Time Series Analysis for Congestion Detection in TCP/AQM Networks

Lin Yang, Jin Zhu, *Member, IEEE*, Wanqing Xie, and Geir E. Dullerud, *Fellow, IEEE*

***Abstract*—This letter aims at modeling the congestion detection issue commonly discussed in the research community. Based on the proposed model, we investigate how the correlation of traffic and measurement noise process affects the detection algorithm's performance by time series analysis approach.**

***Index Terms*—Congestion detection, time series analysis.**

I. Introduction

IT'S well known that congestion detection is essential for higher efficiency of congestion control algorithms in routers, as Active Queue Management (AQM) [1]. The Congestion Indicator (CI) of AQM can be considered as an on-line estimator which tries to discern the congestion information about the traffic in the presence of noise. So far, there have been several CI algorithms proposed according to different modeling methodologies, such as the ones based on Markov-modulated Bernoulli Process [2], wavelet analysis [3], maximum likelihood estimation [4], [5], control theory tools [6] etc., together with the traditional Exponentially Weighted Moving Average (EWMA) algorithm in RED [7]. However, these designs pay little or none attention to the time-correlation of measurements, in other words, these algorithms handle the measurements separately without utilizing their interrelation. Although correlation in high-speed networks is a very important factor for algorithm performance [8], to our best knowledge, there's few work devoted to studying this issue quantitatively. Motivated by that, we mathematically model the queue-based congestion detection issue in routers. Our model enables us to get an in-depth understanding of the time-correlation of traffic and noise process in TCP/AQM networks. Under the proposed model, we also deduce the corresponding optimal CI algorithms in mean square sense by time series approach. As shown in this letter, our work has good scalability for multi-traffic-pattern case, and is easy to implement via recursive design.

Manuscript received July 16, 2014; accepted December 16, 2014. Date of publication December 24, 2014; date of current version March 6, 2015. This work was supported in part by the National Natural Science Foundation of China under Grants 61374073, 60904021, and 61233003; by the Anhui Provincial Natural Science Foundation under Grant 1308085QF107; and by the Fundamental Research Funds for the Central Universities under Grant WK2100000003. The associate editor coordinating the review of this paper and approving it for publication was B. Rong.

L. Yang and J. Zhu are with the Department of Automation, University of Science and Technology of China, Hefei 230027, China (e-mail: ylin@mail.ustc.edu.cn; jinzhu@ustc.edu.cn).

W. Xie is with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China (e-mail: wqxie@ustc.edu.cn).

G. E. Dullerud is with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: dullerud@illinois.edu).

Digital Object Identifier 10.1109/LCOMM.2014.2385774

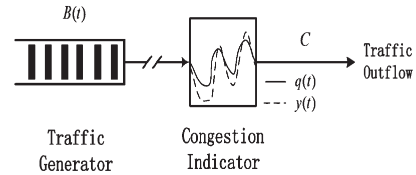

Fig. 1. Traffic monitoring.

II. Model Analysis for Congestion Detection

A. Problem Formulation

In window-based TCP/AQM networks, the real-time traffic load passing the routers can be represented as:

$$B(t)=\sum_{i=1}^{N(t)}\frac{W_i(t)}{R_i(t)} \tag{1}$$

where $N(t)$ is the application flow number, $W_i(t)$ and $R_i(t)$ are the congestion window size and Round Trip Time (RRT) of the ith flow respectively. Then the monitoring for traffic can be represented as the model in Fig. 1. The rate of traffic production $B(t)$ minus the outflow rate C, namely $B(t)-C$, directly determines the evolution of queue size in routers. And the queue-based CI aims at discerning the traffic-load-related information in some way through the observed queue size.

Theoretically, the queue size $q(t)$ should be subject to:

$$\frac{dq(t)}{dt}=B(t)-C. \tag{2}$$

Yet owing to unstable transmission in packet-switched networks, the time that every packet costs in transmission is different, so the observation about $q(t)$, noted as $y(t)$ (indicated by the dotted line in Fig. 1), may couple with some noise $e(t)$, namely, $y(t)=q(t)+e(t)$. Generally speaking, the function of CI is to filter out the noise $e(t)$, meanwhile respond quickly enough for time-varying estimated parameters. For example, RED adopts the low-pass filter algorithm with a "forgetting" factor as its congestion indicator, which is:

$$x_k \leftarrow (1-\alpha)x_{k-1}+\alpha y_k$$

where x_k is called the EWMA of queue size, and y_k is the discrete samples of queue size. The weighting introduced by α leads to an algorithmic memory that dies away into the past in an exponential fashion. However, this kind of heuristic design, though easy to be implemented, fails to take full account of the correlation of data, thus it's unclear how the optimal algorithm parameter α should be quantified.

B. Auto-Regressive CI Model

Obviously the queue size $q(t)$ is time-correlated, and its correlation can be described using an auto-regressive model $AR(l)$, where l is the order. According to (1), (2), the autocorrelation of $q(t)$ closely relies on the congestion window size $W_i(t)$. Assume $W_i(t)$ increases by $f_{\mathbf{I}}(\cdot)$ and decreases by

$f_{\mathbf{D}}(\cdot)$ each time. In general case, $f_{\mathbf{I}}(\cdot)$ and $f_{\mathbf{D}}(\cdot)$ depend on current congestion window size only, then q_k can be proved to be second-order auto-correlative if ignore the influence of TCP flow number, $N(t)$. In following discussion, we denote the discrete form of $q(t)$ as q_k, and the sample time is set as a RTT. Define $\Delta q_{k-1} = q_k - q_{k-1}$, and we have:

$$\begin{aligned} q_k &= q_{k-1} + \Delta q_{k-1} \\ \Delta q_k &= \Delta q_{k-1} + \frac{\sum_{i=1}^{N_k} W_{i,k} - \sum_{i=1}^{N_{k-1}} W_{i,k-1}}{\bar{R}} T \end{aligned}$$

where $\bar{R}$ represents the average of Round Trip Time. Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $\boldsymbol{q}_k = \begin{bmatrix} q_k \\ \Delta q_k \end{bmatrix}$, we can get the state equation in vector form, namely:

$$\boldsymbol{q}_k = \boldsymbol{A}\boldsymbol{q}_{k-1} + \boldsymbol{B}\eta_{k-1} \tag{3}$$

where

$$\eta_{k-1} = \frac{\sum_{i=1}^{N_k} W_{i,k} - \sum_{i=1}^{N_{k-1}} W_{i,k-1}}{\bar{R}} T.$$

Here we assume the flow number is a constant, then η_k, $k = 1, 2, \ldots$ can be considered as an i.i.d. random variable series satisfying:

$$E\{\eta_k\} = 0; \quad Cov(\eta_k, \eta_j) = \sigma^2 \delta_{k,j}$$

where $\delta_{k,j}$ is the Kronecker delta function, and $\sigma^2 = Var\{\eta_k\}$. As N_k is assumed to be a constant, σ^2 is only related with the TCP congestion window regulation scheme (see Appendix A). The measurement equation is:

$$y(k) = \boldsymbol{H}^T \boldsymbol{q}_k + e_k$$

where $\boldsymbol{H}^T = [1 \quad 0]$, and e_k is the discrete form of the measurement noise. For simplicity, in this subsection we ignore the correlation of e_k, then it can be handled as a white noise with constant variance, i.e. $Cov(e_k, e_j) = \sigma_e^2 \delta_{k,j}$. Let $\hat{\boldsymbol{q}}$ be the estimate of $\boldsymbol{q}$ and the Recursive Least Squares (RLS, [9]) estimation of $\boldsymbol{q}$ (in Appendix B, the derivation is revealed in vector form) is:

$$\hat{\boldsymbol{q}}_k = \boldsymbol{A}\hat{\boldsymbol{q}}_{k-1} + \boldsymbol{g}_k^*(y_k - \boldsymbol{H}^T \boldsymbol{A}\hat{\boldsymbol{q}}_{k-1}) \tag{4-a}$$

$$\boldsymbol{g}_k^* = \boldsymbol{P}_{k|k-1}^* \boldsymbol{H}(\sigma_e^2 + \boldsymbol{H}^T \boldsymbol{P}_{k|k-1}^* \boldsymbol{H})^{-1} \tag{4-b}$$

$$\boldsymbol{P}_k^* = \boldsymbol{P}_{k|k-1}^* - \boldsymbol{g}_k^* \boldsymbol{H}^T \boldsymbol{P}_{k|k-1}^* \tag{4-c}$$

where $\boldsymbol{P}_k^*$ and $\boldsymbol{P}_{k|k-1}^*$ are the covariance matrix of the estimation error and a *priori* estimation error respectively.

Remark 1: Correlation is an important issue for algorithm performance. In [2], the traffic is modeled as a Markov process, which implies the queue length process is first-order correlated. Through analysis, we can obtain the second-order model corresponds to the case where $N(t)$'s correlation is ignored, and for the higher-order case, i.e., $l > 2$, (3) can be extended where state will contain more variables, and similar algorithm can then be deduced.

For simplicity, if we only consider first-order correlation of $q(t)$, namely $l = 1$, and then we can obtain the simplest form for this kind model:

$$\begin{aligned} q_{k+1} &= q_k + \tilde{\eta}_k \\ y_k &= q_k + e_k. \end{aligned} \tag{5}$$

The corresponding optimal estimation algorithm for the first-order model in mean square sense is:

$$\hat{q}_k = \hat{q}_{k-1} + g_k^*(y_k - \hat{q}_{k-1}) \tag{6-a}$$

$$g_k^* = p_{k|k-1}^* \left(\sigma_e^2 + p_{k|k-1}^*\right)^{-1} \tag{6-b}$$

$$p_k^* = p_{k|k-1}^* - g_k^* p_{k|k-1}^*. \tag{6-c}$$

For first-order autocorrelation model, there is:

$$p_{k|k-1}^* = p_{k-1}^* + \tilde{\sigma}^2$$

where $\tilde{\sigma}^2$ is the variance of $\tilde{\eta}_k$. According to (6-b) and (6-c), there is also:

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

Let $k \to \infty$, the steady-state solution of these equations can be obtained via $p_k^* = p_{k-1}^* = p^*(\infty)$, and we have:

$$p^*(\infty) = \frac{1}{2}\left(-\tilde{\sigma}^2 + \sqrt{\tilde{\sigma}^4 + 4\tilde{\sigma}^2\sigma_e^2}\right). \tag{7}$$

It is shown that equation (6) works just as an EWMA and the optimal "forgetting" factor α^* in EWMA can be deduced:

$$\alpha^* = g^*(\infty) = \frac{p^*(\infty)}{\sigma_e^2 + p^*(\infty)}. \tag{8}$$

Remark 2: Substitute (7) into (8), the optimal "forgetting" factor is given as

$$\alpha^* = \frac{\sqrt{1+4\gamma} - 1}{2\gamma + \sqrt{1+4\gamma} - 1}$$

which only relies on the noise/signal ratio $\gamma = \sigma_e^2/\tilde{\sigma}^2$. Here σ_e^2 is concerned with the unstable packet transmission discussed in subsection C. $\tilde{\sigma}^2$ weights the change rate of $q(t)$ determined by the packet dropping rate, flow number, flow arrival and lifetime pattern etc. As γ increases, i.e., a high noise/signal ratio, we can find α^* decreases. This means the algorithm handles the instantaneous samples in a more smooth way.

Remark 3: For multi-traffic-pattern case, q can be rewritten as a vector form as well as η_k. For different traffic profile, such as long-term and bursty traffic, the arrival and lifetime pattern are different which will be reflected in different change rates of $N(t)$ and $q(t)$ as well. By specifying the covariance matrix of η_k, we can extract the desired congestion information. In this sense, it's just similar to the function of wavelet coefficients depicted in [3], but our method can take the recursive form as EWMA and thus be easier to implement.

C. Consider the Correlation of Noise e_k

In above discussions, we have studied the congestion detection issue under the assumption that e_k is white noise series. However, e_k is far more complicated than ideal assumption. Consider the packets depart from the sending node at relative fixed interval T, and then they are expected to arrive at the router gateway homogeneously over every T. However due to unstable transmission in packet-switched networks, the real time taken for packets to arrive at the router gateway may be varying. Let $\tau_i(t)$ denote the arrival deviation from expectation of the ith packet. Rationally, $\tau_i(t), i = 1,2,\ldots$ can be represented by a series of zero-mean and i.i.d. variables distributing on the interval $[-MT, NT]$. Let $F_\tau(t)$ denote the distribution function of $\tau_i(t)$, and the measurement noise of queue size owing to packet delay jitter can be represented as:

$$\begin{aligned}\varepsilon_k &= \sum_{i=k-M}^{k+N} \left\{ \mathbf{1}_{\{\tau_i(t)<(k-i)T\}} - E\left\{ \mathbf{1}_{\{\tau_i(t)<(k-i)T\}} \right\} \right\} \\ &= \sum_{i=k-M}^{k+N} \left\{ \mathbf{1}_{\{\tau_i(t)<(k-i)T\}} - F_\tau\left[(k-i)T\right] \right\}\end{aligned}$$

where $\mathbf{1}_{\{\cdot\}}$ is indicative function. And the autocorrelation sequence of ε_k is:

$$\begin{aligned}R_\varepsilon(k,k+s) = R_\varepsilon(s) &= E\{\varepsilon_k \varepsilon_{k+s}\} \\ &= \sum_{i=k-M+s}^{k+N} \left\{ E\left\{ \mathbf{1}_{\{\tau_i(t)<(k-i)T\}} \right\} \right. \\ &\qquad - F_\tau\left[(k-i)T\right] \\ &\qquad \left. \cdot F_\tau\left[(k-i+s)T\right] \right\} \\ &= \sum_{i=0}^{M+N-s} \left\{ F_\tau\left[(i+N)T\right] \right. \\ &\qquad - F_\tau\left[(i+N)T\right] \\ &\qquad \left. \cdot F_\tau\left[(i+N+s)T\right] \right\}.\end{aligned}$$

In this letter, we introduce a linear shaping filter to describe the time-correlated noise. From above equation, we can see that ε_k is a stationary series, and for $s > M+N$, $R_\varepsilon(s) = 0$. According to Wold Lemma, ε_k can be decomposed into a r-order moving average process $MA(r)$, $r = M+N$, which is:

$$\varepsilon_k = \mu + \sum_{j=0}^{r} \theta_j \upsilon_{k-j}$$

where υ_k is white noise series whose covariance is denoted as σ_υ^2. Thus we can get:

$$R_\varepsilon(s) = \begin{cases} (\theta_0\theta_s + \theta_{s+1}\theta_s + \cdots + \theta_r\theta_{r-s}) \cdot \sigma_\upsilon^2 & s = 1,2,\cdots,r \\ 0 & s > r. \end{cases} \quad (9)$$

By solving the non-linear simultaneous equations (9), we can get the numerical value of coefficient θ_j. Consequently, the measurement equation can be represented as:

$$y_k = \boldsymbol{H}_k^T \boldsymbol{q}_k + \sum_{j=0}^{r} \theta_j \upsilon_{k-j}.$$

Define $(r+1)$-dimension state vector $\boldsymbol{X}_k$ which subjects to:

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}\boldsymbol{X}_k + \boldsymbol{L}\upsilon_{k+1}$$

where $\boldsymbol{L} = [0 \quad 0 \quad \cdots \quad 0 \quad 1]^T$, and $\boldsymbol{F}$ is a $(r+1)\times(r+1)$ dimension one-step shift matrix, namely:

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{(r+1)\times(r+1)}.$$

Then there is:

$$\varepsilon_k = [\theta_r \quad \theta_{r-1} \quad \cdots \quad \theta_0]_{1\times(r+1)} \boldsymbol{X}_k$$

Let $\boldsymbol{G} = [\theta_r \quad \theta_{r-1} \quad \cdots \quad \theta_0]$. Accordingly, the extended state equations are:

$$\begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{X}_k \end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{A} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{F} \end{array}\right] \begin{bmatrix} \boldsymbol{q}_{k-1} \\ \boldsymbol{X}_{k-1} \end{bmatrix} + \left[\begin{array}{c|c} \boldsymbol{B} & 0 \\ \hline \mathbf{0} & \boldsymbol{L} \end{array}\right] \begin{bmatrix} \eta_{k-1} \\ \upsilon_k \end{bmatrix}$$

and the measurement equation is:

$$y_k = \left[\boldsymbol{H}_k^T \,|\, \boldsymbol{G}\right] \begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{X}_k \end{bmatrix}. \quad (10)$$

We find that the equation (10) does not contain the measurement noise item. This is the so-called noise-free or perfect measurements problem in the estimation literature, and corresponding optimal estimation can be obtained (see [10]).

Remark 4: The technique proposed in [4] handles the measurement noise due to unstable transmission by the Maximum Likelihood Estimation approach. Given the particular queue distribution, the likelihood of congestion $\Lambda(q_o)$ respect to instantaneous queue occupancy q_o can be obtained. The likelihood-based CI handles the noise issue in a very sophisticated way, but does not consider the time correlation of noise. In addition, this algorithm only takes the instantaneous queue size into account, consequently estimation smoothness can not be guaranteed especially faced with bursty traffic and non-stationary queue distribution.

III. Conclusion

The time-correlation of traffic process and measurement noise play a critical role in the performance of congestion indicator algorithms. In this letter, we adopt time series analysis approach to deal with this correlation. By Recursive Least Squares estimation, an optimal CI is proposed in mean square sense. Compared with other works in this filed, the introduced mechanisms will have advantages in scalability and computation complexity.

Appendix A
Statistical Characteristic of TCP Window Driving by AQM

The statistical equilibrium of a TCP Reno flow can be approximately obtained [11]. Omitting time-out loss indications, the congestion window size and throughput at equilibrium are:

$$\begin{aligned}W_i(p) &= E\{W_i(t)|p\} = \sqrt{\frac{8}{3bp}} + o(1/\sqrt{p}) \\ B_i(p) &= \frac{1}{R_i(t)}\sqrt{\frac{3}{2bp}} + o(1/\sqrt{p})\end{aligned}$$

where p is the packet dropping probability and b is the number of packets that are acknowledged by a received ACK. Because W_i is adjusted in a AIMD manner according to TCP congestion-avoidance strategy, namely $f_{\boldsymbol{I}}(w) = 1$ and $f_{\boldsymbol{D}}(w) = \frac{w}{2}$, we can describe $W_i(k)$ via a Markov chain. Moreover, because it's impossible for $W_i(k)$ to grow without restriction in physical sense, as an approximation, we assume the state space of $W_i(k)$ is $[1,2,3,\ldots,U]$, where U is the upper bound that $W_i(k)$ can achieve. And correspondingly, the state transition diagram is as shown in Fig. 2.

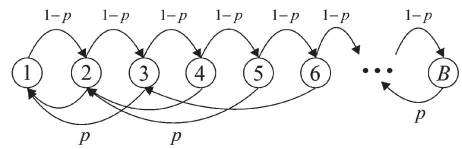

Fig. 2. State transition diagram.

Denoting the one-step transition probability matrix as $\boldsymbol{P}$, the steady state distribution $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_U]^T$ can be obtained:

$$\boldsymbol{\pi} = \boldsymbol{P}^T \boldsymbol{\pi}; \quad \sum_{i=1}^{U} \pi_i = 1.$$

With constant flow number N, easily we can obtain the $Var\{\eta_k\}$ occurring in (3) through follow equation:

$$Var\{\eta_k\} = N(t) \sum_{i=1}^{U} \pi_i \left[1 - p + p\frac{i^2}{4}\right]$$

and similarly, for $\tilde{\eta}_k$ in (5):

$$Var\{\tilde{\eta}_k\} = N(t) \sum_{i=1}^{U} \pi_i \left[i - W_i(p)\right]^2.$$

Appendix B
Recursive Least Squares (RLS) Estimation

Consider the linear regression model in state space form:

$$\begin{aligned} State\ Equations: \boldsymbol{q}_k &= \boldsymbol{A}\boldsymbol{q}_{k-1} + \boldsymbol{B}\boldsymbol{\eta}_{k-1} \\ Measurement\ Equations: y(k) &= \boldsymbol{H}_k^T \boldsymbol{q}_k + e_k \end{aligned}$$

where the $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{H}_k^T$ are the coefficient matrix. The least squares estimate $\hat{\boldsymbol{q}}$ of $\boldsymbol{q}$ is simply that value of recursive average estimation which minimizes a cost function $\mathcal{J}$, defined as the sum of the squares of the differences between y and $\boldsymbol{H}^T\hat{\boldsymbol{q}}$ over the observation period, i.e., in mathematical terms:

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

The objective of RLS is to keep $\mathcal{J}(\hat{\boldsymbol{q}})$ minimum all the time in a recursive way. There is:

$$\frac{\partial}{\partial \hat{\boldsymbol{q}}} \left(y_i - \boldsymbol{H}_i^T \hat{\boldsymbol{q}}\right)^2 = -2\boldsymbol{H}_i^T \left(y_i - \boldsymbol{H}_i^T \hat{\boldsymbol{q}}\right).$$

Then the gradient of $\mathcal{J}(\hat{\boldsymbol{q}})$ with respect to all the elements of $\hat{\boldsymbol{q}}$ is:

$$\frac{1}{2}\nabla_{\hat{q}}(\mathcal{J}) = -\sum_{i=1}^{k} \boldsymbol{H}_i y_i + \left(\sum_{i=1}^{k} \boldsymbol{H}_i \boldsymbol{H}_i^T\right) \hat{\boldsymbol{q}} = 0.$$

As a result:

$$\left(\sum_{i=1}^{k} \boldsymbol{H}_i \boldsymbol{H}_i^T\right) \hat{\boldsymbol{q}}_k = \sum_{i=1}^{k} \boldsymbol{H}_i y_i. \tag{11}$$

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

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

Subsequently, (11) can be rewrote as:

$$\hat{\boldsymbol{q}}_k = \frac{\boldsymbol{P}_k^*}{\sigma_e^2} \bar{\boldsymbol{H}}_k^T \boldsymbol{y}_k. \tag{12}$$

Considering the time-varying parameter $\boldsymbol{q}_k$, the optimal *a priori* estimation about $\boldsymbol{q}_k$ in terms of minimum Mean Square Error (MSE), denoted as $\hat{\boldsymbol{q}}_{k|k-1}$, is:

$$\hat{\boldsymbol{q}}_{k|k-1} = g(\boldsymbol{y}_{k-1}) = E\{\boldsymbol{q}_k | \boldsymbol{y}_{k-1}\}$$

namely $\hat{\boldsymbol{q}}_{k|k-1} = \boldsymbol{A}\hat{\boldsymbol{q}}_{k-1}$. The covariance matrix $\boldsymbol{P}_{k|k-1}^*$ of the *a priori* estimation noises is:

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

where $\boldsymbol{Q}_q$ is the covariance matrix of $\boldsymbol{\eta}$. According to the Orthogonal Projection Theorem, the optimum prediction $\hat{\boldsymbol{q}}_{k|k-1}$ is actually the least-square estimation according to $\boldsymbol{y}_{k-1}$ and $\hat{y}_{k|k-1}$, then combined with (12), we have:

$$\hat{\boldsymbol{q}}_k - \hat{\boldsymbol{q}}_{k|k-1} = \frac{\boldsymbol{P}_k^*}{\sigma_e^2} \bar{\boldsymbol{H}}_k^T \boldsymbol{y}_k - \boldsymbol{A}\hat{\boldsymbol{q}}_{k-1}. \tag{13}$$

Finally, (13) can be transformed into a recursive form and represented as follow:

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

References

[1] R. Adams, "Active queue management: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1425–1476, 2013.

[2] L. Guan, I. Awan, and M. Woodward, "Stochastic modelling of random early detection based congestion control mechanism for bursty and correlated traffic," *Proc. Inst. Elect. Eng.–Softw.*, vol. 151, no. 5, pp. 240–247, Oct. 2004.

[3] M. Kim, T. Kim, Y. Shin, S. S. Lam, and E. J. Powers, "A wavelet-based approach to detect shared congestion," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 763–776, Aug. 2008.

[4] I. Barrera, S. Bohacek, and G. Arce, "Statistical detection of congestion in routers," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 957–968, Mar. 2010.

[5] I. Barrera, G. Arce, and S. Bohacek, "Statistical approach for congestion control in gateway routers," *Comput. Netw.*, vol. 55, no. 3, pp. 572–582, Feb. 2011.

[6] Y. Ariba, F. Gouaisbaut, S. Rahme, and Y. Labit, "Traffic monitoring in transmission control protocol/active queue management networks through a time-delay observer," *IET Control Theory Appl.*, vol. 6, no. 4, pp. 506–517, Mar. 2012.

[7] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, pp. 397–413, Mar. 1993.

[8] W. Zhao, X. Huang, K. Shi, and L. Zhang, "TSBCC: Time series-based congestion control algorithm for wireless network," *J. Netw.*, vol. 8, no. 5, pp. 1058–1064, May 2013.

[9] P. C. Young, *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. New York, NY, USA: Springer-Verlag, 2011.

[10] P. S. Maybeck, *Stochastic Models, Estimation, Control (Volume 1)*. New York, NY, USA: Academic, 1982.

[11] J. Padhye, V. Firoiu, D. Towsley, and J. F. Kurose, "Modeling TCP reno performance: A simple model and its empirical validation," *IEEE/ACM Trans. Netw.*, vol. 8, no. 2, pp. 133–145, Apr. 2000.