End-to-end rate-based congestion control using EWMA for multicast services in IP networks


Abstract: In high-speed communication networks, the determination of a transmission rate using a congestion control scheme is critical for the stability of a closed-loop network system. A new rate-based congestion control scheme is proposed which employs an exponential weighted moving average algorithm. This scheme can be used to develop an efficient feedback control mechanism for congestion avoidance in high-speed communication networks. The newly proposed scheme ensures not only the stability of switch buffers but also higher link utilization of the whole network system.

1 Introduction

In general, congestion in high-speed communication networks occurs when input traffic to a network link is higher than the link capacity. This causes a buffer overflow at the network node and leads to the degradation of quality of service such as data losses and longer transmission delay or even deadlock. For the best-effort service in an internet, the congestion control is usually achieved by transport protocols. To this end, various kinds of window-based end-to-end flow control schemes such as TCP-friendly congestion control [1, 2], TCP-like congestion control [3], binary-feedback algorithms [4, 5], and an equation based method [6, 7] have been developed so far. These algorithms have been proved to work well in the environment where a sender relies on self clocking [2]. However, current real-time streaming applications in the internet [8, 9] typically do not rely on such selfclocking and thereby a rate-based end-to-end flow control is required in this case [10]. Up to the present, several kinds of rate-based end-to-end congestion control schemes have been proposed [10–15] most of which have been limited to max–min fairness cases [11–14]. Hence these schemes have a limitation in that they are only applicable to wide-area networks (WANs). Furthermore, these schemes have a lower utilization of network resources in the underload condition since the transmission capacity of sources is not counted. To overcome this problem a congestion control scheme with a new fairness criterion for multicast services was proposed by Nho and Lim [15] in which a bandwidth is allocated according to the capacity of the transmission sources. However, this scheme is only applicable to circuit-switching networks or virtual circuit-switching networks such as an asynchronous transfer mode (ATM) network since it requires feedback information for the traffic rate control.

To alleviate this problem and further extend its applicability to packet-switching networks, the status information of the network traffic must be transferred from downstream nodes to the source nodes. In ATM networks, the available bit-rate (ABR) service is based on a feedback mechanism, i.e. the network status is transferred to the ABR source by a resource management (RM) cell. Such cells contain the traffic information of downstream nodes for the traffic rate control. However, the traffic status of downstream nodes cannot be directly transferred to the source node for most of the streaming applications of IP-based networks in the current internet environment and thereby a rate-based end-to-end congestion control scheme is required by making use of the packet loss as the only available feedback information from the network.

Motivated by these practical situations, we present a simple but effective rate-based congestion control scheme using the exponential weighted moving average (EWMA) algorithm [16], which estimates the available bandwidth for multicast connections in IP-based communication networks. The presented control scheme is an extended version of the algorithm in [15] by improving its scalability and extending its applicability to IP-based networks. In particular, we apply a proportional-integral (PI) control concept instead of proportional (P) control to eliminate the steady-state error in the buffer level control. A comparative simulation study is provided to verify its efficiency compared with the previous work in [15].

2 Congestion network model

Figure 1 shows the structure of a switch considered in this paper [17, 18]. To simplify the development of a congestion control scheme to be proposed we postulate the following assumptions.

(i) The switch is a multicast virtual output queuing (MC-VOQ) switch. In other words, each output port has a virtual output queue capable of processing $1 \leq X < \infty$ data.

(ii) At each input port there are $1 \leq M < \infty$ FIFO queues dedicated to the input traffic.

(iii) An input packet arriving at the input module is queued into one of the $M$ queues and then switched from the
(iv) The switch fabric is a bufferless crossbar.

Based on the switch structure in Fig. 1 we consider the following model of network congestion. In Fig. 1, \( u_{ij} \) represents the \( i \)th unicast source dedicated to the \( j \)th destination, \( m_i \) denotes the \( i \)th multicast source, and \( x_i \) indicates the occupied level of the \( i \)th virtual output queue.

Let \( x_i(n) \) further denote the occupied buffer level of the \( i \)th congested link at the \( n \)th time-slot. Then we have

\[
\begin{align*}
x_i(n+1) &= \max\{0, x_i(n) + f_i(n) - c_i\} \\
i &= 1, 2, \ldots, L
\end{align*}
\]

(1)

where \( c_i \) is the available bandwidth (packets per slot) of the \( i \)th congested link and \( f_i(n) \) is the amount of input data in the buffer of the \( i \)th congested link during \([n, n+1]\). In [14], \( f_i(n) \) is obtained as follows:

\[
f_i(n) = \sum_{j=1}^{J_i} s_{ij}^u(n - \tau_{ij}^u) + \sum_{k \in M_i} s_{ik}^m(n - \tau_{ik}^m)
\]

(2)

where \( J_i \) is the number of unicast connections passing through the \( i \)th congested link, \( M_i \) is the set of multicast connections passing through the \( i \)th congested link, \( \tau_{ij}^u \) is the round-trip delay between the \( i \)th congested link and the source of the \( j \)th unicast connection passing through the congested link, \( \tau_{ik}^m \) is the round-trip delay between the branch point of the congested switch and \( k \)th multicast connection, \( s_{ij}^u(n) \) denotes the transmission rate of the \( j \)th unicast connection passing through the \( i \)th congested link at time instant \( n \), and \( s_{ik}^m(n) \) indicates the transmission rate of the \( k \)th multicast connection at time instant \( n \). Note that these transmission rates are all computed at the congested link.

We consider the following transmission rate of each source:

\[
s_{ij}^u(n) = \max\left(0, \min\left(r_{ij}/R_i, \tilde{c}_i - \sum_{k \in \Theta_i} \min\left(d_i/R_i\right)\right) - x_i(n) - \tilde{c}_i\right)
\]

\[
s_{ik}^m(n) = \min\left(d_i/R_i\right)
\]

(3)

(4)

where \( \Theta_i \) is the set of congested links through which the \( k \)th multicast connection passes, \( r_{ij} \) is the maximum transmission rate of the \( j \)th unicast connection passing through the \( i \)th congested link, \( R_i = \sum_{j=1}^{J_i} r_{ij} \), \( R_l \) is the sum of maximum transmission rates of the \( k \)th connection, \( d_i \) is the current traffic rate of the \( k \)th connection, \( \tilde{c}_i \) is the estimated available bandwidth of the \( i \)th congested link, \( x_i \) is the desired buffer level, and \( x_i \) is the control parameter to be chosen. Then the closed-loop equation can be described as follows:

\[
x_i(n+1) = x_i(n) - x_i\sum_{j=1}^{J_i} \left[x_i(n - \tau_{ij}^u) - X_i\right] + \tilde{c}_i - c_i
\]

(5)

In the case of the ABR service in ATM network, \( \tilde{c}_i \) is equal to \( c_i \). However, the available bandwidth \( c_i \) of the network varies according to the status of downstream nodes and the
traffic status cannot be directly transferred to the source node in packet switching networks. To overcome this problem and to prevent any oscillatory behaviour of the network states we employ the EWMA algorithm to estimate the available bandwidth. So we can formulate

\[ \hat{c}_i(n+1) = (1 - \beta)\hat{c}_i(n) + \beta AC_i \]

(6)

where \( \hat{c}_i(n) \) is the estimated available bandwidth of the \( i \)th connection over \( n, n+1 \) and \( 0 \leq \beta < 1 \) (optimal value of \( \beta \) can be chosen experimentally). In (6), \( AC_i \) can be replaced by \( k \times RTT \) since the available bandwidth of a network is inversely proportional to \( RTT \) of the network (\( k \) is a proportional factor to be chosen experimentally); \( RTT \) is usually transferred to the switch by an internet control message protocol (ICMP). Using this approach, the available bandwidth \( c_i \) can be elevated to the estimated available bandwidth \( \hat{c}_i \). Moreover, this scheme can ease the effect of the burst traffic on the overall network performance. So we have

\[ x_i(n + 1) = x_i(n) - xi \sum_{j=1}^{L} x_i(n - t_{ij}^u) - \hat{x}_i \]

(7)

Subsequently the characteristic polynomial of (7) is obtained as follows:

\[ \Phi_i(z) = z^{D_i + 1} - z^{D_i} + \sum_{j=1}^{L} z^{D_i - t_{ij}^u} \]

(8)

where \( D_i = \max \{ t_{ij}^u \} \).

The case of unicast connections is illustrated in Fig. 2, where the buffer levels of the links cannot be regulated to their desired buffer level since \( x_i \) has a constant value. This operating mechanism corresponds to a P control scheme which cannot eliminate a steady-state error. To guarantee the desired buffer level, i.e. to eliminate the steady-state error, \( x_i \) should be chosen based on a PI (proportional and integrative) control scheme. The proposed congestion control scheme based on the EWMA algorithm conceptually corresponds to the PI control scheme. From [19], the stability condition of \( x_i \) is obtained as follows:

\[ 0 < \frac{2}{J_i} \sin \left( \frac{\pi}{4D_i + 2} \right) \quad i = 1, 2, \ldots, L \]

(9)

where \( J_i \) is the number of unicast connections that pass through the \( i \)th congested link.

Fig. 2 Unicast connection model

3 Simulation studies

To exemplify the proposed congestion control scheme in this Section we show simulation studies for a sample network model based on numerical simulations by using the functional blocks (Fig. 2) of Simulink (Matlab Version 6.5.0.180913a R13 from Mathworks) with a Pentium IV 2.4 GHz computer. The sample network model is illustrated in Fig. 3 where \( u_i \) represents the \( i \)th unicast source, \( m_i \) denotes the \( i \)th multicast source, and A, B and C indicate each congested switch. The input traffic is assumed to have a uniform distribution. The characteristics of the input traffic is described in Table 1.

Table 1: Characteristics of input traffic

<table>
<thead>
<tr>
<th>Connection type</th>
<th>Unicast</th>
<th>Multicast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection number</td>
<td>1 2 3 4 5</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Maximum transmission rate [packets per slot]</td>
<td>500 400 200 400 300 300 200 100</td>
<td></td>
</tr>
</tbody>
</table>

The desired buffer level is 400 [packets] and the total bandwidth of the switch A is 1200 [packets per slot]. The transmission rates of unicast and multicast input traffic sources are governed by (3) and (4), respectively. The occupied buffer level of a congested link is described by (7). Figure 4 compares the transmission rates of each source connected to the switch A. Figure 4b demonstrates that a more flexible transmission rate is achieved through the proposed congestion control scheme based on a feedback mechanism considering the network status. In Fig. 4a note that each source has a higher transmission rate than available in the network since the variation of available bandwidths is not considered. This causes a buffer overflow in the network node and results in increased data loss.

Figure 5 shows the comparison of buffer levels in the switch A, where the proposed congestion control scheme exhibits a smaller buffer level than that of [15]. Moreover, the buffer levels of the congested links are still further regulated to their desired level in the case of the proposed control scheme. However, the control scheme in [15] results in a nonzero steady-state error due to the inherent limitation of the proportional control mechanism.

Table 2 shows the comparison of data loss ratios at the switch A. This comparison illustrates the more flexible scalability achieved by the proposed congestion control scheme. The data loss ratios in Table 2 include the steady-state errors of the proposed congestion control scheme.

Fig. 3 Sample network model for simulation studies
For a quantitative comparison of the network performance, we consider the following measure of fairness $J_f$ [20] as follows:

$$J_f = \max_i \left( \frac{I_i - S_i}{I_i} \right) - \min_i \left( \frac{I_i - S_i}{I_i} \right)$$  \hspace{1cm} (10)$$

for $i \in [1, N]$, where $I_i$ is the input traffic at the $i$th source, $S_i$ is the successfully transmitted data traffic at the $i$th source, and $N$ is the number of sources in the network. Table 3 illustrates the comparison of the fairness measure $J_f$ of each scheme. The proposed congestion scheme shows a smaller value of $J_f$ than either that of [15] or the max–min fair scheme of [11]. Hence one can verify that the proposed congestion scheme guarantees fairer utilisation of network resources in the WAN environment.

### 4 Conclusions

We have presented an adaptive congestion control scheme by employing the EWMA algorithm for multicast services in the IP-based network. The proposed congestion control
scheme can reduce data loss ratios and achieves a faster response at the same time. In particular, the proposed congestion control scheme can guarantee fairer operation of the network resources under the variable conditions of the available bandwidth in the network. Moreover, the new scheme is more flexible, being able to cope with an increased number of input sources in the network, which results in improved scalability. It is expected that this scheme is applicable to the internet such as real-time streaming services.

5 Acknowledgments

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Network-based Automation Research Centre (NARC) at the University of Ulsan in Korea, by a grant from a Korea Ministry of Science and Technology (Korean Systems Biology Research grant M10503010001-05N030100111), and also by the 21C Frontier Microbial Genomics and Application Centre Program, Ministry of Science & Technology (grant MG05-0204-3-0).

6 References

8 Microsoft Co., Windows Media Player, Available online: http://www.microsoft.com/windows/mediaplayer/
9 Real Networks, RealPlayer, Available online: http://www.real.com