GRCA : A Gateway-based Rate Control Algorithm for the CBR Traffic over the Wireless Link *

Youngseok Lee, Yung Yi, Youjae Won†, and Yanghee Choi

E-mail : {yslee, yiyung, yhchoi}@mmlab.snu.ac.kr
Dept. of Computer Engineering, Seoul National University, Korea
†E-mail :yjwon@mail.etri.re.kr
†Electronics and Telecommunications Research Institute, Korea

Abstract

Recently continuous media applications in the Internet usually employ their own rate control schemes. In a wireless/mobile network, multimedia applications may experience severe performance degradation by high bit error rate, low bandwidth of the wireless link, and frequent disconnections due to fading or mobility. Especially in the wireless/mobile data network without any bandwidth allocation or admission control, the low bandwidth link should be shared fairly among flows. We investigated the wireless link congestion problem by the RTP-based mobile multimedia applications and proposed a gateway-based rate control algorithm(GRCA) that employs an additive increase/multiplicative decrease control operation with the fast RTCP RR messages. By preventing the congestion over the wireless link the GRCA achieves the fast rate convergence among mobile multimedia streams as well as between RTP streams and TCP connections. It is very important in the low-bandwidth wireless network with frequent handovers to adapt effectively the high bit rate applications which may cause a significant congestion collapse. To provide a fast control mechanism by dealing only with the local wireless link congestion the GRCA is implemented at the intermediate pivot gateway, or the the RTP mixer, instead of the original sender. Therefore the limited wireless link bandwidth is utilized efficiently. We evaluated the proposed GRCA algorithm through intensive simulations with NS-2. The results showed that GRCA provides the fast rate adaptation, utilizes the wireless link bandwidth fully, reduces packet loss rate, and shares fairly the available bandwidth with RTP streams or TCP connections.

Keywords: *rate control, wireless/mobile network, CBR, RTP, handover*

1 Introduction

As the wireless/mobile networks are widely deployed, more applications and hosts use the wireless/mobile computing environment. In general the wireless network is composed of mobile hosts,

*This work is sponsored by Electronics and Telecommunication Research Institute, Korea

base stations, and mobile switching centers(Fig. 1). The wireless/mobile hosts communicate over a single wireless link connected to the high-speed backbone network.

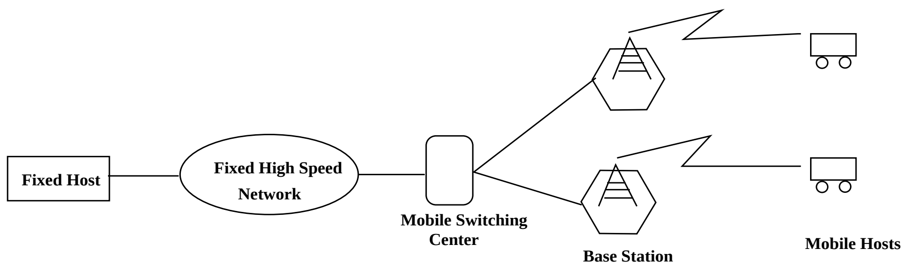


Figure 1: The conventional wireless/mobile network architecture

The current traffic statistics show that the data traffic will surpass the voice traffic in the near future. In particular, the multimedia application traffic such as video streaming is constantly increasing. In the wireless/mobile network, the multimedia applications will also grow rapidly. However there are problems that should be overcome in order to provide smooth multimedia services in the wireless network: the narrow bandwidth and the high loss rate. The reliable data delivery and the fair bandwidth utilization have become key issues because the air medium suffers from high BER(bit error rate), bursty packet loss, and low limited bandwidth. Usually packet loss occurrs due to link errors and the congestion at the wireless link.

- Wireless link error: Usually BER in the wireless link is much higher than in the wired network. And frequent disconnections from the fading or hanover can cause the packet loss in the wireless link.
- Congested wireless link: When several data flows are concentrated on the wireless link, high packet losses occur by the queue overflow at the gateway for the wireless hosts.

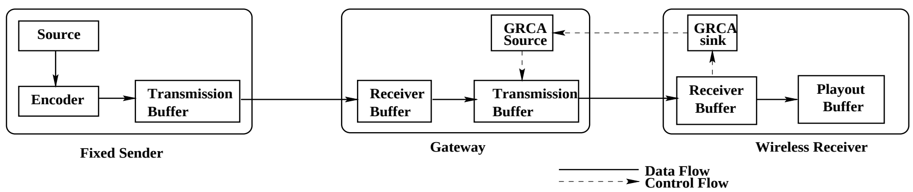
The wireless/mobile network will be an important part of the Internet and applications will not be much different from the wired Internet. Most of the multimedia applications in the Internet are based on RTP(Real-time Transport Protocol)[1] and UDP(User Datagram Protocol) over IP(Internet Protocol). RTP is just a thin protocol for the real-time applications, and it delegates error and rate control functions to the application located at the source and destination node. It has been observed that TCP traffic suffers from unfairness caused by multimedia traffic over RTP/UDP. This is mainly due to adaptive feedback based congestion control found in TCP, and not int RTP.

This situation will occur more frequently in the low-bandwidth wireless link without any QoS(Quality of Service) guarantee. To support fairness between the responsive(TCP) and the unresponsive(RTP) traffic, many end-to-end rate control algorithms[12, 13, 9] and the queue management schemes such as RED(Random Early Detection)[4] have been proposed. In the wired/wireless network the end-to-end control for the multimedia applications might not be so effective, because the bottleneck point would be the wireless link. And it takes a long round-trip time until the appropriate steps against the congestion will be useful. This is unacceptable for the multimedia applications that require real-time data delivery.

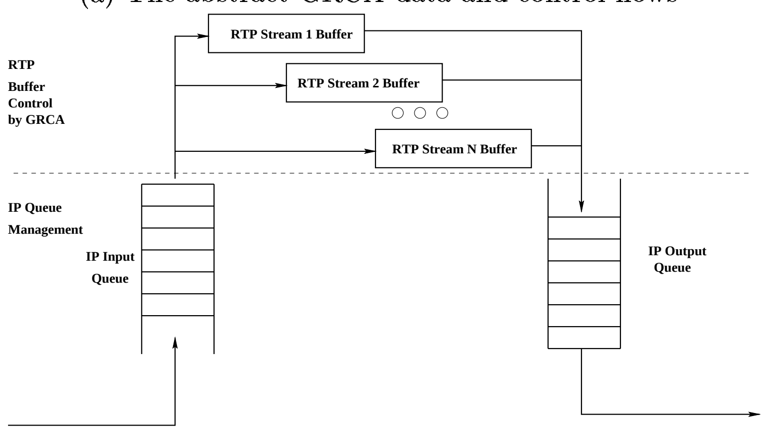
Therefore we focus on the intermediate gateway system which connects the fixed network with the wireless link. The gateway can be located at the base station or at the mobile switching center. And in order to solve the congestion collapse problem in the wireless link, we propose a gateway-based rate control algorithm(GRCA). The GRCA is mainly designed to provide the fast rate adaptation, fair share among streams on the link.

Particularly in order to support fast and smooth handover we must provide a fast rate adaption considering the multimedia sessions during handover. Under this scheme, the end-to-end response time is reduced.

The overall gateway architecture is shown in Fig. 2. The gateway uses the RTP mixer which supports the end-to-end semantics and adjusts the RTP rate with feedback from receivers. As shown in Fig. 2-(b), with the GRCA the gateway controls the RTP buffers situated on top of the IP level input/output queues.



(a) The abstract GRCA data and control flows



(b) Relationship between RTP buffers and IP queues

Figure 2: The gateway architecture

It is noted that additional functions such as the recovery of lost packets by retransmission, the bandwidth adaptation between the high speed fixed network and the wireless network by transcoding are easily incorporated in the mixer.

This paper is organized as follows. Related works about the transport protocols in the wireless network are given in section 2. The proposed gateway-based rate control algorithm will be described in section 3. The performance evaluation results using NS-2 simulator[14] are described next. Finally section 5 concludes this paper.

2 Related Works

There have been many studies[5, 6, 10] to adapt TCP in the wireless network because the successful TCP behavior in the fixed network does not guarantee good performance in the wireless network. When TCP detects packet losses from timeout or duplicated ACKs, it reduces its congestion window size by half to adapt its transmission rate to the current congestion. Packet losses in the wireless network, however, occur due to not only the congestion but also the high BER in the transmission system, fading, or handover. This may lead to the poor TCP throughput. So many studies have tried to enhance the TCP throughput in the wireless network.

Indirect-TCP(I-TCP)[5] separates the end-to-end connection into two in the wired and wireless part. The sender thinks the agent for the mobile host as the receiver, and the agent uses another independent TCP connection to the mobile host by splitting the end-to-end TCP connection. In this mechanism, it is not necessary to modify the sender. But it is complex to support the handover because the agent should keep all the TCP states. Similarly, SNOOP[17] tries to split the end-to-end connection into two, but it doesn't focus on the transport layer but the link level. The base station caches the data from the sender and retransmits them when packet losses on the wireless link occur. SNOOP also supports handover by sending data to the multicast addresses. To come up with the cumulative ACK scheme of the standard TCP and to make the response of the sender fast for the multiple packet lossses within a transmission window, SACK(Selective Acknowledgment)[18] is proposed. Also the explicit loss notification(ELN) option to the TCP can make the sender just retransmit lost data without invoking congestion control.

On the contrary, Mobile UDP(M-UDP)[7] and Mobile RTP(M-RTP)[8] are proposed to adapt non-TCP applications in the wireless/mobile network. Hence, The RTP mixer is used to decrease RTP data rate for the low bandwidth wireless link. M-RTP also focused on reducing RTCP rate at the RTP mixer for the wireless host, because the control data rate may overload the wireless link especially in case of multicasting. Also they proposed an RTP data rate control scheme which uses recoding or intelligent packet discarding with LPTSL(Loss Profile Transport SubLayer)[8]. LPTSL uses a logical data segment structure to discard packets efficiently. Although M-RTP suggests a scheme to control RTP and RTCP traffic at the Supervisor Host(SH), they didn't consider how to adjust its rate to the dynamically changing wireless network state. M-UDP can provide enhanced UDP performance(fewer lost packets) due to the appropriate retransmission schemes using SH buffer against the fade in the wireless link.

Most of the multimedia applications in the Internet adopt the aggressive and non-congestion-controlled transmission strategy based on UDP. So malicious UDP-based applications may penalize the competing TCP connections.

Recently several end-to-end adaptive protocols for the continuous media have been proposed to make RTP and UDP-based continuous media applications TCP-friendly. Loss-Delay based Adjustment Algorithm(LDA)[12] suggests a scheme to control the current source rate according to the congestion level. The LDA algorithm was designed to reduce losses and improve the utilization in a TCP-friendly way. So the rate control method resembles that of TCP; Additive Increase and Multiplicative Decrease(AIMD). It uses the smoothed loss ratio as the congestion sign. It also used the measurement of the bottleneck bandwidth to determine the additive increase ratio dynamically. The Rate Adaptation Protocol(RAP)[13] proposed a fine-grained AIMD algorithm for the streaming applications using inter-packet gap from ACK, and showed that it is fairly TCP-friendly. [9] presented a congestion control algorithm for unicast traffic using the modified version of the TCP-friendly equation. After each round of M time units, the sender estimates the round-trip time. If any packets were lost in that round, then the sender sets the sending rate to that specified by the TCP-friendly equation for the loss rate experienced during that round. If no packets were lost in that round, then the sender doubles the sending rate.

The above works didn't consider the rate adaptation speed in the wireless/mobile network.

3 Gateway-based Rate Control Algorithm(GRCA)

We modified and used RTP mixer as the gateway which behaves as the sender instead of the original source. The RTP mixer can receive QoS information such as RTT, packet loss rate and delay jitter in RTCP RR(Receiver Report) from wireless hosts. In addition, the RTP mixer can perform transcoding among several media formats.

In RFC 1889[1] the maximum RTCP rate is limited up to 5% of the session bandwidth and the interval between two consecutive RTCP messages should be at least 5 seconds. The next RTCP RR packet sending time is determined by using this bandwidth constraint, and is randomized by 0.5 - 1.5 times to suppress synchronous RTCP RR implosion. In order to support frequent session creations, handover and terminations in the mobile network, we modified RTCP of the mobile host to send RTCP RR messages more often when a event occurrs. This is because a 5 second interval between RR messages is too long to obtain stable shortly after events like the frequent handover, short-lived sessions in the wireless environment. And a 5 second interval is more strict condition than the 5% of the total session bandwidth. It means that under the 5 second interval the RTCP rate ratio will be very small compared with 5%. Therefore we relaxed the RTCP RR interval constraint of RFC1889, while keeping the RTCP rate limit condition on the long term. We defined the dynamic RTCP rate, $R_{RTCP}(t)$(% of the total session bandwidth) as a function of an event(new session or handover) start time t, as follows;

$$R_{RTCP}(t) = R_{width} \cdot e^{a \cdot t} + R_{min}, (0 < R_{min}, R_{width} \leq 5, R_{min} + R_{width} < 5, a < 0)$$

The constant a is used to control the decreasing speed of the RTCP rate to the minimum value. The RTCP messages are generated at the maximum rate of $(R_{min}+R_{width})$ at first, and then decreased exponentially to the R_{min}. Under this scheme, the sender can adjust its rate more often just after an event occurrs. But as the link reaches stable state with fair bandwidth sharing among streams, rate adjustment happens less frequently.

The RTCP messages include QoS information like delay, jitter and loss rate. The average loss rate indicates the coarse network state experienced by the receiver, while the delay and jitter information are used as the fine-grained control parameters.

We assumed the CBR video traffic source. The inter-packet gap(IPG) of the CBR traffic will be almost same over the single hop wireless link in a small cell coverage without critical fading or wave blocking. The smoothed IPG value sent in RTCP RR messages by the mobile host is used for the rate control at the gateway. A mobile host computes the smoothed IPG value on every received RTP packet and reports this value in RTCP RR profile-specific extensions field. Because the RTCP RR message informs the gateway of the link QoS with negligible delay, the smoothed IPG is considered to reflect the relative current network state by the low pass filter. The smoothed IPG for the ith received RTP packet is computed as follows;

$$Smoothed\_IPG_i = \alpha \cdot Smoothed\_IPG_{i-1} + (1-\alpha) \cdot Sample\_IPG, \; (0 \leq \alpha \leq 1)$$

The $Sample\_IPG$ is the inter-arrival time between $(i-1)$th and ith RTP packets and $Smoothed\_IPG_i$ is computed with this sample and the previous $Smoothed\_IPG_{i-1}$ as shown in Fig. 3. When a successive RTP packet loss occurrs, the $Sample\_IPG$ becomes the interval from the previously well received packet to the last received one.

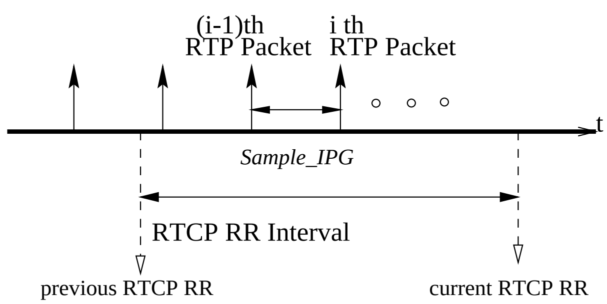


Figure 3: The RTCP RR interval

The smoothed IPG represents the recently experienced network status by the mobile host. Therefore when a gateway receives an RTCP RR message from a mobile host, it classifies the wireless link state as *congested* or *unloaded* by comparing the $Smoothed\_IPG$ with the pre-defined thresholds. The *unloaded* state means that IPG has not been altered much, and in this case the gateway increases its rate optimistically. The *congested* is detected when IPG value exceeds the threshold. The gateway decreases its rate by the multiplicative decrease method against the *congested* state. It increases its rate additively for the *unloaded* state.

Algorithm 1 GRCA: Gateway-based rate control algorithm

```
1: if (Smoothed_IPG_i ≥ Congestion_threshold) then
2:     Multiplicative Decrease
3: else
4:     Additive Increase
5: end if
```

The detailed GRCA behavior is as follows:

- **Additive Increase(AI)**
  GRCA increases the transmission rate additively in a big step, which will lead to reach the maximum available bandwidth rapidly. The transmission rate(R_i) is dynamically increased by the predicted available bandwidth(Δ_i).

$$R_i = \frac{|P_i|}{IPG_i} = R_{i-1} + \Delta_i$$
$$\Delta_i = \Delta \cdot (1 - \frac{r}{b})^{f(n)}$$

  P_i represents the current received packet size, r denotes the current transmission rate and b for the estimated wireless link bandwidth which is computed as the packet size over the minimum inter-packet arrival time[12]. In the additive increase phase the increasing ratio(Δ_i) indicates the predicted available bandwidth, i.e. the maximum sending rate. Increasing sending rate too highly will result in high packet loss. In order to prevent the over-estimated transmission rate, we added a function, $f(n) = c \cdot n$(c is the coefficient for the rate increment ratio and n is the total number of RTP sessions). This function prohibits the gateway from going beyond the maximum available bandwidth too fast. Δ is the maximum increasing rate by one host at once, and is set by considering the total RTP session counts, $N_{session}$.

$$\Delta = k \times \frac{b}{N_{session}}$$

  k represents the average bandwidth usage coefficient by each session(We acquired the appropriate value for k after several simulations. In this paper it is 0.8.).

- **Multiplicative Decrease(MD)**
  MD tries to recover from the congested state by decreasing the transmission rate multiplicatively.

$$R_i = R_{i-1} \cdot MDR$$
$$MDR = m^{-\frac{r}{b}}, (m > 1)$$

MDR makes the gateway reduce its sending rate in proportion to the current rate r, in order to rapidly converge to the fairness state. The reduced rate is bounded to $\frac{1}{m}$ to guarantee the minimum rate and to avoid the abrupt rate reduction.

With the GRCA at the gateway, the RTP sessions over the wireless link can share fairly the limited bandwidth. Especially dynamic AI/MD functions adapt the current sending rate fast enough to the fair state for all sessions.

The protocol stack at the gateway implementing the GRCA algorithm in RTP mixer is illustrated in Fig. 4. The Mobile RTP(MRTP) at the gateway uses the modified RTP protocol for the RTP mixer, and it notifies new events in the RTCP SR message. The MRTP of the mobile host should report the smoothed IPG and the estimated bandwidth of the wireless link in the RTCP RR message. And the application level gateway is used for transcoding and handling hierarchical or multiple coded data received from the source[3]. The GRCA assumes that this application level gateway converts the original data to the appropriate low bit rate media format for the wireless host. Also the mobile host and the gateway can use the mobile IP protocol to provide the IP-level mobility and handover functions. The mobile host uses a modified RTP protocol to report the IPG instead of the RTT value.

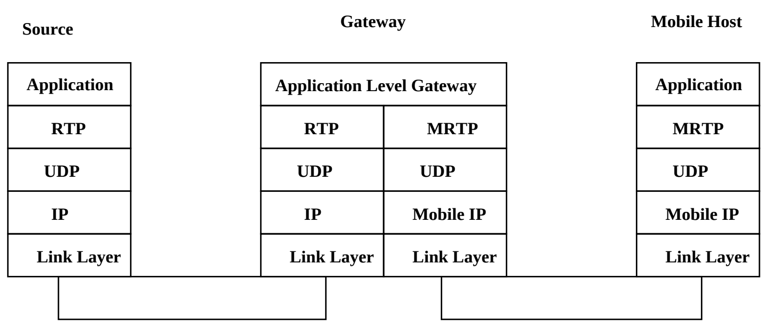


Figure 4: Gateway protocol stacks

4 Performance Evaluation

In this section, the simulation results of the GRCA algorithm with ns2 are given. We have implemented GRCA algorithm in the session-relaying mixer and used two different the packet loss models at the wireless link.

4.1 Packet Loss Models at the Wireless Link

Two packet loss models are considered in our simulation: uniform and two-state Markov distribution. The uniform error model reflects the pure randomness of the wireless link error. The average packet loss rate in the uniform distribution is set to 1.5% in this paper. Two-state Markov model is known to be simple but reflects the burstiness of wireless link error very well[15]. The basic states of two-state Markov model are *error* and *error free* which have its own distribution. When a packet is sent in the *error* state channel, it will be lost. The state transition diagram of two-state Markov model is shown in Fig. 5.

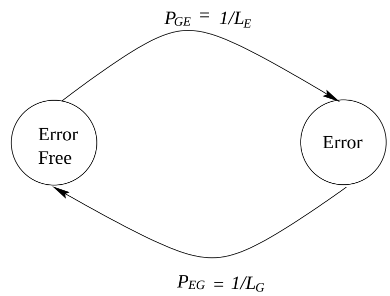


Figure 5: Two-state Markov model

P_{GE} denotes the transition probability from the *error free* state to the *error* state which is defined by the inverse of the state occupancy time L_E, and the transition probability from the *error* state to the *errorfree* state(P_{EG}) with L_G. The Markov property of each state can make us to express transition probability between each state by the state occupancy time. In this paper the state occupancy time is expressed by the number of packets($L_E = 3$,$L_G = 200$). On the other hand, it is assumed that the packet loss probability at the wired link is very low(10^{-9}).

4.2 Simulation Network Configuration

Fig. 6 shows the network topology used for the simulation.

As shown in Fig. 6, the wireless network is a single hop link. We placed various RTP and TCP connections on nodes in the above topology and assigned lower bandwidth, higher packet error rate, and higher delay in the wireless link than in wired one. Therefore the wireless link between $R0$ and $D0$ becomes always the bottleneck. We considered only droptail queue in this paper.

Table 1 explains the RTP session parameters used in the simulation: the start sending rate at the gateway(R_g). R_s and the original sending rate. It is assumed that the gateway sends data in a reduced data rate by transcoding. And the start sending rate values are set differently to examine the effects of the rate control functions. In order to consider the handover we add or delete the sessions dynamically.

4.3 Performance Metrics

The simulation was executed to evaluate performance aspects of the GRCA behavior. At first we examined the fairness among RTP streams, and between RTP streams and TCP connections. With

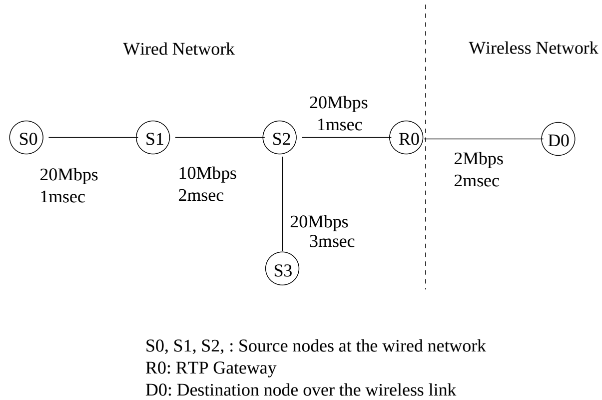


Figure 6: Simulation network topology

Table 1: Simulation parameters: RTP sessions

RTP session	R_s	R_g
RTP0(S0-D0)	1500Kbps	256Kbps
RTP1(S0-D0)	1500Kbps	640Kbps
RTP2(S1-D0)	1500Kbps	640Kbps
RTP3(S1-D0)	1500Kbps	256Kbps
RTP4(S1-D0)	1500Kbps	256Kbps
RTP5(S3-D0)	1500Kbps	640Kbps
RTP6(S3-D0)	1500Kbps	256Kbps
RTP7(S3-D0)	1500Kbps	640Kbps

this metric, it is shown that the GRCA utilizes the wireless links in the mobile network both fairly and fully. And in order to investigate how fast the GRCA can support the handover streams, the rate convergence time to the bounded fairness state was measured. Also the loss rate was used to estimate how well the GRCA adjusts the rate of streams to the network status.

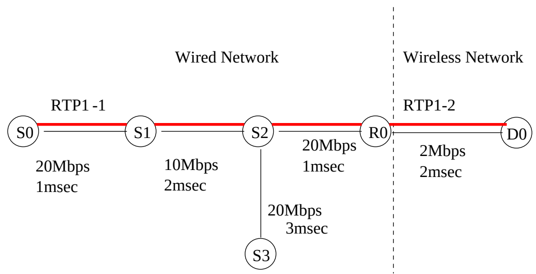
We define the intra-protocol fairness ratio, $F_{intra}(t)$ which is the coefficient of variation(C.O.V.) of the sending rate in the active session i at time t, $R_r^i(t)$.

$$F_{intra}(t) = C.O.V.(R_r^i(t)) = \frac{\sigma(R_r^0(t), R_r^1(t), ..., R_r^N(t))}{mean(R_r^i(t))}$$

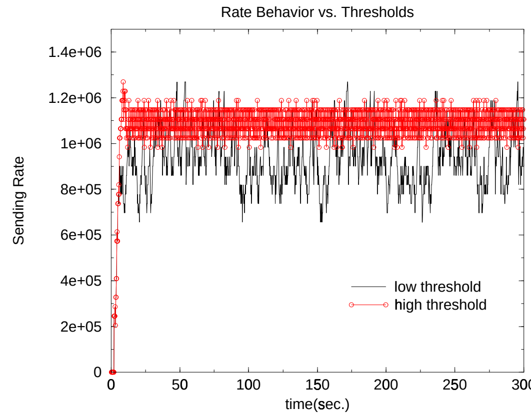
The inter-protocol fairness ratio $F_{inter}(t)$ is the ratio of the average RTP session rate to the average TCP connections rate.

$$F_{inter}(t) = \frac{mean(R_r(t))}{mean(R_t(t))}$$

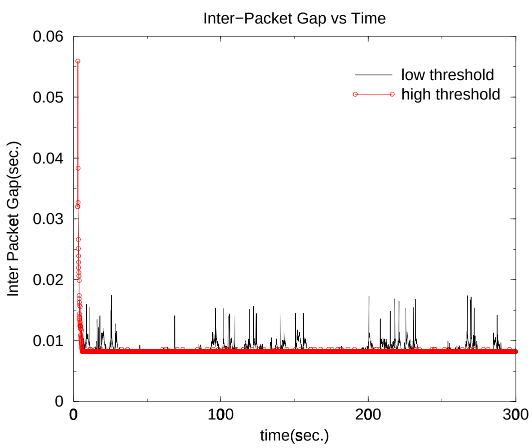
4.4 Rate Control Behavior vs. Thresholds



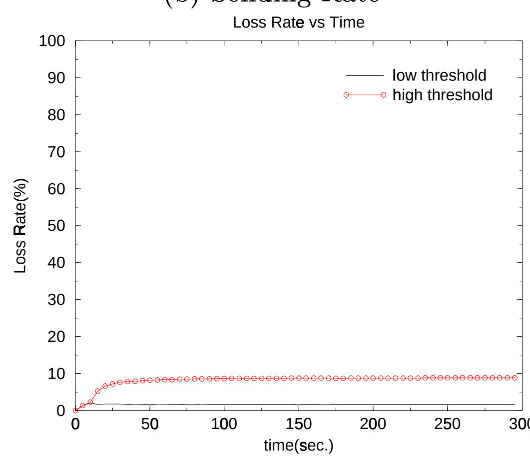
(a) Simulation Network



(b) Sending Rate



(c) IPG values



(d) Cumulative loss rate

Figure 7: Rate control behavior vs. the threshold values: sending rate, loss rate and IPG values

To explain the effects of the threshold values, we tested one RTP session on the given network in the Fig. 7-(a) with two different $congestion\_thresholds$(x= 1.0 or 1.3). The threshold value is defined by $x \times current\ IPG$($x \geq 1$). When the threshold is low, the gateway responds frequently to a slightly increased IPG and tries to reduce its rate to solve the wireless link congestion collapse as shown in Fig. 7-(b), because the receiver sends a congestion sign by the high smoothed IPG value(Fig. 7-(c)). Thus this rate control behavior can be regarded as *active*. The *active* behavior means that it makes the gateway adjust its sending rate strongly to a less congested state, resulting in low throughput at the receiver.

On the other hand, in case of the high threshold, the gateway considers most packet losses due to the simple wireless link error and keeps its rate. Its behavior can be thought of *passive*. Therefore under the low threshold, the mobile host experiences low loss rate(Fig. 7-(d)). The *passive* rate control tries to persist its rate by ignoring most of the packet losses. Therefore this makes the GRCA reach the maximum available bandwidth as soon as possible. The threshold value plays an important role to discriminate the congestion error from the wireless link error. Therefore the threshold value should be set according to the appropriate gateway control policy. In this paper the congestion threshold was set to 1.05 by heuristics.

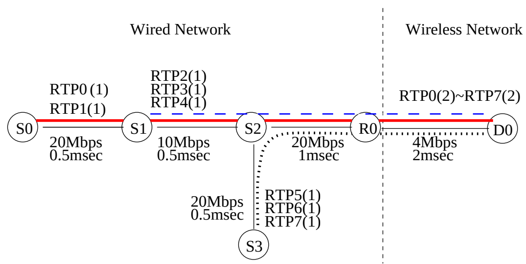
4.5 Intra-protocol Fairness

Fig. 8-(a) describes configured RTP sessions on the simulated network topology to evaluate the GRCA performance among 8 RTP sessions. For convenience, we denote RTP sessions in the fixed network as RTPi(1), and RTPi(2) in the wireless network respectively for the same session RTPi. For the dynamic session behavior, we generated events such as start(or incoming handover)/end(or outgoing handover) as shown at Table 2.

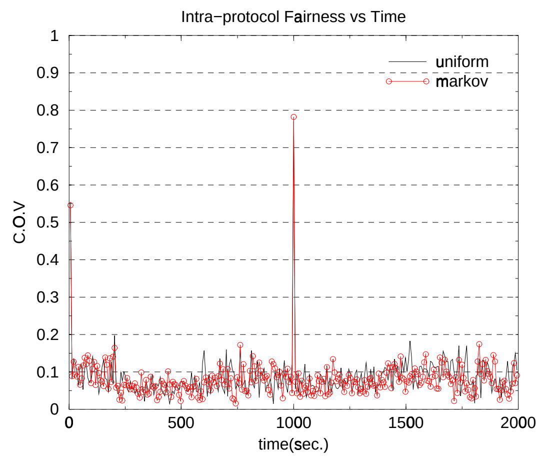
Table 2: RTP Session Events

Time	Events	Sessions
200 sec.	Session End	RTP0, RTP1
600 sec.	Session End	RTP2, RTP3
1000 sec.	Session Start	RTP0, RTP1
1400 sec.	Session Start	RTP2, RTP3
1700 sec.	Session End	RTP0, RTP1, RTP2, RTP3

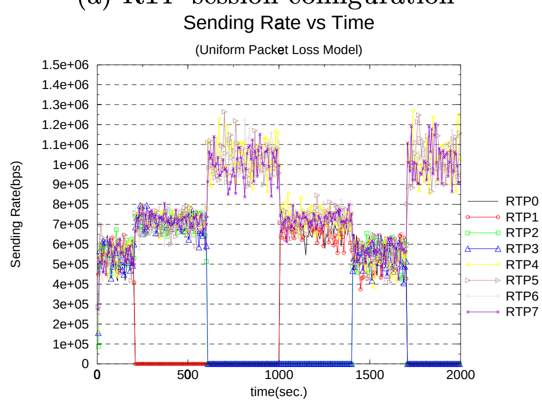
The simulation was performed with this configuration during 2000 seconds. The fairness ratio, which is defined with C.O.V is shown in Fig. 8-(b) under the drop-tail queueing according to each error model. The X axis is time(sec.) and the Y represents C.O.V of rate samples at time t. Fig. 8-(b) explains that all RTP sessions fairly share the bandwidth most of time(average C.O.V is 0.086, 95% of C.O.V are under 0.14 and 99% under 0.2), except a few spike points such as session add or deletion(for example at time 1000). At events, the GRCA adjusts quickly its rate to the fair state by fast event-driven RTCP RR messages. The dynamic adjusted sending rates at the gateway under session handover are shown in Fig. 8-(c), (d) according to each error model. The loss rates under uniform and two-state Markov error model error model are about 6.7% and 7.1% respectively. Moreover the GRCA utilizes the wireless link fully(the average is about 94%). From these results, we can know that the GRCA can adapt its rate in a fair manner in the homogeneous RTP traffic environment, and can adjust RTP rates rapidly in the mobile network with handover.



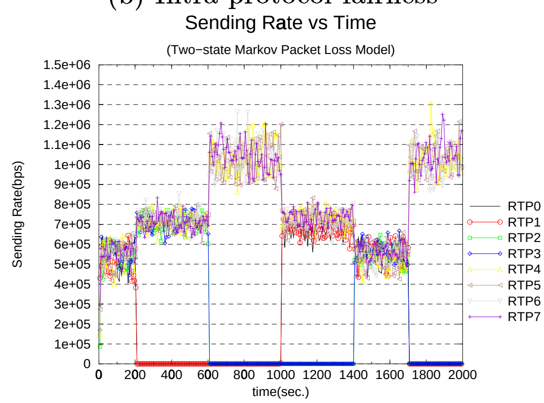
(a) RTP session configuration



(b) Intra-protocol fairness



(c) RTP Sending Rate(uniform)



(d) RTP Sending Rate(Markov)

Figure 8: Intra-protocol fairness

To investigate the effects of fast and event-driven RTCP, we examined the RTCP rate whose sessions(RTP4, 5, 6, and 7) lived during the whole simulation period in Fig. 9. Though the RTCP rate increases when events occur, the each average rate ratio per session bandwidth is kept at about 1%. This can be extended to the multicast session in a cell under the bandwidth limit.

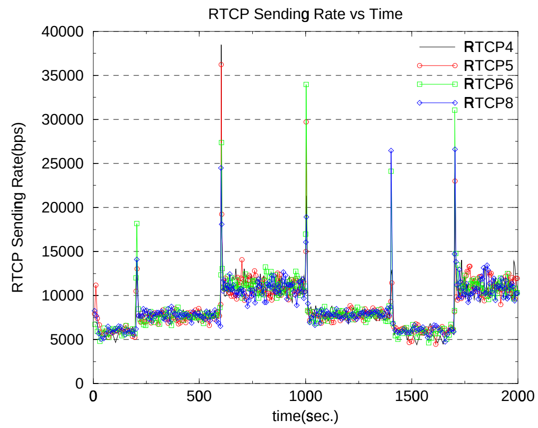


Figure 9: RTCP Rate

4.6 Inter-protocol Fairness

We evaluated the performance of GRCA algorithm on how much it adapts to TCP connections, and how fast it converges. Because TCP traffic dominates the current Internet, the multimedia streams using RTP should coexist fairly with TCP in the best-effort network.

In general TCP responds immediately to the congestion because TCP recognizes one packet loss as a congestion symptom while RTP sender gets the congestion hints from RTCP RR message only after RR interval. In this situation, it is possible that the gateway gets RTCP RR message after the TCP sender decreased multiplicatively its sending rate immediately by reducing its window size. This may cause the gateway to increase the RTP rate further. Although the RTCP RR interval makes the gateway stay in a decreased or increased state for a while compared with TCP's abrupt state transition, the rate of the RTP stream converges to the steady state which shares the bandwidth with TCP connections fairly by adopting AI/MD control method and the fine-tuned threshold.

For simuation the transport connection configuration is composed of 8 RTP streams and 6 TCP connections competing for the link bandwidth dynamically. The session/connection start/end events are as follows(Table 3).

TCP used in this simulation is TCP-Reno deployed widely. The sending behavior of the original RTP and relayed RTP streams is same as RTP1 at the previous case. All TCP flows are "FTP" sessions with an infinite amount of data as shown in Fig. 10-(a). The packet size is 512 bytes and the maximum window size is set to 1000 bytes.

The Fig. 10-(b) shows the GRCA fairness between RTP and TCP connections in the droptail queue according to each error model. Under two packet loss models, the GRCA maintains the fair rate ratio of the RTP over TCP(the average RTP/TCP ratio is 0.97, 0.91 under two-state Markov

Table 3: RTP Session and TCP Connection Events

Time	Events	Sessions
200 sec.	Session End	RTP0, RTP1
600 sec.	Session End	RTP2, RTP3
1000 sec.	Session Start	RTP0, RTP1
1400 sec.	Session Start	RTP2, RTP3
1700 sec.	TCP End	FTP0, FTP1, FTP2

packet loss model). Since the burst packet losses(average loss rate is about 4.45%) will occur more often in the two-state Markov model, the throughput in this case is a little lower than that of the uniform loss model(the average loss rate about 3.9%). The average sending rates of RTP and TCP at the gateway are shown in Fig. 10-(c), (d). There are still some peak points when some events happen. But after this time, the rate ratio is maintained fairly. When events related with the RTP session occur, the GRCA adapts its rate quite well in a fair manner, while it slightly yields to the TCP connections in case of TCP events(after 1700 sec.). This is because the GRCA increases its rate by considering the current RTP session counts. Therefore if the GRCA can be aware of the TCP connections, it will guarantee the inter-protocol fairness quite well.

Through simulation, we evaluated the GRCA performance especially in terms of intra-protocol and inter-protocol fairness. It is shown that the GRCA provides the fast intra-protocol fairness and the full link utilization with low loss rate under dynamic RTP session behavior in the mobile network. Also in the heterogneous environment mixed with the TCP connections, it still has a acceptable fairness ratio between the RTP and TCP.

The mobile hosts experience the quality degradation when the gateway increases or reduces abruptly and freqeuntly. It will, however, be smoothed by the buffering at the receiver[19].

5 Conclusion

In this paper, we proposed a gateway-based rate control algorithm(GRCA) for the CBR multimedia traffic in the mobile network, which employs an additive increase and a multiplicative decrease with the fast event-driven RTCP messages including smoothed inter-packet gap measured at the mobile host. The GRCA provides the rate adaptation over the wireless link and supports handover efficiently by adjusting the RTP rate dynamically. The simulation results of the proposed algorithm with ns-2 showed that the GRCA has good features for the mobile multimedia applications such as fast convergence to the fair link utilization, low error rate, and maximum throughput. Especially the GRCA supports the CBR traffic application handover by ensuring fast rate convergence in the mobile network with the low bandwidth wireless link. Also it provides the maximum utilization of the low bandwidth of the wireless link, and ensures fair sharing of the bandwidth with other streams. The limited wireless network bandwidth is fair shared by homogeneous RTP streams as well as by the RTP streams and TCP connections. Therefore the GRCA can contribute to prevent the congestion collapse at the wireless link.

Though the GRCA is not targeted at the full end-to-end solution, it can coexist with other end-to-end congestion control mechanisms, resulting in the total enhanced throughput and keeping the wireless link robust and stable. For the effective rate control at the gateway, it is necessary that the application level gateway can transcode from high bit rate traffic to the appropriate one, when the source cannot send the hierarchical/multiple coded data. The gateway adapts the continuous rate in the fine-grained level by using the smoothed inter packet gap between the gateway and the mobile host.

In this paper, we assumed only CBR traffic which sends packets periodically. Therefore it is necessary to extend the GRCA for the VBR traffic sources. The threshold used by GRCA is set statically, but it should be dynamically adaptive so as to discriminate the congestion from the wireless link effectively. In order to apply the GRCA at the gateway efficiently, the coarse-grained discrete rate control method for the hierarchical or multiple coding should be devised.

References

[1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *"RTP: A Transport Protocol for Real-Time Applications"*, IETF RFC1889, Jan. 1996

[2] K. Claffy, G. Miller and K. Thompson, *"The Nature of the Beast: Recent Traffic Measurements from an Internet Backbone"*, INET'98, 1998

[3] E. Amir, S. McCanne, and Hui Zhang, *"An Application Level Video Gateway"*, ACM Multimedia'95, 1995

[4] S. Floyd, S., and V. Jacobson, *"Random Early Detection gateways for Congestion Avoidance"*, IEEE/ACM Transactions on Networking, Vol.1 No.4 pp. 397 - 413, August 1993

[5] A. Bakre, B. Badrinath, *"I-TCP: Indirect TCP for Mobile Hosts"*, 15th Intl. Conf. on Distributed Computing Systems, May 1995

[6] K. Brown and S. Singh, *"M-TCP: TCP for Mobile Cellular Networks"*, ACM Computer Communication Review , Vol. 27, No. 5, pp. 19-43, 1997

[7] K. Brown and S. Singh, *"M-UDP: UDP for Mobile Computing"*, SIGCOMM Computer Communication Review, Vol. 26, No. 5, pp. Oct. 1996

[8] K.Brown and S.Singh, *"Extensions to RTP to support Mobile Networking"*, The 3rd Intl. Workshop on Mobile Multimedia Communications, 1996

[9] J. Padhye, J. Kurose, D. Towsley, R. Koodli. *"A TCP-Friendly Rate Adjustment Protocol for Continuous Media Flows over Best Effort Networks"*, UMass-CMPSCI Technical Report TR 98-04, October 1998

[10] H. Balakrishna, V. N. Padmanabhan, S. Seshan, and R. H. Katz, *"A Comparison of Mechanisms for Improving TCP Performance over Wireless Links"*, IEEE/ACM Transactions on Networking, Vol. 5, No. 6, pp. 756-769, Dec. 1997

[11] A. Seneviratne, and B. Sarikaya, *"Cellular Networks and Mobile Internet"*, Computer Communications, Vol. 21, No. 14, pp.1244-1255 , 1998

[12] D. Sisalem, H. Schulzrinne, *"The Loss-Delay Adjustment Algorithm: A TCP-friendly Adaptation Scheme"*, Network and Operating System Support for Digital Audio and Video (NOSSDAV), 1998

[13] R. Rejaie, M. Handley, D. Estrin, *"RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet"*, IEEE INFOCOMM'99, 1999

[14] S. McCanne and S. Floyd, *Ns(network simulator)*, http://www-mash.cs.berkeley.edu/ns

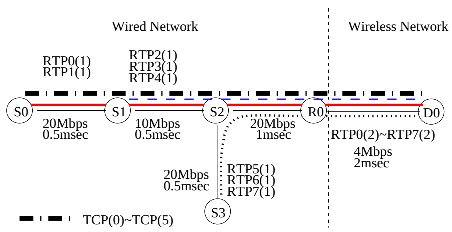
[15] G. T. Nguyen, R. H. Katz, B. D. Noble, and M. Satyanarayanan, *"A Trace-based Approach for Modeling Wireless Channel Behavior"*, in Proc. Winter Simulation Conf., Dec. 1996

[16] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, *"An Architecture for Differentiated Services"*, RFC 2475, 1998

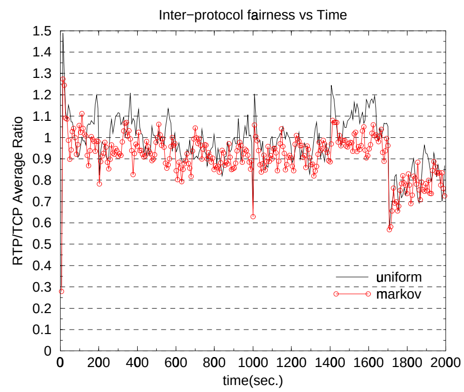
[17] H. Balakrishman, S. Seshan, and R. H. Katz, *"Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks"*, ACM Wireless Networks, Vol. 1, No. 2, pp. 469-481, Dec. 1995.

[18] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, *"Selective Acknowledgement Options"*, RFC2018, 1996.

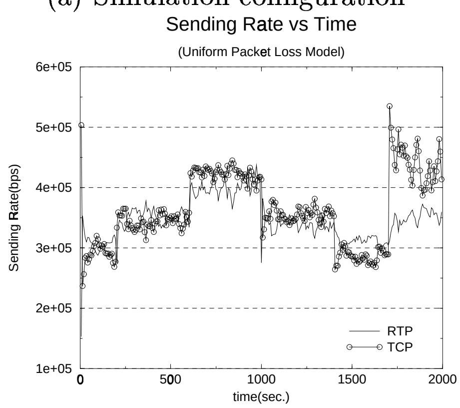
[19] R. Rejaie, D. Estrin, and M. Handley, *"Quality Adaptation for Congestion Controlled Video Playback over the Internet"*, SIGCOMM'99, 1999.



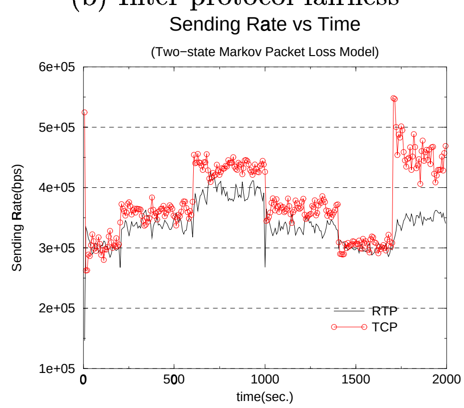
(a) Simulation configuration



(b) Inter-protocol fairness



(c) Sending Rate(uniform)



(d) Sending Rate(Markov)

Figure 10: Inter-protocol fairness ratio