$) distribution during *k*[th] *RTT* is given by M/M/1 with utilization

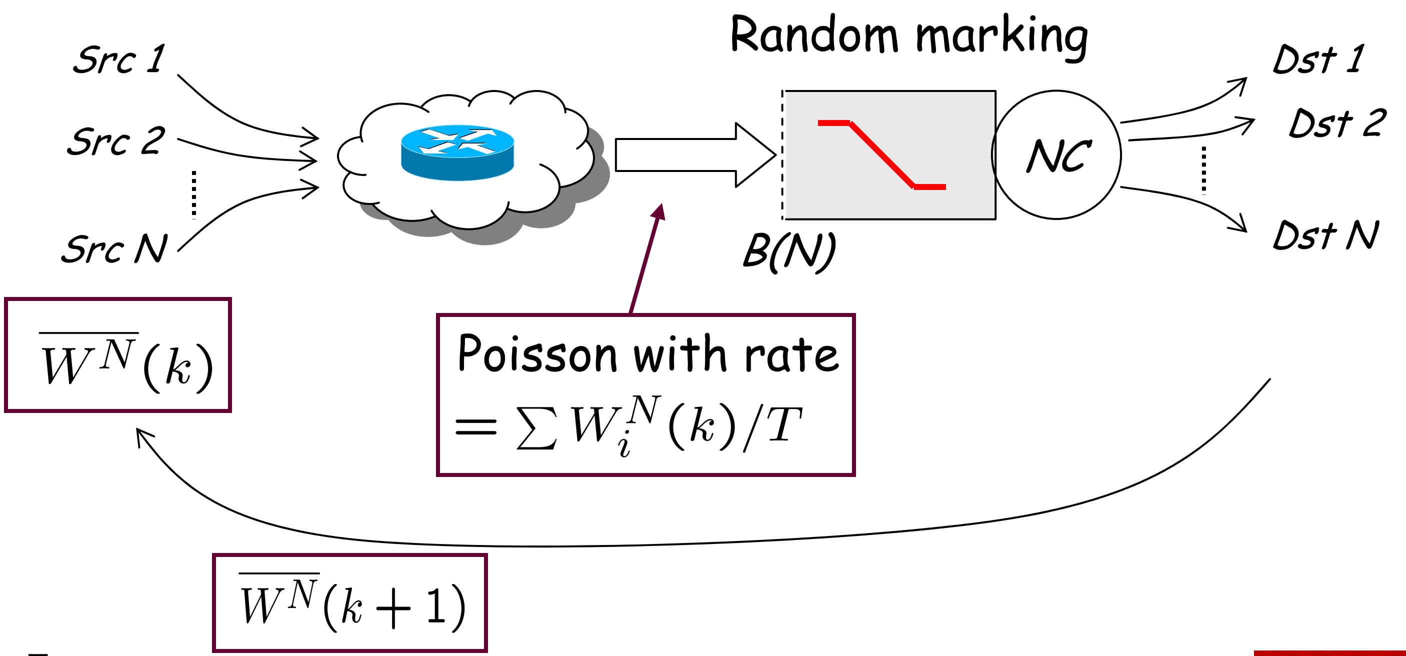
$$\rho_N(k) := \frac{1}{NCT} \sum_{i=1}^{N} w_i^N(k).$$

  - Flow *i* receives no marks with probability : $\mathbb{E}_Q \left\{ \left[1 - p^N(Q_{\overline{w^N}(k)}) \right]^{w_i^N(k)} \right\}$

- Given $\overline{W^N}(k) = \overline{w^N}(k)$ and if $\sum_{i=1}^N w_i^N(k) \geq NCT$, then all *N* flows receive marks and back off

Markov chain model

- Assumption: Given current window sizes for all N flows, the window sizes at the next RTT for flows i, j are independent.
  - Holds true in reality: will verify this later on.
- Then, for any given N, $\left\{\overline{W^N}(k)\right\}_{k\geq 0}$ forms an N-dim. Homogeneous Markov chain



Convergence to a steady-state

- Given *N*, the Markov chain is ergodic, or positive recurrent (In general, apply Foster's criterion)
- There exists a stationary distribution π
- $\left\{\overline{W^N}(k)\right\}_{k\geq 0}$ converges *in variation* to π

$$\Longrightarrow \quad \lim_{k\to\infty} \sum_{\overline{w^N}\in E} \left|\mathbb{P}\left\{\overline{W^N}(k)=\overline{w^N}\right\} - \pi\left\{\overline{w^N}\right\}\right| = 0$$

- Regardless of initial distributions of window sizes, the chain converges to a steady-state where $\overline{W^N}(k)$ has a stationary distribution π

Performance metrics of interest

- System is in steady-state $\Longrightarrow$ $\overline{W^N}(k)$ is stationary in k
- Utilization:

$$\rho(N) := \mathbb{E}\left\{\frac{\sum_{i=1}^{N} W_i^N}{NCT}\right\}$$

- Queue-length distribution: Distribution of $Q_{\overline{W^N}}$

- How to find?
  - Solving balance equation? → computationally infeasible
  - Can still get the results without solving the balance eq.

Probability of flow receiving marks

- **Proposition:** Let $f_i(N)$ be the probability that flow *i* receives at least one marks. Then, for some constants *a*, *b* $\in$ *(0,1)*, and for all *i* and *N*,

$$0 < a \le f_i(N) \le b < 1$$

- There are always some fraction of flows receiving marks:
  - Flows adjust themselves to the marking scale
  - No synchronized behavior !
  - Induce all the good performances as desired

Main results on performance

- **Theorem**: Let the system be in steady-state, and $\rho(N)$ be the utilization. Then,

$$\lim_{N \to \infty} \rho(N) = 1$$

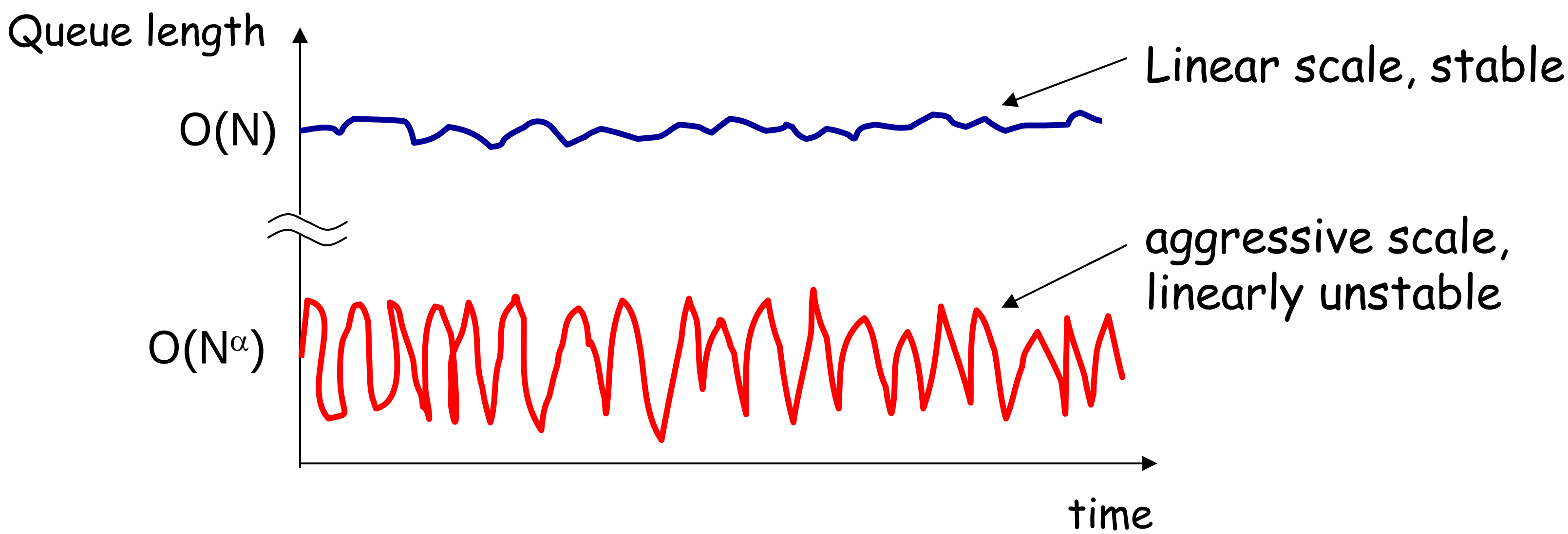
Further, let $\widehat{Q}_{\overline{W^N}}$ be the steady-state queue-length random variable. Then, for any given $\varepsilon > 0$, we have

$$\lim_{N \to \infty} \frac{Q_{\overline{W^N}}}{N^{\alpha+\epsilon}} = 0 \quad \text{in probability.}$$

Scaling TCP/AQM in large systems

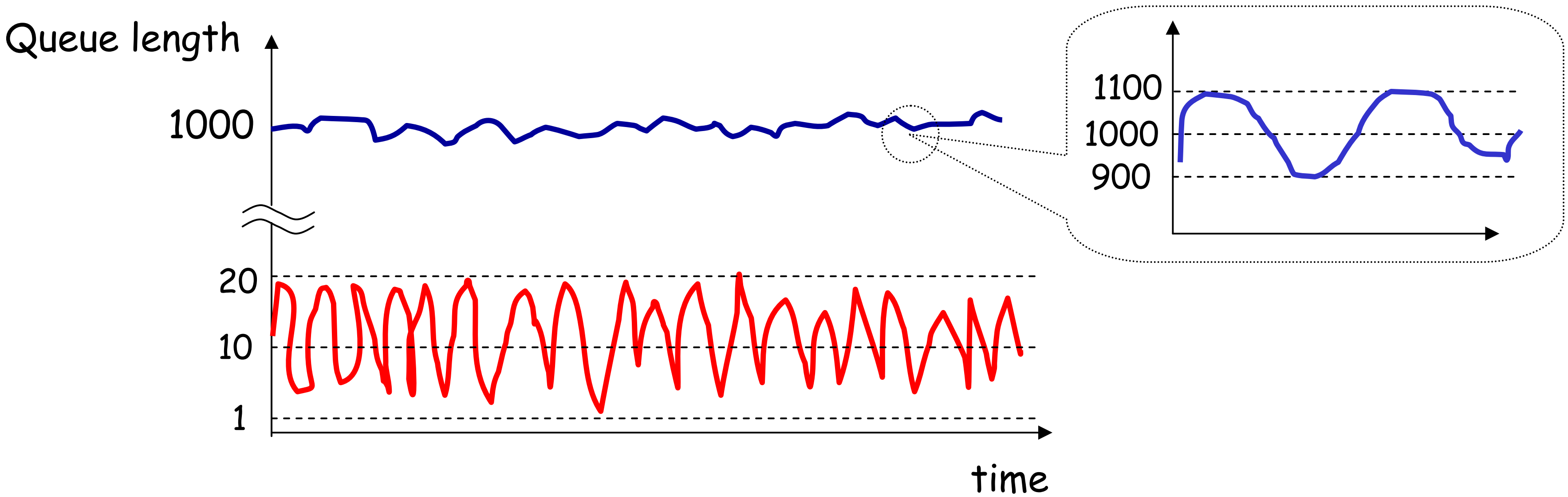
- Aggressive scaling works!
- High utilization and low packet drops
- Queueing delay decreases to zero!
- Buffer size can be much smaller, i.e., $O(N^{\alpha+\varepsilon}) \ll O(N^{0.5})$
- No need to scale less as long as there are many flows
- **Linearly unstable system, but with all the "good" performances**

What is happening at the queue?



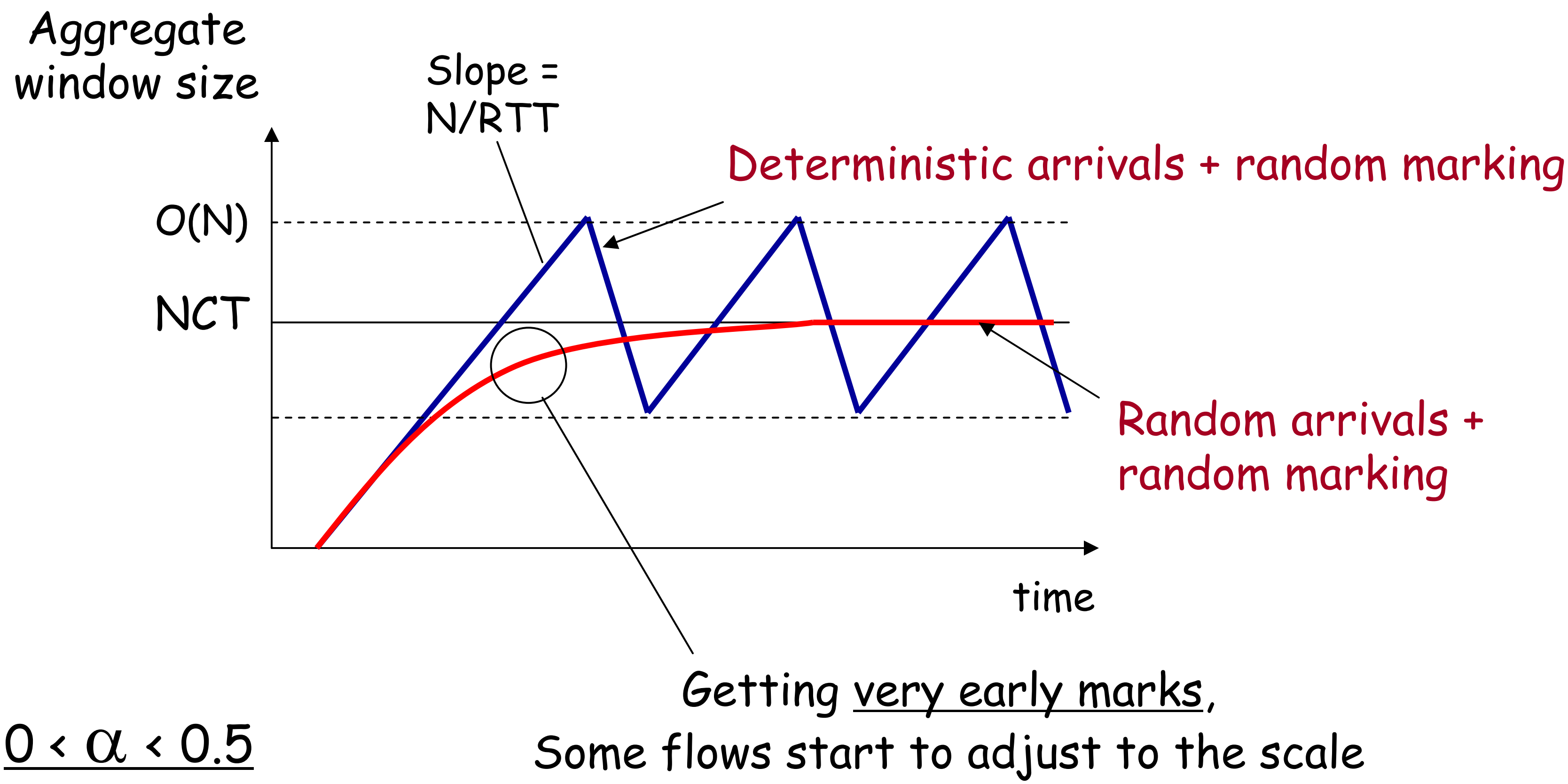
- Wild queue-length fluctuation → “linearly unstable”
- But, **“controlled fluctuation”** with <u>high utilization</u> and <u>almost zero queueing delay</u> (Q^N/NC), etc.

What is happening at the queue?



- Stabilize $Q^N(t)/NC$, but $Q^N(t)$ itself is really large!
- Which is better?
  - Queue-length stays around 1000 packets (with <u>seemingly</u> small, slow fluctuation)
  - Queue-length fluctuates fast between 1 and 20 packets

Total window size ($\sum W_i^N$)



$0 < \alpha < 0.5$

Numerical results (ns-2)

- Consider 5 AQMs: RED, EXP, REM, PI, Drop-Tail
- Simple dumbbell topology with N flows, hetero. RTT,

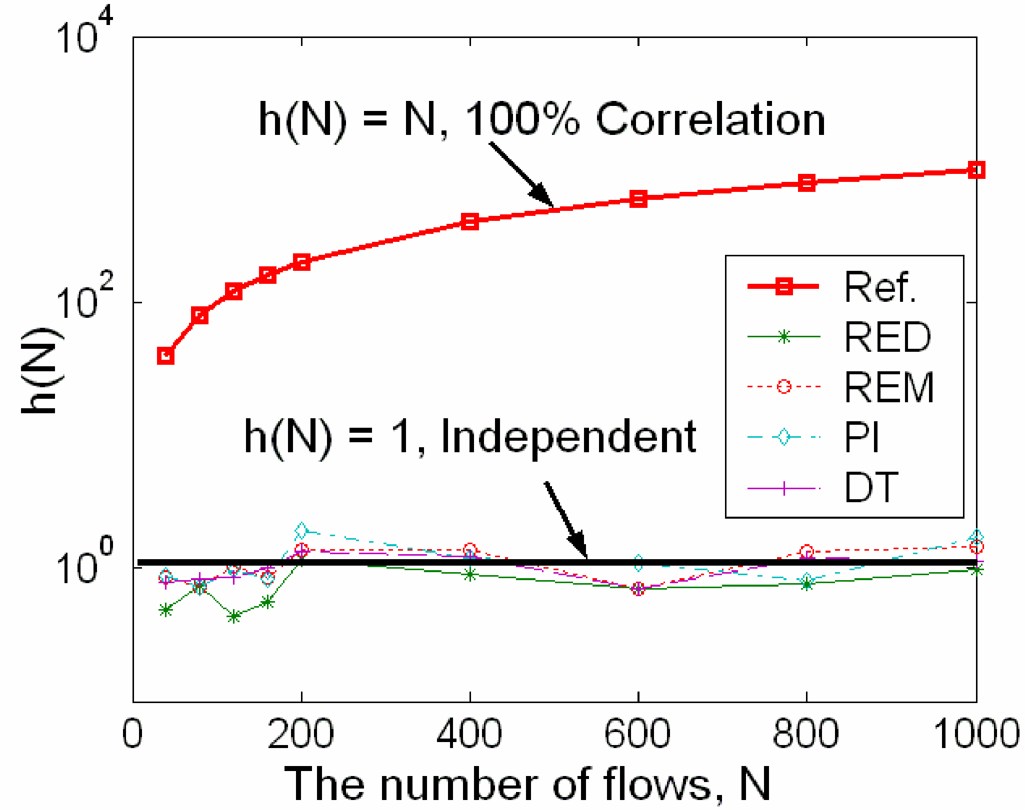
Table 1: AQM parameters for *ns*-2 simulations.

AQM	Parameters
RED	$q_{min}N^{\alpha} = 2N^{\alpha}$, $q_{max}N^{\alpha} = 10N^{\alpha}$ $P_{max} = 0.2$, Buffer Size $B(N) = 12N^{\alpha}$
EXP	$\gamma = -10/\ln(1 - P_{max})$, Buffer Size $= 12N^{\alpha}$
REM	pbo_ $= 2N^{\alpha}$, Buffer Size $B(N) = 12N^{\alpha}$
PI	$q_{ref} = 2N^{\alpha}$, Buffer Size $B(N) = 12N^{\alpha}$
DT	Buffer Size $B(N) = 12N^{\alpha}$

Independence among flows

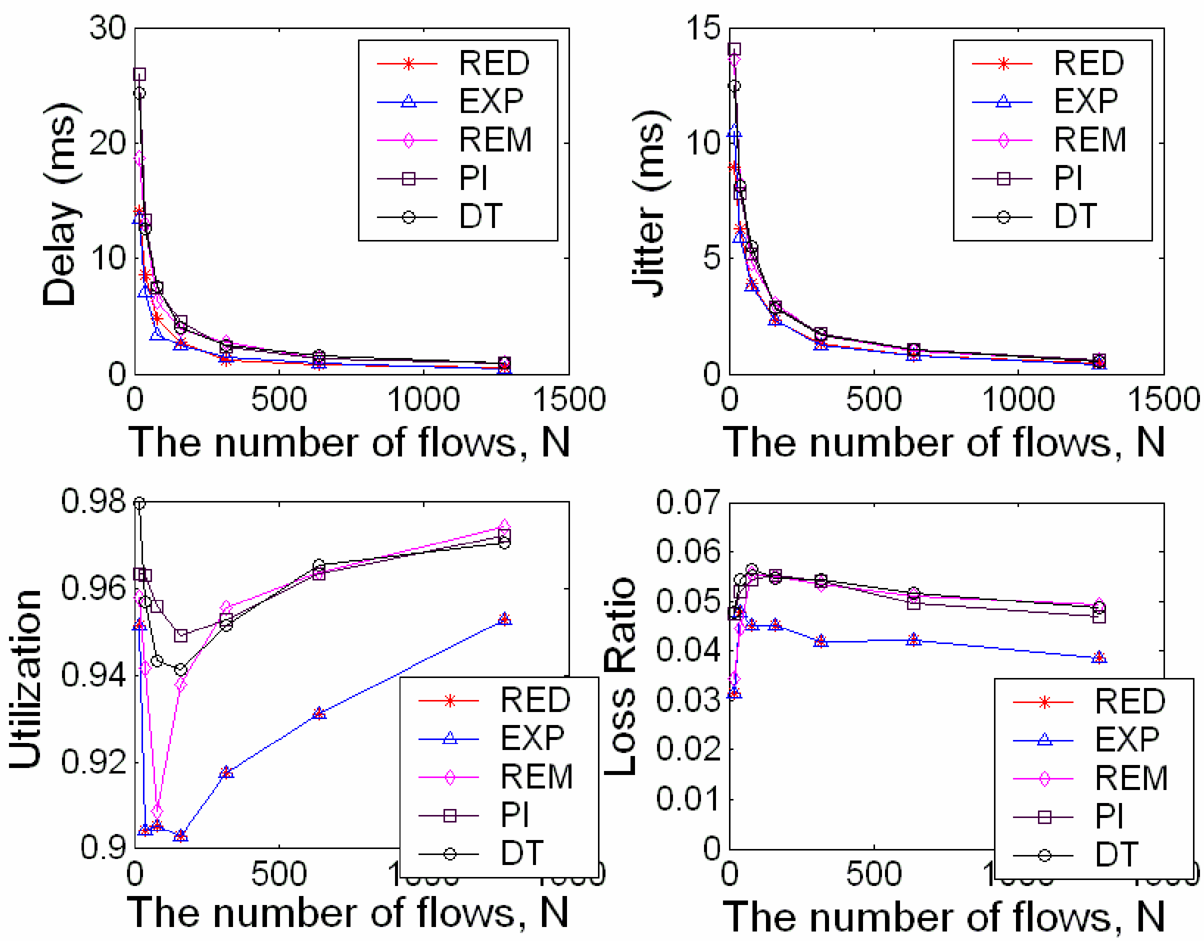
$$h(N) := \frac{\text{Var}\{\sum_{i=1}^{N} W_i^N\}}{\sum_{i=1}^{N} \text{Var}\{W_i^N\}}.$$

Use aggressive marking with $\alpha = 0.2$



- Window sizes are mostly <u>independent under the aggressive scale</u>
  - Our assumption holds good

Performance metrics (2)



Hetero. RTT:
unif [120,180]ms

Buffer size
$B = 12N^{0.2}$

- For N=1000:

α=1 → B =12000

α=0.5 → B =158

α=0.2 → B =48

Conclusions

- Aggressive scaling works well under many flows
- Buffer size can be chosen much smaller!
- Scaling governs the performance regardless of AQM schemes
- Doubly-stochastic models for TCP/AQM:
  - Random packet arrivals + random packet marking
- Traditional fluid-models or any model on a coarser time scale are not suitable
- “Stability” of fluid models can be misleading!

Thank You !

Questions?