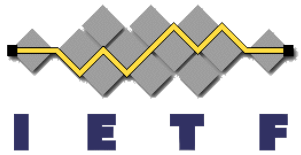
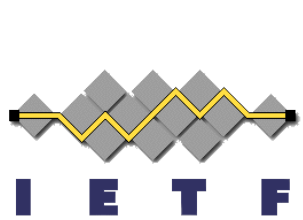
AQM Recommendation

Fred Baker

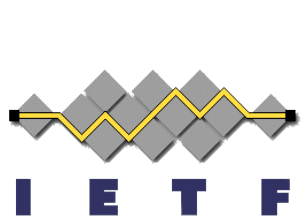


History

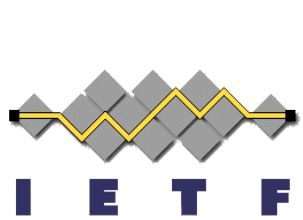


- At IETF 86, TSVAREA decided to update the recommendation of RFC 2309 to not recommend the use of RED
  - Argument: operational utility was low because of difficulty in configuration

My contribution

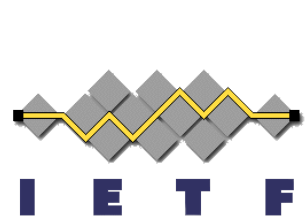


- I started with two approaches:
  - An email and subsequent draft on new recommendations
  - draft-baker-aqm-recommendation, which incorporates recommendations and removes RED
  - Various comments on the list
- Gorry Fairhurst offered to co-author
  - Edited draft-baker-aqm-recommendation-02.txt



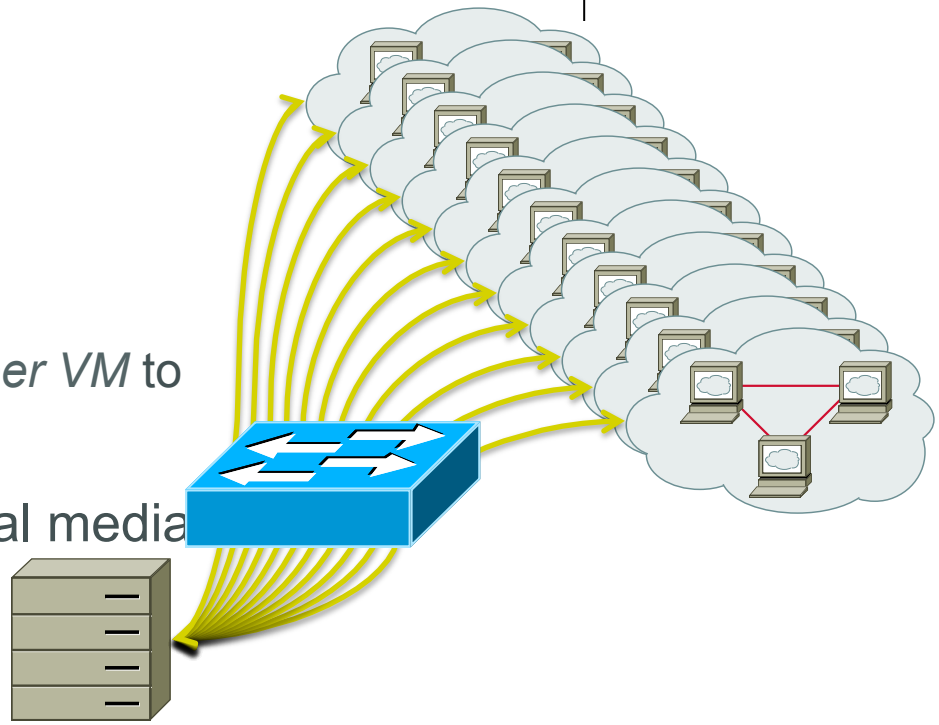
TARGET ISSUES

Persistent Deep Queues



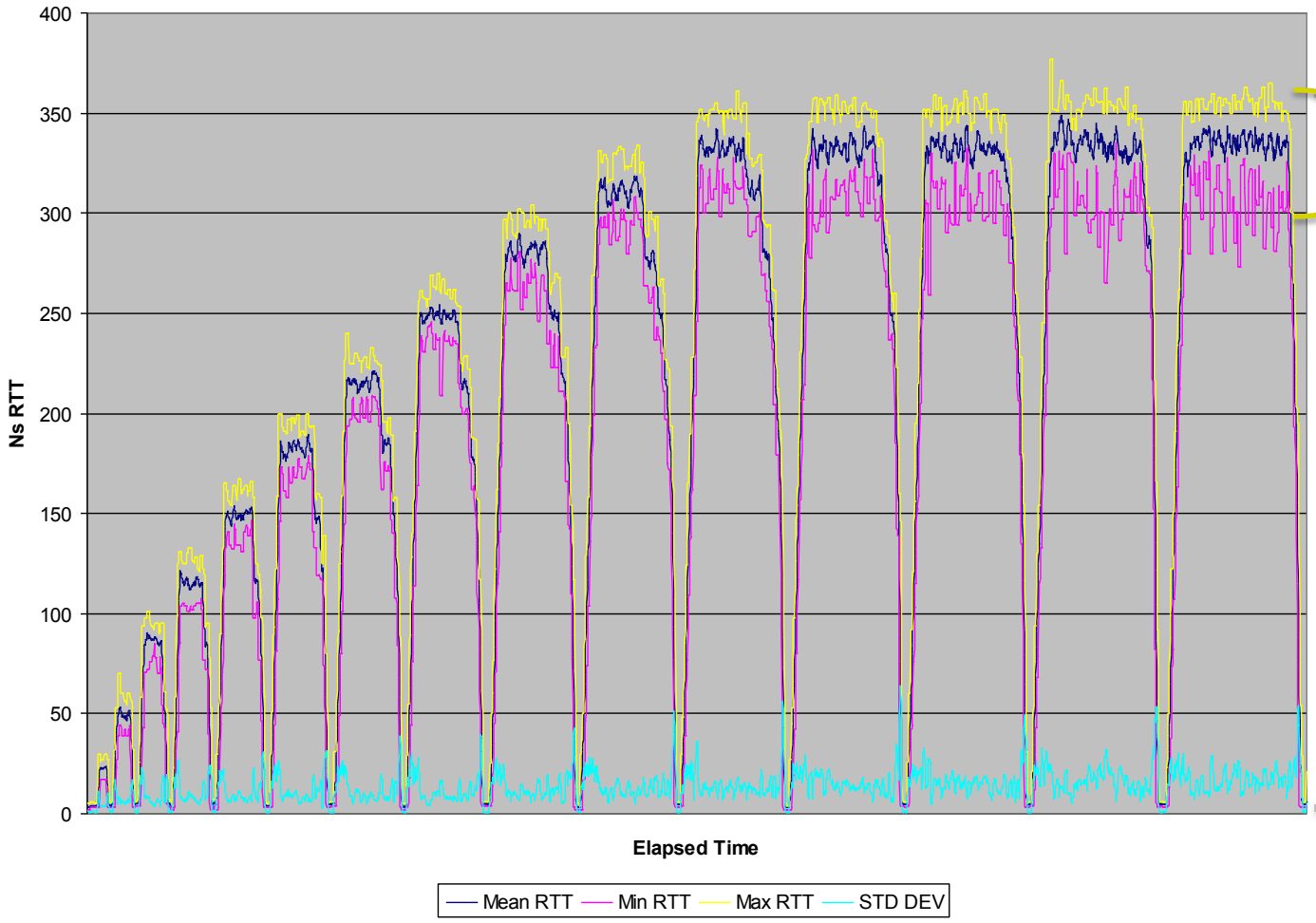
- In access paths (Cable Modem, DSL, Mobile Internet)
  - Generally results from folks building a deep queue with permissive drop thresholds
  - One DSL Modem vendor provides ten seconds of queue depth
- In multi-layer networks (WiFi, Input-queued Switches)
  - ***Channel Acquisition Delay***
  - Systems not only wait for their own queue, but to access network
  - In WiFi, APs often try to accumulate traffic per neighbor to limit transition time
  - In Input-queued switches, multiple inputs feeding the same output appear as unpredictable delay sources to each other
  - In effect, **managing *delay* through queue**, not queue depth

- Names withheld for customer/vendor confidentiality reasons
- Common social networking applications might have
  - $O(10^3)$ racks in a data center
  - 42 1RU hosts per rack
  - A dozen Virtual Machines per host
  - $O(2^{19})$ virtual hosts per data center
  - $O(10^4)$ standing TCP connections *per VM* to other VMs in the data center
- When one opens a <pick your social media application> web page
  - Thread is created for the client
  - $O(10^4)$ requests go out for data
  - $O(10^4)$ 2-3 1460 byte responses come back
  - $O(45 \times 10^6)$ bytes in switch queues **instantaneously**
  - At 10 GBPS, **instant 36 ms queue depth**



Data Center Applications

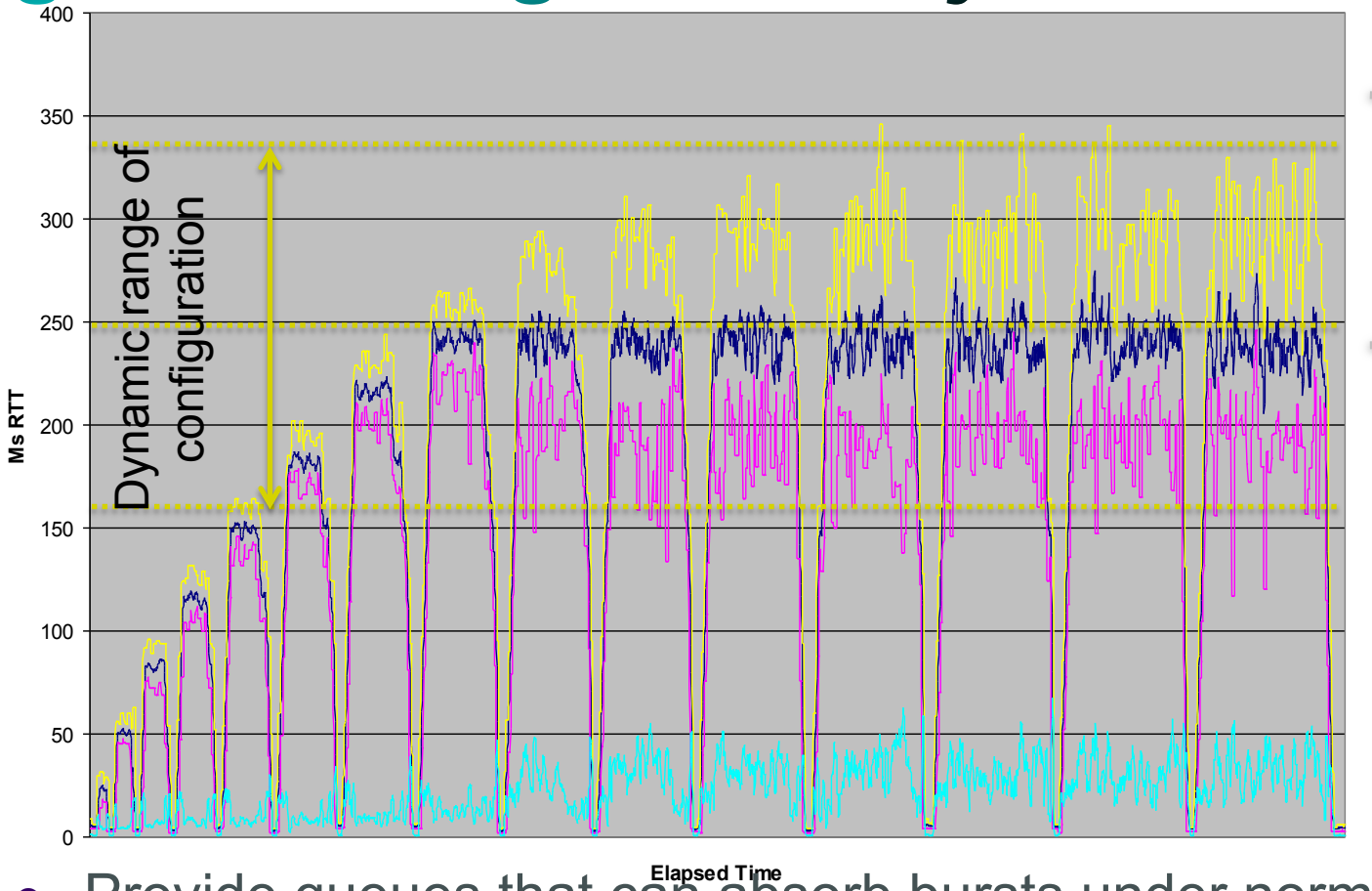
Tail Drop Traffic Timings



Typical variation in delay only at top of the queue

Mean Latency Correlates with Maximum Queue Depth

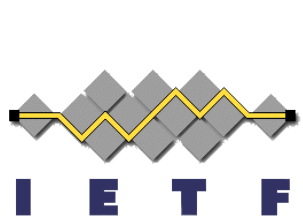
The objective:
generate signals early



Additional Capacity to Absorb Bursts

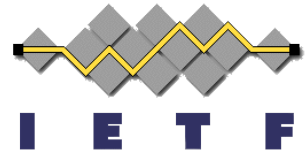
Mean Latency Correlates with target queue depth, min-threshold

- Provide queues that can absorb bursts under normal loads, but which manage queues to a shallow average depth
- Net effect: maximize throughput, minimize delay/loss, minimize SLA issues



DRAFT RECOMMENDATIONS

Conclusions/ Recommendations



1. Network devices SHOULD **implement** some **AQM** mechanism
2. Deployed AQM algorithms SHOULD support Explicit Congestion Notification (ECN) as well as loss to signal congestion to endpoints.
3. The algorithms that the IETF recommends SHOULD NOT require operational (especially manual) configuration or tuning.
4. AQM algorithms SHOULD respond to measured congestion, not application profiles.
5. AQM algorithms SHOULD NOT interpret specific transport protocol behaviours.
6. Transport protocol congestion control algorithms SHOULD maximize their use of available capacity (when there is data to send) without incurring undue loss or undue round trip delay.
7. Research, engineering, and measurement efforts are needed regarding … flows that are unresponsive to congestion notification or are responsive, but are more aggressive than present TCP.