Abstract- In this paper, we design a self-tuning utility-based controller for end-to-end congestion in the IP (Internet Protocol)-based Internet. Multiple controlled sources transmit the packets through a series of AQM (Active Queue Management) routers into their destinations simultaneously and share the limited bandwidth of the Internet. Each AQM router runs the RED (Random Early Detection) algorithm that uses ECN (Explicit Congestion Notification) packet marking strategy to provide the link congestion information through IP packets. A self-tuning utility-based controller is placed in every source node to regulate source transmission rate based on the feedback route congestion information from the AQM routers through ACK packets. The pole placement technique in classical control theory is used to allow the user to achieve good transient network performance. By assigning a proper interval of damping ratio $\zeta$ each controller self-tunes only when the change of network parameters drifts $\zeta$ outside its specified interval. Our simulations demonstrate the stability of the Internet achieved by our self-tuning utility-based congestion controller.

Keywords: Utility-Based Control, Self-Tune, Random Early Detection, Explicit Congestion Notification, Pole Placement

1. INTRODUCTION

Congestion control has been a critical factor in the robustness of the Internet. As a dominant end-to-end congestion control algorithm, TCP (Transmission Control Protocol/IP (Internet Protocol) has been widely implemented for the survival of the Internet [1]. In TCP/IP networks, the router applies AQM (Active Queue Management) algorithms (e.g. [2,3,4]) to implicitly inform the source the network congestion by dropping the packets. However, as the delay-bandwidth product increases, there always exist great fluctuations in the source transmission rates and router queue sizes regardless of the AQM scheme employed. Mathematical analysis and simulation show that no AQM scheme can maintain small fluctuation in the source’s sending rate over very high-capacity or large-delay links [5]. Unfortunately, relatively smooth sending rate is very important for streaming media transmission in the Internet [6], and alternative method must be sought.

There has been a surge of interest in studying how to stabilize the source transmission rates and to utilize the network resources efficiently in the Internet. In order to achieve this goal, a tractable network model was proposed to analyze the stability and fairness of the rate control algorithm, and utility function was applied in the controller design in order to make the optimal use of the network resources [7]. An optimal solution to the resource allocation problem was obtained when the communication delays between the end users are negligible, and a primal algorithm was provided to adjust the users’ sending rates to converge at an equilibrium state [7]. It was shown that this utility-based controlled network is globally stable without any constraints on the control gain. However, the communication delay cannot be neglected over the physical channel especially in high-speed networks. Furthermore, the propagation delay may have an adverse impact on network performance due to the dramatic queuing fluctuation in the bottleneck. Such queuing oscillation may cause data flow fluctuation and reduce the bandwidth utilization of the network.

Utility-based congestion control schemes in [7] can be divided into two classes: primal algorithms and dual algorithms. In primal algorithms e.g. [8], the users regulate their source sending rates dynamically based on the route prices, and the links select a static law to calculate their link prices directly from the arrival rates at the links. On the other hand, in dual algorithms e.g. [9], the links adjust their link prices dynamically based on the link rates, and the users select a static law to compute their source sending rates directly from the route prices and the source parameters [10].

Many attempts have been made in the utility-based control of Internet traffic recently. For example, a simple framework using deterministic fluid model was proposed, on which utility functions, random losses and explicit congestion notification (ECN) were exploited on the design of end-to-end congestion controllers [11]. Linearization approach was adopted to solve the optimization problem and make the source transmission rates to converge to an equilibrium point, where the local stability of the primal algorithm was achieved e.g. [8,10,12-15]. A utility-based controller was design to realize the global stability of the network under the case of a single source and a single link [16]. These papers adopt time domain analysis in the controller design. However, they have very limited performance evaluation on the transient response of control system, which is an important measure of the control system performance. Moreover, these papers only yielded a
fixed solution for the controllers, thus allowing no flexibility to great variations in the load level of the Internet, which may result in the fluctuation of data flow in the network. Therefore, it is practical to introduce adaptive control. Adaptive control has widely applied in industrial process control e.g. [17, 18, 19], which will play an important role in the network traffic control.

In this paper, we introduce a general network model with communication delay (i.e. round-trip time) for the Internet traffic. Considering the network control system by the state-space representation, we propose the closed-loop transfer function for the network control system in the frequency domain. We apply pole placement technique of the classical control approach in the self-tuning utility-based congestion controller design, and focus more on its transient behavior. Instead of a fixed value, an interval for the damping ratio \( \zeta \) is appropriately chosen such that a) the transient response of the closed-loop network control system is satisfied; b) all the poles would lie in the left-half s-plane to guarantee the stability of the network. We have derived a simple equation to obtain utility-based congestion controller as well as the equation for monitoring the damping ratio \( \zeta \) online upon the change of the network parameters. Self-tuning utility-based control self-tunes only when the change of the network parameters causes the damping ratio \( \zeta \) to drift outside its interval, and thus can achieve good network performance in face of any dramatic changes in network environment. To our best knowledge, no other work on utility-based controller design adopts frequency domain technique and introduces adaptive control upon the change of the network traffic.

This paper is organized as follows. Section 2 introduces network utility model for the Internet with communication delays. Pole placement method is applied in self-tuning utility-based congestion controller design in Section 3. Section 4 presents simulations and performance evaluation for utility-based congestion control algorithm to realize the stability of the network control system. Finally some conclusions are given in Section 5.

For the remainder of the paper, the following symbols and notations would pertain.

- \( G(s) \): Closed-loop transfer function of the network control system
- \( K_r \): Proportional gain of the utility-based congestion controller
- \( p_j(.) \): Packet marking rate (i.e. link price) regulated by RED and ECN marking strategy for the packets routed through the AQM router \( j \)
- \( q_j(t) \): Transmission rate of the utility-based controlled source node \( j \)
- \( q_r(t) \): Change of source transmission rate regulated by utility-based controller for the source node \( r \)
- \( w_r \): Price that the user (utility-based controlled source) \( r \) is willing to pay or Weight assigned to the user \( r \) according to the price paid by the user \( r \) (i.e. source price)
- \( \alpha(t) \): Total packet marking rate (i.e. route price) for the packets that are sent by utility-based controlled source \( r \) and routed through a series of AQM routers between the source \( r \) and its destination
- \( \tau_r \): Round trip time for the utility-based controlled source \( r \)
- \( \zeta \): Damping ratio of the closed-loop network control system
- \( \omega_0 \): Natural frequency of the closed-loop network control system

2. NETWORK MODEL FOR UTILITY-BASED CONTROL

Consider a network with a finite set \( J \) of AQM routers (i.e. resources), and let \( C_j \) be the finite link capacity (i.e. service rate) of the router \( j \) that can be shared by a finite set of sources (i.e. users), indexed by \( r \). Let \( R_j \) be a route starting from the source \( r \) and routed through a non-empty set of \( J \), which consists of two paths: (1) the data path through which the source \( r \) transmits its data to its destination; (2) the control path through which the destination sends the ACK (Acknowledge) packets back to the source \( r \). The ACK packets contain route congestion information (i.e. route price) so that the utility-based controller can regulate the source transmission rate to prevent the network from congestion. Let \( R=\{R_j\} \) be the set of all possible routes. For any route \( R_j \), define a 0-1 routing matrix \( A=(A_{jr}, j\in J, R_j\in R) \) such that for any \( R_j\in R \), we have

\[
A_{jr} = \begin{cases} 
1, & \text{if } j \in R_j \\
0, & \text{otherwise}
\end{cases}
\]

Our network model considers the utility function for the Internet traffic, where the round trip delays are non-negligible\(^1\). Let \( \tau_f \) be the forward delay from the utility-based controlled source \( r \) to the AQM router \( j \) and \( \tau_b \) be the feedback delay which is the sum of delay from the AQM router \( j \) to the destination and from the destination to the source \( r \). Let \( \tau_r = \tau_f + \tau_b \) be the round trip time between the source \( r \) and its destination. If we further let \( \sum_{j\in R_j} A_{jr} q_j(t) \) account for the total incoming traffic rate from all active sources

\(^1\) A summary of the utility network model without communication delay [7] is also provided in Appendix A and on which we have extended.
represents the total marking rate of the packets by a probability algorithm [2] and ECN [20] marking strategy to mark packets routed through a series of routers located between the source and its destination, i.e. route price for the source r. In this paper, we will adopt the RED and its destination, i.e. route price r.

To simplify the controller design, we will let

\[ \alpha_r(t) = \sum_{j \in R_r} \left( \sum_{s : j \in R_s} q_s(t - \tau_s) \right) \]

so that we can rewrite Equation (1) as

\[ \dot{q}_r(t) = K_r \left( w_r - q_r(t - \tau_r) \alpha_r(t) \right) . \]  

(2)

As shown in Equation (2), we can treat \( q_r(t) \) as a state variable and \( \alpha_r(t) \) as a varying state feedback gain for the network control system [21]. This is captured by the state space representation of the utility-based network control system in Figure 1. Therefore we can achieve the good transient performance by adjusting the dynamics of the transmission rate \( q_r(t) \) of the source r. The proportional gain \( K_r \) is the focus of our controller design below.

In frequency domain, the closed-loop transfer function of the network control system in Equation (2) can be obtained as

\[ G(s) = \frac{K_r w_r}{s + K_r \alpha_r e^{-s \tau}} . \]  

(3)

So we can obtain the characteristic equation of the closed-loop network control system as

\[ s^2 + \left( \frac{2}{\tau_r} - K_r \alpha_r \right) s + \frac{2K_r \alpha_r}{\tau_r} = 0 . \]  

(5)

Let \( s_1 \) and \( s_2 \) to be the roots of the Equation (5), which are also the poles of the second-order closed-loop control system. To guarantee the stability of the closed-loop control system, \( s_1 \) and \( s_2 \) will be placed in the left-half s plane [21].

In order to achieve a desirable transient response of a second-order control system, we will specify the proper interval of \( \zeta \) of our self-tuning utility-based controller to be 0.6 \( \leq \zeta \leq 1 \), because too small a \( \zeta \) yields excessive overshoot in the transient response, and a system with a large value of \( \zeta \) responds sluggishly [21]. By performance evaluation excises and the Relationship 1, we have determined that 0.6 \( \leq \zeta \leq 1 \) would give a satisfactory transient response of utility-based control system. Therefore, we can specify nominal value of \( \zeta \) to be 0.8 (the mean of 0.6 and 1), so that our self-tuning
utility-based controller can clamp $\zeta$ around the range of $0.6 \leq \zeta \leq 1$ upon the change of network environment.

Relationship 1 [21]

From the classical control theory, it is well known that the relationship between maximum percent overshoot $M_p$ and the damping ratio $\zeta$ is $M_p = e^{-\zeta / \sqrt{1 - \zeta^2}} \pi$ [21]. Then based on the range of $0.6 \leq \zeta \leq 1$, we can obtain the desirable overshot of unit step response $0 \leq M_p \leq 10\%$.

Under the range of $0.6 \leq \zeta \leq 1$, we can obtain a pair of complex conjugate closed-loop poles $(s_1, s_2) = -\zeta \omega_n \pm j \omega_n \sqrt{1 - \zeta^2}$, that is, $|s_1| = |s_2| = \omega_h$. Therefore, from Equation (5), we can obtain

$$\frac{2}{\tau_r} - K_r \alpha_r = 2 \zeta \omega_n \quad ,$$

$$\frac{2 K_r \alpha_r}{\tau_r} = \omega_n^2 \quad .$$

From Equation (7), we can determine the natural frequency as

$$\omega_n = \sqrt{\frac{2 K_r \alpha_r}{\tau_r}} \quad .$$

Considering the condition $\omega_h \tau_r \leq 1.2$ for the Pade approximation, we can obtain $\omega_n \tau_r = \sqrt{2 K_r \alpha_r \tau_r} \leq 1.2$, i.e.

$$K_r \leq \frac{18}{25 \alpha_r \tau_r} \quad .$$

Then substituting Equation (8) back into Equation (6), we can obtain

$$\frac{2}{\tau_r} - K_r \alpha_r = 2 \zeta \sqrt{\frac{2 K_r \alpha_r}{\tau_r}} \quad ,$$

which can be simplified as

$$K_r^2 - \frac{4}{\alpha_r \tau_r} \left(1 + 2 \zeta^2 \right) K_r + \left(\frac{2 \zeta}{\alpha_r \tau_r} + 1 \right)^2 = 0 \quad .$$

Solving Equation (11) based on the condition from Equation (9), we can obtain

$$K_r = \frac{2}{\alpha_r \tau_r} \left[1 + 2 \zeta^2 - 2 \zeta \sqrt{\zeta^2 + 1} \right] \quad .$$

From Equations (8) and (12), we can obtain

$$\omega_n \tau_r = 2 \sqrt{1 + 2 \zeta^2 - 2 \zeta \sqrt{\zeta^2 + 1}} \quad .$$

Considering the specified interval $0.6 \leq \zeta \leq 1$, we have $0.83 \leq \omega_h \tau_r \leq 1.13$ that satisfies the Pade approximation condition $|\omega_h \tau_r| \leq 1.2$.

In order to keep the proportional gain $K_r$ of utility-based controller constant, we have to adjust the damping ratio $\zeta$ whenever the network parameters $\alpha_r$ or $\tau_r$ change. Under this scenario, Equation (10) becomes

$$\zeta = \frac{1}{2} \sqrt{\frac{2}{K_r \alpha_r \tau_r} - \frac{K_r \alpha_r \tau_r}{2}} \quad .$$

Equation (13) will be used to monitor the damping ratio $\zeta$ of the utility-based control system. When the damping ratio $\zeta$ of the utility-based control system falls outside its specified interval due to the dramatic change of the network parameters ($\alpha_r$ and $\tau_r$), the utility-based controller need to self-tune to give a new $K_r$ value.

3.1. Implementation of Utility-Based Control Algorithm

We adopt the RED algorithm for AQM in every router and use the ECN packet-marking strategy to mark our packets based on current link congestion information (i.e. link price). Then we convey the total marking rate $\alpha_r$ (i.e. route price) to the source node $r$ through the ACK packets when the packets reach their destinations. Considering the stochastic characteristics of the network, we can use the average value $\overline{\alpha}_r(t)$ of total marking rate $\alpha_r$ to calculate the proportional gain $K_r$ (through Equation (12)). Then we applied the utility-based control algorithm in Equation (2) to calculate the source transmission rate based on instantaneous feedback network congestion information i.e. $\alpha_r$. We also use $\overline{\alpha}_r(t)$ to compute the damping ratio $\zeta$ of the control system by Equation (13).

**Calculation of the average marking rate $\overline{\alpha}_r(t)$**

We adopt the averaging filter method in RED algorithm [2] to calculate the average value $\overline{\alpha}_r(t)$ based on the instantaneous total packet-marking rate $\alpha_r$, i.e.

$$\overline{\alpha}_r(t) = (1 - W_d) \overline{\alpha}_r(t - 1) + W_d \alpha_r(t) \quad .$$

Considering the stochastic characteristic of the data stream in the Internet and the measurement accuracy, the averaging weight is recommended as $W_d = 0.01$ [2, 24]. The initial value of $\overline{\alpha}_r(t)$ is set as the mean of $\alpha_r$ during the first round-trip time period, i.e.

$$\overline{\alpha}_r_{\text{initial}} = \frac{1}{\tau_r} \sum_{n=0}^{\tau_r} \alpha_r(t) = \frac{1}{\tau_r} \sum_{n=1}^{\tau_r} \alpha_r(n) , \text{ where } \delta \text{ is sampling time.}$$

To be compatible for the current dominant TCP/IP protocol in the Internet, we convert the source transmission rate calculated by Equation (2) into the advertised window size for the controlled source nodes. The similar methods have been well accepted in TCP/IP networks e.g. [5, 25]. Thus the advertised window size $\text{win}_r(t)$ for the controlled source node $r$ is computed as

$$\text{win}_r(t) = \tau_r \cdot q_r(t) \quad .$$

where the window size must be an integer and be restricted between 1 and the maximum window size...
\( \text{win}_{r}^{\text{max}} \) allowed for the controlled source node \( r \), i.e. \( 1 \leq \text{win}_{r}(t) \leq \text{win}_{r}^{\text{max}} \).

We now introduce the behaviors of a controlled source, an AQM router and a destination in our utility-based controller.

**Behavior of Source \( r \)**
- **Step 1:** Specify an interval of \( \zeta \) to be \( 0.6 \leq \zeta \leq 1 \) and nominal \( \zeta = 0.8 \).
- **Step 2:** Estimate the round trip time \( \tau \), by the method in [6]. Record and keep its transmission rate for a round-trip time, according to Equation (2).
- **Step 3:** Extract the instantaneous total marking rate \( \alpha_{r}(t) \) from the ACK packet when it receives an ACK packet.
- **Step 4:** Apply the utility-based control algorithm in Equation (2) to calculate the current transmission rate, based on instantaneous marking rate \( \alpha_{r}(t) \) so as to prevent the network from congestion.
- **Step 5:** Apply Equation (15) to calculate the source window size \( \text{win}_{r}(t) \).

**Real-time Monitoring of Source \( r \)**
- **Step 6:** Apply Equation (12) to compute the proportional gain \( K_{r} \), each of the utility-based controller based on the current information in the source \( r \), i.e. the average marking rate \( \bar{\alpha}_{r}(t) \) and the round trip time \( \tau \).
- **Step 7:** Utilize Equation (14) to compute the average value \( \bar{\alpha}_{r}(t) \) of total packet marking rate.
- **Step 8:** Apply Equation (13) to compute the damping ratio \( \zeta \) of the control system based on the current information in the source \( r \), i.e. the average marking rate \( \bar{\alpha}_{r}(t) \) and the round trip time \( \tau \).
- **Step 9:** (9a) If the damping ratio \( \zeta \) falls outside the specified interval, go to Step 6.
(9b) If the damping ratio \( \zeta \) is still within the interval, the utility-based controller would not self-tune, go to Step 7.

**AQM Router Behavior**
- **Step 1:** Apply RED to calculate the packet marking probability (i.e. link price) based on current congestion situation [2].
- **Step 2:** Use ECN marking strategy to mark the packets [20].

**Destination Behavior**
- **Step 1:** Check whether the received packets are marked or not, when the packets arrive their destination.
- **Step 2:** Calculate the total marking rate \( \alpha_{r}(t) \) (i.e. route price) of packets in a fixed period. (The total marking rate \( \alpha_{r}(t) \) is defined as the ratio of the total number of marked packets and the total number of received packets in the period).

**Step 3:** Integrate the total marking rate \( \alpha_{r}(t) \) into the ACK packet and sent it back to the source.

**Example:**
Considering a utility-based controlled source \( r \), we have \( w = 20 \), \( \mu_{r} = 0.3 \) and \( \tau = 0.6s \). If we specify the interval of \( \zeta \) to be \( 0.6 \leq \zeta \leq 1 \) and nominal \( \zeta = 0.8 \), then we can obtain \( K_{r} = 2.57 \) based on Equation (12). Then we apply utility-based control algorithm in Equation (2) and Equation (15) to regulate the transmission rate of source \( r \) to prevent the Internet from congestion. At the mean time, we apply Equation (13) to monitoring the damping ratio of the control system upon the change of the network parameters online.

### 4. PERFORMANCE EVALUATION

Since our self-tuning utility-based controller can regulate the source transmission rate over time, we are particularly interested in analyzing the transient behavior of every source transmission rate and fairness among all the sources in Section 4.1. Furthermore, we also like to see how our controller will adapt the source transmission rate to the dramatic change of network parameters and guarantee the stability of the network in Section 4.2. In our experiment, we assume that all the source window sizes (calculated by utility-based control algorithm) are smaller than their maximum window sizes and the total marking rate \( \alpha_{r}(t) \) does not vary faster than the system response time.

#### 4.1 The Transient Response of Different Sources

We perform MATLAB [26] simulations to see how well our controller achieves a smooth source transmission rate. We have verified the fairness among the sources with different source prices under the same network environment.

![Figure 3. Utility-based controlled sources with multiple AQM routers](image)

Without loss of generality, we simulate a network with 3 AQM routers located between the source and its destination. Although many sources transmit the packets to their destination through these 3 AQM routers, we also like to see how our controller will adapt the source transmission rate to the dramatic change of network parameters and guarantee the stability of the network in Section 4.2. In our experiment, we assume that all the source window sizes (calculated by utility-based control algorithm) are smaller than their maximum window sizes and the total marking rate \( \alpha_{r}(t) \) does not vary faster than the system response time.
routers, we only present the comparison results on the behaviors of 3 sources with different weight $w_r$ and the packets sent by these 3 sources routed through the same 3 AQM routers, as shown in Figure 3. The weights $w_r$ of the source node $r \ (r=1,2,3)$ are 10, 20 and 30 respectively. The round-trip times $\tau_r$ of the source node $r \ (r=1,2,3)$ are 1s, 0.6s and 0.4s respectively. The initial values of all the source transmission rates are set as zero, i.e. $q_1(0)=q_2(0)=q_3(0)=0$. We assume that the RED and the ECN marking probabilities $p_j$ for the packets in AQM router $j \ (j=1,2,3)$ are uniformly distributed with a mean of 0.05, 0.1 and 0.15 respectively. Thus, the average of total marking rate can be approximated as $\frac{3.01+0.1+0.05}{3}=1.015$. Such control system parameter approximation is well accepted in control system design e.g. [17-19, 23]. When the time $t \rightarrow \infty$, the $\alpha_r(t)$ will be also converge to $\int_0^\infty (p_1(t)+p_2(t)+p_3(t))dt = 0.3$.

Figure 4. The instantaneous source transmission rate (Kbps)

![The instantaneous source transmission rate (Kbps)](image)

Figure 4 shows the instantaneous transmission rates of different sources. We can see that the proper damping ratio $\zeta=0.8$ make all the 3 sources exhibit good transient behavior. It is shown that the transmission rates of the source $r \ (r=1,2,3)$ converge to their equilibrium points $q_1^* = w_1 / \alpha_1^* = 33.33$Kbps , $q_2^* = w_2 / \alpha_2^* = 66.67$Kbps , and $q_3^* = w_3 / \alpha_3^* = 100$Kbps respectively, as anticipated. The fairness among these 3 sources is achieved when their transmission rates approach their respective equilibrium points, as we can observe from Figure 4.

### 4.2 Self-tuning Utility-Based Controller versus Fixed Utility-Based Controller

We want to demonstrate the importance of adaptive control [17] in the network traffic control by making the comparison between self-tuning utility-based controller and fixed utility-based controller. The proportional gain of a fixed utility-based controller is fixed no matter how large the change of the network parameters, while that of a self-tuning utility-based controller should be re-computed if the damping ratio $\zeta$ drifts outside its specified interval.

Figure 5. Utility-based controlled sources with 5 AQM routers

![Utility-based controlled sources with 5 AQM routers](image)

We consider a network with 5 AQM routers in Figure 5 that we have implemented fixed utility-based control algorithm and self-tuning utility-based control algorithm for Source $r$ respectively. In this network, from the time $t=0s$ to 100s, the packets are transmitted by Source $r$ and routed through 2 AQM routers (Router 1 and Router 2) before they reach the destination. At time $t=100s$, we turn off Router 2 (e.g. the system crashed), so that the packets have to by-pass Router 2 and have found a new route through Router 3, 4 and 5.
(i.e. the new path is now Routers 1, 3, 4 and 5) before they arrive the destination. The interval of damping ratio $\zeta$ is specified as $0.6 \leq \zeta \leq 1$ and nominal value of $\zeta$ is specified $\zeta = 0.8$.

Figure 6 Total packet marking rate of Source $r$
(a) instantaneous total packet marking rate $\alpha_r(t)$
(b) average total packet marking rate

Figure 7 Instantaneous transmission rate of Source $r$
(a) regulated by fixed utility-based control algorithm
(b) regulated by self-tuning utility-based control algorithm

Like the system in Section 4.1, it is assumed that RED and ECN marking probabilities $p_j$ for the packets in AQM router $j$ ($j=1,2,3,4,5$) are uniformly distributed with a mean of 0.05, 0.1, 0.15, 0.1 and 0.05 respectively. The round-trip time $\tau_r$ between the source $r$ and its destination is 1s if the packets are routed through Routers 1 and 2, while the round trip time $\tau_r$ is 1.5s if the packets are routed through Routers 1, 3, 4 and 5. At the beginning of the simulation, the initial value of average total packet marking rate $\bar{\alpha}_r$ is calculated as the mean of instantaneous total packet marking rate $\alpha_r(t)$ during the first round-trip time period, as discussed in Section 3.1. Figure 6 shows the instantaneous total packet marking rate $\alpha_r(t)$ and the average marking rate $\bar{\alpha}_r$ (calculated by Equation (14)) before and after Router 2 shuts down.

Figure 8. The instantaneous damping ratio monitored by Equation (13)
(a) for fixed utility-based control algorithm
(b) for self-tuning utility-based control algorithm

Figure 9. The instantaneous proportional gain $K_r$
(a) of fixed utility-based control algorithm
(b) of self-tuning utility-based control algorithm

Figure 7 shows the instantaneous source transmission rate by fixed utility-based control algorithm and self-tuning utility-based control algorithm respectively. Figure 8 shows the instantaneous damping ratio monitoring by Equation (13) for the fixed utility-based control algorithm and the self-tuning utility-based control algorithm respectively. Figure 9 shows instantaneous proportional gain $K_r$ of fixed utility-based control algorithm and self-tuning utility-based control algorithm respectively.
From the time $t=0s$ to $100s$, the total packet marking rate is uniformly distributed on $[0, 0.3]$, and the average value of the total packet marking is clamped around 0.15, as shown in Figure 6. The damping ratio $\zeta$ has only a little fluctuation around 0.8, and falls within the interval of $0.6 \leq \zeta \leq 1$, so the self-tuning utility-based controller remains unchanged. Both source transmission rates converge to the equilibrium point $q^*_s = w_s / \alpha_s = 10 / 0.15 = 66.67$ Kbps at time $t=5s$, as we anticipated. At time $t=100s$, Router 2 shuts down, the packets have to be routed through Routers 3, 4 and 5. The instantaneous total packet-marking rate for Source $r$ has great changes and is uniformly distributed in $[0, 0.7]$, as shown in Figure 6. The round-trip time becomes longer. The damping ratio $\zeta$ goes outside the interval of $0.6 \leq \zeta \leq 1$, so the self-tuning utility-based controller need to self-tune. After $t=101s$, the damping ratio $\zeta$ has little oscillation around $\zeta \approx 0.71$, which falls inside the interval of $0.6 \leq \zeta \leq 1$. So self-tuning utility-based controller remains unchanged. The source transmission rate converges to a new equilibrium point $q^*_s = w_s / \alpha_s = 10 / 0.35 = 28.57$ Kbps at time $t=110s$, as shown in Figure 7(b). Fixed utility-based controller is fixed, so the damping ratio $\zeta$ reduces to around 0.1, which causes the large oscillation of the source transmission rate, as shown in Figure 7(a).

5. CONCLUSION

We have developed a self-tuning utility-based controller for Internet traffic and prevent the end-to-end congestion in the Internet. Pole placement method from the classical control theory has been employed in the controller design. By selecting a proper interval of the damping ratio $\zeta$, we can achieve a satisfactory transient response of the network control system. The simple equation for obtaining the utility-based controller as well as the equation for monitoring system damping ratio has been derived. Self-tuning utility-based controller self-tunes only when the change of the network parameters (the round-trip time and total packet marking rate) causes the damping ratio $\zeta$ to drift outside the interval. We have used the RED algorithm and the ECN packet marking strategy in the AQM routers to provide the link congestion information (i.e. link price). The utility-based controller is located in the every source and it can calculate the source transmission rate based on the source price and the instantaneous feedback route congestion information (i.e. route price) through ACK packets. We have demonstrated by simulations that with our utility-based congestion controller applied, the Internet is stable and can achieve desirable network performance in terms of source transmission rate. Furthermore, our utility-based control algorithm can provide a relatively smooth transmission rate, thus it is quite suitable for rate control of multimedia transmission such as voice over IP in the Internet [6].

ACKNOWLEDGEMENT

We would like to appreciate many constructive suggestions from anonymous reviewers. The financial support from NSERC Research Grant (#RGPIN42878) is also acknowledged.

REFERENCES


APPENDIX A. SUMMARY OF UTILITY NETWORK MODEL

Following Kelly’s work [7], we would associate a route r with a source, and then assign a utility function $U_r(q_r)$ to each source r with respect to its transmission rate $q_r(t)$, so that we can formulate the resource allocation problem as a utility maximization problem:

$$\max_{q \geq 0} \sum_{r \in R} U_r(q_r) \quad \text{subject to } y \leq c$$

where $q$ is the vector of source transmission rates, $c$ is the vector of link capacities, and $y \equiv Ag$. The constraint $y \leq c$ means that the aggregate incoming traffic rate $y = \sum_{r \in R} A_r q_r$ at the router $j$ cannot exceed the link capacity $c_j$. Then we can obtain the Lagrangian [27] for the problem in Equation (A-2) as

$$L(q, p) = \sum_{r \in R} U_r(q_r) + \sum_{j \in J} p_j (c_j - y_j).$$

where $p_j$ is the Lagrange multiplier (or called the shadow price) associated with router $j$. We can interpret $p_j$ as a price charged by the router $j$, i.e. per unit flow through the router $j$, on the total flow $y_j$ through the router $j$. In network terminology, $p_j$ is the network congestion information of the router $j$ (i.e. link price), which is delivered by ACK packets back to inform the sources.

As the router $j$ does not know the utility functions of its sources, the optimal solution to the problem in Equation (A-2) is generally not tractable. So we assign a logarithmic utility function $U_r(q_r) = w_r \log(q_r)$ to each source $r$, where $w_r$ is the price the source is willing to pay or a weight assigned to the source according to the price paid by the source. Therefore, we can update the problem in Equation (A-2) as

$$L(q, p) = \max_{q \geq 0} \sum_{r \in R} w_r \log(q_r) - \sum_{j \in J} \left( \sum_{i : j \in s_i} \alpha_i p_j \right) dy_j.$$  

The optimal solution to the problem in (A-3) can be shown to be $q^*_r = w_r / \alpha^*_r$ with $\alpha^*_r = \sum_{j \in J} p_j$ [7].

Furthermore, a primal algorithm has been proposed to regulate the transmission rate of the source $r$ to converge to $q^*_r = w_r / \alpha^*_r$ in the presence of no communication delays, i.e.

$$q_r(t) = K_r \left( w_r - q_r(t) \sum_{j \in R} p_j \left( \sum_{i : j \in s_i} A_i q_i(t) \right) \right),$$

where $K_r$ is the proportional gain, and

$$\sum_{j \in R} p_j \left( \sum_{i : j \in s_i} A_i q_i(t) \right)$$

represents the total marking rate of the packets routed through a series of routers located between the source $r$ and its destination, i.e. route price for the source $r$ [7]. It means that every source $r$ adjusts its sending rate based on feedback information provided by the routers in the network in such a way that at an equilibrium point the source price $w_r$ equals to its aggregate cost [13]. It is shown that with the utility-based control algorithm in Equation (A-4) applied, the network system without communication delays is globally stable if the marking function is $p_j$ continuous, non-decreasing, non-negative and not identically zero [7]. Normally the communication delays in the Internet are not negligible. Therefore, we need to extend the primal algorithm in Equation (A-4) by including various delay components.
APPENDIX B. THE CONTROL THERORETICAL BACKGROUND

In this Appendix, we will briefly review control theoretical background for pole placement method with the purpose of summarizing the notation and terminology used in this paper.

![Complex conjugate poles of the closed-loop control system G(s)](image)

Figure B-1. Complex conjugate poles of the closed-loop control system G(s)

We consider a second-order control system with a closed-loop transfer function expressed in the standard form as

\[ G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \quad (B-1) \]

where \( \zeta \) and \( \omega_n \) are the damping ratio and the natural frequency of the control system respectively. These parameters can be used to describe the dynamic behavior of the second-order control system [21]. In the undamped case \( 0 < \zeta < 1 \), the transient response is oscillatory and the closed-loop poles are a pair of complex conjugates \( s_1, s_2 = -\zeta \omega_n \pm j\omega_n \sqrt{1-\zeta^2} \) in the left-half \( s \) plane, as shown in Figure B-1. The control system is called undamped if \( \zeta = 0 \), and the transient response does not die out. The system is called critically damped if \( \zeta = 1 \), and over-damped system if \( \zeta > 1 \) [21]. All the closed-loop poles must be in the left-half \( s \) plane in order to guarantee the stability of the control system [21], a basis for the design rule our utility-based controller.

The transient response of a practical control system often exhibits damped oscillations before reaching the steady state. For the second-order closed-loop control system as expressed in Equation (B-1), the transient specifications such as maximum percent overshoot and settling time are determined by \( \zeta \) and \( \omega_n \). For example, the maximum percent overshoot enlarges as the damping ratio \( \zeta \) reduces under the range of \( 0 < \zeta < 1 \), while for a fixed value of \( \zeta \), the settling time increases as the natural frequency \( \omega_n \) decreases [21].

APPENDIX C. RANDOM EARLY DETECTION

The RED (Random Early Detection) algorithm [2] detects link congestion and measures the traffic load level in the queue of the router \( j \) (i.e. packet arrival rates in the link \( j \)) using the average queue size \( \bar{x}_j(t) \). This is calculated using an averaging filter and can be expressed as

\[ \bar{x}_j(t) = (1-W_q)\bar{x}_j(t-1)+W_qx_j(t) \quad , \quad (C-1) \]

where \( W_q \) is averaging filter weight and \( x_j \) is the instantaneous queue size. When the average queue size \( \bar{x}_j(t) \) is smaller than a minimum threshold \( \text{min th} \), no packets are marked. When the average queue size \( \bar{x}_j(t) \) exceeds the minimum threshold \( \text{min th} \), the router \( j \) randomly marks arriving packets with a given marking probability. The packet marking probability \( p_a \) (i.e. link price) for an arriving packet depends on the average queue size \( \bar{x}_j(t) \), the time elapsed since the last packet was marked, and the maximum packet marking probability \( \text{max p} \). If the average queue size is larger than a maximum threshold \( \text{max th} \), all arriving packets are marked.

The packet marking probability \( p_a \) is computed as

\[ p_a = \frac{p_b}{1 - \text{count} \times p_b} \quad , \quad (C-2) \]

with

\[ p_b = \max_p \frac{\bar{x}_j}{\text{max th} - \text{min th}} \quad , \quad (C-3) \]

where \( \text{max p} \) is the maximum value for \( p_a \), and \( \text{count} \) is a variable that keeps track of the number of packets that have been forwarded since the last packet marking.

Citation of this paper: Y. Hong, O.W.W. Yang, “Self-Tuning Utility-Based Controller for End-to-End Congestion in the Internet,” In Proceedings 3rd International Conference on Broadband Communications, Networks and Systems (IEEE BROADNETS), San Jose, California, U.S.A, October 2006.