Design and Performance Analysis of a Sliding-mode Prediction Based Active Queue Management

Zhang Xizheng and Wang Yaonan
School of Computer Science, Hunan Institute of Engineering, Xiangtan 411104, Hunan Province, China

Abstract: As the source congestion in TCP network plays an important role in causing the network instability and deteriorating the quality of service, this study aims to investigate a robust congestion control to solve this problem. A fluid-flow model was formed to represent the nonlinear TCP behavior and the Sliding-Mode Prediction (SMP) approach was applied to design the active queue management controller. The stability and robustness of the SMP scheme are then effective validated under different network scenarios in NS-2. Simulations results verify that SMP scheme can provide smaller queue fluctuations, higher link utilization and low packet loss rate than the traditional controllers.

Key words: Congestion control, active queue management, sliding-mode control, model prediction

INTRODUCTION

With the incredible development of the Internet technique since 2000s, the congestion detection and the congestion control of the complex Internet have attracted a lot of researchers’ interests (Jain, 1990; Low et al., 2002; Floyd and Jacobson, 1993). It has two components: (1) the end-to-end congestion control protocol, such as TCP (Athuraliya et al., 2001) and (2) an active queue management (AQM) scheme implemented in routers (Holot et al., 2002; Jasem et al., 2010). In AQM scheme, the source is supposed to adjust the sending rate and the packets are subject to dropping or marking in case the congestion is detected. Normally, AQM aims to (1) stabilize the buffer queue length at a given target and to (2) minimize the occurrences of queue overflow and underflow. Since the average queue length is kept at a designated value, it can provide more sufficient capacities to deal with abrupt data flows without any dropping.

In recent years, researchers had presented several mathematical models of TCP/AQM system (Kim, 2006). Among those, the fluid-flow model is mostly successful to be used to design or to implement AQM controller. As a well-known approach, RED (Floyd and Jacobson, 1993) adopts the fluid-flow model to design AQM. Actually, RED is a feedback controller which uses the average queue length to measure the congestion effect on Internet. However, the performance of RED is sensitive to traffic loads and the parameter settings. There is not any good method to set the RED parameters yet. REM (Athuraliya et al., 2001) is not easily to be implemented, because the exponential function involves high computing loading on router. Due to issue from the classical control theory, PID scheme has the merits such as simplicity, low complexity and easy implementation when is applied into the TCP/AQM system (Camacho and Bordons, 1999). Unfortunately, compared to RED and REM, PID needs a longer response time and raises the propagation delay which deduces more chattering in queue length.

It's acknowledged that TCP network is a time-varying nonlinear system with various uncertainties such as network parameter variations, multi-flow disturbances. Thus, the mathematical modeling is difficult to establish. Recently, the fluid flow model is extensively used for network performance evaluation and control, particularly for congestion control problems (Kim, 2006). In order to develop the network congestion controller, several model-based schemes have been proposed to provide theoretic analysis for networking problems, but most of them are based on linear control theory. For example, analysis and control design tools are applied to control traffic in asynchronous transfer mode networks (Athuraliya et al., 2001) and analyze the stability of congestion control schemes in TCP/IP networks (Holot et al., 2002; Sun et al., 2006). However, since the mathematical model of the TCP network is inherently complex and nonlinear, it's a difficult and challenging test to design a congestion controller to deal with the whole nonlinearity and uncertainty as well as analyse the control performance.

Corresponding Author: Zhang Xizheng, School of Computer and Communication, Hunan Institute of Engineering, Xiangtan, 411104, Hunan Province, China Tel: +86 13574084952 Fax: +86 731 58688512

1309
As a robust control approach, Sliding-Mode Control (SMC) is proved to be efficient to control dynamic nonlinear systems with uncertainties (Yufeng et al., 2011; XiZheng et al., 2011). Some researchers have reported to apply SMC into AQM system (Yan and Yue, 2010; Wang et al., 2008; Wang et al., 2010; Zhang et al., 2009). However, any sliding-mode is a mode of motions on the discontinuity set of a discontinuous dynamic system and chattering phenomena still limits the application of SMC strategy. Thus, it’s necessary and vital to remove the chattering phenomenon for SMC’s application. Meanwhile, Model Predictive Control (MPC) technique is successfully applied in the congestion control for TCP networks (Haeri and Mohsenian Rad, 2006; Wang et al., 2002; Li et al., 2003). MPC has been found to be robust for systems with time-delays, even uncertain ones. By utilizing this characteristic of MPC, an efficient congestion control has been proposed by Misra et al. (2000). In the existed congestion control schemes based on MPC technique, most of them utilize the linear ARMA (Li et al., 2003) model to obtain the prediction of the queue buffer. According the predictive level, the source planned the traffic rates in a best-effort way. However, since the propagation and the processing delay cannot be precisely obtained, the delay and other parameters’ variation inevitably affected the control performance.

To solve the aforementioned problems, a robust Sliding-Mode-Prediction (SMP) based AQM scheme is proposed in this paper. The SMP possesses the inherited merit of the sliding-mode control such as fast response, parameter insensitivity and strong robustness. At the same time, because of receding horizon optimization, control signal can be optimized continuously and on-line. The principle purpose of this paper is to present a robust TCP/AQM scheme to stabilize the queue size and minimize the number of the discarded packets. There are two main contributions. First, we present a linear delay-embedded transformation to convert the time-delay model into the delay-free one; Second, we combine the SMC and MPC to develop a novel AQM scheme, SMP which makes the AQM algorithm more robust towards network parameters variations, noises, modeling errors and model nonlinearities.

**DYNAMIC DELAY-FREE MODEL OF TCP/AQM**

**Nonlinear model of TCP/AQM dynamics:** A fluid-flow model of TCP behavior was developed in (Kim, 2006) based on stochastic differential analysis. Simulation result demonstrated that this model accurately capture the dynamics of TCP. A simplified version of the model ignoring the TCP time out mechanism can be described by the following coupled, non-linear equations:

\[
\begin{align*}
\dot{w}(t) &= -\frac{1}{\tau(t)} w(t) - \frac{w(t)}{2\tau(t)} p(t) \\
Q(t) &= \frac{w(t)}{\tau(t)} N(t) - C + d(t)
\end{align*}
\]  

where, \(w(t)\) is the TCP window size, \(Q(t)\) is the instantaneous queue length of the router buffer, \(R(t)\) is the Round Trip Time (RTT) and can be expressed as \(R(t) = Q/C + T_p\). \(C\), \(T_p\) and \(N\) are parameters related to the network configuration and represent the transmission capacity of the router, the propagation delay and the number of TCP sessions respectively. The variable \(p(t)\) is the packet-dropping probability of a packet and \(0 \leq p(t) \leq 1\) which is the control input used to reduce the sending rate and to maintain the bottleneck queue.

In the mathematical model (1), the first equation gives the adaptation of the traffic rate under the control input \(p(t)\) and the second equation describes the change rate of the queue length on the router. It also should be noticed that the time delay effects on \(R(t)\), \(N(t)\) and \(p(t)\) have been approximated by other terms. In the model (1), we have introduced an additional signal \(d(t)\) which models cross traffics through the router and filling the buffer. These traffics are not TCP based flows (not modeled in TCP dynamic) and can be viewed as perturbations since they are not reactive to packets dropping (for example, UDP based traffic).

A linearization and some simplifications of (1) was carried out to allow the use of traditional control theory approach. The linearized fluid-flow model of TCP is as follow:

\[
\begin{align*}
\dot{w}(t) &= -\frac{N_s}{CR_0} (\Delta w + \Delta w(t - h(t))) - \frac{N_s}{CR_0} (\Delta Q - \Delta Q(t - h(t))) \\
\dot{Q}(t) &= \frac{C R_0}{2N_s^2} \Delta p(t - h(t)) \dot{Q}(t) - \frac{N_s}{R_0} \Delta w(t) - \frac{1}{R_0} \Delta Q(t) + d(t)
\end{align*}
\]  

where, \(\Delta w = w - w_o\), \(\Delta Q = Q - Q_o\) and \(\Delta p = p - p_o\) are the perturbation variables about the operating point. The operating point \((w_o, Q_o, p_o)\) is defined by:

\[
\begin{align*}
w(0) &= p_o w_0 \rightarrow 2 \\
Q(0) &= w_0 + \frac{1}{R_0} T_p
\end{align*}
\]  

The input of the model (2) corresponds to the drop probability of a packet. This probability is fixed by the AQM. This latter has for objective to regulate the queue size of the router buffer.
In this study, this regulation problem is addressed in Section 3 with the design of a stabilizing state feedback for time delay systems. Indeed, an AQM acts as a controller and in order to design it, we have to solve a synthesis problem.

From (2), it's shown that the dropping probability \( p(t) \) plays an important role in the avoidance of network congestion, the AQM program needs to adjust \( p(t) \) according to the arriving packets rate. If we design a good control rule of \( p(t) \), then the network can be free from congestion.

**Linear delay-free model of TCP/AQM:** In this study, we choose to model the dynamics of the queue and the congestion window as a time delay system. Indeed, the delay is an intrinsic phenomenon in networks and taking into account its characteristic should improve the precision of our model with respect to the TCP behavior. After the process of discretization and linearization, the discrete TCP fluid model (2) can be rewritten as the following time delay system:

\[
\begin{align*}
\begin{cases}
x(k+1) = & Ax(k) + A_d x(k-k_1) + Bu(k-k_2) + B_d d(k) \\
\end{cases}

s_i(t) = \phi(t), \text{with } t \in [h_0, 0]
\end{align*}
\]

(4)

with

\[
\begin{align*}
A &= \begin{bmatrix}
-\frac{\Delta w}{\Delta t} & -\frac{1}{\Delta t} \\
\frac{1}{\Delta t} & -\frac{1}{\Delta t}
\end{bmatrix} \\
A_d &= \begin{bmatrix}
-\frac{\Delta Q}{\Delta t} & 0 \\
0 & \frac{1}{\Delta t}
\end{bmatrix} \\
B &= \begin{bmatrix}
-\frac{\Delta p}{\Delta t} \\
0
\end{bmatrix} \\
B_d &= \begin{bmatrix}
0 \\
1
\end{bmatrix}
\end{align*}
\]

where \( x(t) = [\Delta w(k), \Delta Q(k)] \) is the state vector and \( u(k) = \Delta p(k) \) is the control input. \( \phi(t) \) is the initial condition. \( k_1 \) and \( k_2 \) are the steps of state delay and control delay, respectively.

For a linear system without any delay, it's easy and effective to design sliding-mode controller to achieve global stability. While for the controllable system (4) with both control and state delay, the conventional sliding surface would be a functional and is difficult to design. In this paper, through a particular linear transformation, the original time-delay system is first transformed into a delay-free system. Then, the sliding-mode controller is designed based on the resulted system.

Consider the following linear transformation:

\[
y(k) = A^h w(k) + \sum_{i=1}^{k_1} A_i x(k-k_1+i) + \sum_{i=0}^{k_2} A_{d,i} x(k-k_2+i)
\]

(5)

By the transformation in Eq. 5, the time-delay system is converted into the following:

\[
y(k+1) = Ay(k) + A_d Bu(k) + D\{A^h w(k) + B_d d(k)\}
\]

(6)

where, \( D \) is the resulted matrix with proper dimension. In Eq. 6, the original states \( x(k) \) and disturbance \( d(k) \) are both dealt as perturbation of the system, thus, it is no need to deduce the exact expression of \( D \).

Equation (5-6) provides a design of the control of the dropping probability \( p(t) \), the key is how to employ the information of the buffer queue, since the router has no other way to get more information of the status of the network. Thus, an efficient AQM algorithm depends on the design of the control of the dropping probability \( p(t) \). Different from many existing results, our scheme is based on nonlinear difference Eq. 5-6 which makes us be capable of using digital technique (software) to implement the control of the network.

**SLIDING-MODE PREDICTION CONTROLLER DESIGN**

**Design of sliding-mode prediction (SMP) model:** It's well known that an effective AQM scheme should be in feedback control form. This section will illustrate the proposed SMP control scheme for TCP/AQM system. In fact, the system (6) can be deemed as a single input-single output system and is a suitable object for sliding-mode predictive controller.

For system (6), the model reference system is defined as:

\[
y(k+1) = Ay(k) + Bu(k)
\]

(7)

Accordingly, the reference sliding function is defined as:

\[
s_i(k+1) = \sum_{i=1}^{k} \varepsilon_i y_i(k)
\]

(8)

where, the constants \( \varepsilon_i \) is chosen according to the pole placement rule such that the resulting quasi-sliding mode motion is stable.

In order to handle the perturbation in (6), the time-delay sliding-mode compensator is designed as:

1311
\[ s_k + \tau = \frac{1}{\sigma_k} \left[ CAy(k - \tau - 1) + CA^T B v(k - \tau - 1) \right] \]  

(9)

where, \( \tau = \max(k, k_s) \) is a integer and \( C \) is the selected constant matrix according to the pole placement rule.

Thus, the sliding-mode prediction model is obtained as:

\[ s_k + \tau = s_k + \xi \]  

(10)

In fact, the output \( s_k \) of the sliding-mode prediction model is deduced from the nominal model and the predictive model.

In this case, the compensator output of time \( k \) on time \( k+1 \) can be deduced from (6):

\[ s_k + \tau = \frac{1}{\sigma_k} \left[ CAy(k - \tau - 1) + CA^T B v(k - \tau - 1) \right] \]  

(11)

In practice, SMP model will inevitably have some errors because of time-variability, nonlinearity, or disturbances and so forth. Therefore, the model output will not be the same as the real output. In predictive control method, a common way to solve this problem is to employ the output error feedback correction approach. That is, firstly, at time \( k \), calculate the error between the real value and the model output value; secondly, add the error to model output value at time \( k+1 \); then take the sum as closed-loop output predictive value at time \( k+1 \).

Here, the error between practical sliding mode value \( s_k \) and SMP model predictive value \( s_k(k-1) \) is used to make feedback correction for SMP model output \( s_k \) and the closed-loop output of SMP \( s_k(k+1) \) is given as follows:

\[ s_k(k+1) = s_k(k+1) + \xi (s_k(k) - s_k(k-1)) \]  

(12)

where, \( \xi \) is a correction coefficient. From the viewpoint of practice, usually \( \xi = 1 \) is simple and suitable. That is, the SMP model uses the information of tracking error to adjust sliding mode function. Specially, when \( k \to \infty \), there exists:

\[ s_k(k+\tau) = s_k(k+\tau) \]  

(13)

Since the stability of sliding-mode motion has been guaranteed, the system represented by SMP model (12) is stable with respect to \( k \to \infty \).

**Design of control law:** For the model prediction based control approach, it’s essential to adopt proper performance index to calculate the control input. In this paper, a traditional quadratic cost function is designed as following:

\[ J = \sum_{k=1}^{\infty} \left( s_k(k+1) - s_k(k+1) \right) ^T \Omega \left( s_k(k+1) - s_k(k+1) \right) \]

\[ + \sum_{k=1}^{\infty} \left( y(k+1) - y(k+1) \right) ^T \Xi \left( y(k+1) - y(k+1) \right) \]

\[ + \sum_{k=1}^{\infty} \left( u(k) - u(k) \right) ^T \Psi \left( u(k) - u(k) \right) \]  

(14)

where \( \Omega, \Xi \) and \( \Psi \) are the weight coefficient matrices which adjusts the relations between the weighting of the control system error and the control signal. Generally, the larger \( \Psi \), is chosen, the slower will be the resulting states response.

For the sake of clarity, define following matrix vectors:

\[ \begin{bmatrix} s_k(k+1) = [s_k(k+1), s_k(k+2), \ldots, s_k(k+\tau)]^T \\ s_k(k+1) = [s_k(k+1), s_k(k+2), \ldots, s_k(k+\tau)]^T \\ E_k = [e_k(k-1), e_k(k-2), \ldots, e_k(k-\tau)]^T \\ U_k = [u_k(k), u_k(k+1), \ldots, u_k(k+\tau)]^T \\ Y_k = [y_k(k), y_k(k+1), \ldots, y_k(k+\tau)]^T \end{bmatrix} \]

(15)

and \( (\ast) = \text{diag}[\ast, \ast, \ldots, \ast] \), where the notation \( (\ast)' \) denotes the matrices \( \Omega, \Xi, \Psi \) and \( \zeta \).

Thus, the cost function (9) can be rewritten as:

\[ J = \left( s_k(k+1) - s_k(k+1) \right) ^T \Omega \left( s_k(k+1) - s_k(k+1) \right) + U_k ^T (\Xi + \Psi)^T Y_k + Y_k ^T \Psi Y_k \]  

(16)

Optimizing the cost function (10), i.e.,:

\[ \frac{\partial J}{\partial \Theta} = 0 \]  

(17)

The corresponding variable structure control law can be obtained:

\[ U\Theta(k) = (H \Theta(\Theta + \Psi)^T \Theta(GY(k) + E(k) - S_y(k+1)) \]  

(18)

Where:

\[ G = \begin{bmatrix} CA \\ CA^T \end{bmatrix}, H = \begin{bmatrix} CB & 0 & 0 & 0 \\ CAB & CB & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ CAB^\ast B & CAB^\ast B & \ldots & CB \end{bmatrix} \]

From (18), the control input from the time instant \( k \) to \( k+\tau - 1 \) can be calculated. While for prediction, it merely needs the control input of the next instant to reduce the computation complexity. By now, the controlled packet-dropping probability of a packet \( u(t) \) is obtained as well as the input \( p(t) = p_y + u(t) \) of the system (4).
PERFORMANCE EVALUATION OF SMP

We extend the NS-2 simulator with our SMP algorithm to check its performance. SMP is compared with other exist AQM schemes, e.g. REM (Athuraliya et al., 2001) and PI (Sun et al., 2006), in the same conditions to show its superiority. Also, the simulation results will verify the analysis in Section 3.

As a widely adopted numerical illustration extracted from the reference (Hollot et al., 2002), the network topology is the dumb-bell (Fig. 1): there is a single bottleneck link form router 1 to router 2 with capacity 15 Mbps and delay 20 ms. Many source and destination pairs are connected to the router 1 and router 2 and the capacity of access link is 10 Mbps and delay 50 ms. Consider the case where \( Q_b = 175 \) packets, \( T_p = 0.2 \) second and \( C = 3750 \) packets/s (corresponds to a 15 Mb/s link with average packet size 500 bytes). Then, for a load of \( N = 60 \) TCP sessions, we have \( w_p = 15 \) packets, \( p_0 = 0.008 \) and \( R_0 = 0.246 \) seconds. According to the synthesis criteria presented in Section 3, the state feedback matrices can be calculated as following.

We mainly focus on performance measures such as Link Utilization (LU), the ratio of output traffic rate to the service rate of a link, Average Queue Length (AQL) and Packet Loss Rate (PLR) of the congested links.

**Experiment 1 Performance with long-lived TCP flows:** To closely simulate the real conditions in the Internet, this experiment assumes that the bottleneck link of the network is shared only by long-lived TCP flows. We increase the number of TCP flows sharing the bottleneck link from 2 to 300 to yield heavy data flow. The queue lengths of bottleneck link and the performance indexes for three AQM are depicted in Fig. 2. It is observed from the Fig. 1 SMP keeps AQL low and remains constant about 30 when the TCP number increases; (2) although REM keeps AQL slightly lower than SMP, it suffers lower about 9%, (3) while keeping low queue length and high link utilization, SMP has lower PLR than REM and slightly higher PLR than PI (less than 0.9%). From Fig. 3, it shows that SMP has fast transient response (less than 3 s) and keep stable at \( N = 150 \) all the time. From the experimental results, one can find the tradeoff among performance measures; that is, increasing in queue length comes with higher link utilization and lower PLR and SMP can achieve a satisfying balance among different performance measure.

**Experiment 2 Robustness to short-lived TCP flows:** Short-lived TCP flows affect the performance of AQM schemes significantly due to its uncontrollability. In this experiment we investigate the robustness of SMP scheme to short-lived flows. The number of long-lived flows is set to be \( N = 150 \) and the number of short-lived flows increases from \( t = 8 \) to \( t = 50 \) sec and each sends only ten packets to simulate the Web-like flows. From Fig. 4, it is easy to find that traffic self-similarity is caused by aggregated Web flows. The mixture of long-lived flows and short-lived flows makes the traffic burst and more realistic. As the number of short-lived flows increases, the PLR increases and the LU decreases for all the three algorithms. However, it shows in Fig. 4 that SMP has the fluctuation of the queue length less than 10 and keeps the queue length more stable than PI, REM which demonstrates that the robustness of the scheme to the unresponsive short-lived flows is distinctly enhanced. Figure 5 shows that the transient queue length for SMP converges to \( N = 150 \) quickly and at most of the time, the tracking error is less than 10 TCP sessions. While at \( t = 8 \) sec, since the short-lived flows are added, the tracking error suddenly increases to about 140 TCP sessions.

Fig. 1: Dumbbell network topology for NS-2 simulations
Fig. 2(a-c): Performance with long-lived TCP flows under different schemes, (a) Average Queue Length, (b) Link Utilization and (c) Packet loss rate

Fig. 3: Transient response of queue length under SMP scheme with long-lived TCP flows
Fig. 4(a-c): Robustness with short-lived TCP flows under different scheme: (a) Average Queue Length, (b) Link Utilization and (c) Packet Loss Rate

Fig. 5: Transient response of queue length under SMP scheme with short-lived TCP flows
DISCUSSION

Compared with existed congestion control scheme, the presented SMP-based AQM controller possesses the following merits and novelties:

- Unlike the traditional fluid-model-based controller, for the first time, our scheme constructs a delay-free state model by utilizing a particular linear transformation which makes the sliding-surface design of SMP easier and more practical than that of the traditional controller.
- By integrating the MPC technique into the sliding-mode controller designing, for the first time, we successfully derived an SMP-based AQM controller which can decrease the tracking error and optimize the control signal by receding horizon optimization continuously and on-line.
- Different from many existing results, our scheme is based on nonlinear difference equations (5-6) which makes us be capable of using digital technique (software) to implement the control of the network.

CONCLUSIONS

In this study, we analyzed a nonlinear model of TCP by using linearization technology. Based on the sliding-mode and model prediction control approach, by solving the optimal control input, an active queue management scheme is developed to solve the problem of the Internet congestion for the uncertain nonlinear systems with time delays. The existence of sliding-mode enables the closed-loop system having strong robustness and control signal can be optimized continuously for the receding horizon optimization. The stability and robustness of the proposed control scheme are then effectively validated under different network scenarios in NS-2. Simulation results have confirmed that the proposed scheme can provide small queue fluctuations, higher link utilization and low packet loss rate than the traditional controllers.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and to the support of the Key Project of Chinese Ministry of Education (No.212122) and of the Scientific Research Fund of Hunan Provincial Education Department (No.11B031).

REFERENCES


