논문 2006-43TC-12-5

이동통신망에서의 모바일 컨텐츠 서버 통신을 위한 최적의 TCP 세그먼트 길이

(Optimal TCP Segment Size for Mobile Contents Server Access over Wireless Links of Cellular Networks)

이구연*, 정충교*, 김화종*, 이용**

(Goo Yeon Lee, Choong Kyo Jeong, Hwa Jong Kim, and Yong Lee)

요 약

이동통신망을 통한 인터넷 접속은 제한된 대역폭 및 무선 접속 링크에서의 높은 bit 오류율로 인하여 병목현상을 야기시킨다. 이러한 병목현상에 따른 TCP 통신의 성능저하를 막기 위해 TCP 연결의 여러 파라미터를 최적화 시키는 작업은 필수적이다. 본 논문에서는 모바일 컨텐츠 서버와의 통신시에 무선링크의 이용 효율을 최대화 시키기 위한 최적의 TCP 세그먼트 길이에 대한 알고리즘을 제시한다. 또한 본 논문에서는 제시된 알고리즘에 대한 몇 가지의 예를 제시한다. 제시된 예에서 보듯이 TCP 세그먼트 길이는 TCP 윈도우 사이즈가 어느 임계값 이상이 되면 더 이상 변하지 않고 고정된 값을 갖음을 알 수 있다. 본 논문의 결과를 이용하면, 단순히 TCP 세그먼트 길이만을 조정하므로서, 모바일 컨텐츠 서버 접속시에, 값비싼 무선 링크의 이용 효율을 최대화 시킬 수 있을 것이다.

Abstract

Internet access from mobile phones over cellular networks suffer from severe bandwidth limitations and high bit error rates over wireless access links. Tailoring TCP connections to best fit the characteristics of this bottleneck link is thus very important for overall performance improvement. In this work, we propose a simple algorithm in deciding the optimal TCP segment size to maximize the utilization of the bottleneck wireless TCP connection for mobile contents server access, taking the dynamic TCP window variation into account. The proposed algorithm can be used when the product of the access rate and the propagation time is not large. With some numerical examples, it is shown that the optimal TCP segment size becomes a constant value when the TCP window size exceeds a threshold. One can set the maximum segment size of a wireless TCP connection to this optimal segment size for mobile contents server access for maximum efficiency on the expensive wireless link.

Keywords : optimal segment size, cellular network, mobile contents server access

Ⅰ. Introduction

In recent years cell phone multimedia services have proliferated in several countries, where the telecommunications operators' revenues from these services are beginning to take a considerable share among the total revenues. This trend is spreading all over the world.

Transfer of the pictures or video taken by cell

---

* 정회원, 강원대학교 컴퓨터학부
(Dept. of Computer Eng. Kangwon National University)
** 정회원, 삼성전자 정보통신총괄 통신연구소
(Samsung Electronics, TN R&D Center)

※ 본 연구는 정보통신연구진흥원의 대학 IT연구센터 (ITRC) 육성사업(ITAC1090060300350001000100100)의 지원을 받아 이루어 졌으며 또한 강원대학교 정보통신연구소의 지원을 받아 연구되었습니다.
접수일자: 2006년5월3일, 수정완료일: 2006년12월12일

phone's embedded camera, download of MP3 music files, video or movies are typical examples of the wireless multimedia applications. Key characteristics of these services can be summarized as follows:

- Cell phone multimedia applications follow the server-client model rather than the peer-to-peer model. Users do not want their cell phones to utilize peer-to-peer networks because of the lack of power, bandwidth, and memory in cell phones. Cell phone users usually receive multimedia contents through the services provided by the telecommunications operators.
- The cell phone multimedia application servers normally use proprietary protocols at the application layer while using TCP connections for the transport layer. Recently, the WAP Forum thoroughly revised its specifications to adopt TCP as its transport protocol in order to get convergence with Internet standards[1-2].
- The locations of the cell phone multimedia application servers are carefully selected. The servers are usually connected to the mobile switching center directly via a high speed link to avoid possible congestions on the wired network which will cause the under-utilization on the valuable wireless link. Therefore a TCP connection between the server and the cell phone is composed of a low-speed error-prone wireless path and a high-speed congestion-free wired path.
- Cell phones use only one TCP connection for most of the time. This is the result of the functional simplicity of applications, which is imposed by limited user interface capabilities. Availability of parallel TCP connections might be limited by design because of the limited hardware resources.
- The data transmission rate of cell phones is still very low compared to the wired Internet access rate, but steadily increasing.

Generally, a TCP connection over a wireless network shows serious performance degradation due to retransmissions on the wireless link, in which the bit error rate is high. Various schemes have been proposed to solve this problem, including Ack suppression and local retransmissions on the wireless link[3], and splitting the TCP connection between the wireless link and the fixed network[4]. These approaches, however, require fairly large modifications to the conventional TCP protocol or break the end-to-end semantics of the transport layer protocol. Several minor optimizations or recommended settings for TCP in wireless environment have been well documented by IETF[5], where optimizations and settings can be used without big changes from the traditional TCP protocol.

Other schemes to improve wireless TCP connections include path maximum transfer unit (MTU) discovery, limited transmit capability, increased initial window, sufficient receiver and sender buffer, selective acknowledgement option, and ECN(Explicit Congestion Notification) support[3]. It is also suggested to maintain a large IP MTU by way of layer two ARQ and transparent link layer fragmentation features. Without these features in the link, smaller MTU is preferred because smaller packet size increases the chance of successful transmission. For a fixed bit error rate, the probability of packet loss increases exponentially and so does packet retransmission probability, as the packet length gets larger. However, too small a packet size would result in a low efficiency for a given packet overhead.

In this work, we analyze the relation between the TCP segment size and the wireless link utilization in an environment in which cell phones access mobile contents service through wireless TCP connections as describes above. A formula that gives an optimum TCP segment size is derived in this analysis. The telecommunications operators would be interested in the result because it enables them to accommodate more users just by a carefully adjusting the segment size and thus maximizing the wireless link utilization. The result shows that a change in the TCP segment size makes a considerable difference in the wireless link utilization.

Following this introduction, Section Ⅱ describes the model used in the work and analyzes the relation between the TCP segment size and the utilization on the wireless access link under the condition that the TCP window size is fixed. In Section Ⅲ, we investigate on the applicability of the analysis results to the real world, where the TCP window size changes dynamically. Conclusion follows in Section Ⅳ.

Ⅱ. Model and Aalysis

Wireless access from a cell phone over a TCP connection to a mobile contents server located close to the mobile switching center is modeled in this section. Most of mobile Internet application servers are located close to the mobile switching center (in terms of delay) and connected via a sufficiently wide bandwidth link since the telecommunications operators do not want to waste the expensive wireless link bandwidth due to wired network congestion. Therefore the wireless TCP connection is usually composed of a slow wireless access link and high speed non-congested wired network. The wideband and non-congested wired part results in a small and almost fixed delay.

We assume that cell phones normally use only one TCP connection. Then a TCP connection can use the entire physical transmission bandwidth on the wireless access link connecting the cell phone to the wired network. How well the TCP connection utilizes this physical bandwidth decides the economics for the telecommunications operators.

Under the above assumptions, a TCP connection between a cell phone and the mobile contents server has the following characteristics:

- The delay in the wireless path is determined by the distance between the cell phone and the base station. The high speed wired path does not suffer from congestion and thus generates a fixed amount of delay. If we assume that the response time of the server is fixed, we have a constant end-to-end round-trip time. With a constant round-trip time the TCP retransmission timer behaves just like a negative acknowledgement. In this case a TCP connection can be adequately modeled by a selective repeat ARQ.
- A TCP connection is different from a link layer selective repeat ARQ connection in that the window size varies during the lifetime. However the real transmission capacity of a TCP connection is limited by the link transmission capacity, not by the window size, if the access link speed is very low compared to the transmission speed of the wired part and there is only one TCP connection on the access link. In other words, the transmission capacity of a TCP connection is fixed and almost the same as the link capacity except when the window size is very small. This is in contrast to the usual wired network cases, where the TCP transmission rate is normally limited by the window size.

Our main concern lies in the analysis of the bottleneck utilization of the wireless TCP connection, where the bottleneck utilization is defined by the utilization of the bottleneck link, which is monopolized by the TCP connection. In this work, we focus on the effective bottleneck utilization defined by the average transmission rate of the application data excluding TCP/IP and link overheads divided by the total link capacity. We will derive an expression of the steady state effective bottleneck utilization for the wireless TCP connection, and suggest a guideline to get an optimal TCP segment size in order to maximize the effective bottleneck utilization.

The effective bottleneck utilization of the wireless TCP connection is affected by the retransmission probability, packet overhead, and window size. We assume that a lost packet is retransmitted after a round-trip time. That is, the time needed to detect a packet loss is ignored. Note that we have assumed the round-trip time variance is very small and, therefore, the timeout value of the TCP retransmission timer can be set tight. Slow start is not considered in our steady state analysis.

In the following analysis, we assume that the window size is constant for the time being. We will investigate the effect of the varying window size in Section 3.

1. Optimal TCP segment Size

In a selective repeat ARQ model which is widely used in modeling a link layer, the link utilization, U, is well known[6] and is given by

$$U = 1 - P_e \text{ when } N \geq 2a+1 \tag{1}$$

$$U = \frac{N}{2a+1}(1-P_e) \text{ when } N \leq 2a+1 \tag{2}$$

where N is the window size (in transmission frames), P_e is the frame-error probability, and a is the normalized propagation delay defined by

$$a = \frac{\text{propagation time}}{\text{transmission time}} \tag{3}$$

We now derive the effective bottleneck utilization of the wireless TCP connection which is composed of wireless and wired paths using (1) and (2). The effective bottleneck utilization of a wireless TCP connection can be expressed as the product of the wireless TCP connection utilization and the link efficiency on the bottleneck wireless link.

The TCP connection utilization can be calculated from (1) or (2) if we use the selective ARQ model for the TCP connection. In this calculation P_e should be interpreted as the probability that a TCP segment fails to be delivered to the destination. This probability can be approximated by the probability that a TCP segment fail to go through the wireless link without an error. In this work, we use the term frame error probability in this sense, i.e., the frame error probability means the probability that a TCP segment fails to go through the wireless link because of errors on the wireless link which are not recoverable by the link layer protocol.

The link efficiency is defined by $\frac{L_P}{L_P + L_O}$, where L_P and L_O is defined in Fig. 1.

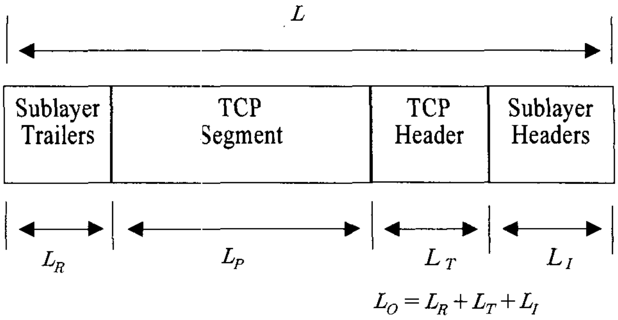


그림 1. 무선 링크에서의 프레임 형식

Fig. 1. Frame format on the wireless link.

The normalized propagation delay a can be written as $a = \frac{t_p}{t_r}$, where t_p is the end-to-end delay and t_r is the TCP Protocol Data Unit (TPDU) transmission time.

If we assume that the TPDU is transmitted entirely embedded in a frame, not segmented into several frames, the time required to transmit a TPDU is the same as the time to transmit a frame, which is given by $\frac{L}{R}$, where L is the frame length, and R is the transmission rate on the wireless link. Therefore, we have

$$a = \frac{t_p}{t_r} = \frac{t_p}{L/R} = \frac{R \cdot t_p}{L}. \tag{4}$$

We insist that the effective bottleneck utilization measures the key characteristics of the wireless TCP connection, i.e., the TCP connection's efficiency and the link overheads on the bottleneck wireless link. This effective bottleneck utilization is heavily dependent on the TCP segment size as will be shown in this work. We will find the optimum TCP segment size that maximizes the effective bottleneck utilization.

The effective bottleneck utilization, U_A, can be expressed as

$$U_A = \frac{L_P}{L}U = \frac{L_P}{L_P + L_O}U. \tag{5}$$

And, the window size N in equation (1) and (2) can be written as

$$N=\frac{W}{L_P}, \tag{6}$$

where W is the sender window size (in bytes), which is the smaller value between the congestion window and the receiver window advertised by the receiver.

From equations (4) and (6), the condition for the equation (1), $N \geqq 2a+1$, can be expressed as

$$\frac{W}{L_P} \geqq \frac{2R \cdot t_p}{L}+1. \tag{7}$$

Since $L=L_P+L_O$, (7) can be written as

$$L_P^2+(2Rt_p+L_O-W)L_P-L_O W \leqq 0. \tag{8}$$

Because L_P is positive, we should have $0 \leqq L_P \leqq L_B$, where

$$L_B=\frac{W-2Rt_p-L_O+\sqrt{(2Rt_p+L_O-W)^2+4L_O W}}{2}. \tag{9}$$

For $N \leqq 2a+1$, on the other hand, we have $L_P \geqq L_B$.

To summarize,

$$0 \leqq L_P \leqq L_B \text{ for } N \geqq 2a+1 \tag{10}$$

$$L_P \geqq L_B \text{ for } N \leqq 2a+1 \tag{11}$$

Now let us consider the frame error probability P_e in equations (1) and (2). Generally, the frame error probability of the wireless link is much higher than that of the wired link. Therefore, we approximate the frame error probability on the route by the frame error probability of the wireless link.

P_e is determined by the bit error probability of the wireless link B_e, the coding scheme used on the wireless link, and the frame size L. B_e is a parameter determined by the wireless link, which we assume to be fixed. P_e is a function of L_P, where L_P is the segment size which is determined by subtracting the overhead from the frame size$L$. Generally the curve of P_e versus L_P changes according to the coding scheme used on the wireless link. However, regardless of the coding scheme, P_e converges always to 1 as L_P increases to infinity. In numerical notation, $\lim_{L_P \to \infty} P_e = 1$. Therefore, we can say that $\lim_{L_P \to \infty} U_A = 0$ from (1) and (2). Furthermore, we notice that $\lim_{L_P \to 0} U_A = 0$ from (5). Since U_A is always non-negative, U_A has a maximum at some point of L_P as L_P increases. That point corresponds to the optimal TCP segment size that maximizes the effective bottleneck utilization.

We consider a simple case where there is no coding on the wireless link layer in this work. Then, the frame error probability is given by

$$P_e = 1-(1-B_e)^L = 1-(1-B_e)^{L_P+L_O} \tag{12}$$

When a coding scheme is used, once the relation between P_e and L_P is given, the remaining procedure can be applied the same way.

First let us consider the case of $N \geqq 2a+1$. By combining equations (1), (5), and (12), we have

$$\begin{aligned} U_A &= \frac{L_P}{L_P+L_O}U = \frac{L_P}{L_P+L_O}(1-P_e) \\ &= \frac{L_P}{L_P+L_O}(1-B_e)^{L_P+L_O}. \end{aligned} \tag{13}$$

U_A in (13) is always positive and has a unique maximum value. The maximum value and the corresponding value of L_P is obtained from

$$\frac{dU_A}{dL_P}=0. \tag{14}$$

Or, we have

$$L_P^2+L_O L_P+\frac{L_O}{\ln(1-B_e)}=0. \tag{15}$$

(15) has only one solution since L_P should be positive. We call the positive one L_{P1}, which is given by

$$L_{P1} = \frac{-L_O + \sqrt{L_O^2 - 4 \cdot \frac{L_O}{\ln(1-B_e)}}}{2} \qquad (16)$$

The maximum U_A can be obtained from (16) and (13). However, we note that (16) should satisfy the condition $L_{P1} \leqq L_B$ in (10). If $L_{P1} \geqq L_B$, the maximum U_A is obtained when $L_P = L_B$, since U_A is monotonically increasing in the region. To summarize, the maximum U_A for the case of $N \geqq 2a+1$, $U_{A,\max 1}$, is given by

$$U_{A,\max 1} = U_A|_{L_P = L_{P1}} \text{ when } L_{P1} < L_B \qquad (17)$$

$$U_{A,\max 1} = U_A|_{L_P = L_B} \text{ when } L_{P1} \geqq L_B. \qquad (18)$$

Next we consider the case of $N \leqq 2a+1$. Combining (2), (4), (5), (6) and (12), we have

$$\begin{aligned} U_A &= \frac{L_P}{L} \cdot \frac{N}{2 \cdot \frac{R \cdot t_p}{L} + 1} \cdot (1-B_e)^L \\ &= \frac{N \cdot L_P}{2 \cdot R \cdot t_p + L} \cdot (1-B_e)^L \\ &= \frac{N \cdot L_P}{L_P + L_O + 2 \cdot R \cdot t_p} \cdot (1-B_e)^{L_P + L_O} \\ &= \frac{W}{L_P + L_O + 2 \cdot R \cdot t_p} \cdot (1-B_e)^{L_P + L_O} \qquad (19) \end{aligned}$$

U_A in (19) is always positive and decreases as L_P increases. Thus, U_A has the maximum value at $L_P = L_B$ from the condition $L_P \geqq L_B$. To summarize, the maximum U_A for the case of $N \leqq 2a+1$, $U_{A,\max 2}$, is given by

$$U_{A,\max 2} = U_A|_{L_P = L_B} \qquad (20)$$

From (17), (18), and (20), the maximum effective bottleneck utilization, $U_{A,\max}$ is given by

$$U_{A,\max} = U_A|_{L_P = L_{P1}} \text{when } L_{P1} < L_B \qquad (21)$$

and

$$U_{A,\max} = U_A|_{L_P = L_B} \text{ when } L_{P1} \geqq L_B \qquad (22)$$

which is

$$U_{A,\max} = U_A|_{L_P = \min(L_{P1}, L_B)} \qquad (23)$$

and its corresponding L_P will be the optimal TCP segment size.

2 Numerical Results and Discussion

As a numerical example we consider the effective bottleneck utilization of the wireless TCP connection when t_p = 15ms, R = 9.6Kbps, L_O = 60bytes, W = 60Kbytes. In this case, L_B is calculated to be 59,964 bytes from (9). For segments smaller than this, the condition $N \geqq 2a+1$ holds as can be seen in (10). The optimal segment size maximizing the effective bottleneck utilization is obtained from (16) and the effective bottleneck utilization with the segment size is decided by (21) and (22). The effective bottleneck utilization versus the segment size is shown in Fig. 2. We can see from the figure that the effective bottleneck utilization and the optimal segment size decrease as the bit error rate on the wireless link gets large. The low effective bottleneck utilization at a small segment size is due to the relatively large frame overhead. Note that, in Fig. 2, TCP segment size is always smaller than L_B and thus $N \geqq 2a+1$ holds, which means that the data flow is limited by the transmission rate.

Fig. 3 shows the effective bottleneck utilization on the wireless TCP connection when t_p=90ms, R =2Mbps, L_O=60bytes, W =40Kbytes. In this case, L_B is calculated to be 437 bytes. For the TCP segment size smaller than 437 bytes, $N \geqq 2a+1$ holds, while $N \leqq 2a+1$ holds for larger TCP segment size. Note that $N \leqq 2a+1$ corresponds to the case that the data flow is limited by the window, where the effective bottleneck utilization decreases monotonically as the segment size increases as seen in equation (19).

From Fig. 2 and Fig. 3 we see that the effective bottleneck utilizations are fairly large (more than 0.8) for a broad range of segment size when the bit error rate is 10^{-6}. On the contrary, when B_e > 0.00001,

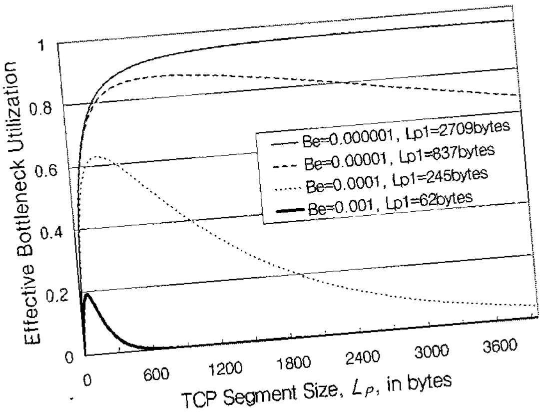


그림 2. TCP 세그먼트 길이와 유효 병목 링크 효율과의 관계 (t_p=15ms, R=9.6Kbps, L_O=60bytes, W =60Kbytes 일 때이며 그 경우의 L_B =59964bytes)

Fig. 2. Effective bottleneck utilization versus TCP segment size (t_p=15ms, R=9.6Kbps, L_O =60bytes, W =60Kbytes, and then L_B =59964bytes).

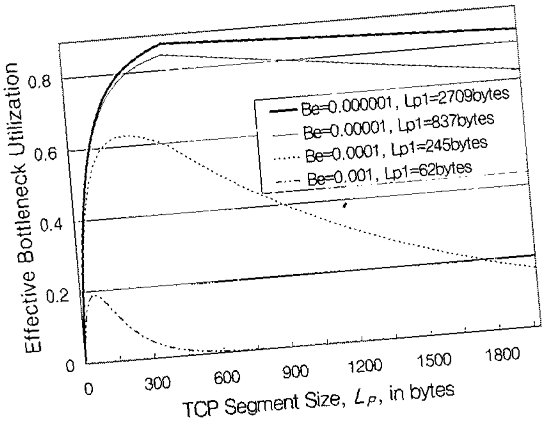


그림 3. TCP 세그먼트 길이와 유효 병목 링크 효율과의 관계 (t_p=90ms, R=2Mbps, L_O=60bytes, W =40Kbytes 일 때이며 그 경우의 L_B =437bytes)

Fig. 3. Effective bottleneck utilization versus TCP segment size(t_p = 90ms, R = 2Mbps, L_O = 60bytes, W = 40Kbytes, and then L_B= 437bytes).

the maximum effective bottleneck utilization decreases fast as the segment size increases. Therefore decision of the segment size is very important when the bit error rate is relatively large, as in the wireless link.

Fig. 4 and Fig. 5 show the optimal TCP segment size and the maximum effective bottleneck utilization of the wireless TCP connection, respectively for varying bit error rate. In Fig. 4, the optimal segment size decreases as the bit error rate increases. Dotted lines correspond to the case of Fig. 2, where the data flow is limited by the wireless link. The solid lines correspond to the case of Fig. 3, where the data flow is limited by the wireless link for large error rates and is limited by the window for small error rates.

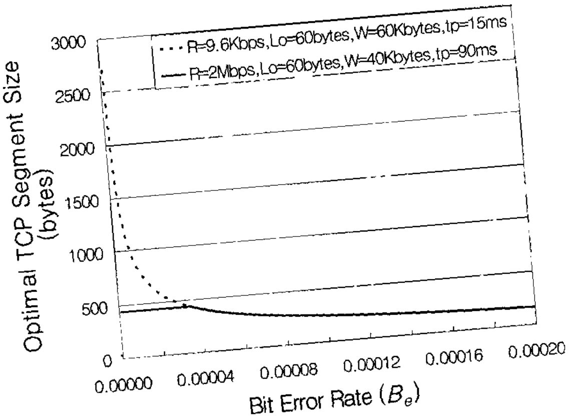


그림 4. 비트 오류율과 최적의 TCP 세그먼트 길이와의 관계 (굵은 실선중 평평한 부분은 B_e <0.000035 인 경우임)

Fig. 4. Optimal TCP segment size vs. bit error rate (solid line flat when B_e<0.000035).

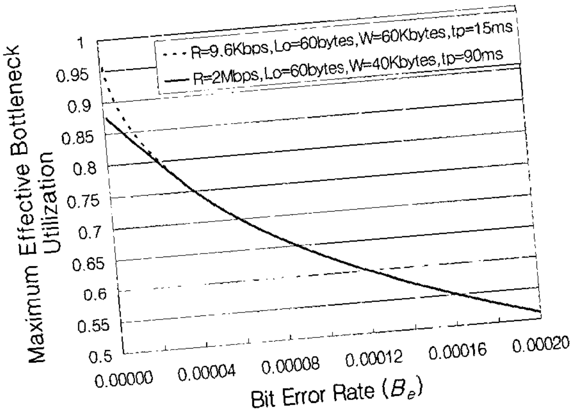
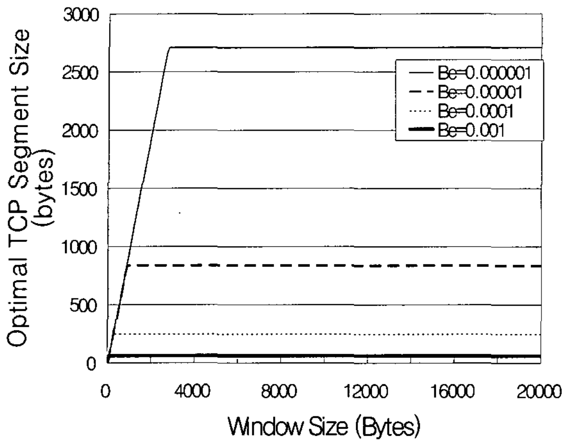


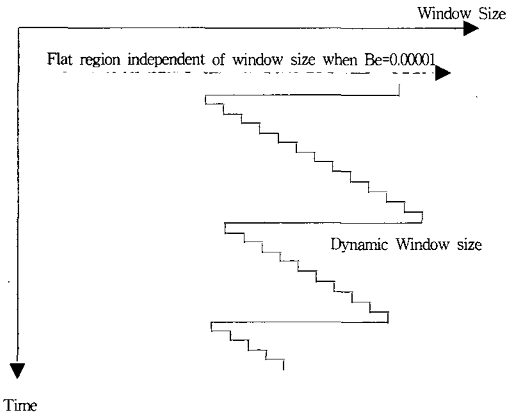
그림 5. 비트 오류율과 최대 유효 병목 링크 효율과의 관계 (굵은 실선은 B_e< 0.000035 일 때 거의 직선의 형태를 띰)

Fig. 5. Maximum effective bottleneck utilization vs. bit error rate (solid line is almost linear when B_e< 0.000035).

Fig. 4 and Fig. 5 show that the optimal segment size and the maximum effective bottleneck utilization increase sharply as the bit error rate approaches zero for the case of Fig. 2. For the case of Fig. 3, the



(a) Optimal TCP segment size versus window size



(b) A possible TCP dynamic window size variation and flat region independent of window size when Be=0.00001.

그림 6. (a) TCP 윈도우 사이즈와 최적의 TCP 세그먼트 길이와의 관계 (t_p=15ms, R=9.6Kbps, L_O =60bytes) (b) 동적인 TCP 윈도우 사이즈 변화의 예

Fig. 6. (a) Optimal TCP segment size versus window size (t_p=15ms, R=9.6Kbps, L_O=60bytes) and (b) possible TCP dynamic window size variation.

optimal segment size has a fixed value in the window-limited region (i.e., when bit error rate is low). It is noted that equation (13) can be approximated as a linear function for small value of B_e :

$$U_A = \frac{L_B}{L_B + L_O}(1 - B_e)^{L_B + L_O}$$

$$\cong \frac{L_B}{L_B + L_O} - L_B \cdot B_e, \qquad (24)$$

which we can observe in the region B_e<0.000035 of Fig. 5.

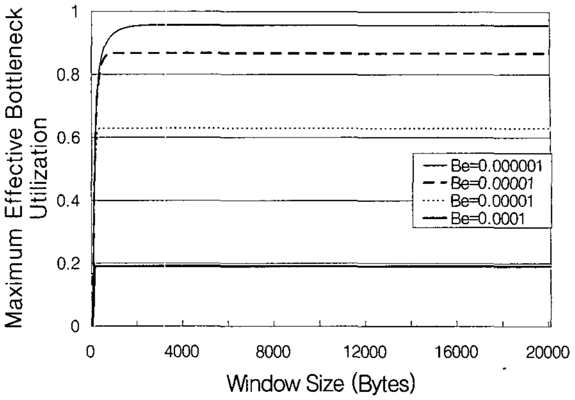


그림 7. TCP 윈도우 사이즈와 최대 유효 병목 링크 효율과의 관계 (t_p=15ms, R=9.6Kbps, L_O =60bytes)

Fig. 7. Maximum effective bottleneck utilization versus window size (t_p=15ms, R=9.6Kbps, L_O =60bytes).

Fig. 6 (a) and Fig. 7 show the optimal TCP segment size and the corresponding maximum effective bottleneck utilization for varying window size. When the window size is small, the effective bottleneck utilization and the optimal TCP segment size increase as the window size gets large. This is the window-limited region. However, when the window size grows over a threshold, both of the maximum effective bottleneck utilization and the optimal segment size curves remain flat at fixed values, in which region the data flow is limited by the wireless link bandwidth. In other words, the optimal TCP segment size is independent of the window size as shown in Fig. 6 (b). It can be said that the wireless TCP connection over the low transmission rate and high error rate link will normally remain in the access link-limited region in a steady state. Thus we can set the MSS (Maximum Segment Size) of the TCP connection at the optimal value obtained from Fig. 6 to have the highest effective bottleneck utilization.

III. Optimal TCP Segment Size in Dynamic TCP Window Size

1. Average Window Operation Region

In a TCP connection, the congestion window is

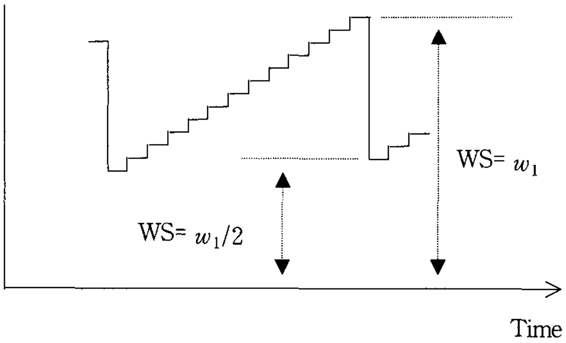


그림 8. 동적인 TCP 윈도우 사이즈 (WS)
Fig. 8. Dynamic TCP window size (WS).

increased by one segment at every round trip time (RTT) in the steady state if there is no transmission error. However when a transmission fails, the congestion window is decreased by half. Fig. 8 shows an example of this process.

In the Fig. 8, the congestion window before a frame error occurs is denoted by a random variable w_1 (in segments). The average value of w_1, $\overline{w_1}$, for the frame error probability P_e is given by $\frac{3}{8}\overline{w^2_{1+\frac{3}{4}\overline{w_1}=\frac{1}{P_e}}}$, where $\overline{w_1}$ should be positive[7].

From $P_e = 1-(1-B_e)^{L_P+L_O}$, we can have $\overline{w_1}$ as follows

$$\overline{w_1} = -1+\sqrt{1+\frac{8}{3(1-(1-B_e)^{L_P+L_O})}}\,. \tag{25}$$

$\overline{w_1}$ depends on L_P and is given in unit of frames. Therefore we can get the window size, $W_1(L_P)$, by multiplying $\overline{w_1}$ with segment size L_P, i.e., $W_1(L_P) = \overline{w_1} \cdot L_P$. The average window operation region is then given from $W_1(L_P)/2$ to $W_1(L_P)$.

2. Optimal TCP Segment Size in Dynamic TCP Window Size

The congestion window of a TCP connection varies dynamically. In the former analysis, however, we used fixed value of TCP window W. As is shown in Fig. 6 (a), the optimal TCP segment size increases as the window size increases for small values of W. However, the optimal TCP segment size does not change at window size over a threshold. When the window size W is small, L_B is less than L_{P1} and we have an optimal TCP segment size when $L_P = L_B$. As W increases, L_B also increases. Crossing a threshold, we have $L_B > L_{P1}$ and the optimal TCP segment size is given when $L_P = L_{P1}$. The analysis in Section 2.1 where we assumed a fixed TCP window size can now be applied when the TCP window size lies on the flat region. From the result of Section 3.1, the sufficient condition that the TCP average window size remains in the flat region is given by:

$$L_B\Big|_{W=\frac{W_1(L_{P1})}{2}} > L_{P1} \tag{26}$$

From (26) we have,

$$R \cdot t_P < \frac{W_1(L_{P1})}{4} \cdot \left(\frac{L_O}{L_{P1}}+1\right) - \frac{L_{P1}+L_O}{2} \tag{27}$$

As a result, the optimal TCP segment size can be determined as follows:

①Obtain t_p from measured RTTs.
②Calculate L_{P1} and $W_1(L_{P1})$.
③Check if (27) is satisfied. If it is satisfied, set the segment size to be L_{P1}.

Obtaining the optimal TCP segment size when (27) is not satisfied is for further study. Even in that case, one may choose L_{P1} in the above procedure for the TCP segment size.

IV. Conclusion

The maximum effective bottleneck utilization of a wireless TCP connection for mobile contents server access can be obtained by setting an optimal TCP segment size which can be calculated from the proposed algorithm. The algorithm can be applied when the product of the access rate and the propagation time is not large. The possibility of increasing the effective bandwidth of a wireless TCP connection just by adjusting the segment size is very attractive. The result of this work will be very useful for the telecommunications operators who provides access to multimedia servers over the cell phone wireless links.

References

[1] Wireless Application Protocol, "WAP Specifications", 2002. Available: http://www.wapforum.org

[2] Wireless Application Protocol, "WAP Wireless Profiled TCP", WAP-225-TCP-20010331-a, April 2001. Available: http://www.wapforum.com/what/technical.htm

[3] H. Balakrishnan, S. Seshan, and R. Katz, "Improving reliable transport and handoff performance in cellular wireless networks," ACM Wireless Networks, 1(4), pp. 469-481, Dec 1995.

[4] A. Bakre and B. Badrinath, "I-TCP: Indirect TCP for mobile hosts," in Proc. 15th Intl. Conf. on Distributed Computing Systems (ICDCS), Vancouver,Canada, May 1995.

[5] H. Inamura, G. Montenegro, R. Ludwig, A. Gurtov, and F. Khafizov, "TCP over Second (2.5G) and Third (3G) Generation Wireless Networks," IETF Network Working Group RFC 3481, February 2003.

[6] W. Stallings, High-speed Networks, TCP/IP and ATM Design Principles, Upper Saddle River, NJ: Prentice-Hall, 1998, pp. 223-235.

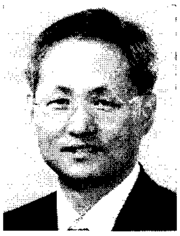
[7] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Reno performance: a simple model and its empirical validation," IEEE/ACM Trans. Networks. 8(2), pp.133-145, 2000.

[8] MyungSeon Ryou, HongSeong Park, SooHee Han and WookHyun Kwon, "Maximum Frame Size Control Based on Predicted BER in Wireless Networks," IEICE Trans. on Commun., vol.E88-B, no.7, pp3065-3068, July 2005.

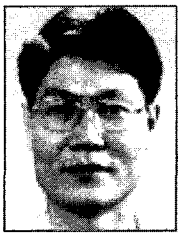
[9] T. Ikegawa and Y. Takahashi, "Analysis of Mean Frame Size of Bernoulli Wireless Links with Reliable-Transmission Window-Based Protocol, "WiOpt'04: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, pp. 402-403, University of Cambridge, March 2004.

저자소개

이구연(정회원)
1988년 KAIST 전기 및 전자공학과(석사)
1993년 KAIST 전기 및 전자공학과(박사)
1993년~1996년 디지콤정보통신 연구소
1996년 삼성전자
1997년~현재 강원대학교 컴퓨터학부 교수
<주관심분야 : 이동통신, 네트워크보안, 초고속통신망, ad-hoc 네트워크>



김화종(정회원)
1984년 KAIST 전기 및 전자공학과(석사)
1988년 KAIST 전기 및 전자공학과(박사)
1988년~현재 강원대학교 컴퓨터학부 교수
<주관심분야 : 데이터 통신, 컴퓨터네트워크, 네트워크 프로그래밍>



정충교(정회원)
1984년 KAIST 전기 및 전자공학과(석사)
1989년 KAIST 전기 및 전자공학과(박사)
1989년~1995년 LG정보통신(주) 책임연구원
1995년~현재 강원대학교 컴퓨터학부 교수
<주관심분야 : 인터넷, 통신프로토콜, 통신망성능분석>

이용(정회원)
1997년 연세대학교 컴퓨터과학과(석사)
2001년 연세대학교 컴퓨터과학과(박사)
1993년~1994년 디지콤정보통신 연구소
2001년~2003년 한국정보보호진흥원 전자서명인증관리센터
2004년~2005년 코넬대학교 방문연구원
2005년~현재 삼성전자 통신연구소
<주관심분야 : 이동통신, Wireless PKI, Mobile Ad Hoc 네트워크, Mobile Ad hoc 네트워크 보안>