Design and Analysis of a Multiscale Active Queue Management Scheme

Qi-Jin Ji (纪其进) and Yong-Qiang Dong (董永强)

School of Computer Science and Engineering, Southeast University, Nanjing 210096, P.R. China

Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University Nanjing 210096, P.R. China

E-mail: engine.cn@gmail.com; dongyq@seu.edu.cn

Received March 23, 2005; revised May 15, 2006.

Abstract Since Internet is dominated by TCP-based applications, active queue management (AQM) is considered as an effective way for congestion control. However, most AQM schemes suffer obvious performance degradation with dynamic traffic. Extensive measurements found that Internet traffic is extremely bursty and possibly self-similar. We propose in this paper a new AQM scheme called multiscale controller (MSC) based on the understanding of traffic burstiness in multiple time scale. Different from most of other AQM schemes, MSC combines rate-based and queue-based control in two time scales. While the rate-based dropping on burst level (large time scales) determines the packet drop aggressiveness and is responsible for low and stable queuing delay, good robustness and responsiveness, the queue-based modulation of the packet drop probability on packet level (small time scales) will bring low loss and high throughput. Stability analysis is performed based on a fluid-flow model of the TCP/MSC congestion control system and simulation results show that MSC outperforms many of the current AQM schemes.

Keywords active queue management, multiscale traffic burstiness, fluid-flow model, stability analysis

1 Introduction

The congestion control mechanism of Transmission Control Protocol (TCP) is essential to make sure the Internet operating in a stable regime[1]. But when TCP senses congestion and takes action accordingly, the network has been congested or is experiencing congestion and suffers significant performance degradation. Since congestion occurs at nodes inside the networks, e.g., routers in the Internet, it is expected that congestion avoidance at the network nodes is more efficient than congestion control at the end systems. IETF (Internet Engineering Task Force) recommended active queue management (AQM[2]) rather than Tail Drop (TD), which is employed in the traditional Internet, in this background.

AQM algorithms randomly drop or mark packets before the occurrence of congestion to make a good tradeoff between link utilization and queuing delay. The routers are responsible for detecting incipient congestion and control actions are taken timely if congestion occurs. Sources, or applications based on TCP, will cooperate with the control actions, i.e., adjust their sending rate or window size to avoid the congestion. There are two approaches to measuring congestion: 1) queue based, and 2) rate based. In queue based AQMs congestion is indicated by (smoothed) queue occupancy. The drawback of this approach is the backlog between congestion behavior and the queue length dynamics, which will result in undesirable control sluggishness and large delay jitter. Rate based AQMs, on the other hand, detect congestion and take action primarily based on the packet arriving rate. For these schemes, backlog, and all its negative implications, can be eliminated effectively.

Although many AQM schemes have been proposed, they are only efficient in some scenarios in terms of specific performance measures[3]. Network performance is closely related to the traffic characteristics according to the traffic theory. Numerous measurement results show that traffic in packet networks is bursty at both burst level and packet level[4] and self-similarity (Refer to [5] and references therein) was utilized to model such behavior. We present in this paper a novel AQM scheme combining rate-based and queue-based control in two time scales based on the understanding of multiscale burstiness pattern of Internet traffic, which is called multiscale controller (MSC). In MSC, packet dropping① is used as a control mechanism and the packet drop probability (PDP) is mainly determined by the burst-level rate of the traffic arriving at the bottleneck link. Further, this probability is modulated by the packet-level fluctuations of the incoming traffic. While the rate-based packet dropping on burst level (large time scales more than RTT) determines the packet drop aggressiveness and is responsible for low and stable queuing delay, good robustness and responsiveness, the queue-based modulation of the PDP on packet level (small time scales less than RTT) may reduce unnecessary packet loss and results in high throughput. Stability analysis is performed based on

Regular Paper

Supported by the National Grand Fundamental Research 973 Program of China under Grant No. 2003CB314801, the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No. 20040286001 and the National Natural Science Foundation of China under Grant No. 90604003.

① Packet marking is compatible with MSC with the support of end systems, which decreases packet loss at the cost of larger delay jitter.

a fluid-flow model of the TCP/MSC congestion control system and simulation results show that MSC outperforms many of the current AQM schemes. Since the design of MSC takes the traffic burstiness in multiple time scale in consideration and the PDP is updated far less frequently than packet arrival, it enjoys low computation complexity and scales well.

The remaining of the paper is organized as follows. Related work is examined in Section 2; analysis and design of MSC are presented in Section 3; in Section 4, fluid modeling and stability analysis of TCP/MSC system are performed; performance evaluations of MSC are described in Section 5 and conclusions are provided in Section 6.

2 Related Work

Since Floyd's proposal of RED algorithm[6], there has been a surge of interest in the design of AQM schemes. RED utilizes average queue length as both the congestion indication and performance objective. Adaptive RED (ARED[7]) is a new version of RED, which adjusts its parameters adapting to the queue length dynamics. BLUE[8] is another adaptive AQM scheme that adjusts the packet drop probability implicitly based on the load level of the bottleneck links. In BLUE, the load level is inferred from the packet loss and link idleness events, while it is explicitly estimated in MSC. AVQ[9] is a rate based AQM scheme but is only efficient when combined with explicit congestion notification (ECN[10]) and suffers high packet loss rate and low throughput in bursty traffic environment.

We have noted that there are some AQM algorithms resorting to the combination of queue length and traffic arriving rate in recent years, such as REM[11], PI controller[12], LDC[13], PAQM[14] and SFC[15]. These algorithms were designed in different background and ideas, and were claimed to perform better than previous methods. Compared with these algorithms, the most significant difference of MSC lies in the idea of time-scale decomposition, i.e., MSC combines rate-based and queue-based control in two different time scales while the others make the combination at the same time, e.g., on packet arrival or queue length sampling.

The authors of REM pointed out high throughput and low delay could be obtained simultaneously only when the performance and congestion measures were decoupled. Price is defined in REM and is updated according to the rate mismatch and queue length mismatch. Although designed in the framework of control theory, PI got similar result as REM. Both REM and PI use queue length difference to capture the rate variation and this does *not really* decouple the performance and congestion measures. In addition, both algorithms will result in slow response to abrupt traffic changes due to the integral action for control stability. More recently, PAQM uses traffic arriving rate estimation as congestion indication but it is load adaptive RED in essence. Multiple step prediction is required in PAQM to calculate packet drop/mark probability, which may result in non-ignorable error. A more detailed comparison of AQM algorithms can be found in [16].

3 Design of MSC

3.1 Design Rationale

As performance measures, both queue length and packet loss rate are related to but derive from the load level. We believe that load level of the bottleneck links is a better congestion indicator and rate based control schemes are more effective than only queue based ones.

The question now is whether it is *feasible* to get the load information conveniently. Contrary to traditional Poisson assumption, the Internet traffic is better characterized with self-similarity[5], which presents similar bursty behavior in multiple time scales ranging from tens of milliseconds to seconds and beyond. Considering TCP applications are the main contributor to today's Internet traffic, *typical* round trip time (RTT)[5] of TCP congestion control was identified to be a critical time scale separating bursty behavior into two different time scales corresponding to burst level and packet level congestion phenomena respectively.

Although the scale-invariant burstiness implies the existence of concentrated periods of contentions and idleness, the long-range dependence associated with self-similar traffic means significant correlation structure at large time scales and leaves open possibility to be exploited for the traffic control purpose. Burstiness in large time scales or at burst levels means significant low-frequency dynamics in the frequency domain and a low-pass filter or other proper signal processing techniques can be used to extract the low frequency component from the traffic[17]. A simple and robust traffic rate estimation method was suggested and successfully used in [18]. Self-similarity of Internet traffic was considered as a critical factor for designing congestion control schemes, and MTSC[19] was designed as an improved congestion control strategy. Also, the multiscaling behavior of Internet traffic is exploitable for designing AQM schemes.

While it is feasible to get low frequency traffic rate for designing traffic control schemes, the traffic is bursty in smaller time scales or at packet levels. Similarly, burstiness in small time scales means high frequency dynamics in frequency domain. For an accurate control, these deviations of traffic rate in small time scales should be compensated in some ways. Since the traffic fluctuations will cause queue length variation, a queue-based compensation is an appealing approach. It is preferable that at the same load level, packets should be dropped less aggressively when the buffer occupancy is low and somewhat more aggressively when the buffer occupancy is high to make sure that fewer packets are lost.

In summary, two main components should be comprised in MSC: a rate based control scheme working in large time scale greater than typical RTT with load level as a direct and robust congestion indicator and a queue based modulator in small time scale less than the RTT to accommodate the packet level burstiness.

3.2 Analysis of Rate Based Packet Drop

Obviously, taking load level (the ratio of *input* traffic rate to the service rate of a link) as the congestion indicator is compatible with the idea of decoupling in REM. Now the second question is how to transform the congestion measure (load level) to performance measure (queue length or packet loss), which is employed to regulate the control action directly. Random early packet dropping has been validated to be an effective control mechanism for AQM schemes and the packet drop probability (PDP) is essential for such mechanism. We present an analytical relationship between load level and PDP based on queuing theory.

Assuming that the link is busy and the queue is work conserving, the queue length dynamics can be described as

$$Q(t+\Delta t) = [\widehat{A}(t+\Delta t)(1-P_a) - C]\Delta t + Q(t) \quad (1)$$

where $\widehat{A}(t)$ is the estimated traffic arriving rate and Δt is the time scale of rate estimation. P_a is the average PDP in Δt. It can be seen from (1) that the queue length increases monotonously with the arriving traffic rate $\widehat{A}(t)$. Considering the congestion is onset, let $\widehat{A}(t) = \rho C$ ($\rho > 1$), and ρ is the load level of the bottleneck link. Then (1) can be rewritten as

$$Q(t+\Delta t) = [\rho(1-P_a) - 1]C\Delta t + Q(t). \quad (2)$$

We can see from (2), queue length dynamics is a function of load level and current queue length. In steady state, $Q(t+\Delta t)$ should equal to $Q(t)$ approximately. In addition, if we want the queue length is independent of traffic and network conditions (ρ and $C\Delta t$ in (2)), we must have $\rho(1-P_a) - 1 = 0$, i.e.,

$$P_a = 1 - 1/\rho. \quad (3)$$

According to the above analysis, we argue that to obtain stable queue length independent of the traffic dynamics and network conditions, the PDP should not be a constant but a function of load level at the bottleneck link. Further, a load adaptive PDP as in (3) can reduce unnecessary packet loss for underutilization link and result in higher throughput.

3.3 MSC Algorithm

Packet drop according (3) means that the mean rate of packet admitted by the link is never greater than the link capacity and no queue will be formed. Since traffic is also bursty at packet level or in the time scales less than RTT, a small queue is helpful to absorb such burtiness and improve the throughput.

For stable queue length, only one reference queue length should be defined in advance[11], and the mismatch between the instantaneous queue length and reference queue length is a good measure of the packet level burstiness. Denote the instantaneous queue length as $Q(t)$, and reference queue length as $Q_{\rm ref}$. The queue length mismatch can be defined as $\Delta_q = Q(t) - Q_{ref}$. Then the PDP of MSC is $P_d = \min(P_a + \alpha_1.\Delta_q, 1)$, where P_a is adapted to the traffic rate in large time scale greater than RTT as defined in (3) and $\alpha_1 = \alpha/Q_{ref}$ ($0 < \alpha \leqslant 1$, which is a weighting factor determining the significance of queue based modulator).

Recall that (3) is only valid for $\rho > 1$. When the link load level is less than 1, i.e., $0 \leqslant \rho \leqslant 1$, congestion will not occur in large time scale and we set P_a to be a small constant (0.01 as default) or zero. In contrary to ARED, P_a is adaptive to the load level other than average queue length at the congested links. Another significant difference between MSC and RED/ARED is that the sampled instantaneous queue length is used instead of average queue length to capture the traffic burstiness in small time scales.

Assume the traffic rate can be estimated as in [18] but with samples in a large time scale (Nyquist sampling law should be followed in choosing the sampling time scales) and the pseudo code of MSC algorithm is presented as follows.

MSC Algorithm

```
Variables:  rho, Q, p_a, p_d,
  smpl_bytes, lastset_q, last_set_r
Parameters:  qref, alpha, ts_s, ts_l, r_w
Initialization:
  rho = 1.0, smpl_bytes = 0;
  Q = 0, p_a = 0.01, alpha = 1.0;
For each packet arrival:
  smpl_bytes += arv_bytes;
  if (arv_time − lastset_r ⩾ ts_l){
    //estim.  the arriving traffic rate A(t);
    r = smpl_bytes /(arv_time − lastset_r);
  A(t) = r ∗ r_w + A(t) ∗ (1 − r_w);
  rho = A(t)/C;
  if (0 < rho ⩽ 1.0 )
    p_a = 0.01;
  else
    p_a = 1 − 1/rho;
  smpl_bytes = 0;
  lastset_r = arv_time;
}
if (arv_time − lastset_q ⩾ ts_s){
  lastset_q = arv_time;
p_q = alpha∗(Q-qref)/qref;
//Calculate PDP with p_q:
p_d = min(p_a + p_q, 1);
}
Operate on the incoming packet with the PDP;
```

3.4 Complexity Analysis

AQM schemes are implemented in the routers and they should be as simple as possible for scaling well in high-speed networks. The computation complexity of MSC comes mainly from the traffic rate estimation and the calculation of PDP. While the introduction of traffic rate prediction increases the complexity of MSC, queue length smoothing is removed from it. Hence, MSC has the equivalent complexity as RED in the worst case, i.e., estimating incoming traffic rate on every packet arrival with exponentially weighted smoothing. In fact, MSC does not need update PDP on every packet arrival, but at the frequency of queue length sampling in small time scales. Assuming the sampling interval is 10ms, compared with RED/ARED, the computation of PDP is reduced several to tens times in terms of link capacity. Since the rate-based update of $P_a = 1 - 1/\rho$ is much less frequent than queue based one, the involved computation is very limited. In contrary, PAQM has to perform at least two steps of rate prediction with LMMSE predictors and accordingly consumes more computation even with the same sampling interval as MSC.

4 Fluid Modeling and Stability Analysis of MSC

Intuitively, since control in large time scale is dominant in the control system, this system will work stably. In this section, we first present a fluid flow model of the previous combined TCP and AQM congestion control system, based on the leading work in [20, 21]. Then the stability of the system is analyzed to validate the design rationality, which also works as criteria for setting the system parameters, e.g., the time scale of arrival traffic rate estimation.

A fluid-flow model of TCP behavior was developed in [20] based on stochastic differential analysis. Simulation result demonstrated that this model accurately capture the dynamics of TCP. A simplified version of the model ignoring the TCP time out mechanism can be described by the following coupled, non-linear equations:

$$\dot{W}(t) = \frac{1}{R(t)} - \frac{W(t)W(t-R(t))}{2R(t-R(t))}p(t-R(t))$$

$$\dot{Q}(t) = \frac{W(t)}{R(t)}N(t) - C$$

where W is the expected TCP window size (packets), Q is the expected queue length (packets), R is the round trip time (s), C is the capacity of congested link, N is the number of active TCP sessions and p is the probability of packet mark/drop. Let T_p denote the propagation delay (s), then $R(t) = Q(t)/C + T_p$.

The model was linearized in [21] using small signal method as follows:

$$\delta\dot{W}(t) = -\frac{2N}{R_0^2 C}\delta W(t) - \frac{R_0 C^2}{2N^2}\delta p(t-R_0)$$

$$\delta\dot{Q}(t) = \frac{N}{R_0}\delta W(t) - \frac{1}{R_0}\delta Q(t)$$

where (W_0, Q_0, p_0) is the operating point of the system and accordingly $R_0 = Q_0/C + T_p$. At the operating point we have $A(t) = NW(t)/R_0$, and the above model can be modified as

$$\delta\dot{A}(t) = -\frac{2N^2}{R_0^3 C}\delta A(t) - \frac{C^2}{2N}\delta p(t-R_0), \tag{4.1}$$

$$\delta\dot{Q}(t) = \delta A(t) - \frac{1}{R_0}\delta Q(t). \tag{4.2}$$

According to the previous analysis, PDP of MSC is the function of both estimated traffic rate and instantaneous queue length,

$$p(t) = \frac{\widehat{A}(t) - C}{\widehat{A}(t)} + \alpha\frac{Q(t) - Q_{\rm ref}}{Q_{\rm ref}}. \tag{5}$$

At equilibrium state ($\widehat{A}(t) \approx C$), (5) can be described as an increment derivative form as follows:

$$\delta\dot{p}(t) = \frac{1}{C}\delta\dot{\widehat{A}}(t) + \frac{\alpha}{Q_0}\delta\dot{Q}(t). \tag{6}$$

When the traffic arriving rate is estimated with exponentially weighted moving average as in [18], we have the following equation describing the behavior of $\widehat{A}(t)$[20],

$$\dot{\widehat{A}}(t) = K\widehat{A}(t) - KA(t)$$

and K is determined by weight and sampling frequency.

Taking Laplace-transform on (4) and (6), the block-diagram representation of the linearized control system is given in Fig.1.

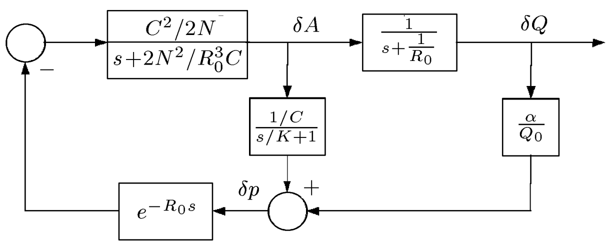

Fig.1. Block-diagram of TCP/MSC congestion control system.

Sequentially, the open-loop transfer function of the feedback control system is

$$G(s) = \frac{C^2/2N}{s + 2N^2/R_0^3 C}\left(\frac{1/C}{s/K+1} + \frac{\alpha/Q_0}{s+1/R_0}\right)e^{-R_0 S}. \tag{7}$$

Since queue variation is trivial near operating point at equilibrium state, the $\delta\dot{Q}(t)$ item in (6) can be ignored and (7) can be modified as

$$G(s) \approx \frac{C^2/2N}{s + 2N^2/R_0^3 C}\frac{1/C}{s/K+1}e^{-R_0 S}. \tag{8}$$

In frequency domain, this approximation can be interpreted as that $\frac{\alpha/Q_0}{s+1/R_0}$ corresponds to high-frequency residual and has insignificant impact on system stability[22]. Obviously, $-2N^2/R_0^3C$ is a negative eigenvalue of the linearized system.

Theorem 1. *Let* K *satisfy* $\frac{(R^+)^3C^2}{4(N^-)^3} \leqslant \sqrt{\frac{\omega_g^2}{K^2}+1}$, *where* $\omega_g = \beta \min\left\{\frac{2(N^-)^2}{(R^+)^3C}, \frac{1}{R_0}\right\}$ *and* $\beta \ll 1$.

Then the feedback control system in Fig.1 is stable for all $N \geqslant N^-$ and $R \leqslant R^+$.

Proof. Based on Nyquist Stability criteria, the proof is similar with the proof of Proposition 1 in [21]. □

5 Performance Evaluation of MSC

We extend the NS2 simulator[23] with our MSC algorithm to check its performance. MSC is compared with other AQM schemes, e.g., REM, PI, ARED and AVQ, in the same conditions to show its superiority. Also, the simulation result will verify the analysis in Sections 3 and 4.

Two types of network topologies are employed in our simulations. The first is the commonly used dumb-bell: there is a single bottleneck link from router 1 to router 2 with capacity 15Mbps and delay 20ms. Many source and destination pairs are connected to the router 1 and router 2, and the capacity of access link is 100Mbps. And another topology for simulation has multiple bottleneck links and the set of link bandwidth and delay

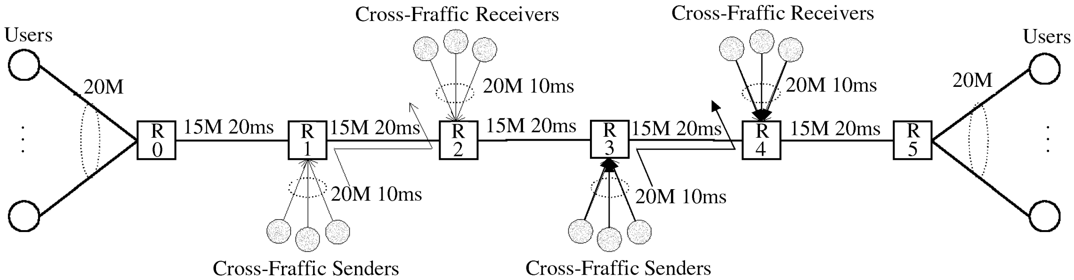

Fig.2 Multiple bottleneck network topology for simulation.

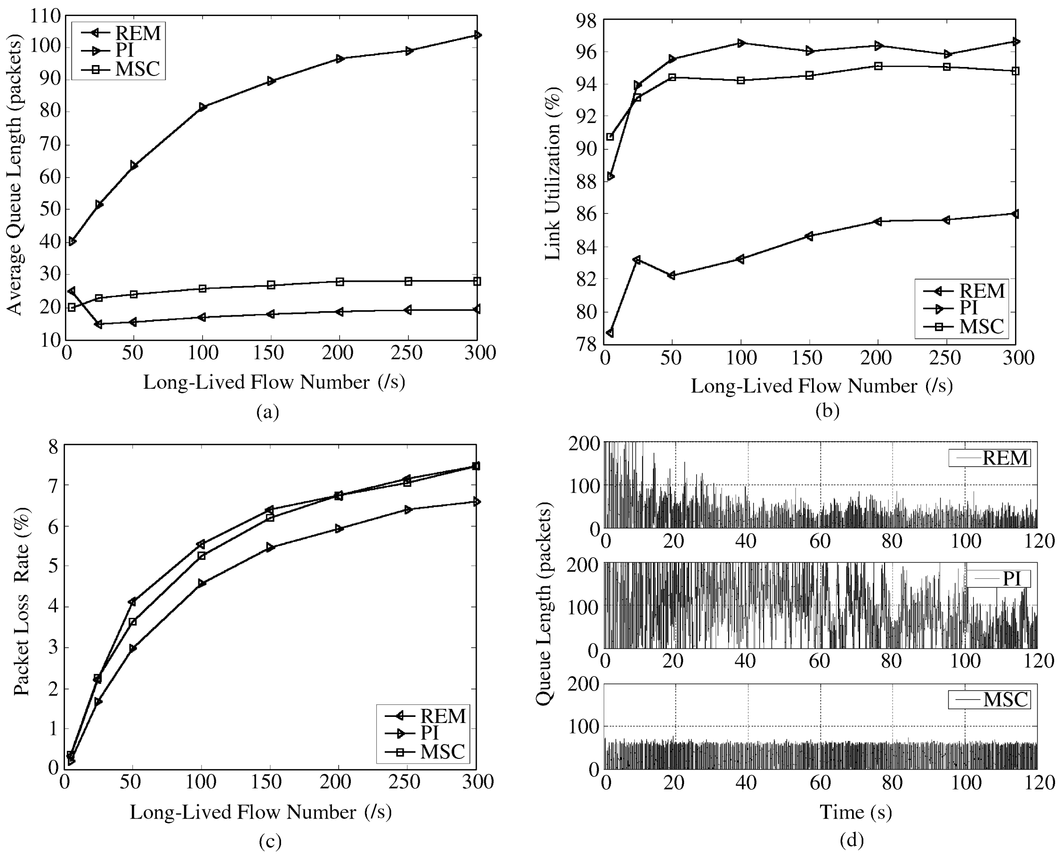

Fig.3. Simulation result of Experiment 1. (a) Average queue length vs. the number of long-lived flows. (b) Link utilization vs. the number of long-lived flows. (c) Packet loss rate (PLR) vs. the number of long-lived flows. (d) Queue length evolution (150 long-lived TCP flows).

is as shown in Fig.2. Different AQM schemes are configured in the bottleneck links respectively and Tail Drop is used for other links. The buffer limit of bottleneck link is mostly 200 packets with size of 500 bytes. While Experiment 1 to Experiment 4 is conducted with dumbbell topology, the Experiment 5 is performed with the multiple bottleneck topology.

To construct a more realistic experimental environment, *reverse-path traffic* and *heterogeneous RTT* are configured in all the simulations[3]. The reverse path traffic is lighter than the forward path and the delay of these links is variable from milliseconds to hundreds of milliseconds. We mainly focus on performance measures such as link utilization (the ratio of *output* traffic rate to the service rate of a link), average packet delay and packet loss rate (PLR) of the congested links.

Experiment 1. Performance with Long-Lived TCP Flows.

Since TCP traffic is dominant in the Internet, in most situations, the load level of a link is related to the number of TCP flows sharing the link. In this experiment, we assume that the bottleneck link of the dumbbell topology is shared only by long-lived TCP flows, such as those of FTP applications. The reference queue ($Q_{\rm ref}$) of REM, PI and MSC is 40 packets corresponding to about 10ms delay. The other parameters of REM and PI are set to the default values in NS2.

The number of long-lived TCP flows sharing the bottleneck link is increased from 5 to 300. The simulation result is shown in Fig.3. It is observed from the figure: 1) MSC keeps average queue length low and stable when the TCP number increases; 2) although REM keeps average queue length slightly lower than MSC, it suffers rather lower link utilization; 3) while keeping low queue length and high link utilization, the MSC algorithm has lower PLR than REM and slightly higher PLR than PI. Fig.3(d) shows the queue length evolution of different AQM schemes when the number of long-lived flow is 150. Obviously, MSC converges to low queue length very fast and keeps stable all the time.

From the experimental results, it is interesting to find the tradeoff among performance measures. Generally, increasing in queue length comes with higher link utilization and lower PLR (this explains why the link utilization of PI is higher than MSC). We believe that the *better* performance tradeoff of MSC mainly benefits from the rate based congestion detection and packet drop.

Experiment 2. Robustness to Short-Lived TCP Flows (Intensity).

Short-lived TCP flows affect the performance of AQM schemes significantly due to its uncontrollability. In this experiment we check the robustness of MSC to short-lived flows. The number of long-lived flows is fixed to 150, and the short-lived flows simulating the Web-like flows increase from 0 to 50 per second with step 5. Each

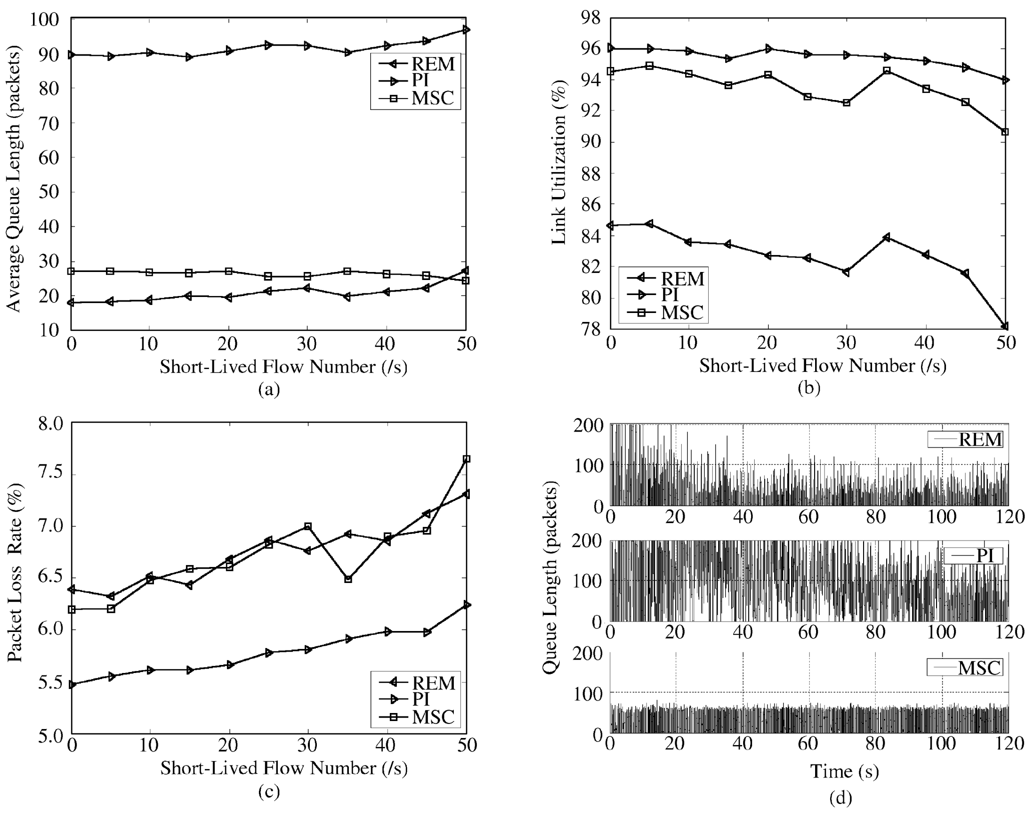

Fig.4. Simulation results of Experiment 2. (a) Average queue length vs. the number of short-lived flows. (b) Link utilization vs. the number of short-lived flows. (c) Packet loss rate (PLR) vs. the number of short-lived flows. (d) Queue length evolution (150 long-lived TCP flows and 25 short-lived TCP flows).

short-lived flow sends only ten packets. It has been found that traffic self-similarity is caused by aggregated Web flows[24]. The mixture of long-lived flows and short-lived flows makes the traffic very bursty and more realistic. The results of this experiment are pre- sented in Fig.4. We find that the PLR increases and the link utilization decreases for all the algorithms with the increase of the number of short-lived flows. However, it shows in Figs.4(a) and 4(d) that MSC keeps the queue length stabler than the other algorithms, which demonstrates that the rate-based scheme is robust to the unresponsive short-lived flows.

Experiment 3. Responsiveness in Dynamic Cross Traffic Situation.

The objective of this experiment is to examine response of MSC to abruptly changing traffic conditions. Instead of adding and reducing TCP flows, dynamic cross traffic is introduced to simulate the extreme burstiness of Internet traffic. The cross traffic is constructed by aggregated exponential On-Off traffic with average rate 5000Kbps and average interval 0.5s of On and Off duration. The cross traffic is added at 10s and stopped at 45s and added again at 75s and stopped until the simulation is over. In addition to the unresponsive cross traffic, there are 150 FTP flows and some Web flows in the bottleneck link and the load is quite heavy. ARED is also compared in this experiment and the minimum and maximum queue length thresholds of ARED are $Q_{ref}/2$ and $3Q_{ref}/2$ packets respectively. To show clearly the high variance of queue length evolution, the buffer limit was set 400 packets in this experiment.

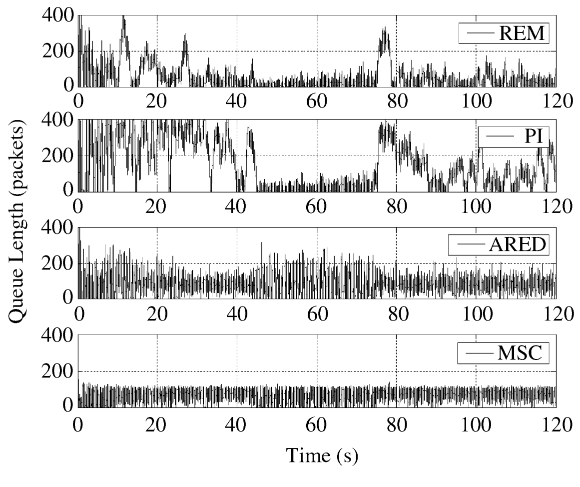

Fig.5. Queue length evolution (150 long-lived TCP flows and 20 short-lived TCP flows).

Since the queue length evolves with the change of load level, good responsiveness can be expected to MSC with rate-based congestion detection. As shown in Fig.5, MSC manages to keep the queue length stable timely with the emerging and disappearing of cross traffic while the queue occupancy of the other algorithms presents significant fluctuations before entering into new steady state after quite a long time.

Experiment 4. Comparison of MSC with Adaptive Virtue Queueing (AVQ). AVQ is a well-know rate-based active queue management algorithm, which is considered to be effective for low delay and high link utilization. However, AVQ is rarely examined in two-way traffic condition in the literature. We try to compare AVQ with MSC in two-way traffic condition in this experiment. To make a fair comparison, we configure in the experiment the reference queue ($Q_{\rm ref}$) of MSC to a small queue (20 packets), which is the same as the queue limit of AVQ. AVQ is configured with and without ECN respectively. The result is presented in Fig.6. From the figure, we observe that AVQ can obtain small queue but

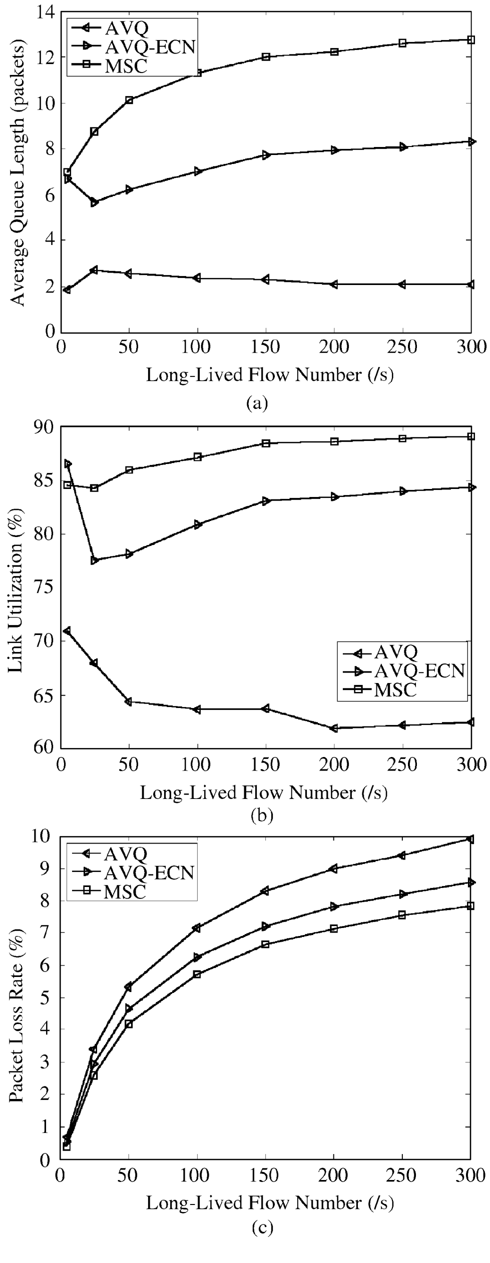

Fig.6. Simulation results of Experiment 4. (a) Average queue length vs. the number of long-lived flows. (b) Link utilization vs. the number of long-lived flows. (c) Packet loss rate (PLR) vs. the number of long-lived flows.

with the cost of rather lower link utilization, especially without ECN and much higher PLR in two-way traffic condition than MSC. This experimental result also shows that the queue-based modulation is indispensable to reduce packet loss and increase link utilization.

Experiment 5. Performance of MSC in Multiple Bottleneck Network Topologies.

In a multiple bottleneck network topology, the offered load comes from multiple sources and the aggregated traffic dynamics is more complex than in single bottleneck topologies. Further, the packet may be lost in multiple congestion points. There are two bottlenecks in the network topology in Fig.2, which are the link Q12 between router R1 and R2 and the link Q34 connecting router R3 and R4. As the traffic from the sources passes Q12 first in the forward path and it will be constrained to some degree, the load at Q34 is lighter than that at Q12 with the same cross traffic configuration. This inference was confirmed by experiments. We examine average queue length, link utilization and packet loss rate of Q12 as Experiment 1. We find that the results are very similar to Experiment 1 and we only present the queue length dynamics in Fig.7, which demonstrates the robustness to multiple bottlenecks of MSC. We believe that if the transient congestion at Q34 is heavy, it may degrade the performance of any type of AQM schemes at Q12 due to larger RTT and more packet loss.

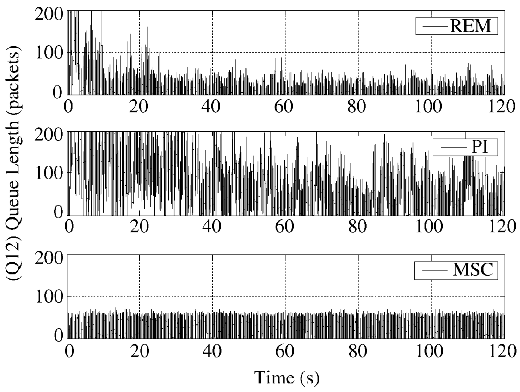

Fig.7. Queue length evolution of Q12 in the multiple bottleneck topology.

6 Conclusion

Many AQM schemes have been proposed since the proposal of RED in 1993, but they are only efficient in some scenarios in terms of specific performance measures. Based on the understanding of traffic bursty behavior in multiple time scale, we designed and analyzed in this paper a multiscale active queue management scheme named multiscale controller (MSC) to make better performance tradeoff. The packet drop probability (PDP) of MSC is mainly determined by the burst-level rate of the traffic arriving at the congested link; in addition, this probability is modulated by the packet-level fluctuations of the incoming traffic. While the rate-based packet dropping at burst level determines the packet drop aggressiveness and is responsible for low and stable queuing delay, good robustness and responsiveness, the queue-based modulation of the PDP at packet level is helpful for low packet loss and high link utilization. A fluid flow model was constructed for MSC and stability analysis was performed based on the model. Simulation results showed that MSC outperformed many of the current AQM schemes, including REM, PI, ARED and AVQ, in almost all the desired performance. A more accurate critical time scale rather than typical RTT for traffic rate prediction is under investigation now.

As a final comment, we argue that the objective of AQM, low delay, low loss and high link utilization[2], can only be achieved with control mechanism making the aggregated TCP traffic smooth.

Acknowledgments The authors would like to thank Professor Guan-Qun Gu for his supervision and Professor Jun Shen for his comments on an early draft of this paper.

References

[1] Jacobson V. Congestion avoidance and control. In *Proc. ACM SIGCOMM'88*, Stanford, USA, Sept. 1988, pp.314–329.

[2] Braden B, Clark D, Crowcroft J *et al.* Recommendations on queue management and congestion avoidance in the Internet. IETF RFC 2309, April 1998.

[3] Floyd S, Kohler E. Internet research needs better models. *ACM Comput. Commun. Rev.*, Jan. 2003, 33(1): 29–34.

[4] Hui J Y. Resource allocation for broadband networks. *IEEE J. Select. Areas in Commun.*, Dec. 1988, 6(9): 1598–1608.

[5] Riedi R H, Willinger W. Toward an Improved Understanding of Network Traffic Dynamics. Self-Similar Network Traffic and Performance Evaluation, New York: John Wiley & Sons, 2000.

[6] Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans. Net.*, Aug. 1993, 1(4): 397–413.

[7] Floyd S, Gummadi R, Shenker S. Adaptive RED: An algorithm for increasing the robustness of RED's active queue management. Aug. 2001, http://www.icir.org/floyd/papers/adaptiveRed.pdf.

[8] Feng W C, Kandlur D, Saha D *et al.* Blue: A new class of active queue management algorithms. Technical Report, CSE-TR-387-99, UM, 1999.

[9] Kunniyur S, Srikant R. Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management. In *Proc. ACM SIGCOMM'01*, San Diego, CA, Aug. 2001, pp.123–134.

[10] Floyd S. TCP and explicit congestion notification. *ACM Comput. Commun. Rev.*, Oct. 1994, 24: 10–23.

[11] Athuraliya S, Low S, Li V *et al.* REM: Active queue management. *IEEE Network Magazine*, May 2001, 15: 48–53.

[12] Hollot C V, Misra V, Towsley D *et al.* On designing improved controllers for AQM routers supporting TCP flows. In *Proc. IEEE INFOCOM'01*, Anchorage, 2001, 3: 1726–1734.

[13] Kwon M, Fahmy S. A comparison of load-based and queue-based active queue management algorithms. In *Proc. SPIE ITCom*, Aug. 2002, 4866: 35–46.

[14] Gao Y, He G, Hou J. On exploiting traffic predictability in active queue management. In *Proc. IEEE INFOCOM'02*, New York, USA, June 2002, pp.1360–1369.

[15] Gao Y, Hou J C. A state feedback control approach to stabilizing queues for ECN-enabled TCP flows. In *Proc. IEEE INFOCOM*, San Francisco, CA, 2003, 3: 2301–2311.

[16] Ji Q, Dong Y. Performance evaluation of active queue management schemes: A comparative study. *Journal of Communications*, Nov. 2005, 26(11): 100–105, 111. (In Chinese)

[17] Sang A, San-qi Li. A predictability analysis of network traffic. In *Proc. INFOCOM 2000*, Tel Aviv, Israel, March 26–30, 2000, 1: 342–351.

[18] Stoica I, Shenker S, Zhang H. Core-stateless fair queuing: Achieving approximately fair bandwidth allocations in high speed networks. In *Proc. ACM SIGCOMM'98*, Vancouver, Sept. 1998, pp.118–130.

[19] Tuan T, Park K. Multiple time scale congestion control for self-similar network traffic. *Perf. Eval.*, 1999, 36: 359–386.

[20] Misra V, Gong W B, Towsley D. Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED. In *Proc. ACM/SIGCOMM*, Stockholm, Sweden, 2000, pp.151–160.

[21] Hollot C, Misra V, Towsley D *et al.* A control theoretic analysis of RED. In *Proc. IEEE INFOCOM*, Anchorage, 2001, 3: 1510–1519.

[22] Hollot C, Misra V, Towsley D *et al.* Analysis and design of controllers for AQM routers supporting TCP flows. *IEEE Trans. Automatic Control*, Jun. 2002, 47: 945–959.

[23] NS-2 Network Simulator. http://www.isi.edu/nsnam/ns/.

[24] Crovella E, Bestavros A. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Trans. Networking*, Dec. 1997, 5(6): 836–846.

Qi-Jin Ji received his M.E. degree from Nanjing University of Posts and Tele-comunictions, China, in 2002 and Ph.D. degree in the Department of Computer Science and Engineering, Southeast University, Nanjing, China in 2005 respectively. His current research interest is in the areas of Internet congestion control and multimedia networking,

Yong-Qiang Dong received his B.E. and M.E. degrees from Nanjing University of Science and Technology in 1994 and 1997 respectively. Now he is a Ph.D. candidate in Southeast University, China. His research interests include network QoS, congestion control and multimedia communications.