Received: 23 July 2017 | Revised: 22 December 2017 | Accepted: 30 January 2018
DOI: 10.1002/acs.2863

SHORT COMMUNICATION

WILEY

An average queue-length-difference–based congestion detection algorithm in TCP/AQM network

Jin Zhu[1] | Tong Luo[1] | Lin Yang[2] | Wanqing Xie[3] | Geir E. Dullerud[4]

[1]Department of Automation, University of Science and Technology of China, Hefei, China
[2]Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong
[3]School of Information Science and Technology, University of Science and Technology of China, Hefei, China
[4]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Correspondence
Wanqing Xie, University of Science and Technology of China, Hefei 230027, China.
Email: wqxie@ustc.edu.cn

Present Address
Wanqing Xie, University of Science and Technology of China, Hefei 230027, China.

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61374073 and 61503356; Anhui Provincial Natural Science Foundation, Grant/Award Number: 1608085QF153

Summary

Congestion detection in transmission control protocol/active queue management networks remains a challenging problem in which the choosing of congestion signal is one of the most important factors. Exponentially weighted moving average of queue length, the most widely used congestion signal, is facing difficulties in detecting incipient congestion and quantifying the optimal forgetting factor. Aiming at these 2 disadvantages, we propose an average queue-length-difference–based congestion detection algorithm where exponentially weighted moving average of queue-length difference is chosen as the congestion signal with the theoretical optimal forgetting factor deduced. First, by defining the queue-length difference as the state, the corresponding state equation is derived from the fluid model. Second, we prove that the inflow traffic in state equation is a discrete-time martingale, which can be transformed to a Wiener process according to the martingale representation theorem. Noticing that the observation of state will be coupled with noise because of the unstable transmission, the state estimation is then derived with the application of recursive least squares filter. The filter gain of state estimation, which is a function of the noise-signal ratio, corresponds to the optimal forgetting factor in average queue-length-difference–based congestion detection algorithm. Simulation results in NS-3 and MATLAB illustrate the effectiveness of the proposed algorithm.

KEYWORDS

average queue-length difference, congestion detection, discrete-time martingale, optimal forgetting factor, recursive least squares

1 | INTRODUCTION

The congestion detection algorithm, which is responsible for acquiring congestion signal precisely, is critical to the active queue management (AQM) algorithms in transmission control protocol/Internet protocol (TCP/IP) network. With the acquired congestion signal, inflow rate and outflow rate can be adjusted according to AQM algorithms such that congestion control can be achieved.[1] Surrounding the detailed expressions of congestion signals, different measurable network parameters are introduced, such as instantaneous queue length,[2] average queue length,[3-5] average sending rate,[6] arrival rate,[7,8] link utilization,[9] and buffer occupancy.[10,11] Compared with other expressions, the average queue-length–based AQM algorithms have advantage in robustness while dealing with bursty traffic or transient congestion and require less observation information in addition. Taking the widely used random early detection (RED) algorithm for example, exponentially weighted moving average (EWMA) of queue length is adopted as the congestion signal, where the forgetting factor plays an important role. This sort of approach, however, raises 2 disadvantages. First and foremost, if the inflow rate itself is quite high and the queue length is small, the average queue-length–based AQM algorithms will drop a much smaller fraction of packets than it should and, thus, makes the algorithm inept at detecting incipient congestion.[12] Second, the performance of the average queue-length–based AQM algorithms is dependent on the value of the forgetting factor, which is often given by experiences.[13] The former disadvantage can be alleviated by adopting the arrival rate as the congestion signal such that a faster response speed can be obtained although it may cause transient fluctuations simultaneously.[12] In order to filter out transient fluctuations and smoothen the congestion signal, an EWMA of sending rate is then chosen in the work of Lovewell.[6] This algorithm, however, faces the same challenge that has been raised as the second disadvantage, ie, how to determine the optimal forgetting factor. If the forgetting factor is set large, the algorithm will fail to smoothen abnormal data caused by bursty traffic or network noise, whereas on the contrary, it will result in overreliance on historical data.

Motivated by the aforementioned analysis, we propose an improved congestion detection algorithm where the congestion signal is represented by the EWMA of queue-length difference. Moreover, we deduce the optimal forgetting factor that is expressed as a function of the noise-signal ratio. First, according to the fluid model and inflow/outflow rate, we model the number of flows as a discrete-time martingale and further prove the inflow traffic in a round trip time (RTT) to be a discrete-time martingale also. Second, by defining the queue-length difference as the system state, we derive the corresponding state equation and the discrete-time martingale in it can be transformed to Wiener process via the martingale representation theorem. Meanwhile, considering the existence of random delay jitters (namely unstable transmission) during packets deliver in network,[14] the observation of state may be coupled with noise. We model the unstable transmission as the observation noise and then derive the observation equation. Combining the state equation and observation equation, recursive least squares (RLS) filter is adopted to deduce the optimal estimation of state, which is represented as the EWMA of queue-length difference. Correspondingly, the optimal forgetting factor here is the filter gain of the RLS filter, which is a negatively related function of the noise-signal ratio. The quantitative relationship of this function shows that for a severer unstable transmission, the noise-signal ratio will increase and the optimal forgetting factor will then decrease such that the historical value has a greater weight. Furthermore, this theoretical analysis is consistent with the actual network situations. To sum up, the advantages of our average queue-length-difference–based congestion detection algorithm lie in 2 aspects, ie, detecting incipient congestion and revealing the quantitative relationship between the optimal forgetting factor and practical network parameter.

2 | MODEL ANALYSIS

Let $C(t)$ and $O(t)$ represent the traffic inflow rate and outflow rate, respectively. Consider there are $N(t)$ flows in the router and $w_i(t)$ and $R_i(t)$ represent the congestion window size and the RTT of the ith flow, then the real-time traffic load passing the router can be calculated as

$$C(t) = \sum_{i=1}^{N(t)} \frac{w_i(t)}{R_i(t)}. \tag{1}$$

Without loss of generality, $O(t)$ is equal to a constant O, which is determined by link bandwidth. Therefore, the queue length $q(t)$ should satisfy

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = C(t) - O. \tag{2}$$

In this paper, the sample time for queue length is set as a constant T, which has the same magnitude with the empirical value of RTT. Let q_k be the discrete sample of queue length $q(t)$, and we denote the congestion window size of the ith flow at time slot k as $w_{i,k}$ and the average RTT as $\bar{R}$. According to Equations (1) and (2), q_k satisfies second-order autocorrelation

$$q_k = q_{k-1} + \Delta q_{k-1}; \tag{3}$$

$$\Delta q_k = \Delta q_{k-1} + \frac{\sum_{i=1}^{N_k} w_{i,k} - \sum_{i=1}^{N_{k-1}} w_{i,k-1}}{\bar{R}} T. \tag{4}$$

Define queue-length difference Δq_k as the system state x_k, then Equation (4) can be rewritten as

$$x_k = x_{k-1} + \frac{\sum_{i=1}^{N_k} w_{i,k} - \sum_{i=1}^{N_{k-1}} w_{i,k-1}}{\bar{R}} T. \tag{5}$$

Much existing work assumes the flow number N_k remains constant[1,13,15]; however, because of the stochastic property of the requests from users, N_k should be a stochastic process instead of a constant. On the basis that N_k together with $w_{i,k}$ are finite for the constraint of link bandwidth, we propose the following assumption.

Assumption 1. Both flow number N_k and congestion window size $w_{i,k}$ of the ith flow are discrete-time martingales that satisfy

$$\begin{aligned}
&\text{(i)} \quad E\{|N_k|\} < +\infty, \quad E\{N_{k+1}|N_k\} = N_k(\text{a.e.}) \\
&\text{(ii)} \quad E\{|w_{i,k}|\} < +\infty, \quad E\{w_{i,k+1}|w_{i,k}\} = w_{i,k}(\text{a.e.}).
\end{aligned}$$

Here, $k \in N^+ = \{1, 2, \ldots\}$, and a.e. means this equality establishes with probability 1.

Define the inflow traffic in an RTT as $Y_k = \sum_{i=1}^{N_k} w_{i,k}$ and we have the following theorem.

Theorem 1. *The inflow traffic* $\mathrm{Y_k}$ *is a discrete-time martingale, ie, there is*

$$E\{|Y_k|\} < +\infty, \quad E\{Y_{k+1}|Y_k\} = Y_k(\text{a.e.}).$$

Proof. Noticing that the congestion window size has an upper bound $w_{\max}$, according to Assumption 1, the expectation of Y_k satisfies

$$E\{|Y_k|\} = E\left\{\sum_{i=1}^{N_k} w_{i,k}\right\} \le E\left\{\sum_{i=1}^{N_k} w_{\max}\right\} < +\infty.$$

Let $w_{1,k-1} = m_{1,k-1}, w_{2,k-1} = m_{2,k-1}, \ldots, w_{N_{k-1},k-1} = m_{N_{k-1},k-1}$ denote the congestion window size and $N_{k-1} = n$ denote the flow number at time slot $k-1$, respectively, then the inflow traffic $Y_{k-1} = y = \sum_{i=1}^{n} m_{i,k-1}$. Moreover, bearing in mind that N_k and $w_{i,k}$ are discrete-time martingales, the condition expectation of Y_k with $Y_{k-1} = y$ is given as

$$\begin{aligned}
E\{Y_k|Y_{k-1} = y\} &= E\left\{\sum_{i=1}^{N_k} w_{i,k}|N_{k-1} = n, w_{i,k-1} = m_{i,k-1}\right\} \\
&= E\left\{\sum_{i=1}^{n} w_{i,k}|w_{i,k-1} = m_{i,k-1}\right\} \\
&= \sum_{i=1}^{n} m_{i,k-1} \\
&= y.
\end{aligned}$$

Thus, we have

$$E\{Y_k|Y_{k-1}\} = Y_{k-1} \quad (\text{a.e.}).$$

□

According to the martingale representation theorem,[16] for martingale Y_k, there exists a unique stochastic process $h(t)$ satisfying

$$Y_k = \int_0^{kT} h(t)\mathrm{d}B(t), \tag{6}$$

where $\{B(t), t \geq 0\}$ subjects to a standard Wiener process or Brownian motion.

Substituting Equation (6) into Equation (5) and defining stochastic process η_{k-1} as

$$\eta_{k-1} = \frac{\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)}{\bar{R}} T,$$

we have the state equation rewritten as

$$x_k = x_{k-1} + \eta_{k-1}. \tag{7}$$

Lemma 1. *The variance of stochastic process η_{k-1} is given as follows:*

$$\sigma_{k-1}^2 = \frac{T^2}{\bar{R}^2} \int_{(k-1)T}^{kT} E\left\{h^2(t)\right\} \mathrm{d}t.$$

Proof. First, we write the discrete form of Ito integral as

$$I_n^{k-1} = \int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t) = \underset{\triangle_n \to 0}{\mathrm{l.i.m}} \sum_{m=1}^{n} h\left(t_{m-1}^{k-1}\right)\left[B\left(t_m^{k-1}\right) - B\left(t_{m-1}^{k-1}\right)\right],$$

where

$$(k-1)T = t_0^{k-1} < t_1^{k-1} < \cdots < t_n^{k-1} = kT$$
$$\triangle_n = \max_{1 \le m \le n} \left\{t_m^{k-1} - t_{m-1}^{k-1}\right\}.$$

Then,

$$E\left\{\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)\right\} = E\left\{\underset{\triangle_n \to 0}{\mathrm{l.i.m}} \sum_{m=1}^{n} h\left(t_{m-1}^{k-1}\right)\left[B\left(t_m^{k-1}\right) - B\left(t_{m-1}^{k-1}\right)\right]\right\}.$$

Since $h(t_{m-1}^{k-1})$ is independent of $B(t_m^{k-1}) - B(t_{m-1}^{k-1})$, the aforementioned equation can be calculated as

$$\begin{aligned} E\{\eta_{k-1}\} &= E\left\{\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)\right\} \\ &= \lim_{\triangle_n \to 0} \sum_{k=1}^{n} E\left\{h\left(t_{m-1}^{k-1}\right)\right\} E\left\{B\left(t_m^{k-1}\right) - B\left(t_{m-1}^{k-1}\right)\right\} \\ &= 0. \end{aligned}$$

Similar to I_n^{k-1}, we rewrite another I_l^{k-1}, and

$$\lim_{\triangle_{n,l} \to 0} E\left\{I_n^{k-1} I_l^{k-1}\right\} = E\left\{\left[\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)\right]^2\right\} = \int_{(k-1)T}^{kT} E\left\{h^2(t)\right\} \mathrm{d}t.$$

Therefore, we can get the variance of η_{k-1}

$$\begin{aligned} \sigma_{k-1}^2 &= E\left\{(\eta_{k-1} - E\{\eta_{k-1}\})^2\right\} \\ &= \frac{T^2}{\bar{R}^2}\left\{E\left\{\left[\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)\right]^2\right\} - \left(E\left\{\int_{(k-1)T}^{kT} h(t)\mathrm{d}B(t)\right\}\right)^2\right\} \\ &= \frac{T^2}{\bar{R}^2} \int_{(k-1)T}^{kT} E\left\{h^2(t)\right\} \mathrm{d}t. \end{aligned}$$

□

Due to the existence of random delay jitters, the time that every packet costs in transmission is different and the observation of x_k will be coupled with noise e_k. Thus, observation y_k is given as

$$y_k = x_k + e_k. \tag{8}$$

It can be seen from Equation (8) that we can not get x_k directly. In the next section, we will focus on seeking the optimal estimation of x_k.

3 | AVERAGE QUEUE-LENGTH-DIFFERENCE–BASED CONGESTION DETECTION ALGORITHM

In Section 2, we have deduced the corresponding state equation Equation (7) and the observation equation Equation (8), and now, we go on to derive the estimation of state.

Let e_k be a Gaussian white noise, namely, $\mathrm{Cov}(e_k, e_j) = \sigma_e^2 \delta_{k,j}$, where $\delta_{k,j}$ is Kronecker function (actually, for colored noise, this problem can also be handled with a forming filter and this will add no mathematical difficulties). Then, $\hat{x}_k$, the estimation of x_k, is given by the RLS method[17] as

$$\hat{x}_k = \hat{x}_{k-1} + g_k^*(y_k - \hat{x}_{k-1}), \tag{9}$$

where g_k^*, p_k^*, and $p_{k|k-1}^*$ are filter gain, the variance of estimation error, and the variance of 1-step prediction error, respectively, satisfying

$$\begin{aligned} g_k^* &= p_{k|k-1}^*\left(\sigma_e^2 + p_{k|k-1}^*\right)^{-1} \\ p_k^* &= p_{k|k-1}^* - g_k^* p_{k|k-1}^* \\ p_{k|k-1}^* &= p_{k-1}^* + \sigma_{k-1}^2. \end{aligned} \tag{10}$$

For the detailed proof, please refer to the appendix.

Notice that Equation (9) can be rewritten as

$$\hat{x}_k = \left(1 - g_k^*\right)\hat{x}_{k-1} + g_k^* y_k, \tag{11}$$

which means the estimation $\hat{x}_k$ is also given in the form of EWMA and the filter gain g_k^* is then the corresponding forgetting factor.

Remark 1. Recall that the RED algorithm adopts the EWMA of queue length as the congestion signal with a forgetting factor α

$$x_k^R = (1-\alpha)x_{k-1}^R + \alpha y_k^R, \tag{12}$$

where y_k^R is the discrete sample of queue length and x_k^R is the average queue length. The smooth process that the influence of the past sample points decays in an exponential fashion has the capacity of ignoring the effect of bursty traffic and transient congestion. Nevertheless, the optimal forgetting factor is often given by experience. From Equation (12), we know that the EWMA of queue length can be included as a specific case of our algorithm and the forgetting factor corresponding to g_k^* if we define the queue length as the system state. Compared with the EWMA of queue length, we have the capacity of detecting incipient congestion by exploiting higher-order information about queue length, and the forgetting factor is no longer a constant but self-adaptive according to the network environment.

According to lemma 1, the variance of η_{k-1} in state equation is only related with congestion window, flow number, average RTT, and sample time. As we know, the establishment of a link in the TCP network contains slow start and congestion avoidance, and we focus on the period of congestion avoidance for data flows. Similar to the concept of steady state in control systems, the network environment in congestion avoidance period is called "steady." In steady network environment, the variation of window size can be described as a 2-reflecting-barrier Markov chain with stationary distribution because of physical restriction. Consequently, the variance of η_{k-1} can be treated as a constant approximately and we can obtain the quantitative relationship between the optimal forgetting factor and network parameters.

Theorem 2. *Optimal forgetting factor and noise-signal ratio are negatively related.*

Proof. According to Equation (9), there is

$$p_k^* = p_{k|k-1}^* - \frac{p_{k|k-1}^{*2}}{\sigma_e^2 + p_{k|k-1}^*}. \tag{13}$$

For the steady network environment, there is $\lim_{k\to\infty} p_k^* = \lim_{k\to\infty} p_{k-1}^* = p^*(\infty)$, $\lim_{k\to\infty}\sigma_k = \sigma$, we have

$$p^*(\infty) = p^*(\infty) + \sigma^2 - \frac{\left(p^*(\infty) + \sigma^2\right)^2}{\sigma_e^2 + p^*(\infty) + \sigma^2},$$

Keeping in mind that $p^*(\infty)$ is positive definite, there is

$$p^*(\infty) = \frac{1}{2}\left(-\sigma^2 + \sqrt{\sigma^4 + 4\sigma^2\sigma_e^2}\right). \tag{14}$$

Finally, we have

$$g^* = \lim_{k\to\infty} \frac{p^*_{k|k-1}}{\sigma^2 + p^*_{k|k-1}} = \frac{p^*(\infty) + \sigma^2}{\sigma_e^2 + p^*(\infty) + \sigma^2}.$$

Here g^* is the value of g_k^* in steady network environment. Substitute Equation (14) into the aforementioned equation, and the optimal forgetting factor is then given as

$$g^* = \frac{\sqrt{1+4\gamma}+1}{2\gamma + \sqrt{1+4\gamma}+1}, \tag{15}$$

where $\gamma = \sigma_e^2/\sigma^2$ denotes noise-signal ratio. Therefore, it is obvious that noise-signal ratio is negatively related to optimal forgetting factor. □

Remark 2. According to Equation (15), the optimal forgetting factor is a function of noise-signal ratio $\gamma = \sigma_e^2/\sigma^2$. Figure 1 shows the update of g^* with γ, and it is clear that g^* will converge to 0 with the increasing of γ, $\gamma \in [0, +\infty)$. If the variance of observation noise σ_e increases, which means the delay jitters of packets become severer and result in more negative effects to the accuracy of observation, then the value of g^* will decrease for better smoothness such that the estimation of queue-length difference at current slot will be dependent on the historical value with a larger weight and on the current observation value with a smaller weight. On the contrary, if the variance of observation noise σ_e decreases, which means the delay jitters of packets become more gentle, the value of g^* may increase and the estimation will depend more on the current observation. This theoretical analysis is consistent with the actual network situation.

Remark 3. In the previous discussion, by defining the EWMA of queue-length difference as the congestion signal, we possess the capacity of detecting incipient congestion, and we prove that the optimal forgetting factor in our algorithm is negatively related with noise-signal ratio. For the existing AQM algorithms (which consist of 5 main parts, ie, congestion detection, parameter tuning, flow differentiation, control function, and feedback signal), they can be compared in terms of operation mechanisms that include the type of congestion detection algorithm used, the manner in which their parameters are tuned, the methods by which they perform flow differentiation if any, their control function, and the nature of their feedback signal to the source algorithms.[1] In this paper, the EWMA of queue-length difference is chosen as the congestion signal and congestion detection algorithm, 1 part of AQM algorithms, is then improved, whereas other operations in AQM algorithms remain unchanged. Consequently, we can adopt the congestion detection algorithm proposed in this paper and still adopt existing mechanism in other parts of AQMs, such as

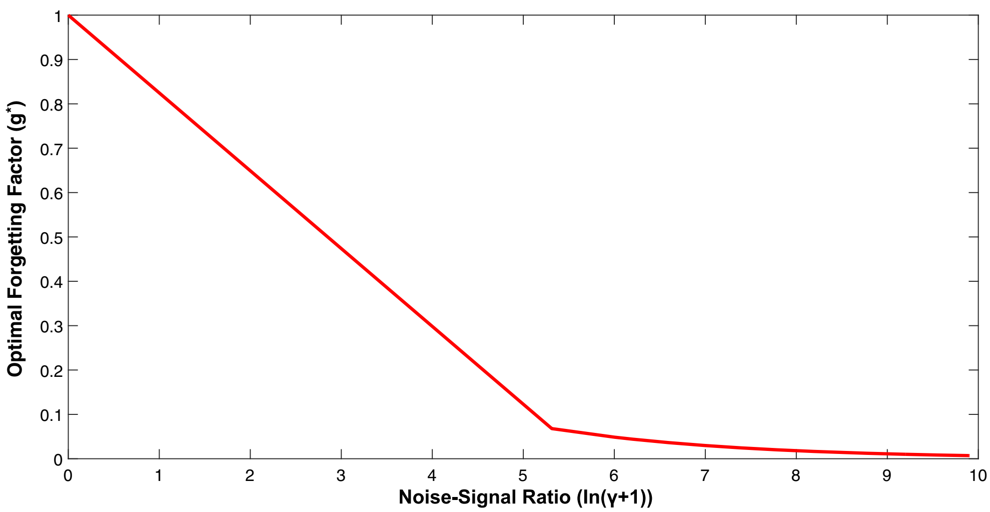

FIGURE 1 Optimal forgetting factor update with time-varying noise-signal ratio [Colour figure can be viewed at wileyonlinelibrary.com]

static for RED or dynamic for Green in parameter tuning, equation based for RED or nonequation-based for modified RED in control function, etc, and a new AQM algorithm can then be obtained. Recently, wireless networks are arousing much attention and future work can be carried out on congestion detection algorithm in wireless network. Compared with wired networks where packet dropouts are mainly caused by congestion, wireless networks will have more complexities such as link error and estimation for multiple time-varying signals.[18-21]

4 | SIMULATION

In this section, we provide simulation results using NS-3 and MATLAB to complement the analysis in the previous sections.

4.1 | EWMA of queue-length–based and queue-length-difference–based congestion algorithms

First, we compare the proposed algorithm with the original RED algorithm[3] in NS-3 platform. The network topology is shown in Figure 2. The simulation contains 2 FTP connections with the bandwidth and time delay given, respectively. Source 1 starts sending packet at time 0, and Source 2 starts after 3 seconds under “TcpNewReno” protocol. Here, RED gateway parameters are set as follows: $X_{\min} = 5$, $X_{\max} = 15$, $\alpha = 0.002$, QL (QueueLimit) = 25, and LI (LInterm) = 50. As for our algorithm, if queue-length difference in 0.1 second is greater than $X_c = 2$, the dropping probability is set as $\lambda = 0.7$. Figure 3 shows the queue length evolution under the 2 different algorithms by NS-3, where X-axis is time and Y-axis is the number of packets.

Second, we present results in MATLAB. The router in TCP/AQM network is with fixed outflow rate $O = 3000$ and stochastic inflow rate where the initial value of each flow i is $c_{i,0} = 30 + \nu$ and ν is the fluctuation of inflow rate taking values in $[-5, 5]$ uniformly. Set the initial flow number $n_0 = 100$ and initial queue length $X_0 = 5000$ with sample time 1s. The variation of flow number and inflow rate are set as follows: suppose at time k, the flow number is n_k and inflow rate is $c_{i,k}$ for each flow; then, at time $k+1$, the flow number n_{k+1} will be $n_k - 1$ or $n_k + 1$ both with probability 0.15, $n_k - 2$ or $n_k + 2$ both with probability 0.05, and the inflow rate $c_{i,k+1}$ will be $c_{i,k} + 3$ or $c_{i,k} - 3$ both with probability 0.1. Suppose bursty traffics happen at $t_1 = 20$ s, $t_2 = 90$ s, and $t_3 = 140$ s with inflow rate 1000, 1200, and 1500 whose duration time are $\tau_1 = 30s$, $\tau_2 = 20s$, and $\tau_3 = 30s$, respectively. According to the RED algorithm, the probability of dropping packet is set to 0 if the queue length q_k is less than $X_{\min}$, increase linearly if q_k locates between $X_{\min}$ and $X_{\max}$, and 1 if q_k is greater than $X_{\max}$. As for the average queue-length-difference–based congestion detection algorithm, on the basis of RED algorithm, if queue-length difference is greater than X_c, the dropping probability is λ. In this simulation, we set $X_{\max}$, $X_{\min}$, X_c, and λ as 24 000, 12 000, 800, and 0.7, respectively. The evolution of queue length is shown in Figure 4.

Figures 3 and 4 show that, when bursty traffic happens, the traditional queue-length–based algorithm cannot detect congestion until the queue length is greater than $X_{\min}$, so the queue length increases rapidly. As a comparison, the average

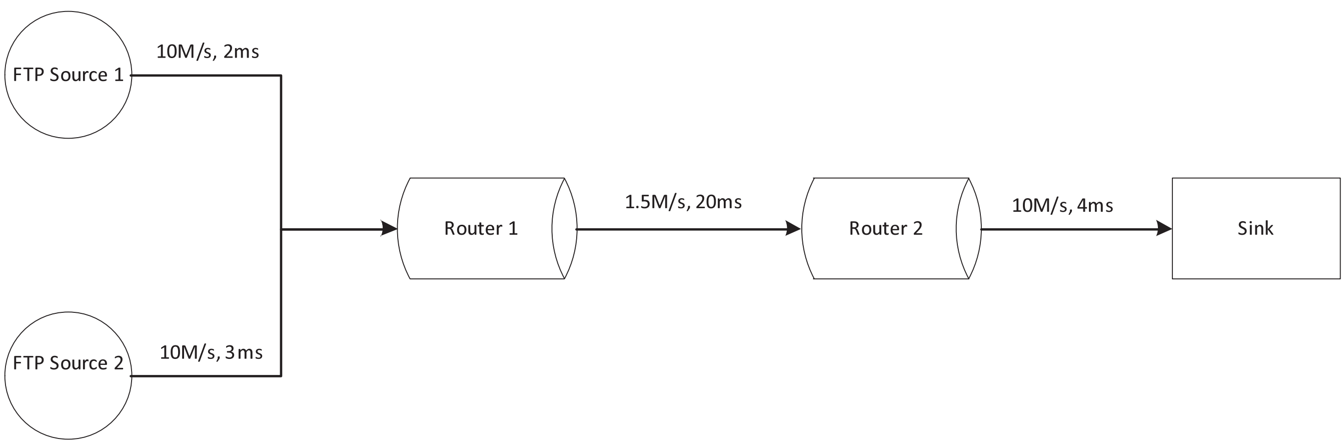

FIGURE 2 Simulation network topology. FTP, File transfer protocol

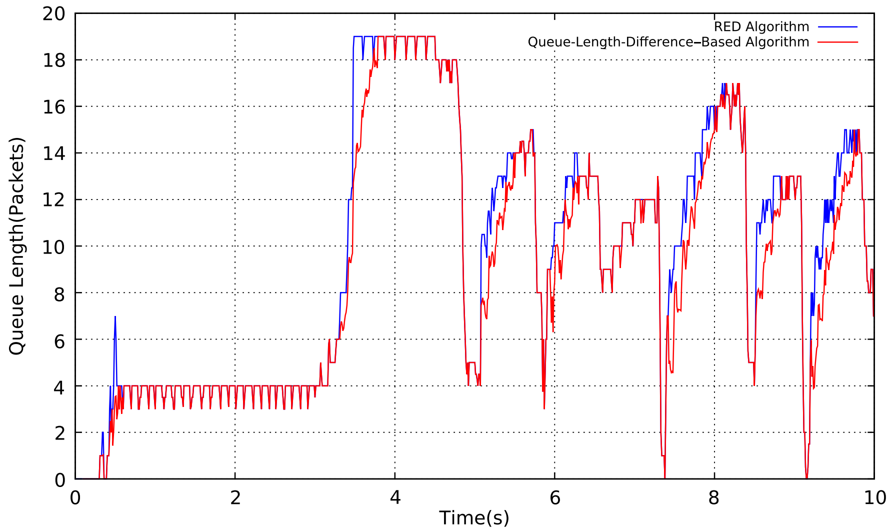

FIGURE 3 Evolution of queue length by NS-3. RED, Random early detection [Colour figure can be viewed at wileyonlinelibrary.com]

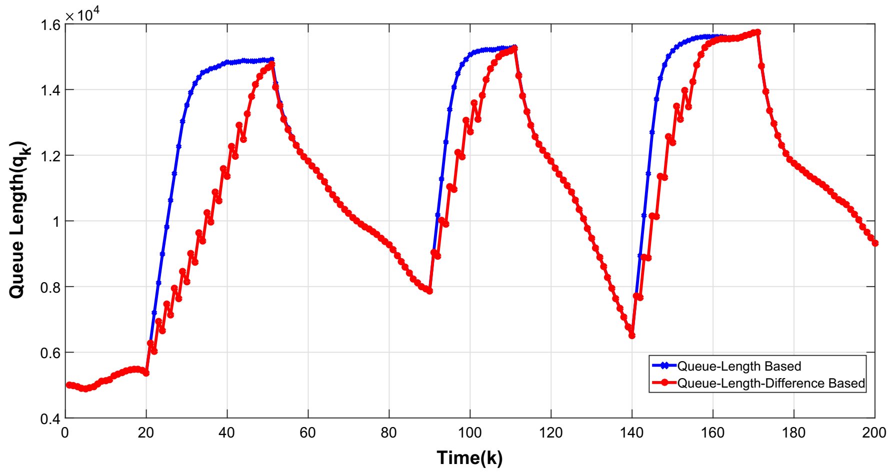

FIGURE 4 Evolution of queue length by MATLAB [Colour figure can be viewed at wileyonlinelibrary.com]

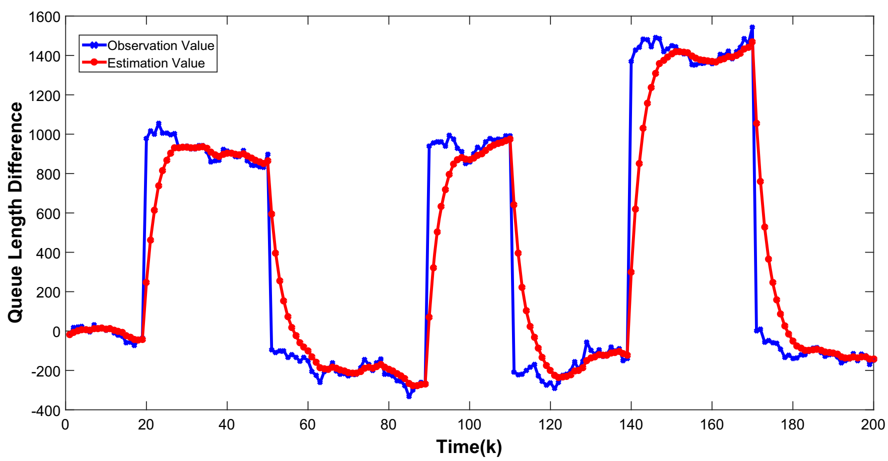

FIGURE 5 Observation and estimation of queue-length difference (σ_e^2=90) [Colour figure can be viewed at wileyonlinelibrary.com]

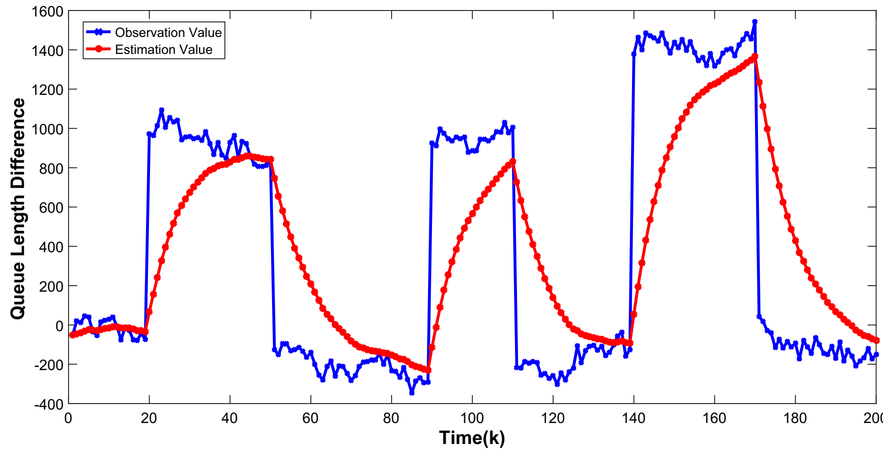

FIGURE 6 Observation and estimation of queue-length difference (σ_e^2=900) [Colour figure can be viewed at wileyonlinelibrary.com]

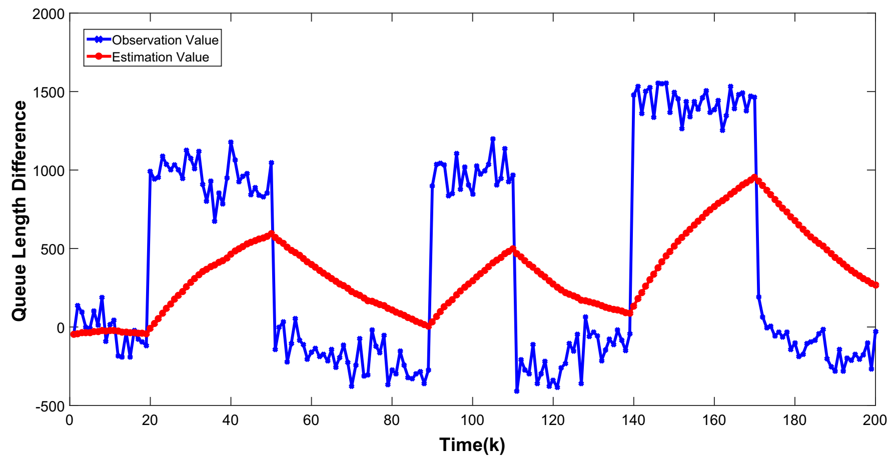

FIGURE 7 Observation and estimation of queue-length difference (σ_e^2=9000) [Colour figure can be viewed at wileyonlinelibrary.com]

queue-length-difference–based algorithm can detect incipient congestion once the queue-length difference exceeds the desired threshold X_c, the increase of queue length will be slower.

4.2 | Effects of observation noise

In this section, we investigate the effects of unstable transmission. Let the network parameters be the same as the above subsection and fix $\sigma^2 = 10$. Figures 5 to 7 show the comparison of observation and estimation of queue-length difference, where σ_e^2 are set as 90, 900, and 9000, respectively. It can be seen that the EWMA of queue-length difference is powerful to smoothen transient fluctuations of queue-length difference. With the decreasing of σ_e^2, the noise-signal ratio will decrease, whereas the optimal forgetting factor increases, then the estimation has better performance in tracking system observation, which is consistent with theoretical analysis.

5 | CONCLUSIONS

In this paper, we have proposed an average queue-length-difference–based congestion detection algorithm with the existence of unstable transmission in the TCP/AQM network. By defining queue-length difference as the state and modeling the variation of flow number along with congestion window size as the discrete-time martingale, state equation is then deduced according to the fluid model. Meanwhile, considering the fact that observation of state will be coupled with noise because of unstable transmission, the RLS method is applied to get the estimation of queue-length difference with optimal forgetting factor obtained. Moreover, we give the quantitative relationship between optimal forgetting factor and noise-signal ratio. Theoretical analysis and simulation show that this algorithm has advantages in detecting incipient congestion and quantifying the optimal forgetting factor.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under grants 61374073 and 61503356 and by the Anhui Provincial Natural Science Foundation under grant 1608085QF153.

ORCID

Jin Zhu http://orcid.org/0000-0002-6038-4339

REFERENCES

1. Adams R. Active queue management: a survey. *IEEE Commun Surv Tutor*. 2013;15(3):1425-1476.
2. Ott TJ, Lakshman TV, Wong LH. SRED: Stabilized RED. Paper presented at: Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'99); March 21-25, 1999; New York, NY.
3. Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. *IEEE/ACM Trans Networking*. 1993;1(4):397-413.
4. Park E-C. Analysis and design of the virtual rate control algorithm for stabilizing queues in TCP networks. *Comput Netw*. 2004;44(1):17-41.
5. Yang L, Zhu J, Xie W, Dullerud G. Time series analysis for congestion detection in TCP/AQM networks. *IEEE Commun Lett*. 2015;19(3):331-334.
6. Lovewell R. Packet-scale congestion control paradigm. *IEEE/ACM Trans Networking*. 2017;25(1):306-319.
7. Habachi O, Shiang H-P, van der Schaar M, Hayel Y. Online learning based congestion control for adaptive multimedia transmission. *IEEE Trans Signal Process*. 2013;61(6):1460-1469.
8. Nyirenda CN, Dawoud DS. Multi-objective particle swarm optimization for fuzzy logic based active queue management. Paper presented at: 2006 IEEE International Conference on Fuzzy Systems; July 16-21, 2006; Vancouver, Canada. ISBN 9780780394889.
9. Long C, Zhao B, Guan X. The Yellow active queue management algorithm. *Comput Netw*. 2005;47(4):525-550.
10. Feng W, Shin KG, Kandlur DD. The BLUE active queue management algorithms. *IEEE/ACM Trans Networking*. 2002;10(4):513-528.
11. Sojoudi S, Low SH, Doyle JC. Buffering dynamics and stability of internet congestion controllers. *IEEE/ACM Trans Networking*. 2014;22(6):1808-1818.
12. Kwon M, Fahmy S. Comparison of load-based and queue-based active queue management algorithms. Paper presented at: Proceedings of the Convergence of Information Technologies and Communications (ITCom'02); July 29-August 02, 2002; Boston, MA.
13. Barrera ID, Bohacek S, Arce GR. Statistical detection of congestion in routers. *IEEE Trans Signal Process*. 2010;58(3):763-776.
14. Kim MS. A wavelet-based approach to detect shared congestion. *IEEE/ACM Trans Networking*. 2008;16(4):2851-2864.
15. Firoiu V, Borden M. A study of active queue management for congestion control. Paper presented at: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies; March 26-30, 2000; Tel Aviv, Israel. ISSN 0743-166X.
16. Oksendal B. *Stochastic Differential Equations*. Berlin, Germany: Springer-Verlag; 2003.
17. Young PC. *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Berlin, Germany: Springer-Verlag; 2011. eBook ISBN 978-3-642-21981-8.
18. Zhao Y, Liu Y, Duan Z. Distributed average computation for multiple time-varying signals with output measurements. *Int J Robust Nonlinear Control*. 2016;26(13):2899-2915.
19. Liu Y, Zhao Y, Chen G. Finite-time formation tracking control for multiple vehicles: a motion planning approach. *Int J Robust Nonlinear Control*. 2016;26(14):3130-3149.
20. Zhao Y, Liu Y, Li Z. Distributed average tracking for multiple signals generated by linear dynamical systems: an edge-based framework. *Automatica*. 2017;75:158-166.
21. Zhao Y, Liu Y, Wen G. Distributed optimization of linear multi-agent systems: edge- and node-based adaptive designs. *IEEE Trans Autom Control*. 2017;62(7):3602-3609.

How to cite this article: Zhu J, Luo T, Yang L, Xie W, Dullerud GE. An average queue-length-difference–based congestion detection algorithm in TCP/AQM network. *Int J Adapt Control Signal Process*. 2018;32:742–752. https://doi.org/10.1002/acs.2863

APPENDIX : RLS ESTIMATION

Consider the state equation and the measurement equation

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{B}\boldsymbol{\eta}_{k-1} \\ y_k &= \boldsymbol{H}_k^T \boldsymbol{x}_k + e_k, \end{aligned} \tag{A1}$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{H}_k^T$ are the coefficient matrix. We characterize the minimum variance estimation of $\boldsymbol{x}$ by $\hat{\boldsymbol{x}}$, then it satisfies minimizing $J(\boldsymbol{x})$, where $J(\boldsymbol{x})$ is the sum of square of the differences between y and $\boldsymbol{H}^T\boldsymbol{x}$, and we write $J(\hat{\boldsymbol{x}})$ in the following form:

$$J(\hat{\boldsymbol{x}}) = \sum_{i=1}^{k} \left(y_i - \boldsymbol{H}_i^T \hat{\boldsymbol{x}}\right)^2. \tag{A2}$$

Consequently, the gradient of $J(\hat{\boldsymbol{x}})$ should be zero

$$\frac{1}{2}\nabla_{\hat{\boldsymbol{x}}}(J) = -\sum_{i=1}^{k} \boldsymbol{H}_i y_i + \left(\sum_{i=1}^{k} \boldsymbol{H}_i \boldsymbol{H}_i^T\right)\hat{\boldsymbol{x}} = 0, \tag{A3}$$

namely,

$$\sum_{i=1}^{k} \boldsymbol{H}_i y_i = \left(\sum_{i=1}^{k} \boldsymbol{H}_i \boldsymbol{H}_i^T\right)\hat{\boldsymbol{x}}. \tag{A4}$$

Let estimation noises $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_k$ and the covariance matrix of estimation noises as $\boldsymbol{P}_k^*$. Furthermore, let $\boldsymbol{y}_k = [y_1, y_2, \ldots, y_k]^T$, $\bar{\boldsymbol{H}}_k = [\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_k]^T$, $\boldsymbol{e}_k = [e_1, e_2, \ldots, e_k]^T$, then we deduce $\boldsymbol{y}_k = \bar{\boldsymbol{H}}_k \boldsymbol{x}_k + \boldsymbol{e}_k$ and the following is supportable:

$$\begin{aligned} \boldsymbol{P}_k^* &= E\left\{\tilde{\boldsymbol{x}}_k \tilde{\boldsymbol{x}}_k^T\right\} \\ &= E\left\{\left(\bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\right)^{-1} \bar{\boldsymbol{H}}_k^T \boldsymbol{e}_k \boldsymbol{e}_k^T \bar{\boldsymbol{H}}_k \left(\bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\right)^{-1}\right\} \\ &= \sigma_e^2 \left(\bar{\boldsymbol{H}}_k^T \bar{\boldsymbol{H}}_k\right)^{-1} \end{aligned} \tag{A5}$$

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{P}_k^*}{\sigma_e^2} \bar{\boldsymbol{H}}_k^T \boldsymbol{y}_k. \tag{A6}$$

According to Equation (A1), given the estimation of $\hat{\boldsymbol{x}}_{k-1}$, we can write a prediction equation and 1-step optimal prediction

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} \tag{A7}$$

and the prediction error is

$$\tilde{\boldsymbol{x}}_{k|k-1} = \hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{x}_k. \tag{A8}$$

The covariance matrix of 1-step prediction noises is

$$\boldsymbol{P}_{k|k-1}^* = \boldsymbol{A}\boldsymbol{P}_{k-1}^* \boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^T, \tag{A9}$$

where $\boldsymbol{Q}$ is the covariance matrix of $\boldsymbol{\eta_k}$. According to rectangularly-weighted past algorithm,[17] we can get the following recursive form of $\hat{\boldsymbol{x}}_k$:

$$\begin{aligned} \hat{\boldsymbol{x}}_k &= \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{g}_k^* \left(y_k - \boldsymbol{H}_k^T \hat{\boldsymbol{x}}_{k|k-1}\right) \\ \boldsymbol{g}_k^* &= \boldsymbol{P}_{k|k-1}^* \boldsymbol{H}_k \left(\sigma_e^2 + \boldsymbol{H}_k^T \boldsymbol{P}_{k|k-1}^* \boldsymbol{H}_k\right)^{-1} \\ \boldsymbol{P}_k^* &= \boldsymbol{P}_{k|k-1}^* - \boldsymbol{g}_k^* \boldsymbol{H}_k^T \boldsymbol{P}_{k|k-1}^*, \end{aligned} \tag{A10}$$

which can be simplified to 1-dimensional situation as

$$\begin{aligned} \hat{x}_k &= \hat{x}_{k-1} + g_k^* \left(y_k - \hat{x}_{k-1}\right) \\ g_k^* &= p_{k|k-1}^* \left(\sigma_e^2 + p_{k|k-1}^*\right)^{-1} \\ p_k^* &= p_{k|k-1}^* - g_k^* p_{k|k-1}^*. \end{aligned} \tag{A11}$$