Analysis and Design of an Adaptive Virtual Queue (AVQ) Algorithm for Active Queue Management*

Srisankar Kunniyur† R. Srikant‡

ABSTRACT

Virtual Queue-based marking schemes have been recently proposed for AQM (Active Queue Management) in Internet routers. We consider a particular scheme, which we call the Adaptive Virtual Queue (AVQ), and study its following properties: stability in the presence of feedback delays, its ability to maintain small queue lengths and its robustness in the presence of extremely short flows (the so-called *web mice*). Using a mathematical tool motivated by the earlier work of Hollot et al, we present a simple rule to design the parameters of the AVQ algorithm. We then compare its performance through simulation with several well-known AQM schemes such as RED, REM, PI controller and a non-adaptive virtual queue algorithm. With a view towards implementation, we show that AVQ can be implemented as a simple token bucket using only a few lines of code.

1. INTRODUCTION

In the modern day Internet, there has been a strong demand for QoS and fairness among flows. As a result, in addition to the sources, the links are also forced to play an active role in congestion control and avoidance. Random Early Discard (RED) [4] was originally proposed to achieve fairness among sources with different burstiness and to control queue lengths. RED allows for dropping packets before buffer overflow. Another form of congestion notification that has been discussed since the advent of RED is Explicit Congestion Notification (ECN)[3]. ECN has been proposed to allow links to help in congestion control by notifying users when it detects an onset of congestion. The links on detecting incipient congestion set a bit in the packet header that notifies the user that a link on its route is experiencing congestion. The user then reacts to the mark as if a packet has been lost. Thus, the link avoids dropping the packet (thereby enhancing goodput) and still manages to convey congestion information to the user.

To provide ECN marks or drop packets to provide fairness and control queue lengths, the routers have to select packets intelligently in a manner that conveys information about the current state of the network to the users. Algorithms which the routers employ to convey such information are called *Active Queue Management (AQM)* schemes. An AQM scheme might mark or drop packets depending on the policy at the router. In this paper, we use the term "marking" more generally to refer to any action taken by the router to notify the user of incipient congestion. The action can, in reality, be ECN-type marking or dropping (as in RED) depending upon the policy set for the router. As in earlier work on studying AQM schemes [14, 7, 6], this distinction is blurred in the mathematical analysis to allow for the development of simple design rules for the choice of AQM parameters. However, our simulations consider marking and dropping schemes separately.

Designing robust AQM schemes have been a very active research area in the Internet community. Some AQM schemes that have been proposed include RED [4], a virtual queue-based scheme where the virtual capacity is adapted [11, 12], SRED [15], Blue [2], Proportional Integral (PI) controller [7], REM [1], a virtual queue based AQM scheme [5] (which we refer to as the Gibbens-Kelly Virtual Queue, or the GKVQ scheme) among others. While most of the AQM schemes proposed detect congestion based on the queue lengths at the link (e.g., RED), some AQM schemes detect congestion based on the arrival rate of the packets at the link (e.g., virtual queue-based schemes) and some use a combination of both (e.g., PI). Also, most of the AQM schemes involve adapting the marking probability (as noted before we use the term *marking* to refer to both *marking* and *dropping*) in some way or the other. An important question is how fast should one adapt while maintaining the stability of the system? Here the system refers jointly to the TCP congestion controllers operating at the edges of the network and the AQM schemes operating in the interior of the network. Adapting too fast might make the system respond well to changing network conditions, but it might lead to large oscillatory behavior or in the worst case even instability. Adapting it too slowly might lead to sluggish behavior

*Research supported by NSF grants NCR-9701525, ANI-9813710 and DARPA grant F30602-00-2-0542.

†S. Kunniyur is with the Department of Electrical and Computer Engineering and Co-ordinated Science Lab, University of Illinois at Urbana-Champaign. Email: kunniyur@comm.csl.uiuc.edu

‡R. Srikant is with the Department of General Engineering and Co-ordinated Science Lab, University of Illinois at Urbana-Champaign. Email: rsrikant@uiuc.edu

SIGCOMM'01, August 27-31, 2001, San Diego, California, USA..

and more losses or marks than desired which might lead to a lower throughput.

In this paper, we start by presenting an implementation of a virtual-queue AQM scheme, namely the Adaptive Virtual Queue (AVQ). We then discuss a methodology for finding the fastest rate at which the marking probability adaptation can take place given certain system parameters like the maximum delay and the number of users so that the system remains stable. We note that, the marking probability in the AVQ is implicit, no marking probability is explicitly calculated and thus, no random number generation is required. On the other hand, we replace the marking probability calculation with the computation of the capacity of a virtual queue. Motivated by the success of the analysis and design of other AQM schemes in [14, 7, 6], we consider a single router accessed by many TCP sources with the same round-trip time (RTT) and use a control-theoretic analysis to study the stability of this system. However, unlike [14, 7, 6], we make no assumptions regarding the dynamic behavior of the linearized system. For instance, the authors in [7] neglect the delay in the flow control dynamics by assuming that the flow rates are large enough which forces their system to have a small number of users. We make no such assumptions in this paper.

The AVQ algorithm maintains a virtual queue whose capacity (called *virtual capacity*) is less than the actual capacity of the link. When a packet arrives in the real queue, the virtual queue is also updated to reflect a new arrival. (This was originally proposed in [11] as a rate-based marking scheme.) Packets in the real queue are marked/dropped when the virtual buffer overflows. The virtual capacity at each link is then modified such that total flow entering each link achieves a desired utilization of the link. It was shown in [12] that a fluid-model representation of the above scheme along with the congestion-controllers at the end-hosts was semi-globally asymptotically stable when the update at the links were done sufficiently slow. A feature of the AVQ scheme that is appealing is in the absence of feedback delays, it is shown in [12], that the system is fair in the sense that it maximizes the sum of utilities of all the users in the network. Combining this with a result in [11] which shows that a TCP user with an RTT of d_r can be approximated by a user with a utility function $\frac{-1}{d_r^2 x_r}$, where x_r is the rate of the TCP user, shows that the network as a whole converges to an operating point that minimizes $\sum_r \frac{-1}{d_r^2 x_r}$. This utility function called the potential delay was introduced as a possible fairness criterion in [13]. The throughput under this utility function is given by $1/d_r\sqrt{p_r}$, where p_r is the loss probability seen by User r which is consistent with the models in [14, 7, 6]. While we use this simplified model for analysis in the paper, our simulations in *ns-2* use TCP-Reno, including slow-start, time-out, fast retransmit, etc. A slightly more refined utility function is used in [9] and the results in this paper can be easily modified to incorporate that utility function.

The starting point of this paper is the fluid-model of the TCP flow-control problem along with the AVQ scheme that was proposed in [11]. However, here we explicitly consider the feedback delay due to the RTT of each user and thus, we obtain a delay-differential equation. We linearize this system and obtain conditions on d, the number of users N, the utilization γ of the link and a smoothing parameter α in the update equation of the AVQ scheme to ensure stability.

The rest of the paper is organized as follows: in Section 2, we present an implementation of the AVQ algorithm and provide design rules for the stability of the AVQ and TCP together. In Section 3, we provide detailed *ns-2* to validate our design rules and also compare the AVQ algorithm with RED, REM, GKVQ and the PI controllers. The PI controller is somewhat similar to AVQ in that it adapts the marking probability in a manner similar to the virtual capacity adaptation in the AVQ scheme, but it depends on the queue size at the link. As a result, for small buffers the system tends to perform poorly. Also, since the marking probability is directly modified and this update has to be slow enough for system stability, the scheme exhibits sluggishness when short flows are introduced. This would be the subject of simulations in Section 3. In Section 4, we provide theoretical justification for the design rules in Section 2. Conclusions are provided in Section 5.

2. THE AVQ ALGORITHM

Let C be the capacity of a link and γ be the desired utilization at the link. The AVQ scheme, as presented in [11, 12], at a router works as follows:

- The router maintains a virtual queue whose capacity $\tilde{C} \leq C$ and whose buffer size is equal to the buffer size of the real queue. Upon each packet arrival, a fictitious packet is enqueued in the virtual queue if there is sufficient space in the buffer. If the new packet overflows the virtual buffer, then the packet is discarded in the virtual buffer and the real packet is marked by setting its ECN bit or the real packet is dropped, depending upon the congestion notification mechanism used by the router.

- At each packet arrival, the virtual queue capacity is updated according to the following differential equation:

$$\dot{\tilde{C}} = \alpha(\gamma C - \lambda), \tag{1}$$

  where λ is the arrival rate at the link. The rationale behind this is that marking has to be more aggressive when the link utilization exceeds the desired utilization and should be less aggressive when the link utilization is below the desired utilization.

We now make the following observations. No actual enqueueing or dequeuing of packets is necessary in the virtual queue, we just have to keep track of the virtual queue length. Equation (1) can be thought of as a token bucket where tokens are generated at rate $\alpha\gamma C$ up to a maximum of C and depleted by each arrival by an amount equal to α times the size of the packet. Define

B = buffer size
s = arrival time of previous packet
t = Current time
b = number of bytes in current packet
VQ = Number of bytes currently in the virtual queue

Then, the following pseudo-code describes an implementation of AVQ scheme:

The AVQ Algorithm

At each packet arrival epoch do

$VQ \leftarrow \max(VQ - \tilde{C}(t-s), 0)$ /* Update Virtual Queue Size */
If $VQ + b > B$
Mark or drop packet in the real queue
else
$VQ \leftarrow VQ + b$ /* Update Virtual Queue Size */
endif
$\tilde{C} = \max(\min(\tilde{C} + \alpha * \gamma * C(t-s), C) - \alpha * b, 0)$ /* Update Virtual Capacity */
$s \leftarrow t$ /* Update last packet arrival time */

We note the following features of the AVQ scheme:

1. The implementation complexity of the AVQ scheme is comparable to RED. RED performs averaging of the queue length, dropping probability computation and the random number generation to make drop decisions. We replace these with the virtual capacity calculation in AVQ.

2. AVQ is a primarily a rate-based marking, as opposed to queue length or average queue length based marking. This provides early feedback, the advantages of which have been explored by Hollot et al [7, 6], which was also mentioned in Kelly et al [10].

3. Instead of attempting to regulate queue length as in RED, PI controller or recent versions of REM, we regulate utilization. As we will see in simulations, this is more robust to the presence of extremely short flows or variability in the number of long flows in the network. The reason is that, when utilization is equal to one, variance introduced by the short flows seems to lead to an undesirable transient behavior where excessively large queue lengths persist over long periods of time.

4. Unlike the GKVQ algorithm [5], we adapt the capacity of the virtual queue. A fixed value of $\tilde{C}$ leads to a utilization that is always smaller than $\tilde{C}/C$ and it could be much smaller than this depending on the number of users in the system. Our marking mechanism is also different in that we do not mark until the end of a busy period after a congestion episode.

5. There are two parameters that have to be chosen to implement AVQ: the desired utilization γ and the damping factor α. The desired utilization γ determines the robustness to the presence of uncontrollable short flows. It allows an ISP to trade-off between high levels of utilization and small queue lengths. Both the parameters α and γ determine the stability of the AVQ algorithm and we provide a simple design rule to choose these parameters.

The starting point for the analysis of such a scheme is the fluid-model of the TCP congestion-avoidance algorithm as proposed in [11]. A theoretical justification of how a stochastic discrete-time equation can be approximated by a fluid-model is shown in [8]. We will then incorporate the virtual capacity update equation with this model and study the stability of system under linearization.

Consider a single link of capacity of C and let the desired utilization of the link be $\gamma \leq 1$. Let N TCP users be accessing that link and let d be the common round-trip propagation delay of each user. We will model the TCP users using the $\frac{-1}{d^2 x}$ utility function as proposed in [11]. For the sake of simplicity and tractability, we will neglect the slow-start and the time-out behavior when modeling the TCP users. We will later show through simulations that even with slow-start and timeouts, the result holds. The congestion-avoidance algorithm of the TCP users can be written as:

$$\dot{x}_i = \frac{1}{d^2} - \beta x_i(t) x_i(t-d) p(\sum_{j=1}^{N} x_j(t-d), \tilde{C}(t-d)), \quad (2)$$

where $\beta < 1$ and $\tilde{C}$ is the virtual-capacity of the link. A β value of $2/3$ would give us the steady-state throughput of TCP as $\frac{1}{d}\sqrt{\frac{3}{2p^*}}$, where p^* is the steady-state marking probability which is consistent with the results in [16]. Hence, we will use $\beta = 2/3$, in all our calculations. Also, note that on substituting $x_i \approx \frac{W_i}{d}$, where W_i is the window-size of user i, we recover the TCP window control algorithm [11, 14].

The update equation at each link can now be written as:

$$\dot{\tilde{C}} = \alpha(\gamma C - \lambda), \quad (3)$$

where $\lambda = \sum_{j=1}^{N} x_j$ is the total flow into the link and $\alpha > 0$ is the smoothing parameter. Note that α determines how fast one adapts the marking probability at the link to the changing network conditions. We will present a design rule that specifies how to choose α for a given feedback delay (d), utilization (γ) and a lower bound on the number of users (N). In fact, as we will show in Section 4, one can arrive at bounds on any of the four parameters α, γ, N or d given the other three using the same design rule. However, in practice, it would seem most natural to choose α given the other three parameters.

The equilibrium point of the non-linear TCP/AQM model is given by:

$$\begin{aligned} \sum_i x_i^* &= \lambda^* = \gamma C \\ x_i^* &= \frac{\gamma C}{N} \\ p(\gamma C, \tilde{C}^*) &= \frac{N^2}{\beta(d\gamma C)^2} \end{aligned}$$

Let us assume that

$$\begin{aligned} \lambda(t) &= \lambda^* + \delta\lambda(t) \\ \tilde{C}(t) &= \tilde{C}^* + \delta\tilde{C}(t). \end{aligned}$$

The linearized model of the non-linear TCP/AQM model can now be written as:

$$\dot{\delta\lambda} = -K_{11}\delta\lambda(t) - K_{12}\delta\lambda(t-d) + K_2\delta\tilde{C}(t-d) \quad (4)$$

$$\dot{\delta\tilde{C}} = -\alpha\delta\lambda(t), \quad (5)$$

where

$$K_{11} := \frac{N}{\gamma C d^2} \qquad K_{12} := \frac{N}{\gamma C d^2} + \beta\frac{\gamma C^2}{N}\frac{\partial p(\gamma C, \tilde{C}^*)}{\partial \lambda}$$

$$K_2 := \beta\frac{\gamma C^2}{N}\left|\frac{\partial p(\gamma C, \tilde{C}^*)}{\partial \tilde{C}}\right|.$$

We will now state the main result of this paper which serves as the design for the AVQ algorithm. The proof of this result is given in Section 4.

THEOREM 1. *Suppose that the feedback delay $\hat{d}$, number of users $\hat{N}$, and the utilization $\hat{\gamma}$, are given. Find α^* satisfying:*

$$\omega\hat{d} + \arctan\left(\frac{\omega}{K_{11}}\right) = \frac{\pi}{2}, \tag{6}$$

where ω is defined as:

$$\omega(\alpha, d, N, \gamma) = \frac{1}{\sqrt{2}}\sqrt{(K_{12}^2 - K_{11}^2) + \sqrt{(K_{12}^2 - K_{11}^2)^2 + 4K_2^2\alpha^2}}.$$

Then, for all $\alpha < \alpha^*$, the system is stable. Moreover, for every $\alpha < \alpha^*$, the system remains stable for all $N > \hat{N}$, $\gamma < \hat{\gamma}$ and $d < \hat{d}$. ■

We note that (6) can be easily solved using a simple numerical solver such as those found in *Mathematica* or *Matlab*.

3. SIMULATIONS

In this section, we will use the packet-simulator *ns-2* to simulate the adaptive virtual queue scheme. We will show that simulation results agree with the convergence results shown in the previous section. In particular, we will select an α, using Theorem 1 that will ensure stability for a given round-trip delay d, and a lower bound on the number of users, $N_{\min}$. We then present a single set out of many experiments that we did to show that α indeed stabilizes the system even in the presence of arrivals and departure of short connections. We will then compare this scheme with many other AQM schemes.

3.1 Simulation Setup

Throughout this section, we consider a single link of capacity 10 Mbps that marks or drops packets according to some AQM scheme. For AVQ, we always let γ, the desired utilization, be 0.98. We use TCP-Reno as the default transport protocol and assume that packets have an average size of 1000 bytes. Each TCP connection has a propagation delay between 40 ms and 130 ms. The buffer size at the link is assumed to be 100 packets.

In the first four experiments, we assume that the link marks packets and thus, any packet loss is due to buffer overflow. In these experiments, we demonstrate that the AVQ scheme achieves high utilization and low packet loss. Further, the algorithm responds quickly to changing network conditions such as varying number of TCP flows. In the last experiment, we compare the AVQ scheme with other schemes when the link drops packets (as opposed to marking) to indicate congestion. Again, the AVQ scheme is shown to have smaller queue lengths compared to other schemes.

The maximum delay that a packet can incur is $T_p + \frac{qmax}{C}$, where T_p is the round-trip propagation delay. Therefore, we design the AVQ controller for a delay of $d = 210$ ms. Using the design rule in Theorem 1, any $\alpha < 0.17$, will ensure stability. In the experiments, we let α be 0.15. In all experiments, we consider two types of flows: FTP flows that are long lived and short flows of 20 packets each.

Experiment 1:
In this experiment, we study the convergence properties and buffer sizes at the queue for the AVQ scheme alone. The number of FTP flows is 180 while the short flows arrive at the link at the rate of 30 flows per second. To simulate a sudden change in network conditions, we start the experiment with only FTP flows in the system. Short flows are introduced after 100 seconds. Again, the propagation delays of the short flows are distributed in the interval [40, 130] ms. The evolution of the virtual capacity is given in Figure 2. At 100 seconds, there is a drop in the virtual capacity since the AVQ algorithm adapts to the changing number of flows. Beyond 100 seconds, the virtual capacity is lower than it was before 100 seconds since the links marks packets aggressively due to the increased load. The queue length evolution for the system is given in Figure 1. Except during transients introduced by load changes, the queue lengths are small, less than 20 packets. At 100 seconds, the queue length jumps up due to the short flows. However, the system stabilizes and the queue lengths are small once again. Table 1 gives the average and the standard deviation of the queue length before and after the introduction of short flows. We can see

Table 1: Experiment 1. Mean and the standard deviation of the queue size before and after the introduction of short flows.

	Before Short Flows	After Short flows
Avg. Queue Size	13.11	10.39
Std. Deviation	20.44	15.17

the average queue lengths and the standard deviation are almost similar. Another important performance measure is the number of packets dropped due to buffer overflow in the system. Since ECN marking is used, we expect the number of packets lost due to buffer overflow to be small. Indeed only 10 out of roughly 250,000 packets are dropped. These drops are primarily due to the sudden additional load brought on by the short flows. Another performance measure that is of interest is the utilization of the link. The utilization was observed to be 0.9827, which is very close to the desired utilization of 0.98. ◇

We will now compare the AVQ scheme with other AQM schemes that have been proposed. Since there are many AQM schemes in the literature, we will compare the AVQ scheme with a representative few. In particular, we will compare the AVQ scheme with

1. Random Early Discard (RED) proposed in [4]. In our experiments, we use the "gentle" version of RED. Unless otherwise stated, the parameters were chosen as recommended in http://www.aciri.org/floyd/REDparameters.txt.

2. Random Early Marking (REM) proposed in [1]. The REM scheme tries to regulate the queue length to a desired value (denoted by $qref$) by adapting the marking probability. The REM controller marks each packet with a probability p which is updated periodically (say, every T seconds) as

   $$p[k+1] = 1 - \phi^{-\mu[k+1]},$$

   where ϕ is a arbitrary constant greater than one and

   $$\mu[k+1] = \max(0, \mu[k] + \gamma(q[k+1] - (1-\alpha)q[k] - \alpha qref)),$$

   and α and γ are constants and $q[k+1]$ is the queue length at the $k+1$ sampling instant. Since REM is very sensitive to ϕ, we will use the values as recommended in [1]

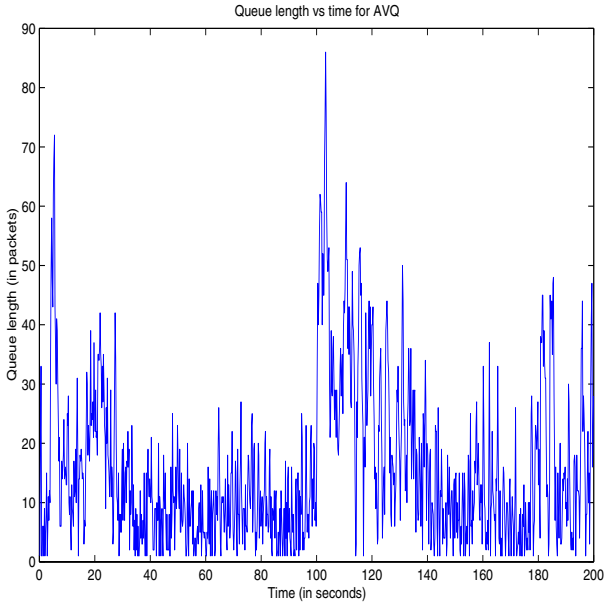

Figure 1: Experiment 1. Queue length vs time for the AVQ scheme

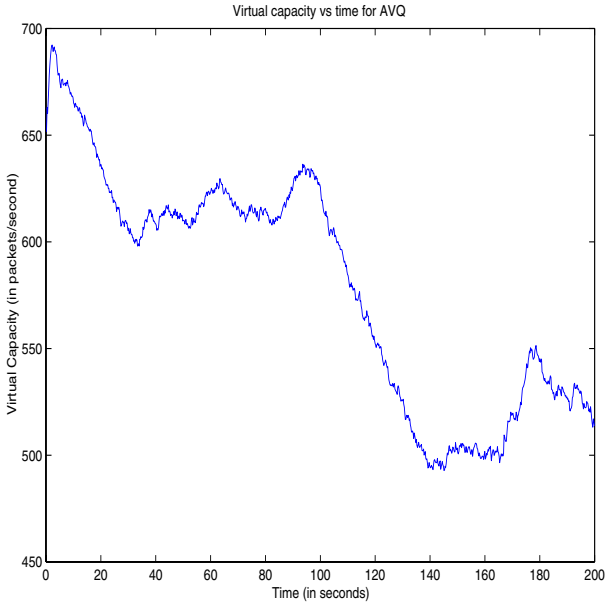

Figure 2: Experiment 1. Evolution of the virtual capacity with time for the AVQ scheme

3. The PI controller proposed in [6]. The PI controller marks each packet with a probability p which is updated periodically (say, every T seconds) as

$$p[k+1] = p[k] + a(q[k+1] - qref) - b(q[k] - qref),$$

where $a > 0$ and $b > 0$ are constants chosen according to the design rules given in [6].

4. The virtual queue based AQM scheme (GKVQ) proposed by [5]. In this scheme, the link maintains a virtual queue with fixed capacity $\tilde{C} = \theta C$, and buffer size $\tilde{B} = \theta B$, where $\theta < 1$, and B is the buffer capacity of the original queue. Whenever the virtual queue overflows, all packets in the real queue and all future incoming packets are marked till the virtual queue becomes empty again. Note that this scheme cannot be used in the case where the link drops the packets instead of marking them because the throughput would be very bad due to aggressive dropping. As in [5], we will use $\theta = 0.90$ in all our simulations using the GKVQ.

Experiment 2:
In this experiment, we compare the performance of the various AQM schemes assuming that the link "marks" packets and in the presence of long-lived FTP flows only. The queue size at the link is 100 packets and we let the desired average queuing delay be between 30ms and 60ms. As a result, the desired queue length for the REM scheme and the PI scheme is set at 50 packets and the minthresh and the maxthresh for the RED (with gentle turned on) scheme is set at 37 and 75 packets respectively. Recall that the desired utilization of the link was set to be 0.98 for the AVQ scheme.

Since we use an average queue length of 50 packets for REM and the PI controller, it is natural to attempt to regulate the queue length to 50 for the AVQ scheme also. However, the AVQ does not directly attempt to control queue size. Thus, for the AVQ scheme, we drop every packet that arrives when there are already 50 packets in the real queue. Note that this is the worst-case scenario for the AVQ scheme, since when ECN marking is used, the natural primary measure of performance is packet loss.

We summarize our simulation results below:

- Packet Losses and Link Utilization: The losses incurred by all the schemes are shown in Figure 3 as a function of the number of FTP flows. The AVQ scheme has fewer losses than any other scheme except the GKVQ even at high loads. The loss rate for GKVQ and AVQ are comparable; however, the GKVQ marks packets more aggressively than any other scheme and thus has lower utilization. Figure 4 shows the utilization of the link for all the AQM schemes. Note that, the utilization of GKVQ is as low as 75%. This can once again be attributed to the aggressive marking strategy of GKVQ. RED also results in a poor utilization of the link. We could have got a higher utilization with RED if we had increased the minthresh to a larger value, but we would have increased the packet drops at the link. REM and PI has an utilization of 1.0 as the queue is always non-empty. For the AVQ scheme, we required a desired utilization of 0.98 and we can see that the AVQ scheme tracks the desired utilization quite well. Thus, the main conclusion from this experiment is that the AVQ achieves low loss with high utilization.

- Responsiveness to changing network conditions: We measure the response of each AQM scheme to different numbers of FTP flows, by letting the number of flows be constant over a 100 second interval and then increasing it. The average queue length (over each 100

second interval) of each scheme as the number of users increase is shown in Figure 5. We see from the figure that PI and REM have higher average queue lengths than the desired queue length. On the other hand, AVQ, GKVQ and RED has smaller queue sizes. This is due to the fact that it takes REM and PI have a long transient period before the queue length converges. The average queue length over each 100 second interval does not delete any transients since one of our goals in this experiment is to study the responsiveness of the AQM scheme to load changes.

Experiment 3:
In this section, we will compare the responsiveness of the AQM schemes when flows are dropped and then introduced later on. Specifically, we only compare REM and the PI controller (since these are only ones among those that we have discussed that attempt to precisely regulate the queue length to a desired value) with the AVQ controller. Unless otherwise stated, all the system parameters are identical to Experiment 2. The number of FTP connections is 140 at time 0.0 At time 100, 105 FTP connections are dropped and at time 150 another 105 FTP connections are established. We will plot the evolution of the queue size for each of the AQM scheme. Figure 6 shows the evolution of the queue size for PI as the flows depart and come. Note that the desired queue length is 50 packets. We can see that the system takes some time to respond to either the departure of the flows or to the new arrivals. On the other hand, the queue in the AVQ scheme in Figure 7 responds quickly to the removal of flows at time $t = 100$, and to the addition of flows at time 150. Figure 8 gives the evolution of the queue sizes for REM. The desired queue level in the REM scheme is 50 packets and REM is very slow to bring the queue level to 50 packets. On removing flows, the queue level drops, but on addition of new flows, there is a large overshoot in REM.

Experiment 4:
Till now we have been comparing AVQ and PI in the absence of short flows. However, a large part of the connections in the Internet comprise of short flows. As a result, it is important to study the performance of an AQM scheme in the presence short flows. In this experiment, we will start with 40 FTP connections that remain throughout the length of the experiment. We also allow the AQM schemes to converge to the optimal solution when there are only 40 FTP connections in the network. We then introduce short flows and study the performance of the AQM scheme as the number of short flows increases. We start with a short-flow arrival rate of 10 per second and gradually increase it to 50 short flows per second. Each short flow transfers 20 packets using TCP/Reno. The round-trip times of the short flows are also distributed uniformly between 40ms and 130ms.

We again study the following performance measures:

- Packet losses and Utilization: The losses incurred by both the schemes are shown in Figure 9. Note that AVQ has lower drops than the RED, REM and the PI schemes. GKVQ incurs no significant packet drops (and hence cannot be seen in the figure) among all the schemes because of its aggressive marking scheme. However, as in Experiment 2, the utilization of GKVQ is poor as seen in Figure 10. We again see that REM and PI have an utilization of one, while RED and GKVQ have poor utilization. Once again, the utilization of RED can be made higher, but this will come at the expense of higher average queue lengths and more packet losses. For the AVQ scheme, the utilization is actually slightly higher than the desired utilization at high loads, but this can be attributed to the short-flows.

- Queue length: The average queue length of each scheme as the rate of the incoming short-flows short connections are increased is shown in Figure 11. We see that the the AVQ controller maintains the smallest queue length.

Experiment 5:
Till now, we have been assumed that the router marks packets upon detecting congestion. Instead one can drop packets when congestion is detected. In this experiment, we use dropping instead of marking when the links detects an incipient congestion event.

Note that, in the case of marking, the main goal of the adaptive algorithm was to match the total arrival rate to the desired utilization of the link. However, in the case of dropping, the link only serves those packets that are admitted to the real queue. As a result, in the case of dropping, one adapts the virtual capacity ($\tilde{C}$) only when a packet has been admitted to the real queue, i.e., only the accepted arrival rate is taken into consideration.

We compare RED, REM and PI controller to the AVQ scheme. We do not use GKVQ as a dropping algorithm as the number of packets dropped on detecting congestion would be very high and it would result in negligible throughput. The buffer limit at the link is set to 100 packets and we require the average queueing delay to lie between 30ms and 60 ms. The users employ TCP NewReno. All the other parameters are as in Experiment 2. However, in this case we simulate the AVQ scheme with both $\gamma = 1.0$ and $\gamma = 0.98$. The reason for using $\gamma = 0.98$ earlier was to have small losses to get the most benefit from ECN marking. Since marking is no longer used, we also study the AVQ under full utilization.

We assume that 40 FTP connections use the link for the entire duration of the simulation. We allow the respective AQM schemes to converge and then introduce short-flows at 100s. Short-flows introduced are TCP-RENO sources with 20 packets to transmit. The rate at which short flows arrive at the link is slowly increased. The average queue length, and the utilization are shown in Figure 12 and Figure 13. The total goodput is shown in Figure 14. By goodput, we mean the number of packets successfully delivered by the link to the TCP receivers. In general, this could be different from the throughput (which is the total number of packets processed by the link) due to TCP's retransmission mechanism. Note that the average queue length, the goodput of each flow and fairness are the three performance objectives that one would use to compare different AQM schemes when dropping is employed as a congestion notification mechanism. In practice, we would like an AQM scheme that maintains a small average queue length with high utilization. However, the AQM scheme should not introduce any additional bias in the rates towards smaller round-trip flows (TCP by itself introduces a bias towards smaller round-trip flows and we do not want to add it). In this experiment, we compared the average queue length and the utilization at the link of AVQ, RED, REM and PI.

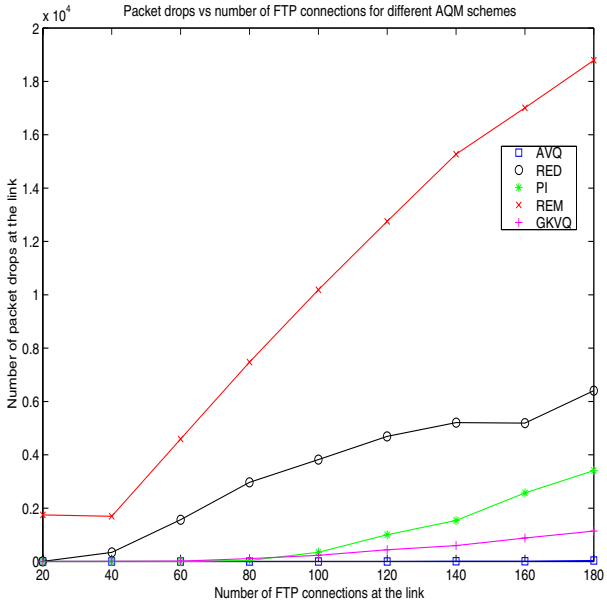

Figure 3: Experiment 2. Losses at the link for varying number of FTP connections for the different AQM schemes

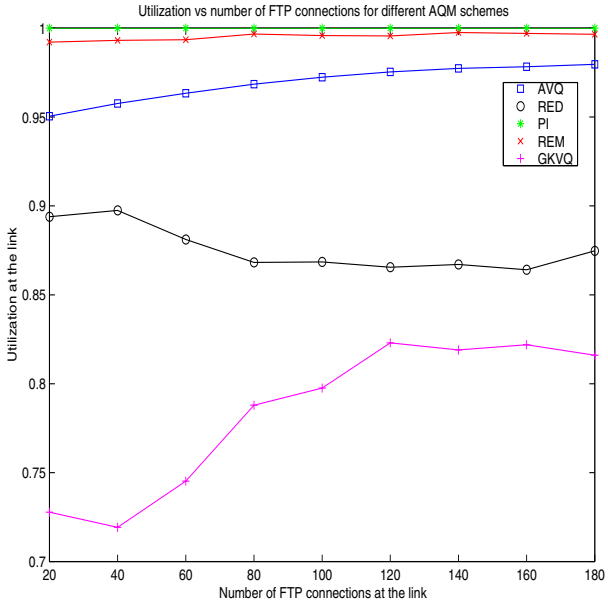

Figure 4: Experiment 2. Achieved Utilization at the link for the different AQM schemes

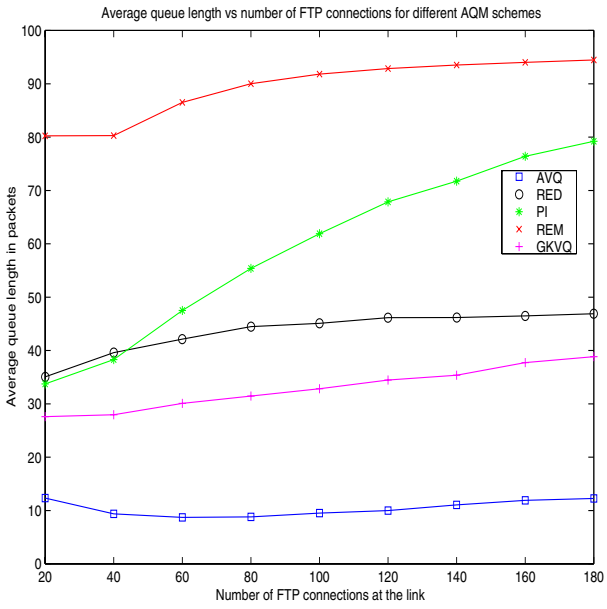

Figure 5: Experiment 2. Queue length at the link for varying number of FTP connections for the different AQM schemes

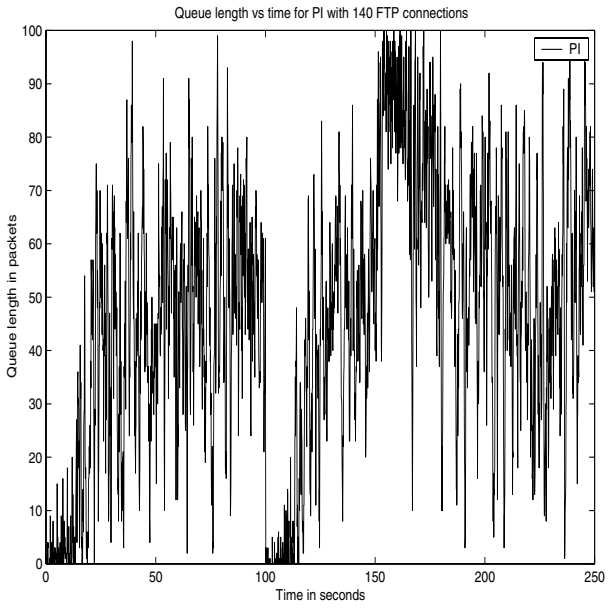

Figure 6: Experiment 3. Evolution of the queue length at varying loads for PI

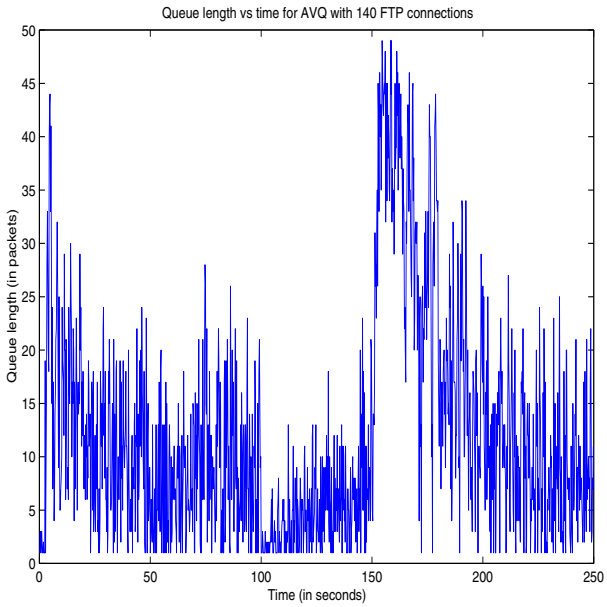

Figure 7: Experiment 3. Evolution of the queue sizes at varying loads for AVQ

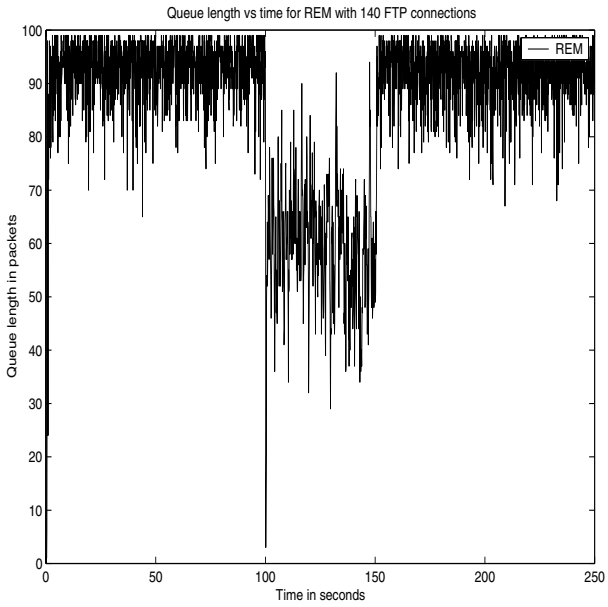

Figure 8: Experiment 3. Evolution of the queue sizes at varying loads for REM

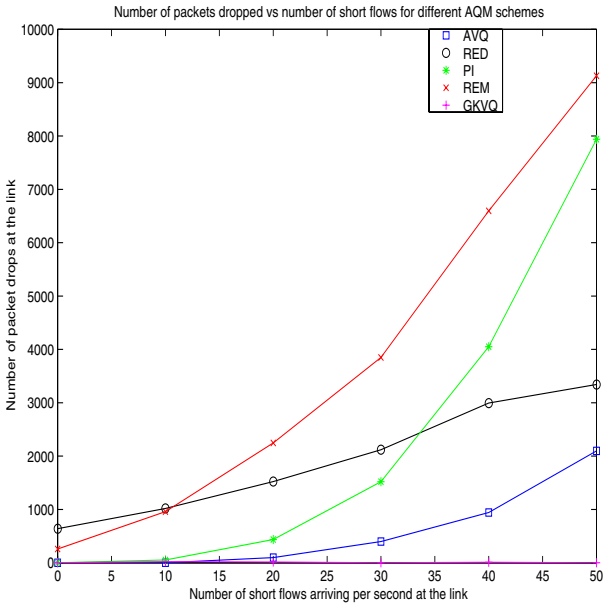

Figure 9: Experiment 4. Packet losses at the link for various AQM schemes

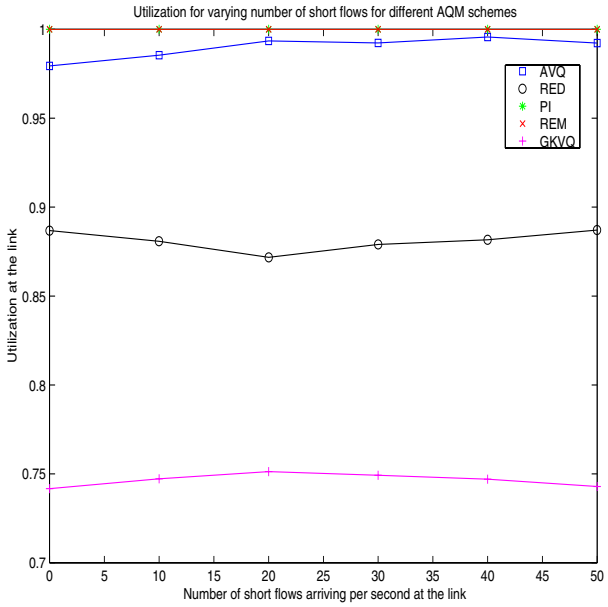

Figure 10: Experiment 4. Utilization of the link for various AQM schemes

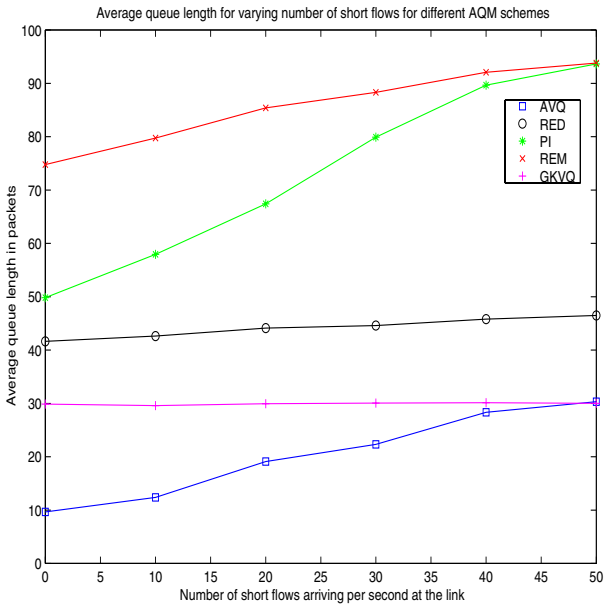

Figure 11: Experiment 4. Average queue length at the link for various AQM schemes

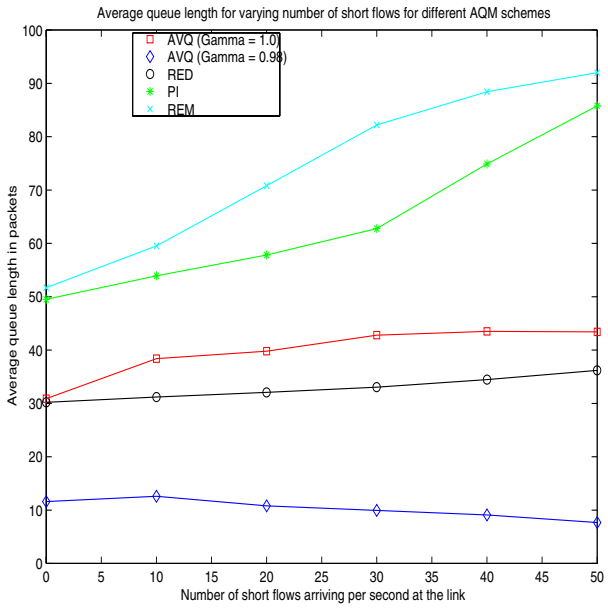

Figure 12: Experiment 5. Average queue lengths for various AQM schemes

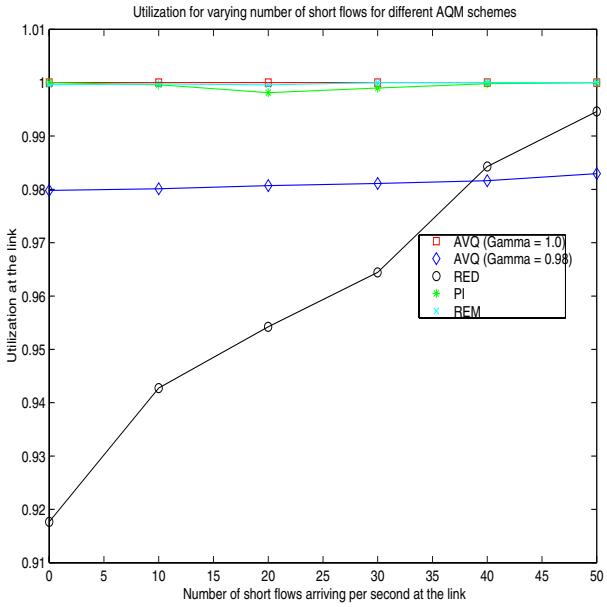

Figure 13: Experiment 5. Utilization for various AQM schemes

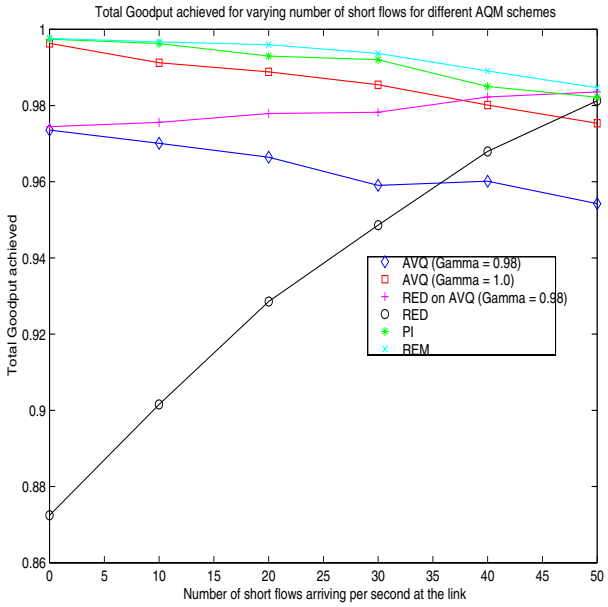

Figure 14: Experiment 5. Total Goodput for the various AQM schemes

The fairness measures of the AVQ algorithm are not shown here due to space limitations.

Note: Instead of marking or dropping a packet in the real queue when the virtual queue overflows, one can mark or drop packets in the real queue by applying RED (or any other AQM algorithm) in the virtual queue. Thus, if there are desirable features in other AQM schemes, they can be easily incorporated in the AVQ algorithm. When marking is employed, our experience is that a simple mark-tail would be sufficient as shown in Experiments 1 through 4. In the case when the link drops the packets, many successive packet drops from the same flow could cause time-outs. To avoid this, one could randomize the dropping by using a mechanism like RED in the virtual queue to prevent bursts of packets of the same flow to be dropped. Our experience has been that, if RED is employed in the virtual queue, the performance of the AQM scheme is not very sensitive to the choice of the RED parameters.

4. STABILITY ANALYSIS OF THE AVQ SCHEME

In this section, we will prove the main result of the paper which was stated in Theorem 1. The starting point of the analysis is the linearized version of the TCP/AQM model derived in (4) and (5). We summarize the main ideas behind the proof:

- The stability of a linear delay-differential equation can be analyzed using its characteristic equation. The characteristic equation of the linear delay-differential equation can be obtained by taking its Laplace Transform. For the linearized system to be stable, its characteristic equation should have all its roots in the left-half plane (i.e., if σ is a root of the characteristic equation, then $\mathrm{Re}[\sigma] < 0$).
- We will first show that for α, N, and γ fixed the system is stable in the absence of feedback delays, i.e., $d = 0$. This implies that all the roots of the characteristic equation lie in the left-half plane. The roots of the characteristic equation are continuous functions of its parameters. Therefore, the roots of the characteristic equation are continuous function of the feedback delay d. By increasing d, one can find the smallest feedback delay d^* at which one of the roots hits the imaginary axis (if there is no such d, then the system is stable for all d.). Hence, for all $d < d^*$, the system has all its roots in the left-half plane and hence it is stable. This is the key idea and we will be using this idea frequently throughout this section.

Recall that the linearized TCP/AQM system was given in (4) and (5). For analytical tractability, we assume that

$$p(\lambda, \tilde{C}) = \frac{(\lambda - \tilde{C})^+}{\lambda}. \tag{7}$$

Note that, while this is not differentiable everywhere in λ or $\tilde{C}$, it is differentiable in the region $\lambda > \tilde{C}$. Substituting for $\frac{\partial p(\gamma C, \tilde{C}^*)}{\partial \lambda}$, and $\frac{\partial p(\gamma C, \tilde{C}^*)}{\partial \tilde{C}}$ and using the fact that $p(\gamma C, \tilde{C}^*) = \frac{N^2}{\beta(d\gamma C)^2}$, we find that

$$K_{11} = \frac{N}{\gamma C d^2} \qquad K_{12} = K_2 = \beta \frac{\gamma C}{N}. \tag{8}$$

Note that $K_{12} > K_{11}$. Let $\Lambda(s)$ denote the Laplace-Transform of $\delta\lambda(t)$ and let $\Psi(s)$ denote the Laplace-transform of $\delta\tilde{C}(t)$. Taking the Laplace-transforms of (4) and (5), we get:

$$s\Lambda(s) = -K_{11}\Lambda(s) - K_{12}e^{-sd}\Lambda(s) + K_2 e^{-sd}\Psi(s) \tag{9}$$

$$s\Psi(s) = -\alpha\Lambda(s). \tag{10}$$

Substituting (10) in (9), we get the so-called characteristic equation

$$s + K_{11} + e^{-sd}\left(K_{12} + \alpha\frac{K_2}{s}\right) = 0. \tag{11}$$

Once again, the key idea in this approach is that roots are continuous functions of the round-trip delay d. As a result, if the system is stable with $d = 0$ for a fixed value of α, then the roots are strictly in the left-half plane. Therefore, we can choose d small enough such that the roots still remain in the left-half plane. This will help us to find the maximum feedback delay for which the system is stable for a given α. We will then show that we can use the same technique to show that given a feedback delay d, one can find the maximum value of α for which the system is stable. We will then discuss the impact of the number of users on the stability of the system. We will formalize these ideas in this section.

When $d = 0$, i.e., there is no feedback delay in the system, the characteristic equation reduces to:

$$s + K_{11} + K_{12} + \alpha\frac{K_2}{s} = 0. \tag{12}$$

Solving the quadratic equation, we get:

$$s = \frac{-(K_{11} + K_{12}) \pm \sqrt{(K_{11} + K_{12})^2 - 4\alpha K_2}}{2}.$$

If $4\alpha K_2 \le (K_{11} + K_{12})^2$, then the system has all real roots which lie strictly in the left half-plane. If $4\alpha K_2 > (K_{11} + K_{12})^2$, then the system has complex roots that also lie strictly in the left half-plane. Thus, for all values of $\alpha > 0$, the system is stable.

The following theorem gives the necessary condition on the RTT for the stability of the system given by (4) and (5).

THEOREM 2. *Fix $\alpha = \hat{\alpha}$, the number of TCP users, N and the utilization γ. Find the smallest $d = \hat{d}$ such that*

$$\omega(\hat{\alpha}, d, N, \gamma) = \frac{1}{\sqrt{2}}\sqrt{(K_{12}^2 - K_{11}^2) + \sqrt{(K_{12}^2 - K_{11}^2)^2 + 4K_2^2\hat{\alpha}^2}} \tag{13}$$

satisfies

$$\omega d + \arctan\left(\frac{\hat{\alpha}}{\omega}\right) + \arctan\left(\frac{\omega}{K_{11}}\right) = (2k+1)\pi, \tag{14}$$

for some $k = 0, 1, 2, \cdots$. Then, the TCP/AQM system given in (4) and (5) is stable for all values of $d < \hat{d}$.

Proof: The characteristic equation of the TCP/AQM system (11) can be rewritten as:

$$1 + \frac{e^{-sd}\left(K_{12} + \alpha\frac{K_2}{s}\right)}{s + K_{11}} = 0. \tag{15}$$

Let $j\omega$ be one of the roots of the characteristic equation at the smallest $d = d^*$ such that the roots hits the imaginary

axis. Since the roots on the imaginary axis are complementary, we will concern ourselves only with $\omega \geq 0$. From (15), we get:

$$\frac{e^{-j\omega d}\left(K_{12}+\alpha\frac{K_2}{j\omega}\right)}{j\omega+K_{11}} = -1.$$

To satisfy the last condition, the following conditions must be met simultaneously:
Condition on magnitude:

$$\left|\frac{e^{-j\omega d}\left(K_{12}+\alpha\frac{K_2}{j\omega}\right)}{j\omega+K_{11}}\right| = 1$$

Condition on angles:

$$\angle\frac{e^{-j\omega d}\left(K_{12}+\alpha\frac{K_2}{j\omega}\right)}{j\omega+K_{11}} = (2k+1)\pi \qquad k=0,\pm1,\pm2.$$

From the condition on magnitude, we get

$$\frac{\sqrt{K_{12}^2+\frac{K_2^2\hat{\alpha}^2}{\omega^2}}}{\omega\sqrt{K_{11}^2+\omega^2}} = 1$$

$$\Rightarrow \omega(\hat{\alpha},d,N,\gamma) = \sqrt{\frac{(K_{12}^2-K_{11}^2)+\sqrt{(K_{12}^2-K_{11}^2)^2+4K_2^2\hat{\alpha}^2}}{2}}.$$

From the condition on angles, we get:

$$\omega d + \arctan\left(\frac{\hat{\alpha}}{\omega}\right) + \arctan\left(\frac{\omega}{K_{11}}\right) = (2k+1)\pi,$$

for $k = 0, 1, 2, \cdots$. Since K_{11} is a decreasing function of d, and K_{12} and K_2 are independent of d, we have $\omega(\hat{\alpha}, d, N, \gamma)$ as an increasing function of d. Therefore, the smallest $d = \hat{d}$ that solves (14) gives the first time that at least one of the roots hit the imaginary axis. Therefore, for all $d < \hat{d}$, the system is locally asymptotically stable. ■

REMARK: Although the above theorem provides a necessary and sufficient condition, it is hard to verify the conditions of the theorem numerically due to the following issues:

- What value of k will yield the smallest d?
- If $\hat{d}$ solves (14), how can we be sure that there exists no $\tilde{d}$, such that $\tilde{d}$ solves (14) and $\tilde{d} < \hat{d}$?

The following theorem solves these issues by giving an easily verifiable sufficient condition for stability.

THEOREM 3. *Fix $\alpha = \hat{\alpha}$, the number of TCP users, N and the utilization γ. Find any $d^* > 0$, such that*

$$\omega d^* + \arctan\left(\frac{\omega}{K_{11}}\right) = \frac{\pi}{2}, \tag{16}$$

where ω is as given in (13). Then for all $d < d^*$ the system is stable. Moreover, d^* is unique.

Proof: Note that K_{12} and K_2 do not depend on d. Therefore, as d increases, K_{11} decreases and ω increases.

Let $\hat{d}$ solve (14) for some k. Then, we claim that

$$d^* < \hat{d}. \tag{17}$$

Suppose not. Since

$$\arctan(\frac{\hat{\alpha}}{\omega}) \leq \frac{\pi}{2},$$

$$\omega(\hat{\alpha},\hat{d},N,\gamma)\hat{d}+\arctan(\frac{\omega(\hat{\alpha},\hat{d},N)}{K_{11}(\hat{d})}) \geq \frac{4k+1}{2}\pi, k=0,1,2\ldots \tag{18}$$

Also, since $\hat{d} < d^*$, $\omega(\hat{\alpha},\hat{d},N,\gamma) < \omega(\hat{\alpha},d^*,N,\gamma)$ and $K_{11}(\hat{d}) > K_{11}(d^*)$. Therefore,

$$\arctan(\frac{\omega(\hat{\alpha},\hat{d},N,\gamma)}{K_{11}(\hat{d})}) < \arctan(\frac{\omega(\hat{\alpha},d^*,N,\gamma)}{K_{11}(d^*)}).$$

Thus,

$$\left\{\begin{array}{l}\omega(\hat{\alpha},\hat{d},N,\gamma)\hat{d}+\\ \tan^{-1}(\frac{\omega(\hat{\alpha},\hat{d},N,\gamma)}{K_{11}(\hat{d})})\end{array}\right\} < \left\{\begin{array}{l}\omega(\hat{\alpha},d^*,N,\gamma)d^*+\\ \tan^{-1}(\frac{\omega(\hat{\alpha},d^*,N,\gamma)}{K_{11}(d^*)})\end{array}\right\} = \frac{\pi}{2}.$$

But this contradicts (18). Hence $d^* < \hat{d}$. Thus, any $d < d^*$ also satisfies $d < \hat{d}$, and therefore, for any $d < d^*$, the system is stable from theorem 2.

Suppose that d^* is not unique. Let d_1^* and d_2^* be two values that satisfy (16). Without loss of generality, let us assume $d_1^* < d_2^*$. Therefore,

$$\left\{\begin{array}{l}\omega(\hat{\alpha},d_1^*,N,\gamma)d_1^*+\\ \tan^{-1}(\frac{\omega(\hat{\alpha},d_1^*,N,\gamma)}{K_{11}(d_1^*)})\end{array}\right\} < \left\{\begin{array}{l}\omega(\hat{\alpha},d_2^*,N,\gamma)d_2^*+\\ \tan^{-1}(\frac{\omega(\hat{\alpha},d_2^*,N,\gamma)}{K_{11}(d_2^*)})\end{array}\right\} = \frac{\pi}{2}.$$

But this is a contradiction. Hence, d^* is the unique solution to (16). ■

EXAMPLE 1. *Consider a single link with 10Mbps capacity and let there be 180 users accessing it. Let the average packet size be 1000 bytes and $\gamma = 1.0$. The capacity of the link can be written as 1250 packets per second. Let $\hat{\alpha} = 0.1$. We are interested in finding d^* such that the system is stable for all $d < d^*$. Using (16) to solve for d^*, we get $d^* = 0.210$. Therefore, for all $d < 0.210$ seconds, the system is locally stable.*

Till now, we have been given a fixed α and a fixed N and we were interested in finding the largest feedback delay for which this system is stable. But, a more practical question is the following: given a feedback delay $\tilde{d}$, and number of users N, how can one design α such that the system is stable? The next theorem gives a method by which one can design α such that system is stable. Note that this theorem is the main result of the paper and is also stated in Section 2. We state it again for convenience.

THEOREM 4. *Fix the feedback delay $\tilde{d}$, the number of users N and the utilization γ. Find α^* satisfying:*

$$\omega\tilde{d} + \arctan\left(\frac{\omega}{K_{11}}\right) = \frac{\pi}{2}, \tag{19}$$

where ω is as given in (13). Then, for all $\alpha < \alpha^*$, the system is stable.

Proof: Fix any $\hat{\alpha} < \alpha^*$. Let $\tau(\alpha)$ be the d satisfying (16). We have to show that, $\tau(\alpha^*) > \tilde{d}$. We know that $\tau(\alpha^*) = \tilde{d}$. For any fixed d, we know that $\omega(\alpha, d, N, \gamma)$ decreases as α decreases. Therefore,

$$\left\{\begin{array}{l}\omega(\hat{\alpha},\tilde{d},N,\gamma)\tilde{d}+\\ \tan^{-1}\left(\frac{\omega(\hat{\alpha},\tilde{d},N,\gamma)}{K_{11}}\right)\end{array}\right\} < \left\{\begin{array}{l}\omega(\alpha^*,\tilde{d},N,\gamma)\tilde{d}+\\ \tan^{-1}\left(\frac{\omega(\alpha^*,\tilde{d},N,\gamma)}{K_{11}}\right)\end{array}\right\} = \frac{\pi}{2}.$$

Since $\omega(\alpha, d, N, \gamma)$ is an increasing function of d, we have,

$$\tau(\hat{\alpha}) > \tilde{d}.$$

Therefore, for all $\alpha < \alpha^*$, the system is locally stable. ■

EXAMPLE 2. *Consider the same setting as in Example 1. Let* $\tilde{d} = 0.21$. *We are interested in finding* α^* *such that the system is stable for all* $\alpha < \alpha^*$. *Using (19) to solve for* α^*, *we get* $\alpha^* = 0.10$. *Therefore, for all* $\alpha < 0.10$, *the system is locally stable.*

Another important design aspect is how stability is affected as the number of users changes. In general, given an $\hat{\alpha}$ and $\hat{d}$, can one find the number of users required to make the system stable. The following theorem gives a lower bound on the number of users required to make the system stable.

THEOREM 5. *Fix the feedback delay* $\hat{d}$, *the smoothing parameter* $\hat{\alpha}$ *and the utilization* γ. *Find* $\hat{N}$ *satisfying:*

$$\omega\hat{d} + \arctan\left(\frac{\omega}{K_{11}}\right) = \frac{\pi}{2}, \tag{20}$$

where ω *is as given in (13). Then, for all* $N > \hat{N}$, *the system is stable.*

Proof: Note that in this case, K_{11}, K_{12}, and K_2 are all functions of N. We can easily show that as N increases, K_{11} increases, K_{12} decreases, K_2 decreases and $\omega(\alpha, d, N, \gamma)$ decreases. Using this and following along the lines of the proof for Theorem 4, we can show that for all $N > \hat{N}$, the system is stable. ■

Remarks: One of the most important applications of Theorem 5 comes in the design of α. The system is initially designed for a low value of $N = \hat{N}$, and a particular d. We can now use Theorem 4 to find the value of α^* that will lead to stability. However, Theorem 5 assures us that the system will still be stable when the number of users increases beyond $\hat{N}$. We can state a similar theorem for the utilization γ of the link.

THEOREM 6. *Fix the feedback delay* $\hat{d}$, *number of users* $\hat{N}$ *and the smoothing parameter* $\hat{\alpha}$. *Find* $\hat{\gamma}$ *satisfying:*

$$\omega\hat{d} + \arctan\left(\frac{\omega}{K_{11}}\right) = \frac{\pi}{2}, \tag{21}$$

where ω *is as given in (13). Then, for all* $\gamma < \hat{\gamma}$, *the system is stable.* ■

5. CONCLUSIONS

In this paper, we presented an easily implementable AQM called the Adaptive Virtual Queue and provided simple design rules to choose its parameters. We then showed through simulations that the AVQ controller performs better than a number of other well-known AQM schemes.

6. REFERENCES

[1] S. Athuraliya, D. E. Lapsley, and S. H. Low. Random early marking for Internet congestion control. In *Proceedings of IEEE Globecom*, 1999.

[2] W. Feng, D. Kandlur, D. Saha, and K. Shin. Blue: A new class of active queue management algorithms. U. Michigan CSE-TR-387-99, April 1999.

[3] S. Floyd. TCP and explicit congestion notification. *ACM Computer Communication Review*, 24:10–23, October 1994.

[4] S. Floyd and V. Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, August 1993.

[5] R. Gibbens and F. Kelly. Distributed connection acceptance control for a connectionless network. In *Proc. of the 16th Intl. Teletraffic Congress*, Edinburgh, Scotland, June 1999.

[6] C. Hollot, V. Misra, D. Towlsey, and W. Gong. A control theoretic analysis of RED. UMass CMPSCI Technical Report 00-41, 2000.

[7] C. Hollot, V. Misra, D. Towlsey, and W. Gong. On designing improved controllers for AQM routers supporting TCP flows. UMass CMPSCI Technical Report 00-42, 2000.

[8] P. Hurley, J.-Y. L. Boudec, and P. Thiran. A note on the fairness of additive increase and multiplicative decrease. In *Proc. of the 16th Intl. Teletraffic Congress*, Edinburgh, Scotland, June 1999.

[9] F. Kelly. Mathematical modeling of the Internet. In *Proc. of the 4th Intl. Congress on Industrial and Applied Mathematics*, Edinburgh, Scotland, July 1999.

[10] F. Kelly, P.Key, and S. Zachary. Distributed admission control. *IEEE Journal on Selected Areas in Communications*, 18, 2000.

[11] S. Kunniyur and R. Srikant. End-to-end congestion control: utility functions, random losses and ECN marks. In *Proceedings of INFOCOM 2000*, Tel Aviv, Israel, March 2000.

[12] S. Kunniyur and R. Srikant. A time-scale decomposition approach to adaptive ECN marking. In *To be published in the Proceedings of INFOCOM 2001*, Alaska, Anchorage, April 2001.

[13] L. Massoulie and J. Roberts. Bandwidth sharing: Objectives and algorithms. In *Proc. INFOCOM*, New York, NY, March 1999.

[14] V. Misra, W. Gong, and D. Towlsey. A fluid-based analysis of a network of aqm routers supporting tcp flows with an application to red. In *Proceedings of SIGCOMM 2000*, Stockholm, Sweden, September 2000.

[15] T. J. Ott, T. V. Lakshman, and L. H. Wong. SRED: Stabilized RED. In *Proceedings of INFOCOM*, New York, NY, March 1999.

[16] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: A simple model and its empirical validation. In *Proceedings of SIGCOMM*, Vancouver, Canada, 1998.