Modeling the Internet Congestion Control Using a Smith Controller with Input Shaping*

Saverio Mascolo
Dipartimento di Elettrotecnica ed Elettronica, Politecnico di Bari, 70125 Bari, Italy. **Email: mascolo@poliba.it**

(*)Paper No: CR 2179**:** *Enhanced version of the paper published at the IFAC '03 Workshop on Time-delay Systems and recommended for inclusion in Control Engineering Practice*.

Abstract

The Internet has shown a great capability of endless growing without incurring congestion collapse. The key of this success lies on its TCP/IP congestion control algorithm. In this paper we use control theoretic analysis to model the Internet flow and congestion control as a time delay system. We show that the self-clocking principle, which is known to be a key component of any stable congestion Internet control algorithm, corresponds to implement a simple proportional controller (P) plus a Smith predictor (SP), which overcomes feedback delays that are due to propagation times. Different variants of TCP congestion control algorithms, such as classic TCP Reno or the recent Westwood TCP, can be modeled in a unified framework by proper input shaping of the P+SP controller structure. Finally, we show that controllers that do not implement the Smith predictor, such as proportional (P) controllers or proportional+ derivative+ integral (PID) controllers, provide an unacceptable sluggish system because they do not implement dead-time compensation.

1. Introduction and related works

The stability of the Internet and in particular the prevention of congestion requires that flows use some form of end-to-end congestion control to adapt the input rate to the available bandwidth [1-3],[7-10]. In fact, after the introduction of the Transmission Control Protocol/Internet protocol (TCP/IP), the network was suffering from congestion collapse until congestion control was introduced into the TCP stack in the late 1980s by Van Jacobson [1].

TCP has two feedback mechanisms to tackle congestion: the *flow control* and the *congestion control*. The TCP *flow control* aims at avoiding the overflow of the receiver's buffer and is based on explicit feedback. In particular, the TCP receiver sends to the source the Receiver's Advertised Window, which is the buffer available at the receiver. The TCP *congestion control* aims at avoiding the flooding of the network and is based on implicit feedback such as timeouts, duplicate acknowledgments (DUPACKs), round trip time measurements. In this case the source infers the network capacity using an additive–increase/ multiplicative - decrease (AIMD) *probing* mechanism [11]. The *increase* phase aims at increasing the flow input rate until the network available capacity is hit and a congestion episode happens. The sender becomes aware of congestion via the reception of duplicate acknowledgments (DUPACKs) or the expiration of a timeout. Then it reacts to light congestion (i.e. 3 DUPACKs) by halving the congestion window (*fast recovery*) and sending again the missing packet (*fast retransmit*), and to heavy congestion (i.e. timeout) by reducing the congestion window to one. Both the flow and congestion control implements the self-clocking principle, that is, when a packet exits a new one enters the network. The described mechanisms form the core of the classic Internet congestion control algorithm known as Tahoe/Reno TCP [1,7,10]. It is interesting to notice that these mechanisms still form the main ingredients of all enhanced and successful TCP congestion control algorithms that have been proposed in the literature.

Research on TCP congestion control is still active in order to improve its efficiency and fairness, especially in new environments such as the wireless Internet [12] or the high-speed Internet [24-26], [32-34]. We briefly summarize the most significant modifications that have been proposed up to now [30].

The New Reno feature is an enhancement of Reno that has been proposed to avoid multiple window reductions in a window of data [13]. TCP Vegas estimates the expected connection rate as $cwnd/RTT_m$ and the actual connection rate as $cwnd/RTT$; when the *difference* between the expected and the actual rate is less than a threshold $\alpha>0$, the *cwnd* is additively increased. When the *difference* is greater than a threshold $\beta>\alpha$ then the *cwnd* is additively decreased. When the difference is between α and β, *cwnd* is maintained constant [14]. Vegas TCP provides the basic ideas behind the new Fast TCP congestion control algorithm, which has been recently proposed by researchers at Caltech [32]. In authors' words, "Fast TCP is a sort of high-speed version of Vegas". At the time of this paper Fast TCP is still in a trial phase and authors do not have released any kernel code or *ns-2* implementation. Being based on *RTT* measurements to infer congestion, it could inherit all drawbacks of Vegas, mainly the incapacity to grab bandwidth when coexisting with Reno traffic or in the presence of reverse traffic [33]. TCP Westwood uses an end-to-end estimation of the available bandwidth to adaptively set the control windows after congestion [15,16,33]. Both Vegas and Westwood preserve the standard multiplicative decrease behavior after a timeout.

TCP Santa Cruz proposes to use estimate of delay along the forward path rather than round trip delay and to reach a target operating point for the number of packets in the bottleneck of the connection [17]. The concept of *generalized advertised window* has been proposed in [6] to provide an explicit indication of the network congestion status.

Recently, non linear stochastic differential equations have been proposed to model the TCP dynamics [16], [18-20]. In these models, the dynamics of the expected value of the *cwnd* is mainly expressed as a function of the packet drop probability through a non-linear differential equation. These models, and their linearized ones, have been used to predict the long-term TCP throughput and to design control laws for throttling the packet drop probability of routers implementing Active Queue Management [19]. In particular, the mentioned nonlinear stochastic differential model of the TCP window has been linearized around the equilibrium to derive a transfer function from the packet drop probability to the bottleneck queue length. The linearized model has been employed to design a control law for the packet drop rate aiming at stabilize the queue average length [19]. It is not clear how effective is the model to deal with real-time dynamics of TCP and in presence of multi-bottleneck topologies.

In this paper we propose a general control theoretic framework to model the TCP flow and congestion control along with its variants such as Reno and Westwood. We derive the following main results: (1) different TCP control algorithms can be modeled using the same control structure, which is a proportional controller plus a Smith predictor for dead-time compensation; (2) the Smith predictor plays the fundamental role of overcoming delays in the feedback loop in order to provide a stable and fast control; (3) the Smith predictor provides the control theoretic explanation of the self-clocking principle; (4) different control algorithms such as Reno, Westwood and so on can be explained in terms of different settings of the controller reference input signal.

The work is organized as follows: Section 2 outlines the TCP flow and congestion control algorithm; Section 3 models the dynamic behavior of a generic TCP flow; Section 4 models the TCP flow and congestion control using transfer functions and a proportional controller plus a Smith predictor; Section 5 shows that the Smith predictor enforces the *self-clocking* principle and provides stability; Section 6 models the TCP Reno and Westwood control algorithms by proper shaping of the set point; Section 7 shows that other simpler controllers, such as proportional-integral-derivative (PID), cannot be used in the Internet because they would provide too sluggish behavior; thus, the classic TCP remains the starting point to be considered when designing any new control algorithm; finally Section 8 draws the conclusions.

2. The TCP/IP flow and congestion control

A TCP connection is a virtual pipe between the send socket buffer and the receive socket buffer (see Fig. 1). The TCP has two feedback mechanisms to tackle congestion: the *flow control mechanism* that prevents the sender from overflowing the receiver's buffer, and the *congestion control mechanism* that prevents the sender from overloading the network.

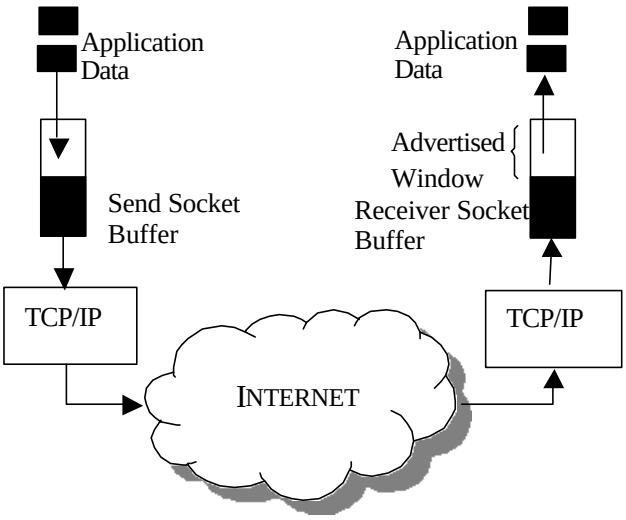

Fig. 1. Schematic of a TCP connection

2.1 The flow control algorithm

The TCP flow control is based on *explicit feedback*. In particular, the TCP receiver sends to the source the Receiver's *Advertised Window*, which is the buffer available at the receiver. Let *MaxRcvBuffer* be the size of the receiver buffer in bytes, *LastByteRcvd* the last byte received and *NextByteRead* the next byte to be read. On the receive side TCP must keep

$$LastByteRcvd - NextByteRead \le MaxRcvBuffer$$

to avoid overflow. Therefore, receiver advertises a window size (*AdWnd*) of

$$AdWnd = MaxRcvBuffer - (LastByteRcvd - NextByteRead)$$

which represents the amount of free space remaining in the receiver buffer. The TCP on the send side computes an *Effective Window W*

$$W = AdWnd - (LastByteSent - LastByteAcked) \quad (1)$$

which limits how much outstanding packets it can send [10].

2.1 The congestion control algorithm

Considering that the network is a "black box" that does not supply any explicit feedback to the source, the issue here is to look for "some measurement" of the network capacity that must be inferred at the end nodes using implicit

feedback received from the networks such as timeout and acknowledgments. Today TCP estimates the best effort capacity of the network using a variable called *congestion window* (*cwnd*). In particular, the TCP learns the appropriate value of the *cwnd* by using an additive-increase/multiplicative-decrease (AIMD) paradigm. The increasing process goes through two phases: the *slow start* and the *congestion avoidance*. During the slow start phase the *cwnd* is exponentially increased until the *slow start threshold* (*ssthresh*) value is reached. This phase is intended to quickly grab available bandwidth. After the *ssthresh* value is reached, the *cwnd* is linearly increased to gently probe for extra available bandwidth. This phase is called *congestion avoidance*. At some point the TCP connection starts to lose packets. After a timeout *cwnd* is drastically reduced to one and the slow start, congestion avoidance cycle repeats. After 3 DUPACKs *cwnd* is reduced by half and the congestion avoidance phase is entered [7,10].

To implement both flow and congestion control, the TCP sender computes the minimum of the congestion window and the advertised window and computes the *Effective Window W* as follows

$$W=MIN(Cwin,AdvWin)-OutstandingPackets \quad (2)$$

where

$$OutstandingPackets=LastByteSent-LastByteAcked$$

represent the in flight packets [10].

3. Modeling a generic TCP flow

In his milestone paper, Van Jacobson (1988) clearly states that: "A packet network is to a very good approximation a linear system made of *gains*, *delays* and *integrators*"[1]. In this paper we propose a detailed model of a TCP/IP connection using (a) integrators to model network and receiver buffers and (b) delays to model propagation times.

A data network is a set of store-and-forward nodes connected by communication links. A generic TCP flow goes through a communication path made of a series of buffers and communication links.

The number of packets of the considered TCP flow that are stored at the generic *i-th* buffer along the communication path is given by the following dynamic equation:

$$x_i(t) = \int_{-\infty}^{t} [u_i(t) - b_i(t) - o_i(t)]dt \quad (3)$$

where $u_i(t)\geq 0$ models the data arrival rate, $b_i(t)\geq 0$ models the data depletion rate, i.e. the used bandwidth, and $o_i(t)\geq 0$ models the overflow data rate, i.e. the data that are lost when the buffer is full and the input rate exceeds the output rate.

The dynamic equation of the generic communication link $(i\text{-}1)$ connecting the $(i\text{-}1)$-*th* buffer to the next (i)-*th* buffer is a pure delay. In particular, letting $b_{i-1}(t)$ be the link input rate at the $(i\text{-}1)$-*th* buffer and $u_i(t)$ be the link output rate at the next (i)-*th* buffer, it results:

$$u_i(t) = b_{i-1}(t - T_{i-1}) \quad (4)$$

where T_{i-1} is the link propagation time.

Starting from the basic equations (3) and (4), we propose to model a generic TCP flow over an IP network as it is shown in Fig. 2. In particular, Fig. 2 shows a functional block diagram made of:

1) The TCP connection receiver buffer of length $x_r(t)$, which is modeled using an integrator with Laplace transfer function 1/s. The receiver buffer receives the inputs $u_r(t)$, $b_r(t)$, $o_r(t)$, which represent the input rate, the depletion rate and the overflow data rate, respectively;
2) The *n-th* buffer that the TCP connection goes through before reaching the receiver buffer, which is modeled using an integrator with output $x_n(t)$. The *n-th* buffer receives the inputs $u_n(t)$, $b_n(t)$, $o_n(t)$, which, again, represent the input rate, the depletion rate and the overflow data rate, respectively. It is important to notice that the depletion rate $b_n(t)$ reaches the next buffer (n+1), which is the receiver buffer, after the propagation time T_n, *i.e.* $u_r(t)= b_n(t\text{-}T_n)$. Moreover, it should be noted that the input rate $u_n(t)$ is equal to the depletion rate $b_{n-1}(t)$ at the previous (n-1)-th buffer, i.e. $b_{n-1}(t\text{-}T_{n-1})= u_n(t)$, where T_{n-1} is the propagation time from the (n-1)-th buffer to the n-th buffer. Depletion rates are unpredictable because they model the best effort bandwidth available for a TCP connection when going over statistically multiplexed IP network.

The series of buffers shown in Fig. 2 can be recursively augmented both in the left direction, to model up to the first buffer node encountered by the TCP connection, and in the right direction to model buffers $n+j$, with $j=2,p$ encountered by ACK packets when going back from the receiver to the sender.

By considering a closed surface that contains the TCP path going from the first to the last buffer modeled by a set of integrators indexed from 1 to $n+p=m$, where the m-th integrator models the last buffer encountered by the TCP along the connection round trip, we can invoke the flow conservation principle for the unique input rate, which is the TCP input rate $u_1(t)$, and the output rates that are: (a) $b_m(t)$, which models the bandwidth used by the TCP connection, i.e. the best-effort bandwidth as viewed by the considered TCP flow through the ACK stream; and (b) the

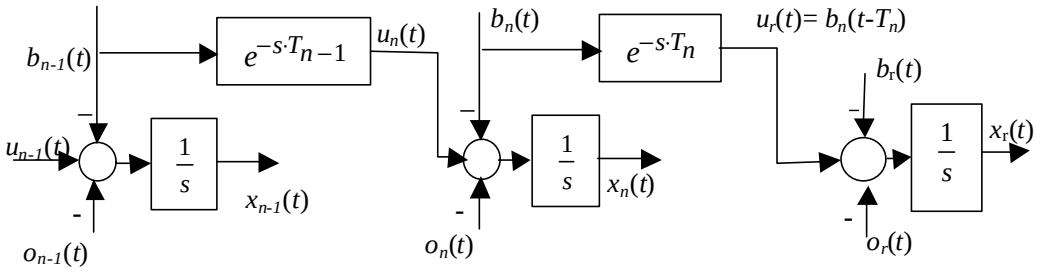

Fig. 2 Dynamic block diagram of a generic TCP/IP flow

overflow rates $o_i(t)$, for i=1,m, which represent packets that are lost at each buffer along the path connection.

In equations, we can write the number $x(t)$ of packets belonging to the considered TCP flow and stored into the network by adding packets stored at each buffer along the path:

$$x(t) = \sum_{i=1}^{m} x_i(t) \qquad (5)$$

Substituting (3) in (5) and considering the (4) it turns out

$$x(t) = \int_{-\infty}^{t} [u_1(t) - b_m(t) - \sum_{i=1}^{m} o_i(t) - \sum_{i=1}^{m-1} (b_i(t) - b_i(t - T_i))]dt$$

that can be rewritten as

$$x(t) = \int_{-\infty}^{t} [u_1(t) - b_m(t) - \sum_{i=1}^{m} o_i(t)]dt - \sum_{i=1}^{m-1} \int_{t-T_i}^{t} b_i(t)dt \qquad (6)$$

Eq. (6) states that the network storage is equal to the integral of the TCP input rate $u_1(t)$ minus the output rate $b_m(t)$ leaving the last buffer of the path, minus the sum of the overflow rates $o_i(t)$, minus the sum of packets that are in flight over each link i.

Since the TCP implements an end-to-end congestion control that does not receive any explicit feedback from the network, it is not possible for the controller to know terms in (6). Thus, we consider the sum of the in *flight packets* plus the *stored packets*, which we call the total network storage x_t:

$$x_t(t) = x(t) + \sum_{i=1}^{m-1} \int_{t-T_i}^{t} b_i(t)dt = \int_{-\infty}^{t} [u_1(t) - b_m(t) - \sum_{i=1}^{m} o_i(t)]dt$$

and the sum of overflow rates o_t:

$$o_t(t) = \sum_{i=1}^{m} o_i(t)$$

Thus, we can write

$$x_t(t) = \int_{-\infty}^{t} [u_1(\tau) - b_m(\tau) - o_t(\tau)]d\tau \qquad (7)$$

By considering that the TCP establishes a "circular flow", i.e. that the data input rate comes back to the sender as an ACK rate, it can be said that $b_m(t)$ models the rate of ACK packets. Thus we can write:

$$b_m(t) = u_1(t - T) - o_t(t) \qquad (8)$$

which says, in mathematical words, that the ACK rate is equal to the input rate, delayed by the round trip time, minus the loss rate. By substituting (8) in (7) it turns out:

$$x_t(t) = \int_{-\infty}^{t} [u_1(\tau) - u_1(\tau - T)]d\tau = \int_{t-T}^{t} u_1(\tau)d\tau \qquad (9)$$

Equation (9) states that the network total storage is equal to the integral of the input during the last round trip time T.

4. Modeling the TCP flow and congestion control

This section aims at showing that the closed loop control system depicted in Fig. 3 implements both the TCP flow and congestion control. In details, the following variables and blocks are shown:

(1) The receiver queue length x_r and the receiver capacity r_1 provide the term $r_1 - x_r$ (i.e. the *Advertised Window*), which reaches the sender after the propagation time T_{fb} that is modelled in the Laplace domain by the transfer function $e^{-sT_{fb}}$;
(2) The set point $r_2(t)$ represents a threshold for the total network storage, which is modeled by the queue $x_t(t)$;
(3) The minimum block takes the minimum between the *Advertised Window* and $r_2(t)$;

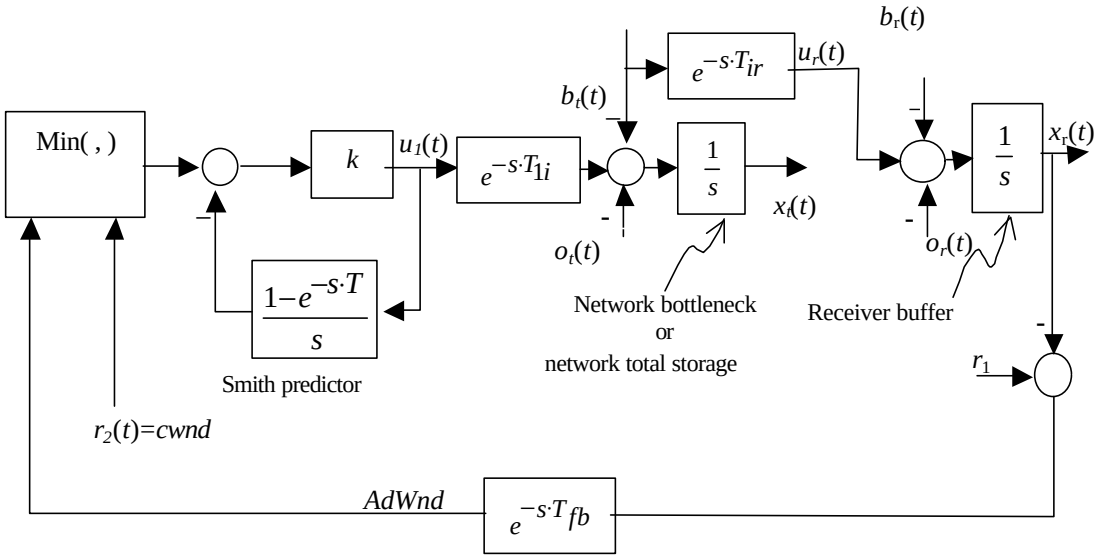

Fig. 3: Functional block diagram of the TCP flow and congestion control

(4) Delays T_{1i} and T_{ir} model the time delay from the sender to the generic node i and from the node i to the receiver, respectively; the forward delay from the sender to the receiver is $T_{fw} = T_{1i} + T_{ir}$;

(5) The controller transfer function

$$G(s) = \frac{k}{1 + \frac{k}{s}(1 - e^{-sT})}, \qquad (10)$$

which contains the proportional gain k and the Smith predictor $(1 - e^{-sT})/s$, where T is the round trip time sum of the *forward* delay T_{fw} and the *backward* delay T_{fb}. Notice that the role of the Smith predictor is to overcome the delay T, which is inside the feedback loop and is harmful for the stability of the closed-loop control system (Mascolo, 1999).

Notice that the buffer x_t in Fig. 3 can model both the total network storage of packets but also it can model the generic buffer x_i that is the bottleneck of the TCP connection at time t; moreover, a moving bottleneck is easily captured by the model through delays T_{1i} and T_{ir} where i is the generic moving bottleneck.

In order to show that the block diagram in Fig. 3 models the TCP/IP flow and congestion control, first we will assume that the bottleneck is at the receiver and then that the bottleneck is inside the network.

4.1 The TCP Flow Control

By assuming that the bottleneck is at the receiver, it results: $\min(Adwnd, r_2(t)) = Adwnd$, $u_r(t) = u_1(t-T_{fw})$ and $o_t(t)=0$. In other words, the connection is constrained by the receiver, and the input rate reaches the receiver after the forward delay without network queuing, i.e. $b_t(t)=u_1(t-T_{1i})$. Under these conditions, Fig. 3 can be transformed into Fig. 4 that models the TCP flow control. The following propositions can be shown.

Proposition 1: The Smith controller (10) implements the TCP flow control equation (1).

Proof: To find the input rate $u_1(t)$ computed by the TCP sender we use standard Laplace techniques, that is, we compute the Laplace transform of the input rate:

$$U_1(s) = [R_1(s) - X_r(s)]e^{-sT_{fb}} \frac{k}{1 + k\left(\frac{1 - e^{-sT}}{s}\right)}$$

that can be written as

$$U_1 = -kU_1\left(\frac{1 - e^{-sT}}{s}\right) + k[R_1 - X_r]e^{-sT_{fb}}$$

By transforming back to time domain it results:

$$\frac{u_1(t)}{k} = r_1(t - T_{fb}) - x_r(t - T_{fb}) - \int_{t-T}^{t} u_1(t)dt \qquad (11)$$

By considering that

$$r_1(T - T_{fb}) - x_r(T - T_{fb}) = \text{Advertised window}$$

and that

$\int_{t-T}^{T} u_1(t)dt = \text{Outstanding packets}$

Equation (11) gives the classic window-based flow control equation (1), where $W = u_1(t)/k$. By considering that $u_1(t) = W/T$ relates the rate and the window of a window-based control, it results $1/k=T$.

Notice that the *outstanding packets* automatically take into account the round trip time T that in general can be time varying due queuing delays. In the case of flow control T is, to a very good approximation, constant since there is no congestion inside the network which implies that network queuing delay is zero and round trip time is pure propagation delay.

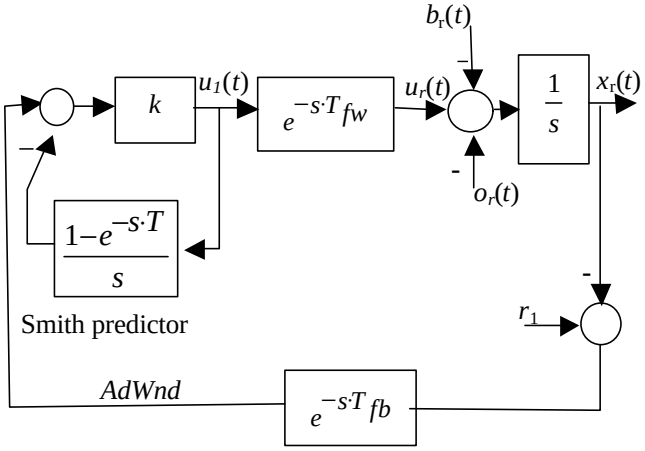

Fig. 4: Functional block diagram of the TCP flow control

Proposition 2: The TCP flow control equation (11) guarantees that the receiver queue is always bounded by the receiver capacity $\bar{r}_1$, i.e.:

$$x_r(t) < \bar{r}_1 \text{ for any } t$$

Proof: The queue length can be computed by exploiting the superposition property of linear systems. In particular, it is easy to compute the input-output transfer function from $R_1(s)$ to the receiver queue length $X_r(s)$ that is:

$$\frac{X_r}{R_1} = \frac{k}{k+s}e^{-sT}$$

and the transfer function from $B_r(s)$ and $O_r(s)$ to $X_r(s)$ that is:

$$\frac{X_r}{O_r + B_r} = -\frac{1}{s} + \frac{k}{s(s+k)}e^{-sT} = -\frac{1}{s} + \frac{e^{-sT}}{s} - \frac{e^{-sT}}{s+k}$$

By assuming $r_1(t) = \bar{r}_1 \cdot 1(t)$, where $\bar{r}_1$ is the receiver buffer capacity and $1(t)$ is the step function that models a connection starting at $t=0$, there results: $R_1(s) = \frac{\bar{r}_1}{s}$. By exploiting the superposition property of linear systems and by transforming back to time domain there results:

$$x_r(t) = L^{-1}\left\{\frac{\bar{r}_1}{s}\frac{k}{s+k}e^{-sT}\right\} - L^{-1}\left\{\frac{B_r + O_r}{s+k}e^{-sT}\right\} -$$

$$- \int_{t-T}^{t}[b_r(t) + o_r(t)]dt$$

which satisfies the condition

$$x_r(t) \le L^{-1}\left\{\frac{\bar{r}_1}{s}\frac{k}{s+k}e^{-sT}\right\} = \bar{r}_1 \cdot \left(1 - e^{-k(t-T)}\right) \cdot 1(t-T) < \bar{r}_1$$

since $o_r(t)$, $b_r(t)$ are always non negative. This concludes the proof.

Lemma 1: Proposition 2 guarantees $o_r(t)=0$ for any t.

Proof: Proposition 2 proves that the receiver queue length is always upper bounded by the receiver queue capacity, which implies that receiver overflow is always avoided, i.e. $o_r(t)=0$ for any t.

4.2 The TCP Congestion Control

By assuming that bottleneck is localized inside the network, there results $\min(\text{Adwnd}, r_2(t)) = r_2(t)$ and we can ignore the outer feedback loop. Therefore, Fig. 3 can be transformed into the equivalent one shown in Fig. 5, which models the TCP congestion control.

Proposition 3: The Smith controller (10) implements the TCP congestion control equation (2).

Proof: By assuming that the bottleneck is inside the network, there results: $\min(\text{Adwnd}, r_2(t)) = r_2(t)$. From Fig. 5, the output of the Smith predictor in the Laplace domain is:

$$Q(s) = U_1(s)\frac{1-e^{-sT}}{s}$$

By transforming back to time domain it results:

$$q(t) = \int_{t-T}^{T} u_1(t)dt = outstandingpackets$$

Therefore the output of the controller is:

$$u_1(t) = k\left(r_2(t) - outstandin\,gpackets\right) \qquad (12)$$

that can be rewritten as

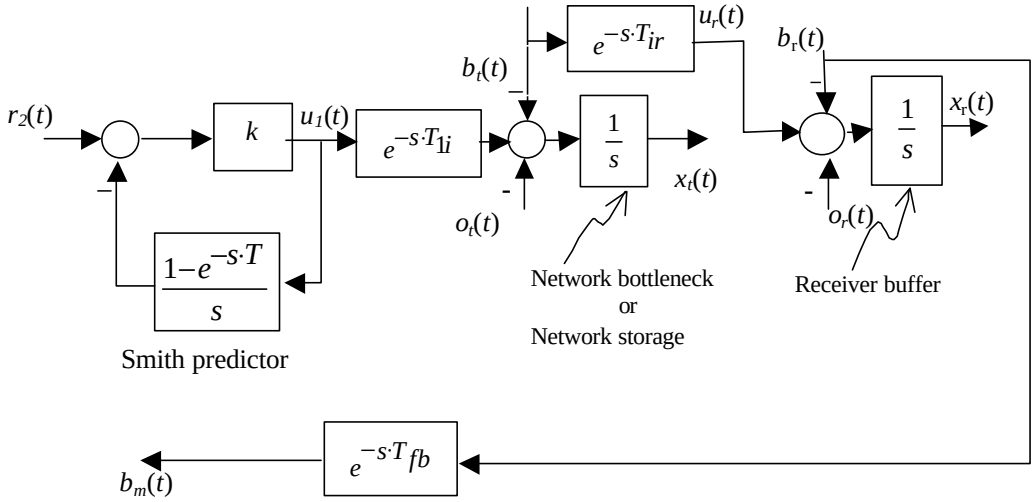

Fig. 5 Functional block diagram of the TCP congestion control

$$\frac{u_1(t)}{k} = r_2(t) - outstandingpackets \qquad (13)$$

Equation (13) gives the classic window-based congestion control equation (2), where $W = u_1(t)/k$, and $r_2(t)$=*cwnd*. This concludes the proof.

Remark 1: It should be noted that (12) and (13) are the rate-besed and window-based versions of the same control equation.

Proposition 4: The TCP congestion control equation (13) guarantees a total network storage x_t that is always bounded by the threshold $r_2(t) > 0$, i.e.:

$$x_t(t) \le r_2(t) \text{ for any } t$$

Proof: From (9), the total network storage is:

$$x_t(t) = \int_{t-T}^{t} u_1(t)dt = q(t)$$

Since $u_1(t)$ and $q(t)$ are always non negative, and $r_2(t)$ is strictly positive, from the control law:

$$u_1(t) = k\left(r_2(t) - \int_{t-T}^{T} u_1(t)dt\right)$$

it turns out $r_2(t) \ge q(t) = x_t(t)$, which concludes the proof.

Lemma 2: If a TCP flow finds in each buffer it goes through a space of c_i packets, where $c_i > r_2(t)$ for any t and i, then the Proposition 4 guarantees $o_t(t)$=0 for any t.

Proof: The proof follows from Proposition 4, which proves that $x_t(t) \le r_2(t)$, and assumptions of Lemma 2.

5. The self-clocking principle

Based on the theoretical control framework outlined in the previous section, we are now ready to show that the *self-clocking* principle can be theoretically explained using the Smith predictor. It is largely known that the *self-clocking* principle is a key feature of the TCP congestion and flow control [1]. This has been recently recognized also in the context of Transport Friendly Rate Control (TFRC) algorithms [21], where it has been shown that algorithms that do not employ the self-clocking principle may exhibit a huge settling time, that is, they may require many RTTs to adapt the input rate to the bandwidth available in the network. As a consequence, to overcome the disastrous effects due to the violation of the self-clocking principle, the original TFRC has been enhanced with the self-clocking mechanism. In this section we show that the *self-clocking* can be mathematically interpreted as the effect of the Smith predictor branch.

Proposition 5. The Smith predictor branch $\left(1-e^{-sT}\right)/s$ enforces the *self-clocking* principle.

Proof: By transforming back to time domain the quantity $q(s) = U_s(s)\left(1-e^{-sT}\right)/s$ it results

$$q(t) = \int_0^t u_s(t)dt - \int_0^{t-T} u_s(t)dt = \int_{t-T}^{t} u_s(t)dt ,$$

where $q(t)$ represents the data that have been sent since the last round trip time T up to now, i.e. the outstanding packets. When the time t advances of Δ, i.e. at time $t+\Delta$, the amount of data

$$q_{acked} = \int_{t-T}^{t-T+\Delta} u_s(t)dt$$

are acknowledged so that the control equation (3) or (5) can send this amount of data in the time interval $[t, t+\Delta]$, that is, the self-clocking principle is enforced.

6. Modeling Reno or Westwood TCP by input shaping

In this section we show that the dynamic model depicted in Fig. 5 is able to model successful variants of TCP congestion control, such as for example Tahoe/Reno [1] or the recent Westwood TCP [15]. Other TCP variants, such as Vegas or Santa Cruz, could also be modeled in the same unified framework.

We have seen that the congestion control algorithm aims at estimating the available bandwidth using a probing mechanism. The classic TCP probing mechanism, which is currently used in all successful variants of the TCP such as Tahoe/Reno, New Reno or Westwood, comprises two mechanisms: the *slow-start* phase, which exponentially increase the congestion window up to the *ssthresh*, and the *congestion avoidance* phase which linearly increase the *cwnd* when $cwnd \geq ssthresh$. Now we show that both these mechanisms can be modeled in the control theoretical framework reported in Fig. 5 by properly shaping the controller input $r_2(t)=cwnd$.

6.1 The Reno Algorithm

The TCP Reno *slow-start* phase can be modeled by setting the reference input $r_2(t)$ as follows:

$$r_2(t) = r_o \cdot 2^{\frac{t}{T}} \qquad while \quad r_2(t) < ssthresh$$

where the initial window r_0 is generally equal to 1 or 2 [19]. TCP Reno enters the *congestion avoidance* phase when $r_2(t)=ssthresh$ at $t_1 = T\log_2(ssthreh - r_0)$. This phase can be modelled by setting the reference input $r_2(t)$ as follows:

$$r_2(t) = ssthresh + \frac{t-t_1}{T} \qquad when \quad r_2(t) \geq ssthresh$$

The TCP probing phase ends when 3 DUPACKSs are received or a timeouts happens, which indicate that the network capacity has been hit. In these cases the *cwnd* behavior can be modeled using the following settings for $r_2(t)$:

After a timeout at t_k

$$ssthresh = \max(\frac{r_2(t)}{2}, r_0)$$

$$r_2(t) = r_0$$

$$r_2(t) = r_0 \cdot 2^{\frac{t-t_k}{T}} \qquad if \quad r_2(t) < ssthresh$$

$$r_2(t) = ssthresh + \frac{t-t_k}{T} \qquad if \quad r_2(t) \geq ssthresh$$

After 3 DUPACKs at t_k

$$ssthresh = \max(\frac{r_2(t)}{2}, r_0)$$

$$r_2(t) = ssthresh + \frac{t-t_k}{T}$$

6.2 The Westwood algorithm

TCP Westwood employs the same probing mechanism of Reno. It differs from Reno because of the behavior after congestion. In fact, Westwood sets the *cwnd* and *ssthresh* using an end-to-end estimate of the network bandwidth $b_m(t_k)$ available at time of congestion. In particular, the Westwood TCP window behavior after congestion can be modeled as follows:

After a timeout at t_k

$$ssthresh = b(t_k) \cdot RTT_{\min}$$

$$r_2(t) = r_0$$

$$r_2(t) = r_o \cdot 2^{\frac{t-t_k}{T}} \qquad if \quad r_2(t) < ssthresh$$

$$r_2(t) = ssthresh + \frac{t-t_k}{T} \qquad if \quad r_2(t) \geq ssthresh$$

After 3 DUPACKs at t_k

$$ssthresh = b(t_k) \cdot RTT_{\min}$$

$$r_2(t) = ssthresh + \frac{t-t_k}{T}$$

7. Why a PID controller is not efficient to control the Internet

We have seen that the Internet flow and congestion control problem reduces to the issue of controlling an integral mode with a time-delay in cascade. It can be said that, in general, congestion control in data networks consists of controlling a time-delay system.

The proportional-integral-derivative (PID) controller is by far the most common control algorithm and performs satisfactorily well in many practical cases [23].

In this section we show that a standard PID cannot satisfactorily control a data network such as the Internet since in order to provide stability for the closed-loop system it is necessary to use a low gain that turns out an unacceptable sluggish system.

To start the discussion we consider the simple proportional controller k shown in Fig. 6. In order to study the stability of this system, we invoke the Nyquist stability criterion. To the purpose, the polar diagram of the open loop-transfer function $\frac{k}{s}e^{-sT}$ is depicted in Fig. 7, which also shows the vertical asymptote of abscise $-kT$ and the circle with unity magnitude.

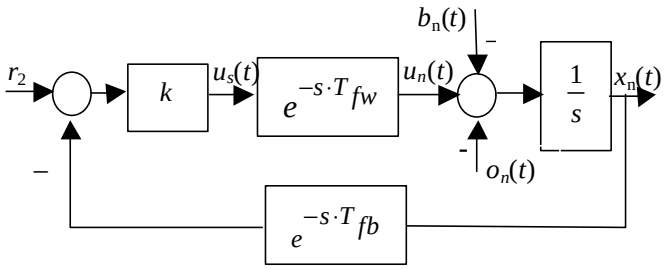

Fig. 6: A data flow controlled by a proportional controller

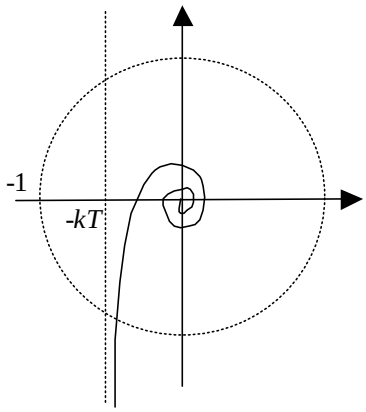

Fig. 7 The Nyquist plot of the open loop-transfer function

For the Nyquist stability criterion, the polar plot must not encircle the point -1. The crossover frequency w_0 (frequency for which the magnitude of the loop frequency response is unity) is obtained from the equality

$$\left|\frac{k}{jw_0}e^{-jw_0T}\right|=1,$$

which turns out $w_0 = k$.

The closed-loop system is stable if the phase margin

$$M_F = p - p/2 - w_0T = p/2 - kT$$

is positive, which turns out the stability condition:

$$k < k_u = p/2T \qquad (14)$$

where k_u is the ultimate gain.

The stability condition (14) states that the gain k must be lower and lower with increasing round trip time T. As a consequence, the proportional controller provides a too sluggish closed-loop behaviour in the case of data networks which are characterized by large propagation delays, such as in the case of high-speed networks, wide area networks or satellite connections.

The use of a PD controller does not help much to improve the promptness of the controller [23]. In fact, when using using a PD, the open-loop transfer function becomes $\frac{k(1+st_D)}{s}e^{-sT}$. The crossover frequency is now $w_1 = k/\sqrt{1-(kt_D)^2} > w_0$. Thus, even though the derivative action adds the positive contribute $arctg(w_1t_D)$ to the phase margin, it must be considered that at the new cross-over frequency w_1 the negative contribute to the phase margin due to the time delay is now augmented of $(w_1 - w_0)T$. Thus, in the presence of large delay T, the lead action of a PD controller may not be useful or may be even pejorative of the stability margin because it may happen that $(w_1 - w_0)T > arctgw_1t_D$.

Finally, the integral action of the standard PID controller is surely not recommended for the system we are considering because it would reduce the phase margin of $p/2$ at any frequency thus making the system stability even more critical.

To get a further insight we compare the proportional controller with a proportional controller plus a Smith predictor using computer simulations. We consider a connection with T=1000 units of time. From (14), the ultimate gain is k_u=0.001571. Following the Ziegler-Nichols rules, we choose the proportional gain $k=k_u/2$=0.000785. We have also tried the Ziegler-Nichols rules for tuning a PI and a PID controller but in this case we have found that they do not provide a stable system, which confirms that it is not easy or efficient to control a system with large delays using a PI or a PID controller.

We set the reference signal r(t)=10000·1(t), which corresponds to setting a queue threshold of 10000 packets at the initial time t=0. We assume an available bandwidth of 4 pkts/unit of time. This pattern is very appropriate to test the promptness of the congestion control algorithm in matching the time-varying available

bandwidth. Fig. 8 shows the input rate dynamics obtained using a proportional controller (P) and a proportional controller plus a Smith predictor (P+SP) both with k=0.000785. It shows that the input rate obtained using the Smith predictor is much faster in reaching the steady state value of 4 pkts/unit of time. Moreover the Smith predictor provides a much smaller band of oscillation for the input rate. Fig. 9 shoes that the Smith predictor provides a much more smaller queue length that is a very important feature for networks since it means much smaller queuing delays.

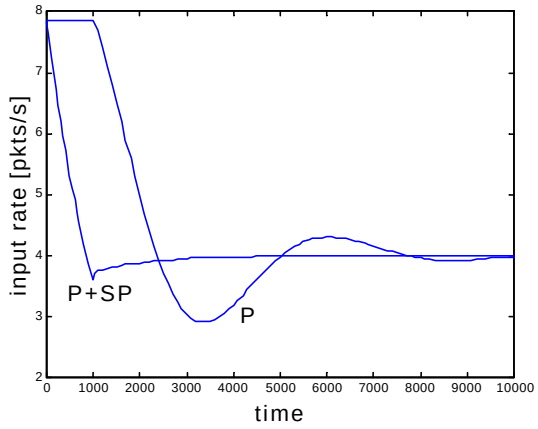

Fig. 8: Input rate using a P or P+SP controller

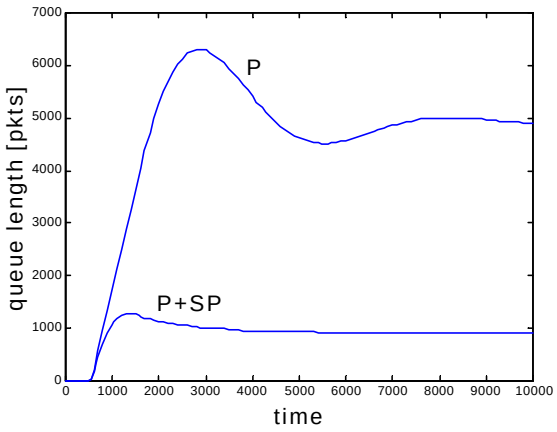

Fig. 9: Queue length using a P or a P+SP controller

It is worth noting that another important advantage of the Smith predictor controller is that system dynamics can be made faster by increasing the proportional gain *k* without risking instability.

To conclude, we look at results of this section from the perspective provided by the Proposition 5. In particular, we observe that a controller without a Smith predictor, such as a PID, does not implement the self-clocking principle. Therefore, as it has been also noted in [21], it may exhibit a huge settling time, that is, it may require many RTTs to adapt the input rate to the bandwidth available in the network.

8. Conclusions

In this paper the TCP congestion and flow control have been modeled as a time-delay system controlled using dead-time compensation. We have shown that a proportional controller plus a Smith predictor provides an exact model of the Internet flow and congestion control. In particular we have shown: (1) enforcing the self-clocking principle corresponds to implement the Smith predictor; (2) the Smith predictor controller guarantees stability and provides efficient congestion control; (3) different TCP congestion control algorithms, such as the classic TCP Reno or the recent Westwood TCP, can be modeled by shaping the reference input. Finally we have shown that controllers that do not implement the Smith predictor, such as PID controllers, provides an unacceptable sluggish systems because they do not implement the self-clocking principle.

References

[1] V. Jacobson, "Congestion Avoidance and Control," ACM Computer Communications Review, 18(4): 314 - 329, August 1988.

[2] S. Mascolo, "Congestion control in high-speed communication networks using the Smith principle", Automatica, vol. 35, no. 12, dec. 1999.

[3] S. Mascolo,"Smith's Principle for Congestion Control in High Speed Data Networks", IEEE Trans. on Automatic Control, vol. 45, no. 2, Feb 2000, pp. 358–364.

[4] O. Smith, "A Controller to Overcome Dead Time", *ISA J.*, vol.6, no.2, pp.28-33, 1959.

[5] K. J. Åström , B. Wittenmark, *Computer controlled systems*, Prentice Hall, Englewood Cliffs, N. J., 1997.

[6] M. Gerla and R. Locigno and S. Mascolo and R. Weng, "Generalized Window Advertising for TCP Congestion Control", European Transactions on Telecommunications, no. 6, Nov/Dec. 2002.

[7] M. Allman, V. Paxson, W. R. Stevens, "TCP congestion control," RFC 2581, April 1999.

[8] D. Clark, "The design philosophy of the DARPA Internet protocols," In Proceedings of Sigcomm' 88 in ACM Computer Communication Review, vol. 18, no. 4, pp. 106 - 114, 1988.

[9] Floyd, S.; Fall, K., "Promoting the use of end-to-end congestion control in the Internet", IEEE/ACM Transactions on Networking, Aug. 1999, vol.7, (no.4): 458-72.].

[10] L. L. Peterson, B. S. Davie, *Computer Networks*, Morgan Kaufmann, San Francisco, CA, 2000.

[11] Dah-Ming Chiu;Jain, R., "Analysis of the increase and decrease algorithms for congestion avoidance in

computer networks", Computer Networks and ISDN Systems, June 1989, vol.17, (no.1), p. 1-14.

[12] T.V. Lakshman and U. Madhow, "The Performance of TCP/IP for Networks with High Bandwidth-Delay Products and Random Loss", *IEEE/ACM Transactions on Networking*, 5(3), June 1997.

[13] S. Floyd, T. Henderson, " NewReno Modification to TCP's Fast Recovery", RFC 2582, April 1999.

[14] Brakmo L. S., O'Malley S. W., and Peterson L. L., "TCP Vegas: End-to-end congestion avoidance on a global Internet," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 13, no.8, pp. 1465-1480, 1995.

[15] S. Mascolo, C. Casetti, M. Gerla, M. Sanadidi, R. Wang, "TCP Westwood: End-to-End Bandwidth Estimation for Efficient Transport over Wired and Wireless Networks", *ACM Mobicom 2001*, July, Rome, Italy and Wireless Networks, vol. 8, no. 5, Sept. 2002.

[16] L.A. Grieco, S. Mascolo, "TCP Westwood and Easy RED to Improve Fairness in High-Speed Networks", *Proc. of the VII International Workshop on Protocols For High-Speed Networks (PfHSN'2002)*, April, 2002 Berlin, Germany. Lecture Notes on Computer Science (Lcns), Springer Verlag.

[17] C. Parsa, J.J. Garcia-Luna-Aceves, "Improving TCP Congestion Control over internets with heterogeneous Transmission media", Proc. IEEE Int. Conf. On Network protocols, Toronto, Oct. 31-Nov. 3,1999.

[18] F. P. Kelly, "Mathematical Modeling of the Internet," *Proc. 4th International Congress on Industrial and Applied Mathematics*, July 1999.

[19] C. V. Hollot, V.Misra, Donald F. Towsley, and Wei-Bo Gong. Analysis and design of controllers for AQM Routers supporting TCP flows. *IEEE Trans. Automatic Control*, 47(6):945–959, June 2002.

[20] S. H. Low, "A duality model of TCP flow control", *Proc. of ITC Specialist Seminar on IP Traffic Measurements, Modeling and Management*, Sept. 2000.

[21] D. Bansal and H. Balakrishnan and S. Floyd and S. Shenker, "Dynamic Behavior of Slowly-Responsive Congestion Control Algorithms", *Proc. of Sigcomm 2001*.

[22] M. Allman, S. Floyd, C. Partridge, "Increasing initial TCP's initial window," RFC 2414, Sept. 1998.

[23] K. Ästrom and T. Hägglund, *PID Controllers: Theory, Design, and Tuning*, ISA, 1995.

[24] V. Jacobson, R. Braden, D. Borman, " TCP Extensions for High Performance ", RFC 1323, May 1992.

[25] Hoe, J., C., "Improving the Start-up Behavior of a Congestion Control Scheme for TCP," *Proc. of ACM Sigcomm'96*, pp. 270-280.

[26] Villamizar, C. and Song C. (1995), "High Performance TCP in ANSNET", *ACM Computer Communication Review*, vol. 24, no. 5, pp. 45-60.

[27] S. Keshav, "A Control-theoretic Approach to Flow Control," Proc. ACM Sigcomm 1991, September 1991.

[28] S. Mascolo, "Modeling and Stability Analysis of the Internet Congestion Control", Technical Report no. S17/03 DEE Politecnico di Bari

[29] S. Mascolo, "The TCP/IP Flow and Congestion Control, Part I: a Dynamic Model", Submitted to CDC 2003.

[30] S.H. Low, F. Paganini, and J. C. Doyle, "Internet congestion control," *IEEE Contr. Syst. Mag.*, vol. 22, pp.28-43, Jan. 2002.

[31] P. Karn and C. Partridge, "Improving round-trip time estimates in reliable transport protocols", *ACM Transaction on Computer Systems*, vol. 9, no. 4, pp. 364-373, November 1991.

[32] Fast TCP at http://netlab.caltech.edu/FAST/

[33] L. A. Grieco, S. Mascolo, "Performance Evaluation and Comparison of Westwood+, New Reno, and Vegas TCP Congestion Control", ACM Computer Communication Review, vol. 34, no. 2, April 2004.

[34] S. Floyd, "HighSpeed TCP for Large Congestion Windows", IETF Internet Draft draft-ietf-tsvwg-highspeed-00.txt, work in progress, July 2003.