Impact of AQM scheme on WLANs using queue thresholds

Lin Guan
Department of Computer Science,
Loughborough University, LE11 3TU, UK
E-mail: L.Guan@lboro.ac.uk

Irfan U. Awan* and Mike E. Woodward
Department of Computing,
University of Bradford, Bradford, BD7 1DP, UK
E-mail: I.U.Awan@Bradford.ac.uk
E-mail: M.E.Woodward@Bradford.ac.uk
*Corresponding author

Xingang Wang
School of Computing,
Communication and Electronics,
University of Plymouth,
PL4 8AA Plymouth, UK
E-mail: xingang.wang@plymouth.ac.uk

Abstract: Explosive growth of applications for wireless networks and increasing number of users will cause significant increase in the traffic over radio channels. Consequently, uplink and downlink channels may suffer serious traffic congestion problems in the absence of effective congestion control schemes. One approach to enhance the performance of wireless services is to employ an AQM scheme for buffers at the downlink of wireless networks. This paper presents an analytical framework for the performance evaluation of AQM scheme based congestion control mechanism in WLANs. It also carries out a comprehensive performance comparison for various dropping strategies to see the impact of AQM scheme on WLANs using queue thresholds.

Keywords: wireless networks; QoS; quality of service; Markov chain; congestion control; queue thresholds.


Biographical notes: Lin Guan received her PhD in Computer Science from University of Bradford, UK, in 2005, where she was also a Post-doc. Research Assistant for one year. She is currently a Lecturer at the Department of Computer Science, Loughborough University, UK. Her research interests include performance modelling/evaluation of computer networks, congestion control mechanisms with QoS constraints, mobile computing and wireless networks.

Irfan U. Awan is currently a Senior Lecturer at the Department of Computing, University of Bradford, UK. He received his BSc from Gomal University, Pakistan (1986), MSc (Computer Science) from Qauid-e-Azam University, Pakistan (1990). He received his PhD from the Department of Computing, University of Bradford, UK (1997). During his PhD studies, he developed cost effective approximate analytical tools for the performance evaluation of complex queuing networks. His current research mainly focused on service and space priorities in order to provide quality of service and to control the congestion in wired and wireless networks.

Mike E. Woodward graduated with a first class honours Degree in Electronic and Electrical Engineering from the University of Nottingham in 1967 and received a PhD Degree from the same institution in 1971. He is currently the Head of the Department of Computing at the University of Bradford. His current research interests include queuing networks, telecommunications traffic modelling, quality of service routing and mobile communications systems and he is the author of two books and over 100 research papers on the above and related topics.
1 Introduction

Wireless Local Area Networks (WLANs) based on the IEEE 802.11 standard have recently been widely deployed to provide wireless data access to mobile host (e.g., PDA, Laptops, etc) in the campuses, hospitals, enterprises, etc. WLANs are usually interconnected using wired backbone networks, and many applications rely on Transmission Control Protocol (TCP) for the end-to-end transport. Because AQM techniques have been successfully applied in network routers or switches to detect congestion in the wired network, an interesting issue is whether AQM can be feasibly implemented in the wireless environment. So it is very interesting and important to study the effect of implementing an appropriate AQM scheme to address the congestion issue in WLANs.

So far, there are only few researchers to investigate the AQM issues in WLANs. Pilosof et al. (2003) identified four different regions of TCP unfairness which depend on the buffer available at the base station. They have also pointed out that the buffer size at the base station plays an important role in allocating the bandwidth among uplink and downlink traffic in WLANs. Xu et al. (2002) presented an AQM scheme in WLAN, but the performance evaluation of the proposed scheme was only limited to one performance measure, i.e., the delay from wired network to WLAN. In Pang et al. (2003) mainly focused on the performance comparison study of TCP Veno (Fu and Liew, 2003) and TCP Reno over 802.11b WLAN and RED router. They concluded that TCP Veno is compatible with RED and that RED helps to improve fairness even it does not improve the throughput. Yi et al. (2004) studied the feasibility of an AQM scheme to handle the congestion at the wireless Access Point (AP) and proposed the proxy AQM scheme, called Proxy-RED, that performs the AQM functionality at the gateway on the behalf of wireless APs considering the current architectural trend of a lightweight APs. Simulation results showed that the proposed Proxy-RED scheme can bring overall performance improvement in WLANs (Deng, 2004). Several studied used AQM techniques to regulate the buffer queue and the bandwidth allocation (cf., Pilosof et al., 2003; Xu et al., 2002; Pang et al., 2003; Fu and Liew, 2003; Yi et al., 2004; Deng, 2004). Most researchers have used simulation tools as their choice of modelling to examine the performance of wireless networks. There is a current lack of analytical models for traffic congestion control using thresholds based AQM scheme over WLANs.

AQM schemes, such as RED drops arriving packets probabilistically depending on setting maximum and minimum thresholds in the queue. This paper proposes a new analytical framework to apply an AQM scheme to limit the input rate at the network gateway in WLANs and presents performance comparison study through typical numerical results by using queue thresholds. Three discrete-time performance models with threshold based arrival rates, namely, one-step, two-step and linear reduction have been set up to investigate how queue thresholds affect the whole system performance. The proposed analytical models are based on AQM principles adapted in WLANs, which take into account the reduction of incoming traffic

Xingang Wang graduated from Heilongjiang University with a first class degree in 2001 and received a PhD from University of Bradford in 2005. He is currently a Lecturer in School of Computing, Communication and Electronics, University of Plymouth. His research interests include performance modelling of computer networks, multiple access protocols for wireless networks and mobile networks integration.
arrival rate due to packets dropped probabilistically and the drop probability increases linearly with the increasing system contents. For an independent Bernoulli stream, the objective functions have been found for the threshold with parameters delay, throughput and loss probability. The typical numerical results clearly demonstrate how different load settings based on queue thresholds can provide different tradeoffs between throughput, loss probability and delay to suit different service requirements. We also point out the performance advantages of using two thresholds compared to the scheme which uses only a single threshold have been clearly demonstrated that the two thresholds can always be adjusted to give a lower delay for the same throughput.

The remainder of the paper is organised as follows: Section 2 gives a brief overview for network topology which involve an AQM scheme in WLANs. In Section 3, we introduce three discrete-time performance models for congestion control mechanism using queue thresholds, and present analytical expressions for various performance measures. Section 4 presents the performance comparison results of three models through typical numerical results. Conclusions are followed in Section 5.

2 Network topology: an AQM scheme in WLANs

WLANs provide wireless network communication over short distances using radio or infrared signals instead of traditional network cabling. WLANs are built by attaching a device called the AP to the edge of the wired network. In WLANs infrastructure, multiple AP allow users to efficiently share local network resources linking the WLAN to wired and long distance Wide Area Networks (WANs). They are usually interconnected using wired backbone networks, and many applications on the networks run on top of the Transmission Control Protocol/Internet Protocol (TCP/IP) (Biaz and Wang, 2004) (cf. Figure 1).

![Figure 1](image1.png)

All mobile devices are communicating with a server in a fixed network connected to the Internet. Access point (AP – also called ‘base station’) relays the frames between the stations and wired backbone and is responsible for scheduling the access to the media. The gateway router is a network element which provides the bulk of networking functionality and also support features such as QoS, security, management and etc. (Pang et al., 2003; Yi et al., 2004; Deng, 2004).

Due to the speed disparity between wired and wireless networks, the congestion of network traffic could easily occur at the joint point (e.g., router gateway) which interconnects these two networks. In this paper, an analytical framework for congestion control mechanism is proposed to evaluate the performance of thresholds based AQM scheme over gateway or AP from the wired backbone network (e.g., internet) to WLAN (downlink stream). A high level architecture (Deng, 2004) is abstracted as shown in Figure 2, in which AQM principles are adapted by using queue thresholds.

![Figure 2](image2.png)

The thresholds are used to control the arrival rate of the external traffic. We take into account the reduction of incoming traffic arrival rate due to packets dropped probabilistically with the drop probability increasing linearly with the increasing system contents. The initial analysis is based on one threshold with the arrival rate having a one-step reduction from $a_1$ to $a_2$ once the queue length reaches the threshold level. The second part of our analysis is based on two thresholds where the arrival rate reduces from $a_1$ to $a_2$ (step-1) once the queue size reaches the first threshold and then from $a_2$ to $a_3$ (step-2) after the second threshold. All the new packets, after the first threshold, arrive at the reduced rate $a_2$ until either the queue reaches the second threshold and then the arrival rate reduces to $a_3$ or the queue size becomes smaller than the first threshold value and then resumes the original arrival rate $a_1$. All new arriving packets are dropped once the queue becomes full. The third part of the analysis is concerned with performance evaluation when the arrival rate is subjected to a linear reduction between two thresholds, and further performance comparison between one-step, two-step and linear reduction models.

3 Performance analysis of analytical models

In this section, we introduce three proposed theoretical models in discrete-time settings and present the analytical framework to be used in the remainder of the paper. In three discrete-time queuing systems, we will assume that a departure always takes place before an arrival in any unit time (slot). Arrivals form an independent Bernoulli process, with $a_n \in \{0, 1\}$, $n = 1, 2, 3, \ldots$, and there is a finite waiting room of $M$ packets, including any in service. The queuing discipline is first-come first-served (Woodward, 1993).
3.1 Model-I: one threshold with arrival rate one-step reduction

Model-I incorporates one threshold to make the arrival process step reduce from arrival rate $\alpha_1$ to $\alpha_2$ once the number of packets in the system has been reached to the threshold value $L_1$. The source operates normally, otherwise. This may thus be considered equivalent to a feedback from the queue to the arrival process to tell the arrival process to change its rate. Alternatively we can view the reduction of incoming traffic arrival rate to be due to packets dropped probabilistically, hence, in addition, the model can alternatively be viewed as the source continuing to send at rate $\alpha_1$ but with arriving packets dropped with probability $1-\alpha_2/\alpha_1$ (cf., Figure 3).

Let the probability of an arrival in a slot be $\alpha_1$ before the number of packets in the system reaches the threshold $L_1$, the probability of an arrival in a slot be reduced to $\alpha_2$ after $L_1$ and the probability of a departure in a slot be $\beta$. We assume that the queuing system is in equilibrium. The state transition diagram is shown in Figure 4, and the queue length process is a Markov chain with a finite state space $\{0, 1, \ldots, M(M = L_1 + J)\}$.

**Figure 3** One-step reduction model with one threshold ($L_1$)

![State transition diagram for one-step reduction model with single threshold ($L_1$)](image)

We assume that $\alpha_1 \neq \beta$, $\alpha_2 \neq \beta$ ($\alpha_1 > \alpha_2$) and the balance equations of the discrete-time finite queue with one threshold ($L_1$) can be expressed as follows:

$$\pi_0 = \pi_0(1-\alpha_1) + \pi_1 [\beta(1-\alpha_1)]$$

(1)

$$\pi_1 = \pi_0 \alpha_1 + \pi_1 [\alpha_1 \beta + (1-\alpha_1)(1-\beta)] + \pi_2 [\beta(1-\alpha_1)].$$

(2)

In general

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_i \beta + (1-\alpha_i)(1-\beta)] + \pi_{i-1} [\beta(1-\alpha_i)] \quad i = 2, 4, \ldots, L_1-2$$

(3)

$$\pi_{i-1} = \pi_{i-2} [\alpha_1 (1-\beta)] + \pi_{i-1} [\alpha_i \beta + (1-\alpha_i)(1-\beta)] + \pi_i [\beta(1-\alpha_i)].$$

(4)

In general

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_2 \beta + (1-\alpha_2)(1-\beta)] + \pi_{i-1} [\beta(1-\alpha_2)] \quad i = L_1, L_1+1, \ldots, L_1+J-1$$

(6)

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_j \beta + (1-\alpha_j)(1-\beta)] + \pi_{i+1} [\beta(1-\alpha_j)] \quad i = L_1+J, M = L_1+J.$$ 

(7)

Solving these equations recursively, and involving after $L_1$ and the probability of a departure in a slot be $\beta$. We assume that the queuing system is in equilibrium. The state transition diagram is shown in Figure 4, and the queue length process is a Markov chain with a finite state space $\{0, 1, \ldots, M(M = L_1 + J)\}$.

**Figure 4** State transition diagram for one-step reduction model with single threshold ($L_1$) (see online version for colours)

We assume that $\alpha_1 \neq \beta$, $\alpha_2 \neq \beta$ ($\alpha_1 > \alpha_2$) and the balance equations of the discrete-time finite queue with one threshold ($L_1$) can be expressed as follows:

$$\pi_0 = \pi_0(1-\alpha_1) + \pi_1 [\beta(1-\alpha_1)]$$

(1)

$$\pi_1 = \pi_0 \alpha_1 + \pi_1 [\alpha_1 \beta + (1-\alpha_1)(1-\beta)] + \pi_2 [\beta(1-\alpha_1)].$$

(2)

In general

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_i \beta + (1-\alpha_i)(1-\beta)] + \pi_{i-1} [\beta(1-\alpha_i)] \quad i = 2, 4, \ldots, L_1-2$$

(3)

$$\pi_{i-1} = \pi_{i-2} [\alpha_1 (1-\beta)] + \pi_{i-1} [\alpha_i \beta + (1-\alpha_i)(1-\beta)] + \pi_i [\beta(1-\alpha_i)].$$

(4)

In general

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_2 \beta + (1-\alpha_2)(1-\beta)] + \pi_{i-1} [\beta(1-\alpha_2)] \quad i = L_1, L_1+1, \ldots, L_1+J-1$$

(6)

$$\pi_i = \pi_{i+1} [\alpha_1 (1-\beta)] + \pi_i [\alpha_j \beta + (1-\alpha_j)(1-\beta)] + \pi_{i+1} [\beta(1-\alpha_j)] \quad i = L_1+J, M = L_1+J.$$ 

(7)

Solving these equations recursively, and involving
The mean throughput of this finite queue given by the fraction of time the server is busy:
\[ S = (1 - \pi_o) \times \beta. \]  
(11)

The mean delay can be obtained from Little’s law for this finite capacity queue as:
\[ W = \frac{P(i)}{S}. \]  
(12)

Another very important performance measure is the probability of packet loss given by:
\[ P_L = \frac{(\alpha_1 - \beta_2) \pi o}{\alpha_1 (1 - \beta)(1 - \gamma_2) (1 - \gamma_2^+)}. \]  
(13)

where \( \pi_o \) is given by the equation (8).

3.2 Model-II: Two thresholds with arrival rate two-step reduction

The second part of our analysis is based on two thresholds where the arrival rate reduces from \( \alpha_1 \) to \( \alpha_2 \) (step-1) once the queue size reaches the first threshold \( L_1 \) and then from \( \alpha_2 \) to \( \alpha_3 \) (step-2) after the second threshold \( L_2 \). All the new packets, after the first threshold, arrive at the reduced rate \( \alpha_2 \) until either the queue reaches the second threshold and then the arrival rate reduces to \( \alpha_3 \) or the queue size becomes smaller than the first threshold value and then resumes the original arrival rate \( \alpha_1 \). All new arriving packets are dropped once the queue becomes full (cf., Figure 5). The operation with the queue length between the two thresholds can be interpreted as either

I. a reduced arrival rate of \( \alpha_2 \) (or \( \alpha_3 \))

II. the original arrival rate of \( \alpha_1 \), but with packets dropped with probability \( 1 - \alpha_2/\alpha_1 \) (or \( 1 - \alpha_3/\alpha_1 \)).

This latter interpretation can be extended to give a linear increase in packet dropping probability between the two thresholds which will be presented in Section 4.

Let the probability of an arrival in a slot be \( \alpha_1 \) before the number of packets in the system reaches the first threshold \( L_1 \); the probability of an arrival in a slot be reduced to \( \alpha_2 \) after the number of packets in the system exceeds the first threshold \( L_1 \) and before reaching the second threshold \( L_2 \); the arrival rate be further reduced to \( \alpha_3 \) at \( L_2 \); and the probability of a departure in a slot be \( \beta \). We also assume that the queueing system is in equilibrium. The state transition diagram is shown in Figure 6, and the queue length process is a Markov chain with a finite state space \{0, 1, ..., \( L_2 + N \) (\( L_2 + N = M \)).

We assume that \( \alpha_i \neq \beta \) (\( \alpha_i > \alpha_{i+1}, \ i = 1, 2, 3 \)), and the final state \( L_2 + N \) (\( L_2 + N = M \)) is the full buffer status. To find the equilibrium probability distribution, first the transition probabilities of arrivals and departures can be defined as:
\[ \lambda_0 = \alpha_1, \lambda_i = \alpha_i (1 - \beta), \mu_i = \beta (1 - \alpha_i), i = 1, 2, 3. \]  
(14)

**Figure 5** Two-step reduction model with two thresholds \( (L_1 \) and \( L_2) \)

![Two-step reduction model with two thresholds](image)

Can be considered as implicit feedback from queue to the arrival process.

Similarly to one-step reduction model described in Section 3, after solving the balance equations of the discrete-time finite queue with two thresholds \( L_1 \) and \( L_2 \) \( (L_2 > L_1) \) recursively, by using the normalising equation \( \sum_{i=0}^{L_2+N} \pi_i = 1, \pi_0 \) can be obtained as follows:
\[ \pi_0 = \left[ \frac{\lambda_0 (1 - \rho_1) + \lambda_1 (\rho_1 - \rho_1^N)}{\lambda_1 (1 - \rho_1)} \mu_1 (1 - \rho_1) + \frac{\lambda_2 (1 - \rho_2)}{\mu_2} \mu_2 (1 - \rho_2) \right]^{-1}. \]  
(15)

Using the generating function of the queue length process for this finite capacity queue,
\[ P(z) = \sum_{i=0}^{L_2+N} \pi_i z^i, \]
and taking the first derivative of \( P(z) \) evaluated at \( z = 1 \), the mean queue length with two thresholds \( L_1 \) and \( L_2 \) can be obtained as follows:
\[ P(i+1) = \pi_0 \left[ \frac{\lambda_0 \rho_1 (1 - \rho_1^N - \rho_1^{N+i} - L_1 (1 - \rho_i))}{\lambda_1 (1 - \rho_1)^2} \right] \]
\[ + \pi_0 \left[ \frac{\lambda_1 \rho_1^{i+1} - \rho_1 N (L_2 - L_1) (1 - \rho_i)}{\lambda_2 (1 - \rho_2)^2} \right]. \]  
(16)
The delay can be obtained using Little’s law for this finite capacity queue as:

\[ W = \frac{\rho_0}{S} \]  

(17)

where \( S \) is the mean throughput of this finite queue given by the fraction of time the server is busy multiplied by the service rate:

\[ S = (1 - \pi_0)\beta. \]  

(18)

Another very important performance measure is the probability of packet loss given by:

\[ P_D = \sum \frac{(\alpha_1 - \alpha_2)(1 - \rho_1^{N-1})}{\alpha\mu(1 - \rho_1)} \]  

(19)

where \( D_i = \frac{\alpha_1}{\alpha^i}, \quad i = 1, 2 \) and \( \pi_0 \) is given by the equation (15).

Figure 6 State transition diagram for two-step reduction model with two thresholds (\( L_1 \) and \( L_2 \)) (see online version for colours)

3.3 Model-III: two thresholds with arrival rate linear reduction

The third part of the analysis is concerned with performance evaluation when the arrival rate is subjected to a linear reduction between two thresholds (cf., Figure 7).

Figure 7 Linear reduction model with two thresholds (\( L_1 \) and \( L_2 \))

Let the probability of an arrival in a slot be \( \alpha_1 \) before the number of packets in the system reaches the first threshold \( L_1 \), the probability of an number of packets in the system reaches the second threshold \( L_2 \), and the probability of a departure in a slot be \( \beta \). When the number of packets in the system is between the first threshold and the second threshold, the arrival rate (probability) will be linearly reduced with some probability which is the function of \( \alpha_1 \), \( \alpha_2 \) and the two thresholds. So the dropping probability increases linearly from 0 to the maximum \( 1 - \alpha_2/\alpha_1 \). This can be considered as implicit feedback from queue to the arrival process in that dropping packets reduces the effective arrival rate into the queue from \( \alpha_1 \) to \( \alpha_1 - \alpha_2 \) with a linear reduction. We assume that the queuing system is in equilibrium. The state transition diagram is shown in Figure 8, and the queue length process is a Markov chain with a finite state space \{0, 1, \ldots, L_2 + N (L_2 + N = M)\}.

As shown in Figure 8, the arrival rate is \( \alpha_1 \) in part I and \( \alpha_2 \) in part III, which are all independent. However in part II (between two thresholds), the arrival rate depends on the state, that means each arrival rate is different with each state and will be linearly reduced by dropping packets. We assume that \( \alpha_1 \neq \beta, \quad \alpha_2 \neq \beta (\alpha_1 > \alpha_2) \) and the final state \( L_2 + N (L_2 + N = M) \) is the full buffer situation.

To find the equilibrium probability, first the transition probabilities of arrivals and departures from state \( L_1 \) to state \( L_2 - 1 \) can be defined as:

\[ \lambda_1 = \alpha_1 (1 - \beta), \quad \mu_1 = \beta (1 - \alpha_1), \quad L_1 \leq k \leq L_2 - 1. \]  

(20)

where

\[ \alpha_1 = \alpha_1 - (k - L_1 + 1) \frac{\alpha_2 - \alpha_1}{L_2 - L_1 + 1}, \quad L_1 \leq k \leq L_2 - 1 \]  

(21)

and the transition probabilities of arrivals and departures in part I and III can also be defined as:

\[ \lambda_2 = \alpha_1 (1 - \beta), \quad \mu_2 = \beta (1 - \alpha_1), \]  

\[ \lambda_3 = \alpha_2 (1 - \beta), \quad \mu_3 = \beta (1 - \alpha_2). \]  

(22)

Similarly with model I and II, after solving the balance equations of the discrete-time finite queue with two thresholds \( L_1 \) and \( L_2 \) (\( L_2 > L_1 \)) recursively, and involving
\[ \rho_1 = \frac{\lambda_1}{\mu_1} \quad \text{and} \quad \rho_2 = \frac{\lambda_2}{\mu_2}, \]

the equilibrium probability \( \pi_i \) can be expressed in terms of \( \pi_0 \), then we use the normalising equations

\[
\sum_{i=0}^{\lambda^+ N} \pi_i = 1,
\]

thus \( \pi_0 \) can be obtained as follows:

\[
\pi_0 = \frac{\lambda_1 (1-\rho_1) + \lambda_2 (\rho_1 - \rho_1^{-1})}{\lambda_1 (1-\rho_1)}
+ \lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_1} + \frac{\lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_2}}{(1-\rho_1)^2}.
\]

Similarly with model I and model II, by using the generating function of the queue length process for this finite queue,

\[
P(z) = \sum_{i=0}^{\lambda^+ N} \pi_i z^i,
\]

and taking the first derivative of \( P(z) \) evaluated at \( z = 1 \), the mean queue length for this finite queue with two thresholds \( L_1 \) and \( L_2 \) can be obtained as follows:

\[
P'(1) = \pi_0 \left[ \frac{\lambda_1 \rho_1 \left( 1-\rho_1^{-1} - \rho_1^{-1} L_1 (1-\rho_1) \right)}{\lambda_1 (1-\rho_1)^2} + \lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_1} + \frac{\lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_2}}{(1-\rho_1)^2}
+ \rho_2 \left( 1-\rho_2^N - N \rho_2^N (1-\rho_2) \right) \right]/(1-\rho_2)^2.
\]

The delay can be obtained using Little's law for this finite capacity queue as:

\[
W = \frac{\pi_0}{S} \left[ \frac{\lambda_1 \rho_1 \left( 1-\rho_1^{-1} - \rho_1^{-1} L_1 (1-\rho_1) \right)}{\lambda_1 (1-\rho_1)^2}
+ \lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_1}
+ \lambda_2 \rho_1^{-1} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_2}
+ \rho_2 \left( 1-\rho_2^N - N \rho_2^N (1-\rho_2) \right) \right]/(1-\rho_2)^2
\]

where \( S \) is the mean throughput of this queue given by the fraction of time the server is busy:

\[
S = (1-\pi_0) \times \beta
\]

\( \pi_0 \) can be found in the equation (23).

Another important performance measure is the probability of packet loss given by the following expression:

\[
P_k = \sum_{i=0}^{\lambda^+ N} \pi_i P_i + \sum_{i=0}^{\lambda^+ N} \pi_i P_{i+1} + \pi_{i+1} \beta P_{i+1} + \pi_{i+1} (1-\beta)
= \frac{1}{\alpha} \lambda_1 \rho_1^{-1} \pi_0 \alpha - \alpha \sum_{i=1}^{\lambda^+ N} \prod_{i=1}^{\lambda^+ N} \left( \frac{\lambda_2}{\mu_i} \right) \frac{1}{\mu_i}
\]

\[
\left( \frac{1-\rho_2^N}{1-\rho_2} \right) \frac{\rho_2}{\alpha_1} + \pi_0 \left( \frac{1-\rho_2^N}{1-\rho_2} \right) \frac{\rho_2}{\alpha_1}.
\]

Figure 8  State transition diagram for linear reduction model with two thresholds (\( L_1 \) and \( L_2 \)) (see online version for colours)
4 Numerical results

This section presents numerical results based on the three proposed models. Performance measurements comparison between Special Case I and II is shown in the first part and among three models are presented in the second part.

4.1 Performance comparison between Special case I and II

This part of presentation outlines two special cases in the one-step and two-step reduction models respectively as described in Sections 3.1 and 3.2, where the source stops sending packets (i.e., $a_2 = 0$ or $a_3 = 0$) instead of a reduced constant arrival rate ($a_2 \neq 0$ or $a_3 \neq 0$) once the number of packets reaches the threshold level (i.e., $L_1$ or $L_2$). We do not consider the loss probability due to the source is blocked after $L_1$ or $L_2$, even though it will be non-zero between the two thresholds. Other performance measures, such as throughput and delay, remain the same as in the case of one-step or two-step reduction model with $a_2 = 0$ or $a_3 = 0$.

The performance advantages of using two thresholds compared to the scheme which uses only a single threshold have been clearly demonstrated for a discrete-time Geo/Geo/1/N+J queue in that the two thresholds can always be adjusted to give a lower delay for the same throughput.

We can compare the results of using one threshold (cf., Section 3.1) with our analysis for two thresholds by setting $L_1 = L_2 > 0$, which represents a single threshold $L_1$ and arrival rate of $a_1$ while the queue is less than or equal to $L_1$ and an arrival rate of zero otherwise. Alternatively we can set $L_1 = 0$ with $L_2 > 0$, which gives an arrival rate of $a_2$ while the queue is less than or equal to $L_2$ and zero otherwise. Figure 9 compares the results with one and two thresholds in terms of the delay-throughput tradeoff. This clearly demonstrates the advantage of using two thresholds by indicating a lower delay for the same throughput values.

Figure 9 Performance comparison between special case I and II (see online version for colours)

4.2 Numerical performance results comparison

Based on the one-step, two-step and linear reduction analytical performance models presented in Sections 3.1–3.3, the mean packet delay $W$, probability of packet loss $P_L$ and throughput $S$ have been expressed as functions of the thresholds and maximum drop probability. This section presents the performance comparison results for these three models through typical numerical results (cf., Figures 11–17).

Model-I is called ‘One-Step Reduction’ since it incorporates one threshold to make the arrival process one-step reduce from arrival rate $a_1$ directly to $a_2$ once the number of packets in the system has reached the threshold value $L_1$; the source operates normally, otherwise. Model-II is called ‘Two-Step Reduction’ since it incorporates two thresholds to make the arrival process two-step reduce from arrival rate $a_1$ directly to $a_2$ (or further to $a_3$) once the number of packets in the system has reached the first threshold value $L_1$ (or the second threshold value $L_2$); the source operates normally, otherwise. Model-III is called ‘Linear Reduction’ since it incorporates two thresholds to make the arrival rate linearly reduce from $a_1$ to $a_2$ with system contents when the number of packets in the system is between two thresholds $L_1$ and $L_2$. The source operates normally with arrival rate $a_1$ before threshold $L_1$, and with arrival rate $a_2$ after the threshold $L_2$ (cf., Figure 10).

Figure 10 One-step, two-step and linear reduction

Figures 11–13 present a numerical results comparison between one-step, two-step and linear reduction models by setting Normalised Throughput (NS), Normalised Delay (ND) and probability of packets loss against same range of load (after the second threshold), respectively. The performance advantages of using two thresholds with linear reduction of arrival rate compared to schemes which use a single threshold with one-step reduction and two thresholds with two-step reduction of arrival rate have been clearly demonstrated in Figures 11 and 13, where the linear reduction model can always give a higher throughput and a lower packet loss probability than the other two models for the same load. However, in Figure 12, step reductions give...
the lower delay against the same load compared with linear reduction. This is because the arrival rate reduces to its minimum soon after the first threshold in Model-I or the second threshold in Model-II, and so allows fewer packets into the queue compared to Model-III. Consequently, the packets under Model-I and Model-II incur less waiting time.

Figure 11 NS Comparison between one-step, two-step and linear reduction models (see online version for colours)

Figure 12 ND comparison between one-step, two-step and linear reduction models (see online version for colours)

Figure 13 Loss probability comparison between one-step, two-step and linear reduction models (see online version for colours)

4.3 Comparison under the same range of normalised throughput (NS)

Figures 14 and 15 demonstrate that for the same loss probability and same normalised throughput, the linear reduction gives a lower normalised delay than the step reductions (one-step and two-step) over that part range of normalised throughput where the three can be compared.

Figure 14 ND vs. NS (same range of NS as in Figure 15) (see online version for colours)

Figure 15 Loss probability vs. NS (same range of NS as in Figure 14) (see online version for colours)

4.4 Comparison under the same range of normalised delay (ND)

Figures 16 and 17 indicate that the linear reduction gives a higher normalised throughput and the lower loss probability than the step reductions (one-step and two-step) over the same range of normalised delay where the three can be compared.

In summary, results presented in Figures 11–17 indicate that the linear reduction model gives marginally better performance than the step ones (one-step and two-step reduction) over the same range of input data where the three can be compared. The results also clearly demonstrate how different parameter settings can provide different tradeoffs
between throughput, loss probability and delay to suit different service requirements.

**Figure 16** NS vs. ND (same range of ND as in Figure 17) (see online version for colours)

![Graph showing NS vs. ND with different threshold models]

**Figure 17** Loss probability vs. NS (same range of ND as in Figure 16) (see online version for colours)

![Graph showing Loss probability vs. NS with different threshold models]

5 Conclusions

A new analytical framework applying AQM scheme based congestion control mechanism to limit the input rate at the network gateway in WLAN has been developed and analysed in this paper. Comparison of performance measurements among three analytical models for the impact of AQM scheme on WLAN by using queue thresholds has also been clearly demonstrated through typical numerical results. Three discrete-time performance models with threshold based arrival rate, namely, one-step, two-step and linear reduction have been set up to investigate how queue thresholds affect the overall system performance. We take into account the reduction of incoming traffic arrival rate due to packets dropped probabilistically with the drop probability increasing linearly with system contents, and also point out the performance advantages of using two thresholds compared to the scheme which uses only a single threshold. It has been clearly demonstrated that the two thresholds can always be adjusted to give a lower delay for the same throughput. Also the linear reduction model can always give a higher throughput and a lower packet loss probability than the other two models for the same load. However, step reductions give the lower delay against the same load compared with linear reduction. This is because the arrival rate reduces to its minimum soon after the first threshold in Model-I or the second threshold in Model-II, and so allows fewer packets into the queue compared to Model-III. Consequently, the packets under Model-I and Model-II incur less waiting time. The performance model developed and analysed enables the best load settings and drop probability to be chosen to suit a given situation; that is, to give an appropriate trade-off among throughput, delay and packet loss probability.

Acknowledgements

This work is sponsored by the Engineering and Physical Science Research Council (EPSRC), UK, under grant GR/S01658/01.

References


