Computer Communications 33 (2010) 227–239

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Implementation of MPC as an AQM controller

Bahram Marami [a], Mohammad Haeri [b,*]

[a] Electrical and Computer Engineering Department, McMaster University, Hamilton, Canada
[b] Electrical Engineering Department, Sharif University of Technology, Tehran 11365-9363, Iran

ARTICLE INFO

Article history:
Received 18 February 2009
Received in revised form 4 August 2009
Accepted 1 September 2009
Available online 6 September 2009

Keywords:
Active queue management
Model predictive control
Neural network approximator
FPGA timing analysis

ABSTRACT

Utilizing model predictive controllers (MPC) as an active queue management scheme is investigated in this paper. Model based prediction of future output and determining optimized value of the control signal have made MPC as an advanced control strategy in various modern control systems. In this paper a new approach is proposed to alleviate the computational complexity of MPC in order to implement in fast dynamics systems like computer networks. Neural network approximation of MPC as an active queue management (AQM) method implemented here not only has less computational burden with respect to the common MPC approaches, but also results in better performance compare to the well-known AQM methods such as random early detection (RED) and proportional-integral (PI) control. The proposed AQM approach is implemented in a field-programmable gate array (FPGA) system and its feasibility is investigated by timing analysis.

1. Introduction

In recent years, computer networks as a new engineering product has been widely used and a lot of research has been carried out to extend and improve its application. Excessive requests for a limited network resources result in congestion in the computer networks; so that, it leads to aggressive losses of data and underutilization of the resources. Transmission control protocol (TCP) is the dominant transfer layer protocol which is mostly used for controlling the traffic behavior of current internet. During the TCP communication over current drop-tail networks, the TCP sources reduce their sending rate only after detecting packet loss due to queue overflow in data buffers. Detection of packet losses is based on the acknowledgement signal which is sent by receivers when they receive a packet successfully. The time between a packet sending from the source and reception of its acknowledgement is called round trip time (RTT). Since considerable time may pass between the packet drop at the router and its detection at the source, a large amount of data may be lost because the sources continue to send data at a rate that the network cannot support. The other problem of TCP congestion control with drop-tail queues is the global synchronization of sources due to packets drop [1]. To mitigate these problems active queue management (AQM) approaches have been introduced.

Active queue management schemes are implemented in routers to circumvent congestion. These techniques attempt to prevent congestion or buffer overflow by dropping (implicit feedback) or marking (explicit feedback if explicit congestion notification (ECN) is enabled) packets as a function of queue length. Whereupon mentioned feedback, congestion occurrence is reported back to sources and they are instructed to regulate their data sending rates. Two main goals are considered in AQM schemes; first, to reduce the average queue length in buffers and thereby to decrease the end-to-end delay experienced by packets, and second, to utilize network resources efficiently by reducing the packet loss that occurs when queues overflow. However, reducing the queuing delay is in conflict with high utilization of resources. Therefore, there has to be a trade-off between these two main objectives. Random early detection (RED) [2], the earlier and well-known AQM method, randomly drops the arriving packets with a probability proportional to its average queue length in order to detect the baseline congestion. So far, RED is the most prominent and widely studied AQM method which can eliminate the flow-synchronization problem and attenuate the traffic outbursts through the control of the average queue length. After introduction of a mathematical model of TCP dynamical behavior using the theory of stochastic differential equations [3], control theory fundamentals have been used to analyze and develop new AQM schemes. In [4,5] control theoretic approaches have been used to determine the RED parameters. However, because of the time-variant dynamics of computer networks, accurate configuration and tuning of RED parameters in order to obtain a good performance under different congestion scenarios is very difficult [6,7]. Furthermore, RED as an intuitive method is

* Corresponding author. Tel.: +98 21 6615964; fax: +98 21 66023261.
E-mail addresses: bahram@grads.ece.mcmaster.ca (B. Marami), haeri@sina.sharif.edu (M. Haeri).

0140-3664/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2009.09.001

not capable to detect incipient congestion effectively and is very sensitive to network-load variations.

In this paper congestion control of computer networks is considered from control theory point of view and we want to use modern control methods like model predictive control (MPC) as an active queue management strategy. Considering high computational complexity of MPC methods, it will be shown that utilizing MPC based AQM methods are possible. We will prove the feasibility of the proposed AQM approach by timing analysis of their FPGA implementation. The paper goes as follows. Section 2 reviews related researches and describe state of the art of the paper. In Section 3, the dynamics of TCP/AQM networks in congestion avoidance mode is described. Model predictive control approach and its approximation with neural networks are described in Section 4. Simulation results and comparison between the performance of the proposed controller with PI [8] and RED [2] are given in Section 5. In Section 6, computational complexity and timing analysis of the proposed AQM scheme via FPGA implementation is discussed. Finally, the paper is concluded in Section 7.

2. Related researches

In recent years, new congestion control strategies have been proposed from control theory point of view and it has been shown that many types of feedback control systems can extremely improve the performance of existing AQM schemes [5,9–12]. For instance, proportional-integral (PI) controller, one of the simple classic control approaches, is proposed in [8] as an AQM scheme to achieve zero steady-state error, better responsiveness and robustness in buffer queue regulation. Hollot et al. in [8] showed that PI can regulate the queue length better than intuitive methods such as RED. However, its performance deteriorates in presence of load and network parameters variations [9,11]. Other controllers such as: PD [13], PID [14], sliding mode [15], coefficient diagram method (CDM) [16], and adaptive schemes [17] have been designed as AQM methods to regulate the queue length at a certain reference value whereas most of these controllers do not deal with the stochastic behavior and the inherent delay of the computer networks. Although it has been proven that modern controllers have good performance in various areas, most of them are not suitable to be implemented as an AQM method in routers because of their high computational complexities and the fast dynamics of computer networks.

Model predictive controllers (MPC) as modern control strategies have been widely used in different industries in recent years [18]. Using these controllers, the control signal is determined by solving an optimization problem on-line during each sampling interval. These controllers are well-known in dealing with systems with time-variant delay and uncertain model parameters. Here, we apply the model based predictive controller for congestion control in the computer networks. These controllers not only can deal with the uncertain delay time due to the stochastic behavior of the computer networks, but also can regulate the buffer queue length at a desired reference value in presence of network parameters and load variations. The only restriction for applying common MPC approaches to fast dynamic constrained systems such as computer networks is the computational burden. In order to achieve a good response, long prediction and control horizon may be required. This imposes computationally demanding optimization problems, which may not be possible to perform on-line. Another source of high computation cost of MPC is that the nonlinear state space equations representing the model of the plant should be solved numerically in order to determine the predicted future outputs. The later problem may be facilitated by small signal linearization of nonlinear equations around operating point, but this approach is not computationally efficient enough especially when the control and prediction horizons are long. Therefore, to implement the MPC as an AQM method in routers, high-speed processors are needed which imposes great financial burden.

So far, many heuristic algorithms have been developed to decrease computational complexity of MPC in different implementations. Bemporad et al. in [19] showed that the control law in linear time invariant systems is piece-wise linear and continuous for both the finite horizon problems (MPC) and the usual infinite time measures (constrained linear quadratic regulation), so the on-line control computation reduces to a simple evaluation of an explicitly defined piece-wise linear function. Also, a heuristic method which presents a polynomial-time-approximation algorithm as well as a semi-definite programming was developed in [20] to approximate the solution of quadratic programming optimization problems. Furthermore, a framework for embedding model predictive control for special systems-on-a-chip applications was presented in [21]. In recent researches on using MPC in computer networks, Mahramian et al. [22] have used adaptive model predictive controller to schedule differentiated buffers in routers and they proposed two new algorithms to reduce the complexity of their scheduler for implementing in high-speed routers [23]. Furthermore, Bigdeli and Haeri [24] have used predictive functional control (PFC) as a new active AQM method in dynamic TCP networks supporting explicit congestion notification (ECN). Considering the low computational load of PFC, they investigated its robustness and scalability and disturbance rejection.

Some researchers have used the parallel processing property of artificial neural networks to reduce the complexity of MPC in special implementations. In [25] an identified recurrent fuzzy neural networks model has been used to predict future plant outputs. In order to solve the large-scale quadratic programming problems in a massively parallel fashion, a structured neural network was proposed to solve the quadratic programming in the constrained MPC [26]. For using the advantages of MPC in congestion control and overcome the computational burden involved in the on-line optimization, in this paper, the MPC strategy is approximated by a neural network. In other words, the suboptimal results obtained by an MPC are trained to a neural network off-line. Then using this function approximator, control signal is determined at each sampling time on-line. In this way, the computational complexity is reduced considerably with respect to the common MPC schemes. Other advantage of using this approximator is that because of the parallel processing property of neural networks, it is possible to implement the controller in parallel form such as FPGA implementation. Hence, the computation time reduces further.

3. Fluid-flow dynamics of TCP/AQM networks

A mathematical model of TCP dynamics was developed in [3] using fluid-flow and stochastic differential equation analysis. Extensive network simulations have shown that this model truly captures the qualitative behavior of TCP traffic flows, and in recent years, it has been used to develop variety of innovative control theory-based AQM schemes. A simplified version of this model which ignores the TCP timeout mechanism is used in this paper. This model relates the average value of key network variables and is described by the following coupled nonlinear differential equations:

$$\dot{W} = \frac{1}{R(t)} - \frac{W(t)}{2}\frac{W(t-R(t))}{R(t-R(t))}p_d(t-R(t)), \quad W \geqslant 0 \tag{1}$$

$$\dot{q} = -C + \frac{N(t)}{R(t)}W(t), \quad q \geqslant 0 \tag{2}$$

where W is the average TCP window size (packets) and q is the average queue length (packets). Also, N, C, Tp and $R(t) = q(t)/C + Tp$ are