A Nonlinear Flow Control Scheme Under Capacity Constraints[1)]

Yi Fan Zhong-Ping Jiang

(*Department of Electrical and Computer Engineering, Polytechnic University, Six Metrotech Center, Brooklyn* 11201 *U.S.A.*)

(E-mail: yfan01@utopia.poly.edu, zjiang@control.poly.edu)

Abstract We present a nonlinear flow control scheme based on a buffer management model with physical constraints. It extends previous result of Pitsillides et al. in [6] by improving the queue length regulation for better service of network traffics. Besides a single node system, we also address the decentralized control of many cascaded nodes. The proposed discontinuous controller asymptotically regulates the buffer queue length at the output port of a router/switch to a constant reference value, under unknown time varying interfering traffics and saturation constraints on control input and states. Its continuous approximation achieves practical regulation with an ultimate bound on the regulation error tunable by a design parameter.

Key words Congestion control, capacity constraints, buffer management, asymptotic regulation

1 Introduction

The large deployment of communication networks has resulted in many interesting challenges such as service differentiation, QoS, and how to counteract network congestion, etc. To address these challenges, many practising engineers and researchers resort to heuristic and emulation/experiment based approaches. Although solutions have been obtained based on engineering intuition, these approaches involve trials and errors and are not reliable. On the other hand, model based approaches have also been largely explored to address networking problems. For example, linear and nonlinear analysis and control design tools prove effective in ABR traffic control of ATM network[1], congestion control in TCP/IP network[2~4], network performance analysis with time delay[5], and many other issues and references cited in recent literature.

Focusing on the application of nonlinear control theory to solve networking problems, we discuss some results that are closely related to the topic of our paper. In [5] the authors proposed a nonlinear congestion control scheme. The design objective is to regulate the buffer queue length to a constant reference value. Using feedback linearization and robust adaptive control ideas, the authors achieved bounded regulation in the face of unknown time varying interfering traffics.

Our work is in part inspired by the above discussion with particular interest to improve the regulation under disturbances and capacity constraints. Except for considering a single network node[2)] as in [6], we also consider the controller design for several cascaded nodes.

The physical constraints are important issues in many control systems. Numerous results have been established on the stabilization of linear systems with control input saturation constraint[7], while less work is known for general nonlinear systems. We address the constraints on the control input and the state variables. One contribution of our paper is that we give quantitative bounds for choosing controller parameters such that the physical constraints of limited capacity and buffer size are satisfied (using "low gain" feedback). We also specify the explicit conditions (not available in present literature) under which asymptotic regulation can be achieved using saturated control law. The conditions are of two aspects:

1) An upper bound satisfied by the disturbance traffics.

2) A "PE" condition that relates to a physical feature of the network traffics.

We develop a sliding mode type controller design scheme to achieve asymptotic queue length regulation under certain assumptions. To eliminate possible undesirable behaviors such as "chattering" due to the controller discontinuity, a continuous approximation of the discontinuous controller is given

1) Supported in part by the National Science Foundation under grants ANI-0081527 and ECS-0093176

2) By "node", we refer to a router/switch in the network for the rest of the paper

Received April 14, 2004; in revised form July 27, 2004

along with stability analysis. Practical regulation is achieved with the continuous controller and the ultimate bound on the queue length is determined by a design parameter.

The rest of the paper is organized as follows: in Section 2, we introduce a differential equation model that follows from previous work on this subject. Our control objectives are given with practical limitation in mind. The controller design for scalar systems is addressed in Section 3. The control law is further extended to a network with many cascaded nodes in Section 4. Our conclusion is summarized in Section 5 with possible future extensions mentioned.

2 Problem formulation and design objectives

In this work, we discuss the queue length regulation problem at the output port of a network node. We will simply use the term "queue length" when referring to that of the output port buffer. The following model is first introduced in [8]. The authors of [6, 9, 10] continue to consider this model for the purpose of network performance evaluation and control under non-stationary conditions. The model uses the conservation law to establish the buffer queue length dynamic equation, as follows:

$$\dot{x}(t) = -\frac{x(t)}{1+x(t)} \cdot C + \lambda(t) \tag{1}$$

$$x(t) \in [0, x_{buffer}] \tag{2}$$

$$C(t) \in [0, C_{server}] \tag{3}$$

In the above equation, queue size x is taken as the state variable. The assigned capacity C is taken as the control input. These variables are subject to physical constraints (2)~(3) with x_{buffer} denoting size of the buffer and C_{server} the maximum available capacity. λ represents the average incoming traffic rate, which is a disturbance input (Here, no regularity assumption is made on λ.). By conservation law, the first term in the above equation represents the average outgoing traffic rate. Based on the physics of node's service process[8,9], some basic principles when choosing a function $\mu(x)$ for representing the outgoing rate of a network node are:

1) $\mu(0) = 0; \mu(x) \to C$ when $x \to \infty$;

2) $\mu(x)$ is a non-negative, strictly concave and increasing function on $[0, \infty)$.

The particular choice $\mu(x) = \frac{x}{1+x}C$ meets the above requirements, as shown in Fig. 1. The model is proposed for the type of traffics whose arrival and service processes satisfy standard $M/M/1$ queueing system assumptions, as an approximation of the general network traffics[11]. The validity of using $\mu(x) = \frac{x}{1+x}C$ to represent the average departure rate of such traffics has been verified by Filipiak through simulation in [11]. As a result, the operating condition of the above differential equation matches the steady state of an $M/M/1$ queue. The model does not depend on a particular environment such as TCP/IP or ATM network since no other conditions are assumed about the characteristics of the incoming traffic. Fig. 2 shows the structure of a single node (scalar) system.

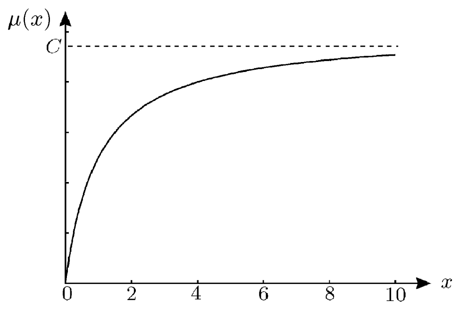


Fig. 1 $\mu(x)$ vs. x

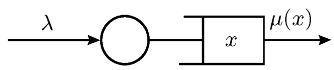


Fig. 2 Single node system

With practical considerations, we assume $0 \leqslant \lambda(t) \leqslant b,\ b > 0$. x_{ref} is introduced as the given reference queue length. It should be chosen such that the switch/router is sufficiently utilized while preserving certain capability to accommodate instantaneous traffic bursts. In practice, an empty or extremely small steady state queue usually leads to link under utilization and is thus undesirable. In this work we suppose the reference value x_{ref} satisfies:

$$\underline{\epsilon} \leqslant x_{ref} \leqslant x_{buffer}$$

The lower bound $\underline{\epsilon} > 0$ may be an arbitrary positive value. Mathematically speaking, our control law is still valid even when $x_{ref} \to 0$. The assumption $x_{ref} \geqslant \underline{\epsilon}$ is due to physical considerations. $\bar{x} := x - x_{ref}$ is introduced to represent the regulation error between the actual queue size and the reference value.

The general objective of controlling network traffics is to suppress congestion and to meet certain performance criteria, including bounded delay, sufficient bandwidth utilization and fairness among traffics, etc. As introduced in [6], these requirements are related to the proper choice of reference value x_{ref}, which is a separate design issue from our work. The design objective of our paper, as well as that of [6] is to accomplish the regulation task when x_{ref} is given. Namely, we want to achieve that $\bar{x} \to 0$ ($x \to x_{ref}$) for all $x(t_0) \in [0, x_{buffer}]$, under the constraint $0 \leqslant C \leqslant C_{server}$, while unknown but bounded time varying disturbance $\lambda(t)$ is present.

When the queue length of a node is below its reference (namely, $x_i(t) < x_i^{ref}$), we consider the node as being under utilized, thus it is unnecessary to assign additional capacity. Based on such consideration, we focus our design mainly on the situation when the node is sufficiently utilized (congested), namely when $x(t) \geqslant x_{ref}$.

3 Controller design for scalar systems

In this section, we use the following control law to achieve the design objective. The same type of controller will be applied to a network composed of many cascaded nodes in the next section. We choose appropriate controller parameters in different cases.

$$C(x) = \begin{cases} 0, & x < x_{ref} \\ C_{server} \cdot sat\left\{ \dfrac{\tilde{C}_d(x)}{C_{server}} \right\}, & x \geqslant x_{ref} \end{cases} \tag{4}$$

where

$$\tilde{C}_d(x) = \frac{1+x}{x}(\alpha\bar{x} + \beta sgn(\bar{x})) \tag{5}$$

where "*sat*" is the commonly used saturation function defined as $sat(y) = \min\{|y|, 1\}sgn(y)$ and "*sgn*" is the standard signum function.

The reason for using the above controller will be clear from the analysis and synthesis shown below. α, β are design parameters to be determined. The subscript "*d*" indicates that this control law is "discontinuous". In the rest of this section, we assume:

Assumption 1. $x_{ref} < \int_{t_0}^{\infty} \lambda(t)\mathrm{d}t \leqslant \infty$ when $x(t_0) < x_{ref}$. For all $t \geqslant t_0, 0 \leqslant \lambda(t) \leqslant b$. $b > 0$ is a constant.

Remark 1. $x_{ref} < \int_{t_0}^{\infty} \lambda(t)\mathrm{d}t \leqslant \infty$ is a "PE"(persistent excitation) requirement. It assumes that there are enough incoming traffics in the long run, such that if $x(t_0) < x_{ref}$, queue length will accumulate until x reaches the reference value.

We complete the controller design as follows. We first propose a "low gain" design. Under certain conditions, the control input is kept from saturation for all initial states. As suggested in [6], since the controller does not deplete the available capacity under this design, the "leftover" capacity can be used to serve instantaneous traffics. To achieve better closed loop system performance such as faster convergence and stronger disturbance rejection, we then propose an improved design using "high gain" ideas. Note that this "high gain" design is not included in our previous work[12].

3.1 "Low gain" feedback without sat

Theorem 1. Consider the system defined in (1)~(3). Suppose $\lambda(t)$ satisfies Assumption 1 where

$$b < \frac{C_{server}}{\left[\dfrac{(x_{buffer} - x_{ref})(x_{buffer}+1)}{x_{ref}^2 + x_{ref}} + 1\right]\dfrac{1}{x_{buffer}} + 1} \tag{6}$$

For all initial queue length $x(t_0) \in [0, x_{buffer}]$ at $t = t_0 \geqslant 0$, x is asymptotically regulated to the reference value x_{ref} using the control law (4)$\sim$(5), if

$$\frac{b}{x_{ref}^2 + x_{ref}} \leqslant \alpha < \frac{C_{server} \cdot x_{buffer}}{(1 + x_{bufer})(x_{buffer} - x_{ref})} - \frac{b}{x_{buffer} - x_{ref}} \tag{7}$$

$$b < \beta \leqslant \min\left\{\alpha(x_{ref}^2 + x_{ref}), \frac{C_{server} \cdot x_{buffer}}{1 + x_{buffer}} - \alpha(x_{buffer} - x_{ref})\right\} \tag{8}$$

Proof. We first show that the parameter ranges (7)$\sim$(8) are valid under the given conditions, followed by showing that such parameter choices guarantee that the control input does not saturate for all initial states. We then analyze the convergence performance of the queue length for the closed loop system.

From (6), we can verify that there exists α satisfying (7). It follows that

$$\min\left\{\alpha(x_{ref}^2 + x_{ref}), \frac{C_{server} \cdot x_{buffer}}{1 + x_{buffer}} - \alpha(x_{buffer} - x_{ref})\right\} > b$$

Thus the choices of α and β are valid. From (6) and (8)

$$\begin{aligned}\frac{\mathrm{d}\tilde{C}_d}{\mathrm{d}x} =& \alpha - (\beta - \alpha x_{ref}) \cdot \frac{1}{x^2} = \alpha - \frac{\beta}{x^2} + \frac{\alpha \cdot x_{ref}}{x^2} \geqslant \\ & \alpha + \frac{\alpha \cdot x_{ref}}{x^2} - \frac{\alpha(x_{ref}^2 + x_{ref})}{x^2} = \alpha\left(1 - \frac{x_{ref}^2}{x^2}\right) > 0, \text{ when } x > x_{ref}\end{aligned}$$

Thus $\tilde{C}_d(x)$ is a monotonically increasing function of x on $(x_{ref}, x_{buffer}]$. The maximum value of $\tilde{C}$ is obtained by equating x to x_{buffer}. $\tilde{C}_d|_{x=x_{buffer}} = C_{max}$ where C_{max} is as follows:

$$C_{\max} := \frac{1 + x_{buffer}}{x_{buffer}}\Big[\alpha(x_{buffer} - x_{ref}) + \beta\Big] \tag{9}$$

From (5),(9) and (8)

$$\tilde{C}_d(x) \leqslant C_{\max} = \frac{1 + x_{buffer}}{x_{buffer}}\Big[\alpha(x_{buffer} - x_{ref}) + \beta\Big] \leqslant C_{server}$$

for all $x \in [x_{ref}, x_{buffer}]$. Thus $C(x)$ is unsaturated. We then analyze the regulation performance of our control law.

a) When $x(t_0) > x_{ref}$

By observing (4)$\sim$(5), we have $C(x) = \tilde{C}_d(x) = 0$ at $x = x_{ref}$. This implies that $x(t) \geqslant x_{ref}$ for all $t \geqslant t_0$ by observation of (1). Consider function $V(x) = \frac{1}{2}\bar{x}^2$. We calculate the derivative of V along the trajectory of the controlled system using $\dot{\bar{x}} = \dot{x}$, plant dynamics (1) and control law (4)$\sim$(5).

$$\dot{V}(t, x) \leqslant -\alpha|\bar{x}|^2 - (\beta - b)|\bar{x}| \leqslant 0 \tag{10}$$

Thus $x(t)$ meets the requirement (2) if $x(t_0) \in [x_{ref}, x_{buffer}]$. Denote by $W(\bar{x}(t)) := \alpha|\bar{x}|^2 + (\beta - b)|\bar{x}|$, using Barbălat's Lemma [10], it can be shown that $W(\bar{x}(t)) \to 0$, thus $|\bar{x}(t)| \to 0$ as $t \to \infty$. (10) ensures that once the trajectory happens to be at the $\{x = x_{ref}\}$, it will be confined at $\bar{x} = 0$ for all future time.

b) When $x(t_0) < x_{ref}$

$\bar{x}(t_0) < 0, C(t) = 0$ as long as $x(t) < x_{ref}$. The dynamic equation is simply $\dot{x} = \lambda(t)$. Thus

$$x(t) = x(t_0) + \int_{t_0}^{t} \lambda(\tau)\mathrm{d}\tau$$

Under the "PE" condition for the incoming traffics, namely $\int_{t_0}^{\infty} \lambda(t)dt > x_{ref}$, there exists $\infty > t_1 > t_0$ such that $x(t_1) \geqslant x_{ref}$ for any $x(t_0) < x_{ref}$. Thus $x(t) \geqslant x_{ref}$ for $\forall t \geqslant t_1$. All the analysis for the case when $x(t_0) \geqslant x_{ref}$ can be extended to the case when $x(t_0) < x_{ref}$.

Combining a) and b), we know that the controller achieves asymptotic regulation of $\bar{x}$ to 0, namely x converges to x_{ref}. □

The above design uses "low gain" feedback control. Under the conditions stated in Theorem 1, the control input does not saturate for all initial states on $[0, x_{buffer}]$. The unused capacity $(C_{server} - C(x))$ can be used to serve instantaneous traffics[6].

A shortcoming of the "low gain" controller is that the available control capacity is not sufficiently utilized and the closed loop system performance (such as convergence rate, disturbance rejection capability) is compromised. We propose another control strategy which improves the utilization of control capacity and achieves better disturbance rejection for the closed loop system.

3.2 "High gain" feedback with sat

Theorem 2. Consider the network model defined in (1)∼(3). Suppose $\lambda(t)$ satisfies Assumption 1 where $0 < b \leqslant \frac{x_{ref}}{1+x_{ref}} C_{server}$. For all initial queue length $x(t_0) \in [0, x_{buffer}]$ at $t = t_0 \geqslant 0$, x is regulated to the reference value x_{ref} asymptotically by the control law (4)∼(5) where $\alpha > 0$ and $\beta \geqslant b$. If in addition $b < \frac{x_{ref}}{1+x_{ref}} C_{server}$ and β is chosen such that $b \leqslant \beta < \frac{x_{ref}}{1+x_{ref}} C_{server}$, the control law is unsaturated when x is close to x_{ref}.

Proof. Before analyzing the regulation of the closed loop system under disturbances, we first introduce a technical lemma to specify the property of the control law (4)∼(5) under saturation constraint.

Due to space limitation, an outline of proof of the lemma is given.

Lemma 1. Define function $g \equiv g(x, \alpha, \beta, x_{ref}, C_{server})$:

$$g = \alpha x^2 + (\alpha + \beta - \alpha x_{ref} - C_{server})x + \beta - \alpha x_{ref} \tag{11}$$

For any given x_{ref}, C_{server} and $\alpha > 0$, if

$$0 < \beta \leqslant \frac{x_{ref}}{1+x_{ref}} C_{server} \tag{12}$$

r_1 and r_2, the solutions of equation $g = 0$, satisfy

$$r_1 \leqslant x_{ref} \leqslant r_2 \tag{13}$$

Furthermore,

1) If $x_{buffer} > r_2$, when $x \in [x_{ref}, r_2]$, the control input $C(x)$, defined in (4), does not saturate; when $x \in (r_2, x_{buffer}]$, $C(x)$ is saturated (namely $C(x) = C_{server}$);

2) If $x_{buffer} \leqslant r_2$, for all $x \in [x_{ref}, x_{buffer}]$, the control input $C(x)$ does not saturate.

Proof of Lemma 1. Apply the parameter choice (12) to equation $g = 0$, we can easily verify that the solutions of the equation are real and satisfy (13). The following two plots in Fig. 3 illustrate the relation of function $g(x)$ and x under the condition $x_{buffer} > r_2$ or $x_{buffer} \leqslant r_2$. According to Fig. 3, we have

1) If $x_{buffer} > r_2$, when $x \in [x_{ref}, r_2]$, $g(x) \leqslant 0$; when $x \in (r_2, x_{buffer}]$, $g(x) > 0$ (see the first plot of Fig. 3).

2) If $x_{buffer} \leqslant r_2$, for $\forall x \in [x_{ref}, x_{buffer}]$, $g(x) \leqslant 0$ (see the second plot of Fig. 3).

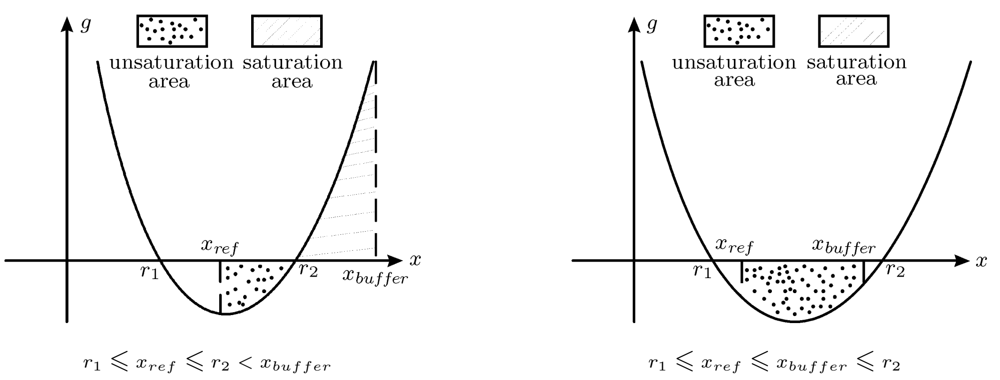


Fig. 3 g vs. x

Using the definition of $g(x)$ in (11) and the above observations, we can arrive at the conclusions of Lemma 1, by means of the definition of $C(x)$ in (4)~(5). □

We then introduce the function $V(\bar{x}(t)) = \frac{1}{2}\bar{x}(t)^2$, and utilize the conclusion of Lemma 1 to prove that asymptotic regulation is achieved for the closed loop system. In the rest of the proof, we only consider the case when $\beta \in [b, \frac{x_{ref}}{1+x_{ref}}C_{server}]$ to simplify the presentation. If instead $\beta \in (\frac{x_{ref}}{1+x_{ref}}C_{server}, \infty)$, we apply similar analysis to arrive at the conclusion.

From the previous analysis, we can assume without loss of generality that $x(t_0) \geqslant x_{ref}$ by observation of (1) and (4)~(5). Thus $x(t) \geqslant x_{ref}$, $\forall t \geqslant t_0$. We consider the two separate situations when $x_{buffer} \leqslant r_2$ or $x_{buffer} > r_2$.

1) If $x_{buffer} \leqslant r_2$, according to Lemma 1, for all $x \in [x_{ref}, x_{buffer}]$, the control input is not saturated. Apparently for all $t \geqslant t_0$, $C(x) = \tilde{C}_d(x)$. We calculate the derivative of $V(\bar{x}(t))$ with respect to t on $[t_0, \infty)$ using $\dot{\bar{x}}(t) = \dot{x}(t)$.

$$\dot{V} \leqslant -\alpha|\bar{x}|^2 - \underbrace{(\beta - b)}_{\geqslant 0}|\bar{x}| \leqslant 0 \tag{14}$$

The above shows that $x(t)$ meets (2) if $x(t_0) \in [x_{ref}, x_{buffer}]$. Furthermore, we define function $W(\bar{x}(t)) := \alpha|\bar{x}|^2 + (\beta - b)|\bar{x}|$. Using Barbălat's Lemma[14], we can prove that when $t \to \infty$, $W(\bar{x}(t)) \to 0$, namely $\bar{x}(t) \to 0$.

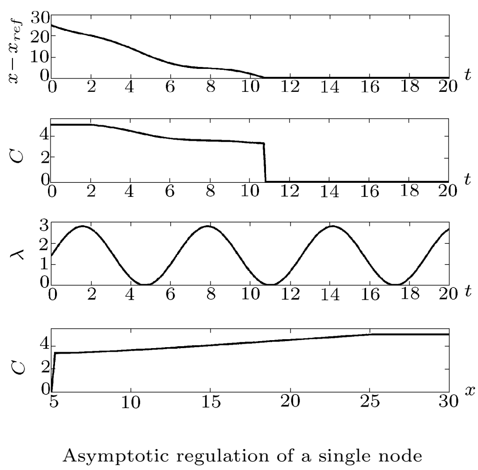
2) If $x_{buffer} > r_2$, we calculate the time derivative of function $V(\bar{x}(t))$ when $x \in (r_2, x_{buffer}]$ or when $x \in [x_{ref}, r_2]$ respectively. When $x \in (r_2, x_{buffer}]$, according to Lemma 1, the control input saturates, namely $C(x) = C_{server}$.

$$\dot{V} = \bar{x} \cdot \dot{\bar{x}} = |\bar{x}|\left(-\frac{x}{1+x}C_{server} + \lambda\right) < |\bar{x}|\left(-\frac{r_2}{1+r_2}C_{server} + \frac{x_{ref}}{1+x_{ref}}C_{server}\right) \leqslant 0 \tag{15}$$

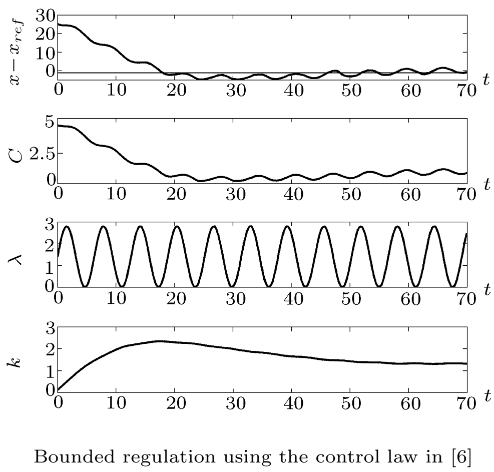
We used the relation $x_{ref} \leqslant r_2$ and that when $x > 0$, $\frac{x}{1+x}$ is an increasing function of x in the above derivation. According to (15), $x(t)$ keeps decreasing until $x \leqslant r_2$.

When $x \in [x_{ref}, r_2]$, the control input is unsaturated, namely $C(x) = \tilde{C}_d(x)$. $\dot{V}(\bar{x}(t))$ satisfies (14). Thus $x(t)$ meets (2) for all $x(t_0) \in [x_{ref}, x_{buffer}]$ and according to Barbălat's Lemma, we conclude that $\bar{x}(t) \to 0$ when $t \to \infty$. □

In the first plot of Fig. 4, we show that the closed loop system achieves asymptotic regulation such that $\bar{x} \to 0$ using the above "high gain" controller. The parameters for the node is $C_{server} = 5$, $x(t_0) = 30$, $x_{ref} = 5$, $x_{buffer} = 30$, $b = 2.8$. Our controller parameters are $\alpha = 0.1$, $\beta = 2.805$. It can be verified that with such choice of controller parameters, the control input does not saturate on $[x_{ref}, r_2]$ while on $(r_2, x_{buffer}]$, the control input saturates. $r_2 = 25.029$. As a comparison, in the second plot of Fig. 4, we show the closed loop system simulation using the control law proposed in [6]. Under the same parameters for the node, the queue regulation error is bounded but not converging. In both simulations, we use sine waves to represent the disturbance traffics.



Asymptotic regulation of a single node



Bounded regulation using the control law in [6]

Fig. 4 Regulation performance of the closed loop system

Remark 2. The above proposed control law is discontinuous at $x = x_{ref}$. This discontinuity raises theoretical as well as practical difficulties. We refer the reader to [13] concerning the existence and uniqueness of solutions for differential equation with discontinuous right hand side. As for practical issues, instead of staying at $x = x_{ref}$ when the trajectory reaches $\{\bar{x} = 0\}$, chattering occurs due to imperfect switching and delay, which is a known phenomenon in sliding mode control. It may excite un-modelled high frequency dynamics and cause instability[14]. This problem is overcome by using a continuous approximation of the discontinuous control law, as shown below.

A continuous controller design

Proposition 1. Consider the system defined in (1)~(3). Suppose $\lambda(t)$ satisfies Assumption 1, where $0 < b \leqslant \frac{x_{ref}}{1+x_{ref}}$. For all $x(t_0) \in [0, x_{buffer}]$ at $t = t_0 \geqslant 0$, the trajectory $x(t)$ meets (2) for all $t \geqslant t_0$ and is ultimately confined to

$$\left\{x_{ref} \leqslant x < x_{ref} + \epsilon\right\} \tag{16}$$

if the following control law is used

$$C = \begin{cases} 0, & x \leqslant x_{ref} \\ \tilde{C}_c(x), & \text{otherwise} \end{cases} \tag{17}$$

$$\tilde{C}_c(x) = \frac{1+x}{x}[\alpha\bar{x} + \beta sat\left(\frac{\bar{x}}{\epsilon}\right)]$$

where $\alpha > 0$ and $\beta \geqslant b$ are constants. ϵ is a design parameter which determines the ultimate bound on queue state x. It is chosen to satisfy

$$0 < \epsilon \leqslant x_{buffer} - x_{ref} \tag{18}$$

and is chosen to be small in practice for a good approximation of the discontinuous control law.

Proof. According to the previous analysis, we consider only the situation when $x(t_0) \geqslant x_{ref}$. Due to that the PE condition in Assumption 1 holds, we can extend the analysis for the case when $x(t_0) \geqslant x_{ref}$ to the case when $x(t_0) < x_{ref}$. As in the discontinuous case, we prove only for the case when $\beta \in [b, \frac{x_{ref}}{1+x_{ref}}C_{server}]$ for simplicity.

First we notice that according to Lemma 1, the solutions to the equation $g = 0$, r_1 and r_2, satisfy (13). Without loss of generality, we only consider the situation when $x_{buffer} > r_2$ and assume that ϵ is chosen small enough such that

$$0 < \epsilon \leqslant r_2 - x_{ref} < x_{buffer} - x_{ref}$$

For any $x(t_0) \geqslant x_{ref}$, it can be shown from (1) and (17) that $x(t) \geqslant x_{ref}$ on $[t_0, \infty)$. Similar to the proof of Lemma 1, we can show that the control law is unsaturated when $x(t) \in [x_{ref} + \epsilon, r_2]$; the control law is saturated when $x(t) \in (r_2, x_{buffer}]$. Consider $V(\bar{x}(t)) = \frac{1}{2}\bar{x}^2(t)$ and calculate $\dot{V}$ along the solutions of the closed-loop system.

a) When $x(t) \in (r_2, x_{buffer}]$, since $C(x(t)) = C_{server}$, along the trajectories of the closed loop system, $\dot{V}$ satisfies (15). This implies that x(or $\bar{x}$) strictly decreases until $x \leqslant r_2$.

b) When $x(t) \in [x_{ref} + \epsilon, r_2]$, the control law is unsaturated, $C(x(t)) = \tilde{C}_c(x(t))$. Simple calculation yields that $\dot{V}$ satisfies (14). This implies that $\bar{x}(t)$ decreases until $x \in [x_{ref}, x_{ref} + \epsilon)$.

The above analysis reveals that the trajectory for any given $x(t_0) \in [x_{ref}, x_{buffer}]$ is bounded and is ultimately confined to the boundary layer defined by (16). □

4 Recursive design for cascaded nodes

In this section, we extend the above controller design for a single node to a system composed of n cascaded nodes, as shown in Fig. 5. n is an arbitrary positive integer. According to the physics of network, the interfering traffics between any two nodes are affected by the activity level of the interfering node, which can be characterized, in part,

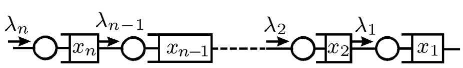


Fig. 5 A network with n cascaded nodes

by the queue length of the interfering node[15]. Under such consideration, we use the following extension of the single node model to represent the system of cascaded nodes[3)].

$$\begin{aligned}\dot{x}_1(t) &= -\frac{x_1(t)}{1+x_1(t)}C_1 + \lambda_1(t, x_2(t)) \\ \dot{x}_2(t) &= -\frac{x_2(t)}{1+x_2(t)}C_2 + \lambda_2(t, x_3(t)) \\ &\vdots \\ \dot{x}_n(t) &= -\frac{x_n(t)}{1+x_n(t)}C_n + \lambda_n(t)\end{aligned} \tag{19}$$

where notations have the same meanings as introduced in model (1)~(3) except for that subscript i denoting the ith subsystem (the i-th node). The model is valid for

$$x_i \in [0, x^{[i]}_{buffer}] \tag{20}$$

$$C_i \in \left[0, C^{[i]}_{server}\right] \tag{21}$$

$$x^{ref}_i \in \left[\underline{\epsilon}, x^{[i]}_{buffer}\right] \tag{22}$$

The disturbance traffics $\lambda_i(t, x_{i+1})$ of each node (subsystem) is unknown and time varying. The function is affected by physical factors such as distance between the nodes, power constraint, whether or not the nodes are directly connected, etc. We assume in this section:

Assumption 2. For each $i = 1, \ldots, n$, at any fixed initial time instant t_0 if $x_i(t_0) < x^{ref}_i$,

$$x^{ref}_i < \int_{t_0}^{\infty} \lambda_i(t, x^{ref}_{i+1})\mathrm{d}t \leqslant \infty \tag{23}$$

For all $t \geqslant t_0$,

$$0 \leqslant \lambda_i(t, x_{i+1}(t)) \leqslant b_i \tag{24}$$

$$\lim_{t\to\infty} \sup \lambda_i(t, x^{ref}_{i+1}) < b_i \tag{25}$$

$b_i > 0, i = 1, \ldots, n$ are constants that satisfy:

$$b_i \leqslant \frac{x^{ref}_i}{1+x^{ref}_i} C^{[i]}_{server}$$

Remark 3. Condition (23) requires that the incoming traffics to node i is "PE" (persistently exciting), in the sense that if the node is initially under utilized, there will be enough incoming traffics in the long run. The utilization of the node, characterized by the queue state x_i, will reach the desired level x^{ref}_i. We will show that under Assumption 2, if $x_i(t_0) < x^{ref}_i$, there exists $t_1 \in (t_0, \infty)$ such that $x_i(t_1) \geqslant x^{ref}_i$.

Theorem 3. Consider a network with n cascaded nodes modelled by (19)~(22). Suppose Assumption 2 holds. If the following control law $C_i, i = 1, \ldots, n$, is used,

$$C_i(x_i) = \begin{cases} 0, & x_i < x^{ref}_i \\ C^{[i]}_{server} \cdot sat\left\{\frac{\tilde{C}^{[i]}_d(x_i)}{C^{[i]}_{server}}\right\}, & x_i \geqslant x^{ref}_i \end{cases} \tag{26}$$

$$\tilde{C}^{[i]}_d(x_i) = \frac{1+x_i}{x_i}\left[\alpha_i \bar{x}_i + \beta_i \mathrm{sgn}(\bar{x}_i)\right] \tag{27}$$

where $\alpha_i > 0$ and $\beta_i \geqslant b_i$, $\forall i = 1, \ldots, n$, the queue length of every node meets (20) and the closed loop system achieves asymptotic regulation. The queue length x_i of every node $i = 1, \ldots, n$ converges to x^{ref}_i in finite time and then stays at $\{x_i = x^{ref}_i\}$. Furthermore, the control input $C_i(x_i)$ meets the requirement (21).

3) The model considers the system of n cascaded nodes where the interfering traffics to node i is dependant of the queue length of its preceding node, $i+1$. It can be extended to the situation where $\lambda_i, i = 1, \ldots, n-1$ is a nonlinear function of the states of all its preceding nodes $i+1, \ldots, n$

Proof. According to conditions (24) and (25) in Assumption 2, we can prove that under the control law (26)$\sim$(27), the queue length x_i of every node, $i = 1, \ldots, n$, meets the requirement (20), using similar analysis as in the proof of Theorem 2. We then prove recursively that the cascaded nodes achieve asymptotic queue length regulation under the saturated control law. We start from node n.

Step 1. For node n, the condition of Assumption 2 reduces to:

1) $x_n^{ref} < \int_{t_0}^{\infty} \lambda_n(t)\mathrm{d}t \leqslant \infty$ if $x_n(t_0) < x_n^{ref}$;

2) $0 \leqslant \lambda_n(t) \leqslant b_n$ for all $t \geqslant t_0$; $\limsup_{t\to\infty} \lambda_n(t) < b_n$.

Since $\limsup_{t\to\infty} \lambda_n(t) < b_n$, there exists t^* (large enough), such that $\lambda_n(t) < b_n$ for all $t \geqslant t^*$. Thus on $[t^*, \infty)$, $\beta_n - \lambda_n(t) \geqslant \varepsilon$ for some $\varepsilon > 0$.

We introduce function $V_n(\bar{x}_n(t)) = \frac{1}{2}\bar{x}_n(t)^2$ to show the convergence of $\bar{x}_n$ on $[t^*, \infty)$. In the rest of the proof, we consider only the case when $\beta_n \in \left[b_n, \dfrac{x_n^{ref}}{1+x_n^{ref}} C_{server}^{[n]}\right]$ to simplify the presentation. If instead $\beta_n \in \left(\dfrac{x_n^{ref}}{1+x_n^{ref}} C_{server}^{[n]}, \infty\right)$, we can use similar argument to arrive at the conclusion. We first notice, similarly with Lemma 1, that $r_1^{[n]}, r_2^{[n]}$, the solutions of equation $g = 0$ (in this case we add superscript (subscript) n to stand for node n), satisfy

$$r_1^{[n]} \leqslant x_n^{ref} \leqslant r_2^{[n]}$$

Furthermore,

1) If $x_{buffer}^{[n]} > r_2^{[n]}$, when $x_n \in \left[x_{ref}^{[n]}, r_2^{[n]}\right]$, the control input $C_n(x)$ does not saturate; when $x_n \in (r_2^{[n]}, x_{buffer}^{[n]}]$, $C_n(x)$ is saturated (namely $C_n(x) = C_{server}^{[n]}$);

2) If $x_{buffer}^{[n]} \leqslant r_2^{[n]}$, for all $x_n \in \left[x_n^{ref}, x_{buffer}^{[n]}\right]$, the control input $C_n(x)$ does not saturate.

From the previous analysis, we can assume without loss of generality that $x_n(t_0) \geqslant x_n^{ref}$. Thus $x_n(t) \geqslant x_n^{ref}$, $\forall t \geqslant t_0$. We consider only the case $x_{buffer}^{[n]} > r_2^{[n]}$ to simplify the presentation. We calculate the time derivative of $V_n(\bar{x}_n(t))$ on $[t^*, \infty)$. We calculate separately when $x_n \in \left(r_2^{[n]}, x_{buffer}^{[n]}\right]$ or $x_n \in \left[x_n^{ref}, r_2^{[n]}\right]$.

Case 1. When $x_n \in (r_2^{[n]}, x_{buffer}^{[n]}]$, the control input is saturated, namely $C_n(x) = C_{server}^{[n]}$.

$$\dot{V}_n = \bar{x}_n \cdot \dot{\bar{x}}_n = |\bar{x}_n|\left(-\frac{x_n}{1+x_n}C_{server}^{[n]} + \lambda_n\right) < |\bar{x}_n|\left(-\frac{r_2^{[n]}}{1+r_2^{[n]}}C_{server}^{[n]} + \frac{x_n^{ref}}{1+x_n^{ref}}C_{server}^{[n]}\right) \leqslant 0$$

We used the relation $x_n^{ref} \leqslant r_2^{[n]}$ and that when $x_n > 0$, $\frac{x_n}{1+x_n}$ is an increasing function of x in the above derivation. We have $x_n(t)$ keeps decreasing until $x_n \leqslant r_2^{[n]}$.

Case 2. When $x_n \in [x_n^{ref}, r_2^{[n]}]$, the control input is unsaturated, namely $C_n(x) = \tilde{C}_d^{[n]}(x)$. $\dot{V}_n(\bar{x}_n(t))$ satisfies

$$\dot{V}_n = -\alpha_n|\bar{x}_n|^2 - (\beta_n - \lambda_n(t))|\bar{x}_n| \leqslant -\alpha_n|\bar{x}_n|^2 - \varepsilon|\bar{x}_n| \leqslant 0$$

The above analysis shows x_n reaches x_n^{ref} in finite time and then stays at $\{x_n = x_n^{ref}\}$.

Step 2. We now analyze the subsystem of node $n-1$. Applying Assumption 2, for $i = n-1$, we find:

1) If $x_{n-1}(t_0) < x_{n-1}^{ref}$ where t_0 is any fixed time instant, $\int_{t_0}^{\infty} \lambda_{n-1}(t, x_n^{ref})\mathrm{d}t > x_{n-1}^{ref}$;

2) For all $t \geqslant t_0$, $0 \leqslant \lambda_{n-1}(t, x_n(t)) \leqslant b_{n-1}, \limsup_{t\to\infty} \lambda_{n-1}(t, x_n^{ref}) < b_{n-1}$.

From Step 1, we know that x_n converges to x_n^{ref} in finite time and then stays at $\{x_n = x_n^{ref}\}$. Since $\limsup_{t\to\infty} \lambda_{n-1}(t, x_n^{ref}) < b_{n-1}$, there exists t^* (large enough) such that for all $t \geqslant t^*$,

$$\lambda_{n-1}(t, x_n(t)) < b_{n-1}$$

Under the control law (26)$\sim$(27), we can prove that queue state x_{n-1} of the $n-1$ subsystem converges to x_{n-1}^{ref} asymptotically. Furthermore, x_{n-1} reaches x_{n-1}^{ref} in finite time and then stays at $\{x_{n-1} = x_{n-1}^{ref}\}$.

Continuing this process, we arrive at Step n where we focus on node 1. By means of similar arguments, we can prove that x_1 meets (20) and converges to x_1^{ref} asymptotically.

We thus conclude that asymptotic regulation is achieved for all n cascaded nodes. Furthermore, x_i converges to x_i^{ref} in finite time and then stays at $\{x_i = x_i^{ref}\}$. □

Remark 4. Using similar analysis as in the proof of Lemma 1 and Theorem 2, we can also show that when $b_i \leqslant \beta_i < \frac{x_i^{ref}}{1+x_i^{ref}} C_{server}^{[i]}$, $C_i(x_i)$ does not saturate on $[x_i^{ref}, \min\{r_2^{[i]}, x_{buffer}^{[i]}\}]$, where $r_2^{[i]}$ is the real solution of equation $g = 0$. g is defined in (11). (In this case since we consider n cascaded nodes, we should add subscript i to every element of (11) to denote the ith node.)

Fig. 6 shows the regulation of two cascaded nodes. The parameters of the two-node system and their respective controller parameters are shown in Table 1.

$$\lambda_1(x_2) = \frac{2C_{server} \cdot x_1^{ref}}{\pi(1 + x_1^{ref})} \arctan(0.3x_2)$$

Note that $\lambda_1(x_2^{ref}) \neq 0$. The traffic disturbances are modelled by sine waves. The controller with the above parameter settings achieves asymptotic regulations of both nodes, as shown in the figure. The controller is unsaturated when the queue lengths are close to the reference values. Continuous approximation can also be used to eliminate possible "chattering" behaviors. It is omitted due to space limitations.

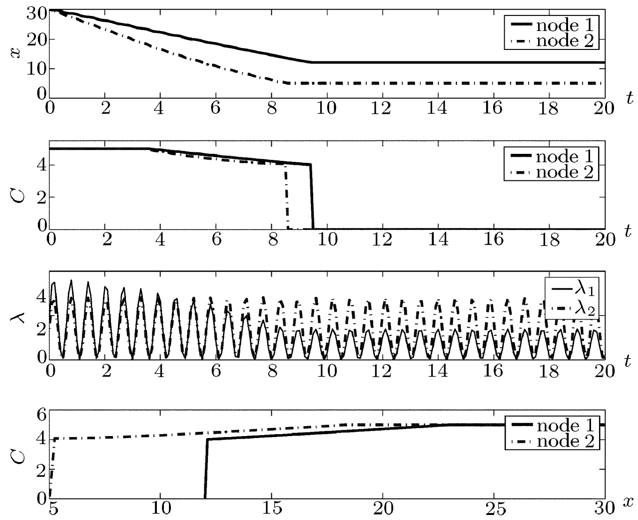


Fig. 6 The regulation of two cascaded nodes

Table 1 Simulation parameters for two cascaded nodes system

	node 1	node 2
Buffer sizes x_{buffer}	30	30
Initial values $x(t_0)$	30	30
Reference values x_{ref}	12	5
Service capacities C_{server}	5	10
Bounds on interferences	$\lambda_1(x_2)$	$b_2 = 2.8$
Controller parameter α	0.1	0.1
Controller parameter β	3.7	2.805

5 Conclusions

Through theoretic analysis and simulation comparison with previous work, we conclude that our sliding mode control law improves the queue regulation, as compared with [6], by achieving asymptotic regulation. Physical constraints on the control input and the state variables have been addressed. The same type of controller has been applied to the system composed of n cascaded nodes. In the future, we will address the following issues:

1) the decentralized control design for the regulation problem of large scale networks with general structure[4)];

4) Note that a preliminary study of decentralized regulation of the large scale system composed of many nodes has been presented in [12] using "low gain" controller

2) other dynamics that affect the traffic patterns in the network such as the dynamics of TCP/IP data sources and wireless mobile units.

References

1 Imer O C, Compans S, Basar T, Srikant R. ABR congestion control in ATM networks, *IEEE Control Systems Magazine*, 2001, **21**(1): 38~56

2 Hollot, C, Misra V, Towsley D, Gong W B. A control theoretic analysis of RED, In: Proceedigns of IEEE Inforcom, 2001, 1510~1519

3 Low S H, Paganini F, Doyle J C. Internet congestion control, *IEEE Control Systems Magazine*, 2002, **22**(1): 28~43

4 Liu S, Basar T, Srikant R. Controlling the Internet: A survey and some new results. In: Proceedings of IEEE Conference on Decision and Control, Maui, Hawaii, 2003, **3**: 3048~3057

5 Alpcan T, Basar T. Global Stability Analysis of End-to-End Congestion Control Schemes for General Topology Networks with Delay, In: Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, 2003, 1092~1097

6 Pitsillides A, Ioannou P, Rossides L. Congestion control for differentiated services using non-linear control theory, In: Proceedings of Sixth IEEE Symposium on Computers and Communications, 2001, 7: 726~733

7 Hu T, Lin Z. Control Systems with Actuator Saturation: Analysis and Design, Birkhauser, Boston, 2001

8 Agnew C. Dynamic modeling and control of congestion-prone systems, *Operations Ressearch*, 1976, **24**(3): 400~419

9 Pitsillides A, Ioannou P, Tipper D. Integrated control of admission, flow rate, and bandwidth for ATM based networks, In: IEEE INFOCOM'96, 15th Conference on Computer Communications,1996, **3**: 785~793

10 Tipper D, Sundareshan M K. Modeling of computer networks for performance evaluation and control under nonstationary conditions, In: Proceedings of INFOCOM '89, 1989, **3**(4): 924~933

11 Filipiak J. Dynamic routing in queueing systems with a multiple service facility, *Operations Research*, 1984, **32**(5): 1163~1180

12 Fan Y, Jiang Z P. Nonlinear flow control for network traffic management with capacity constraints, In: Proceedings of the 2004 American Control Conference, Boston, MA, 2004, 2333~2338

13 Filippov A F. Differential equations with discontinuous right hand side, *Amer. Math. Soc. Transcations*, **42**(2): 191~231

14 Khalil H K. *Nonlinear Systems*, 3rd edition, Prentice Hall, Upper Saddle River, 2002

15 Perkins C, Bhagwat P. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers, In: ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications, 1994, 234~244

16 Brinkmann H W, Klotz E. Linear Algebra and Analytic Geometry, Addison-Wesley Publishing Company, 1971

Yi Fan Received the bachelor degree in 1994 and the master degree in 2001, both from the Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, P.R.China. Currently, Yi Fan is working towards a Ph.D. degree at the Department of Electrical and Computer Engineering, Polytechnic University, U.S.A. His current research interest is the application of modern control theory in communication networks. Yi Fan is an IEEE student member.

Zhong-Ping Jiang Received the bachelor degree in mathematics from the University of Wuhan, Wuhan, P.R.China, in 1988, the master degree in statistics from the Université de Paris-sud, Paris, France, in 1989, and the Ph.D. degree in automatic control and mathematics from the École des Mines de Paris, Paris, France, in 1993. From 1993 to 1998, he held visiting researcher positions in several institutions including INRIA (Sophia-Antipolis), France, the Department of Systems Engineering in the Australian National University, Canberra and the Department of Electrical Engineering in the University of Sydney. In 1998, he also visited several U.S. universities. In January 1999, he joined the Polytechnic University at Brooklyn as an Assistant Professor of Electrical Engineering, and since 2002 he has been a tenured Associate Professor. His main research interests include stability theory, optimization, robust and adaptive nonlinear control, with special emphasis on applications to underactuated mechanical systems and communications networks. He is author of over 60 journal papers in these areas. Currently, Dr. Jiang is a Subject Editor for the International Journal of Robust and Nonlinear Control, an Associate Editor for Systems & Control Letters, and IEEE Transactions on Automatic Control. He served as an Associate Editor for the IEEE CSS Conference Editorial Board from 2000 to 2002. Dr. Jiang is the recipient of a prestigious Queen Elizabeth II Fellowship Award (1998) from the Australian Research Council and a CAREER Award (2000) from the U.S. National Science Foundation. Dr. Jiang is a Senior Member of the IEEE.