On Achieving Weighted Service Differentiation: An End-to-End Perspective

Hung-Yun Hsieh, Kyu-Han Kim, and Raghupathy Sivakumar

GNAN Research Group

School of Electrical and Computer Engineering

Georgia Institute of Technology

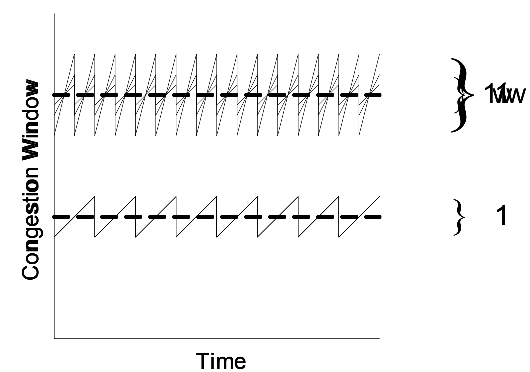
June 4, 2003

Scope and Goals

- Weighted rate differentiation
  - A user (flow) with weight w is provided w times the data rate of one with *unit* weight under the same network conditions (on the same path)
- TCP semantics
  - Reliable and in-sequence delivery
- TCP friendliness*
  - A flow with weight w = sum of w default TCP flows
- End-to-end approach
  - Does not rely on infrastructure support for resource provisioning at the routers

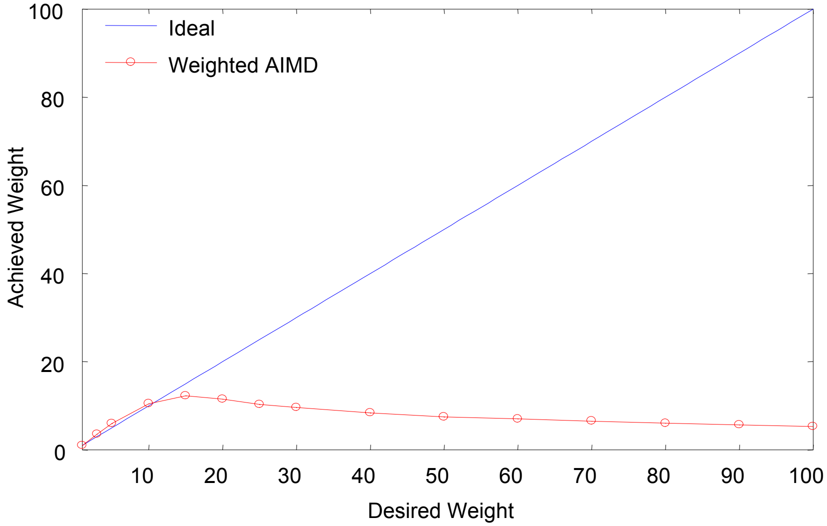
Existing Solutions

- Change the AIMD parameters (α, β) of the default TCP flow
  - Weighted AIMD [Crowcroft 98, Nandagopal 00]
  - $(\alpha, \beta) \rightarrow (\alpha \times f(w), \beta / g(w))$
- To achieve w times throughput
  - Assuming TCP stays in congestion avoidance
  - $f(w) \times [2g(w) - \beta] = w^2(2-\beta)$
    - $f(w) = w^2$, $g(w) = 1$ [bursty]
    - $f(w) = 1$, $g(w) \propto w^2$ [unresponsive]
    - ☞ $f(w) \propto w$, $g(w) \propto w$ [MulTCP]

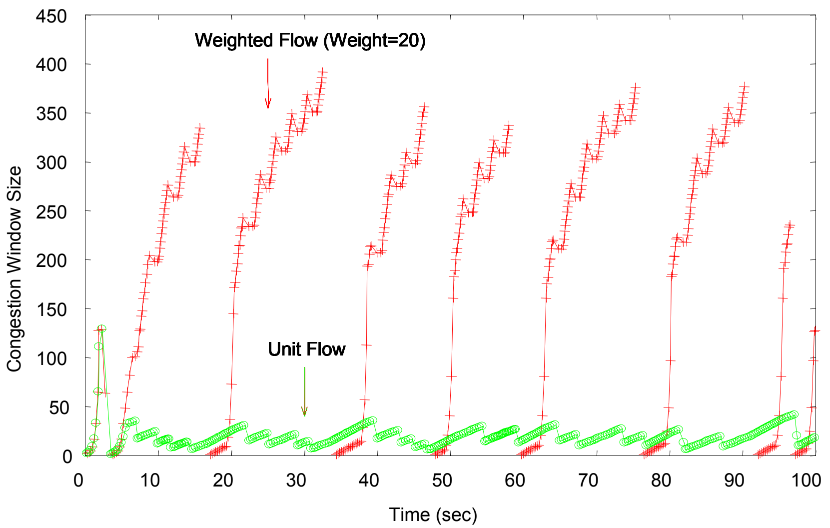


Performance of Weighted AIMD

- Scalability
  - Maximum achievable weight is limited to a small value [Crowcroft 98, Gevros 99, Nandagopal 00]



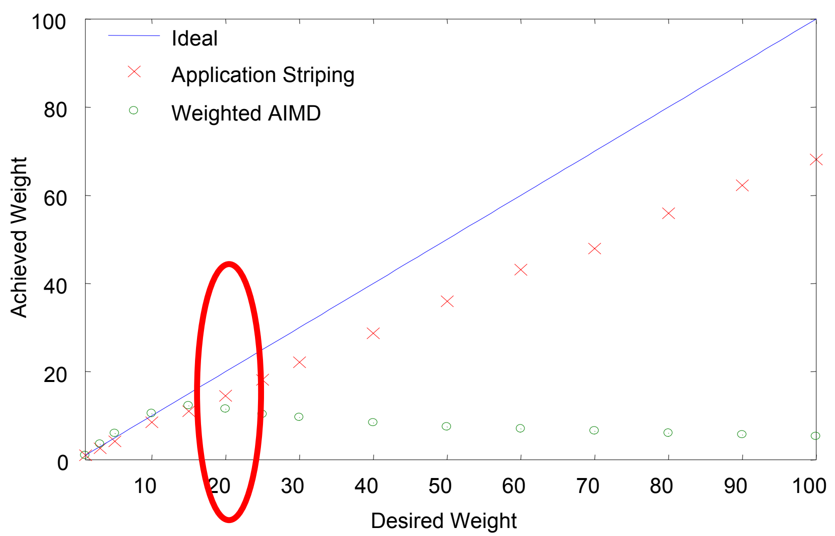
- Burstiness
  - Packet losses
  - Timeouts bring WAIMD out of the weighted region



An Alternative Approach

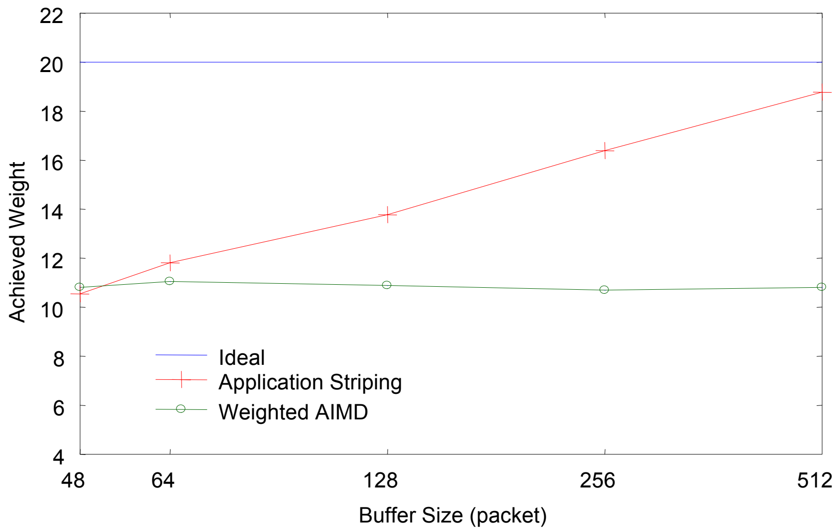
- Use multiple TCP sockets for achieving higher throughput
  - Application striping technique [Allman 96, Sivakumar 00, Hacker 02]
  - Offline reassembly (after all parts are downloaded)
- In-sequence delivery
  - Sending application: round-robin write
  - Receiving application: in-sequence read

✓ Multiple states (TCBs)

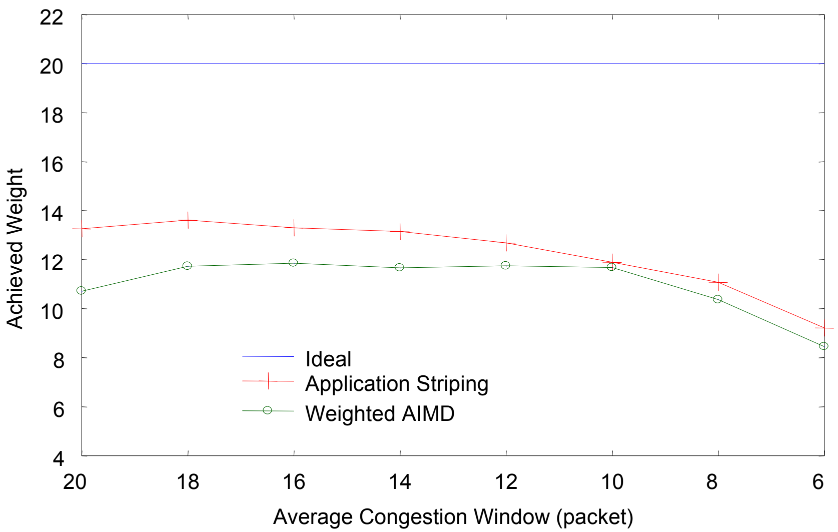


Limiting Factors

- Resequencing buffer
  - Probability of stalling increases with weight
  - Buffer requirement increases with weight



- Congestion window
  - Fair share per socket decreases with weight
  - Timeouts



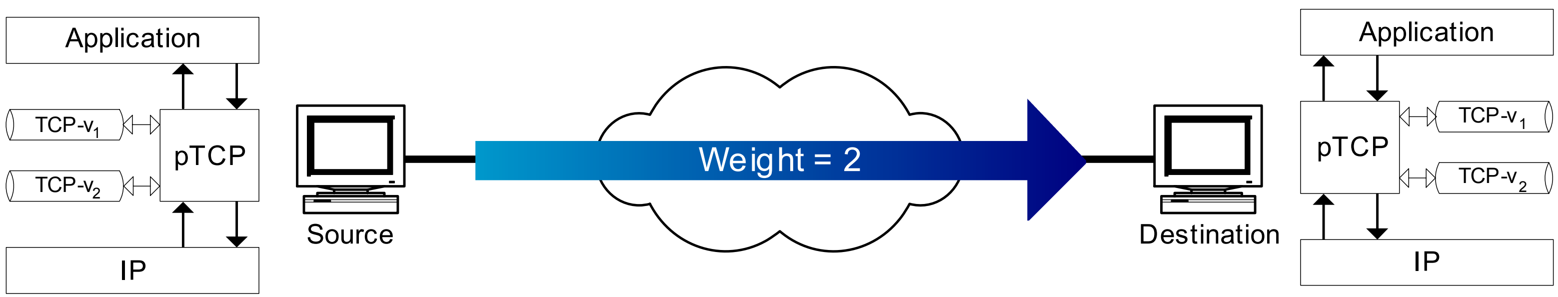
- Head-of-line blocking

pTCP Protocol

- A multi-state transport layer protocol for achieving bandwidth aggregation on a multi-homed mobile host with heterogeneous wireless interfaces [Hsieh 02]
- Addressing HOL blocking
  - Bandwidth differential
  - Delay differential
  - Bandwidth fluctuations
  - Blackouts

☞ Mechanisms used in pTCP can be tailored for achieving weighted rate differentiation

pTCP Key Design Elements



- Decoupling of functionalities
  - TCP-v: virtual packet; pTCP: data packet
  - TCP-v: how much to send; pTCP: what to send
- Congestion window based striping
  - pTCP *binds* a data packet to a pipe only when the TCP-v's congestion window has space
  - No need to explicitly estimate bandwidth available in each pipe for proportional striping

Addressing HOL Blocking (1)

- Dynamic reassignment (re-striping)
  - Avoids packets being held up in one pipe
  - Lost packets detected by individual TCP-v pipes are *unbound* for reassignment to the next available pipe
  - pTCP can detect losses much earlier than individual TCP-v pipes (via out-of-order arrivals)
  - Loss detection at pTCP can remain valid even when fast retransmit at TCP-v fails (e.g. for large weights)

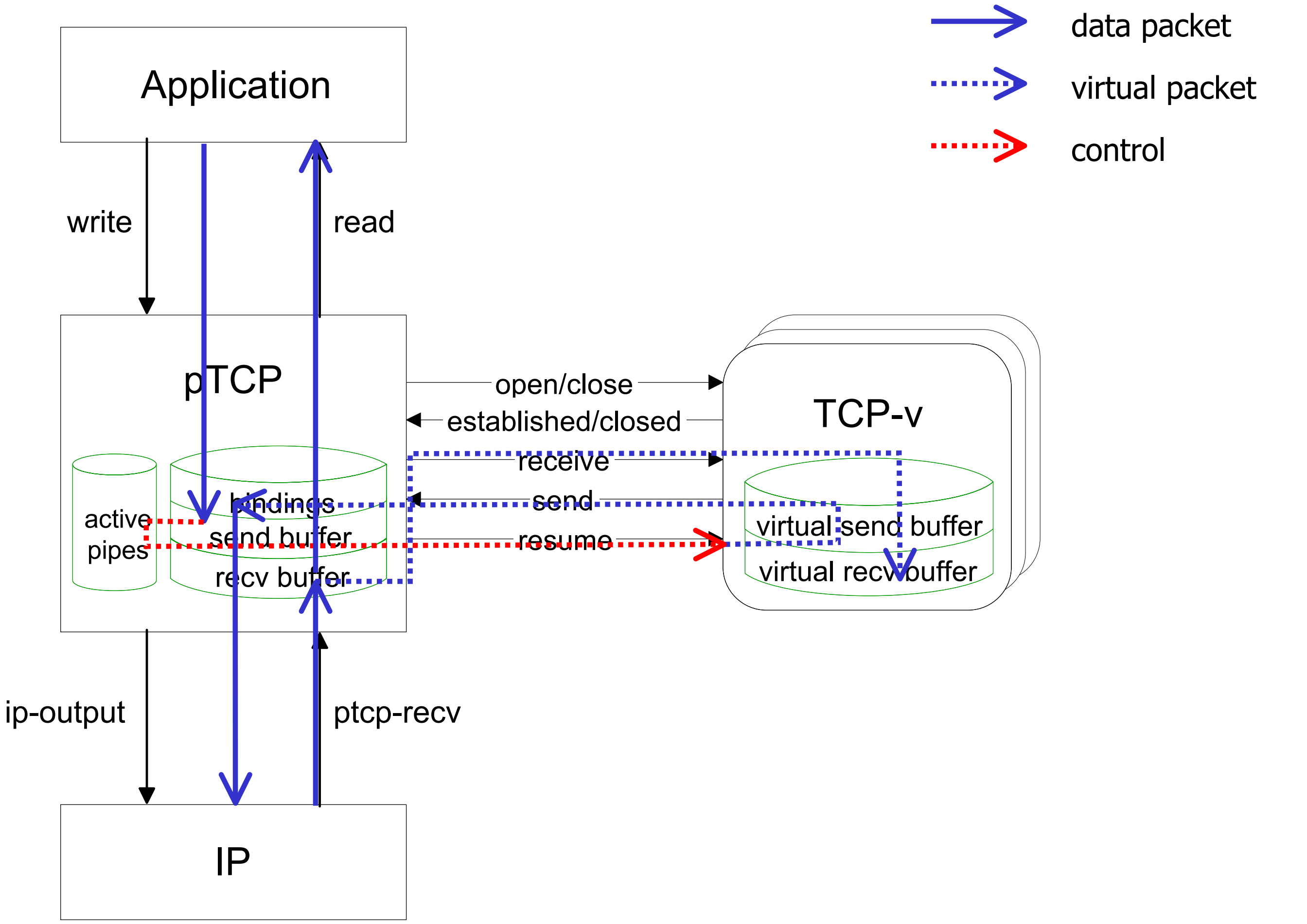
☞ Retransmission in a TCP-v pipe does not necessarily lead to retransmission of the data

Addressing HOL Blocking (2)

- Redundant striping
  - Avoids timeouts in individual pipes from stalling the connection
  - A TCP-v pipe is assigned a data packet *already bound* to another pipe
  - Loss of a retransmitted packet in TCP(-v) is detected only through a timeout
  - Redundant striping of retransmitted packets avoids the potential HOL blocking in pTCP

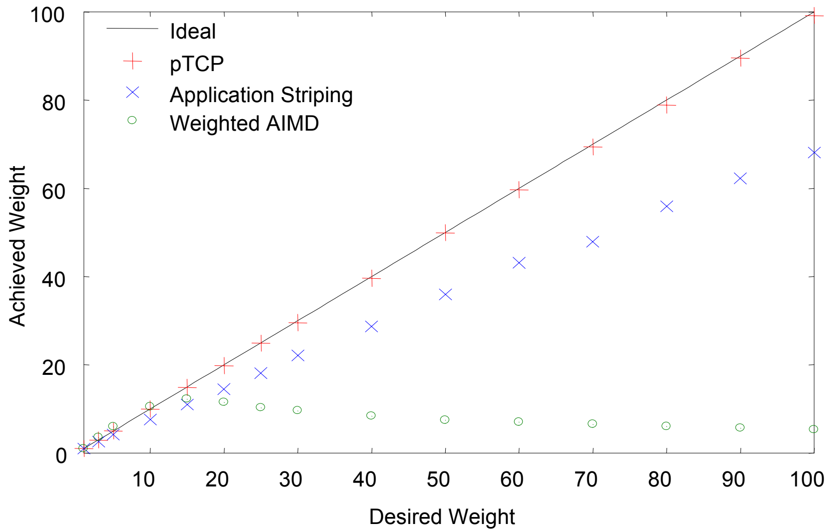
☞ Packet duplicates at the receiving end can be handled using the pTCP sequence number

pTCP Operations

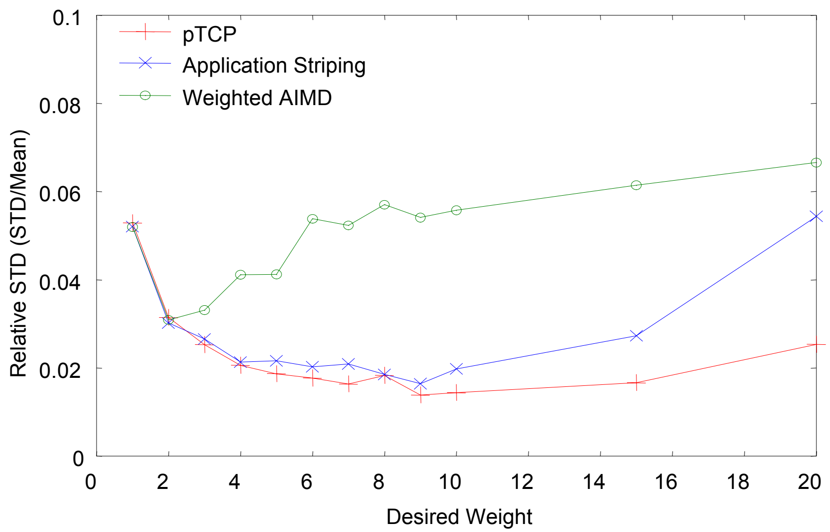


Simulation Results (1)

- Throughput
  - 1 weighted flow + 9 unit flows
  - Achieved weight: throughput ratio

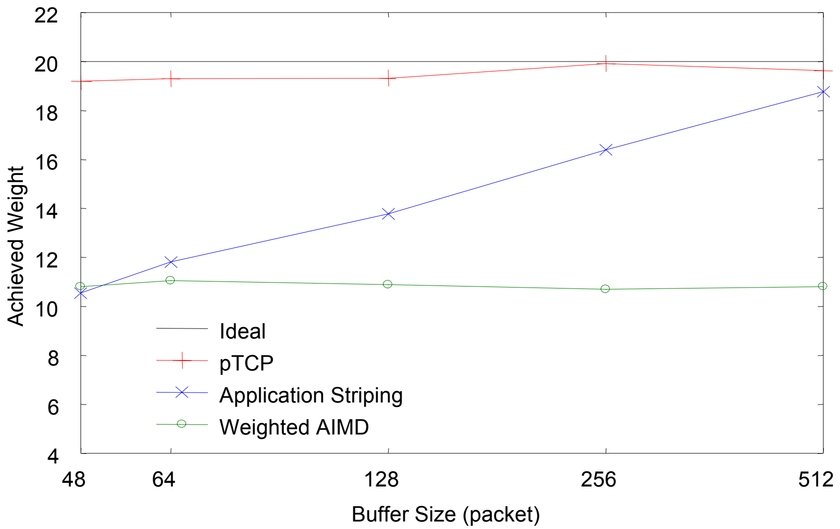


- Fairness
  - 10 weighted flows
  - Unfairness index: coefficient of variance

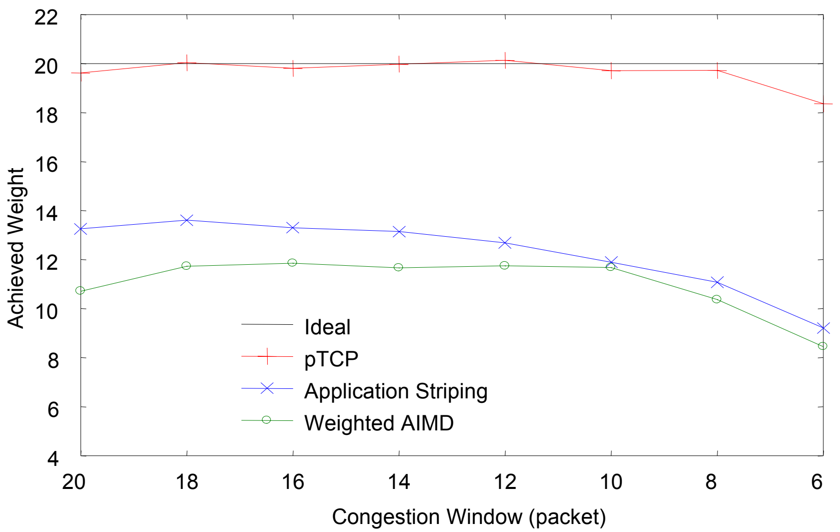


Simulation Results (2)

- Resequencing buffer
  - Congestion window based striping can avoid disproportional striping



- Congestion window
  - Re-striping and redundant striping effectively avoid HOL blocking due to packet losses



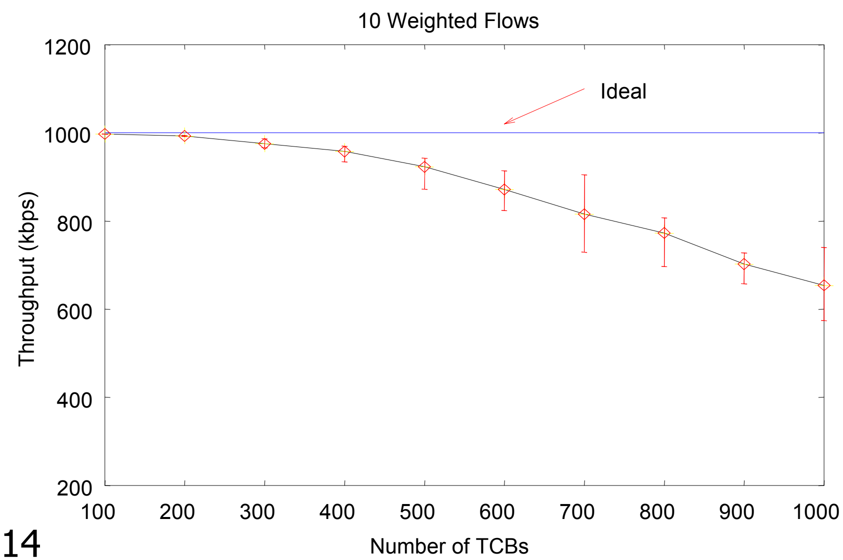
Scalability Limit

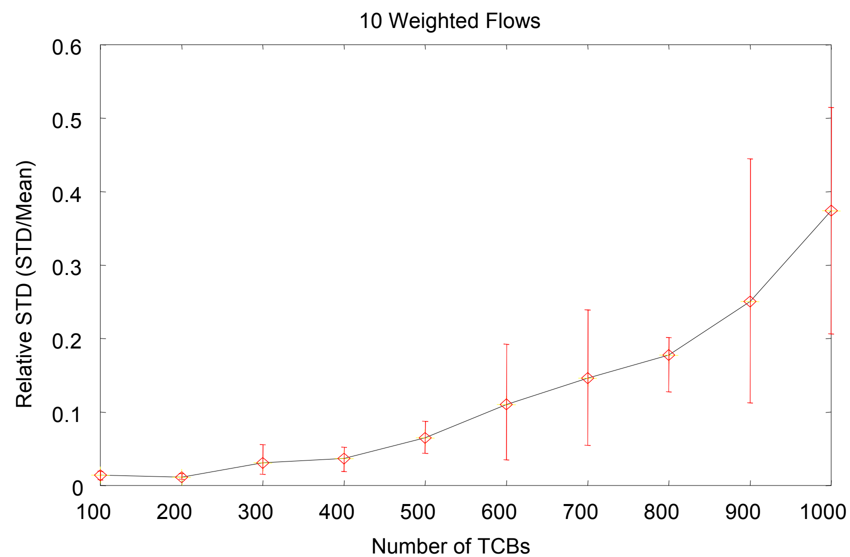
- What is the maximum achievable weight?

☞ $W_{max} = \eta/6$, η: network storage

  - Network storage ≡ BDP + network buffer

- cwnd ≥ 4 to trigger fast retransmit
  - 6 packets per TCP flow [Morris 00]





Summary

- Single state approaches such as weighted AIMD fail to achieve WRD for large weights due to burstiness and timeouts
- Application striping approaches maintain multiple states, but still fail due to head-of-line blocking
- pTCP achieves weighted rate differentiation with scalability to large weights
- ☞ pTCP can support multiple congestion control schemes for use with different QoS models
- ☞ http://www.ece.gatech.edu/research/GNAN