OPTIMAL RETRANSMISSION TIMEOUT SELECTION FOR DELAY-CONSTRAINED MULTIMEDIA COMMUNICATIONS

Jianfei Cai *Wenning Zhan*

School of Computer Engineering
Nanyang Technological University
Email: {asjfcai, pg05238830}@ntu.edu.sg

Zhihai He

Dept. of Electrical and Computer Engineering
University of Missouri-Columbia
Email: HeZhi@missouri.edu

ABSTRACT

The RTO (retransmission timeout) selection is a fundamental problem for any acknowledgment-based transmission. Conventional approaches simply choose RTO according to the estimated round-trip time and do not consider the delay constraint of multimedia applications and the packet loss rates provided by communication networks. In this research, we propose a novel algorithm of optimal RTO selection for delay-constrained multimedia communications based on a network model. By appropriately choosing the RTO, we are able to achieve the optimal tradeoff between the error probability and the rate cost. Experimental results demonstrate the improved performance.

1. INTRODUCTION

In end-to-end communications, many transport protocols rely on retransmission to recover lost packets. For the acknowledgment based reliable transport protocols, such as TCP, a fundamental question is how long a sender should wait until retransmitting in the absence of receiving an acknowledgment (ACK). The waiting time is called retransmission timeout (RTO).

Traditionally, RTO is selected according to the estimated round-trip time (RTT). such as

$$RTO = \beta \cdot RTT, \tag{1}$$

where β is a constant, usually equals to 2 [1], and RTT is commonly estimated based on the receiver's feedback packets [2, 3]. Jacobson [1] further proposed to take into account of the large fluctuation of RTT by updating RTO as

$$RTO = RTT + 4 \cdot Var, \tag{2}$$

where Var is the average deviation of RTT. Although both of the RTO selection schemes work fine for conventional data communications, it may not be able to achieve optimal performance for delay-constrained multimedia communications since they do not consider the delay constraint and packet loss rates provided by communication networks.

To overcome this problem, some research works have been addressed on time-sensitive transmission, such as timelined TCP [4] and soft ARQ [5]. Most of them focused on designing delay-aware congestion control schemes. Recently, Chou et al. [6] proposed a framework for streaming packetized media over a lossy packet network in a rate-distortion optimized way. The proposed framework is able to minimize the end-to-end distortion under a rate constraint by choosing the right packets to transmit at a given transmission opportunity. However, it is very complicate to implement such a framework since it is non-trivial to characterize the distortion importance of each individual multimedia packet and also to perform the proposed optimal packet scheduling.

In this research, we consider a simple case, i.e., assuming all the packets have the same distortion importance and delay constraint. Thus, the problem of optimally transmitting a group of packets can be simplified as that of optimally transmitting one packet. In fact, such a simple case has also been well addressed in [6]. However, unlike that in [6], which has a fixed total number of transmission opportunities and studies whether a packet should be transmitted at each transmission opportunity, we look into the problem of the RTO selection. In other words, in our research, the total number of transmission opportunities is not fixed and the optimal tradeoff between the error probability and the rate cost is achieved through optimal RTO selection.

2. PROBLEM FORMULATION

The performance of transmitting one packet can be described by two parameters: the error probability ϵ and the rate cost ρ. ϵ is defined as the probability that the data packet does not arrive at its destination by the decoder time stamp (DTS), which is the delay deadline. ρ is defined as the expected

---

This research is partially supported by Singapore A*STAR SERC Grant (032 101 0006).

0-7803-8554-3/04/$20.00 ©2004 IEEE.

number of bytes transmitting per source byte [6].

As shown in [6], under different end-to-end loss recovery schemes, the performances of ϵ and ρ are quite different. The simplest scenario is that the sender has only one choice at time t_0, i.e., transmitting a data packet or not. If the data packet is not transmitted, the error probability will be one and the rate cost will be zero. On the other hand, if the data packet is transmitted, the error probability will be $\epsilon = P\{FTT > DTS - t_0\}$, where FTT is the forward trip time, while the cost per source byte is one. Another scenario is to use retransmission for end-to-end loss recovery. Let $t_0, t_1, ..., t_{N-1}$ be N discrete transmission opportunities. If we repeatedly transmit the packet at all N opportunities, it will result in an error probability of $\prod_i P\{FTT > DTS - t_i\}$ and the rate cost equals to N. Chou et al. [6] proposed to smartly select the appropriate opportunities for transmission, which is able to tradeoff between the error probability and the rate cost.

In this research, we consider the scenario of retransmission without a fixed total number of transmission opportunities. In particular, for a packet with a given delay constraint DTS, the total number of transmission opportunities N is determined as $N = \lceil \frac{DTS}{RTO} \rceil$, assuming the first transmission occurs at the time zero. After the sender sends out the packet, it will keep waiting for a period of RTO and then resend the packet if not receiving an ACK. This process would stop after N times of transmissions in the condition of not receiving any ACK.

Similar to [6], the network is modelled as an independent time-invariant packet erasure channel with random delays. When a packet is transmitted at the time t, it could be lost with a probability of ϵ_F. If the packet is not lost, it will arrive at the receiver at the time t', where the forward trip time $FTT = t' - t$ is randomly generated according to a probability density function $p_F(\tau|notlost)$. The lost and delay events of one packet are independent of other packets. As in [6], a Gamma function with parameters of κ_F, n_F and α_F is employed for the delay probability density function (PDF) $p_F(\tau|notlost)$, which can be described as

$$p_F(\tau|notlost) = \frac{\alpha_F}{\Gamma(n_F)}(\alpha_F(\tau - \kappa_F))^{n_F - 1} e^{-\alpha_F(\tau - \kappa_F)}, \tag{3}$$

where $\tau \geq \kappa_F$. In fact, this distribution is the same as a constant κ_F plus the sum of n_F independent identically distributed exponential random variables, each with parameter α_F. It can be interpreted as each packet going through n_F routers, each of which requires a constant processing time κ_F/n_F plus the waiting time in a steady state queue [6]. The mean value of such a distribution is $\mu_F = \kappa_F + n_F/\alpha_F$ and the variance is $\sigma_F^2 = n_F/\alpha_F^2$. Note that FTT with the PDF shown in Eqn. (3) must be larger than κ_F.

By denoting $FTT = \infty$ as the event that the packet is lost, the packet loss probability and the packet delay density can be combined into a single probability measure:

$$P\{FTT > \tau\} = \epsilon_F + \int_\tau^\infty (1 - \epsilon_F) p_F(t|notlost)dt. \tag{4}$$

The backward channel can be setup similarly. In particular, when the receiver sends a packet to the sender, the packet could be lost with a probability of ϵ_B. If the packet is not lost, it will be delayed with a probability density function $p_B(\tau|notlost)$, which is also a Gamma function but with parameters of κ_B, n_B and α_B. Therefore, we can compute the corresponding probability for the round trip time, $RTT = FTT + BTT$, as

$$\begin{aligned} P\{RTT > \tau\} &= \epsilon_F + (1 - \epsilon_F)\epsilon_B + (1 - \epsilon_F)(1 - \epsilon_B) \\ &\times \int_\tau^\infty \int_0^t p_F(t'|notlost) p_B(t - t'|notlost) dt' dt. \end{aligned} \tag{5}$$

The first term of Eqn. (5) is the probability that the packet is lost in the forward trip, the second term is the probability that the ACK is lost in the backward path, and the last term is the probability that the ACK arrives but with a delay exceeding τ [6].

3. PROPOSED OPTIMAL RTO SELECTION

For delay-constrained multimedia communications, it is not so straightforward to select the optimal RTO. If the selected RTO is too small, the sender will not wait long enough for an ACK to arrive at the receiver and thus cause unnecessary retransmission and network throughput decrease. On the other hand, if the selected RTO is too large, it may lose opportunities to recover the lost packet before the delay deadline. In this research, we propose to use the actual throughput as the optimization cost function, which is defined as

$$Actual \quad Throughput = ABW \cdot \frac{1 - \epsilon}{\rho}, \tag{6}$$

where ABW is the available network throughput. Actual Throughput can be viewed as the real throughput for transmitting source media information. Therefore, the optimal RTO problem can be summarized as: given the delay constraint DTS, how to optimally select RTO so that the Actual Throughput can be maximized.

In order to compute the Actual Throughput shown in Eqn. (6), we need to find the relationship between the error probability ϵ and RTO and the relationship between the rate cost ρ and RTO. Let $P\{\text{send i times}\}$ denote the probability that a packet is exactly transmitted for i times. We can formulate the average cost of delivering one packet as

$$\rho = \sum_{i=1}^{N} i \cdot P\{\text{send i times}\}, \tag{7}$$

and the error probability as

$$\epsilon = P\{\text{send N times}\}\prod_{i=1}^{N}\frac{P\{FTT > DTS-(i-1)RTO\}}{P\{RTT > (N-i)RTO\}}. \quad (8)$$

Eqn. (8) means that none of the packets reaches the receiver in time after N transmissions. For $i = 1, 2, \dots, N-1$, we calculate $P\{send \quad i \quad times\}$ by

$$P\{send \quad i \quad times\} = (1 - P\{RTT > i \cdot RTO\}) \times \prod_{j=1}^{i-1} P\{RTT > j \cdot RTO\}, \quad (9)$$

which can be interpreted as the probability that the sender gets the first ACK at the time between the i-th and the $(i+1)$-th transmissions. When $i = N$, $P\{send \quad N \quad times\} = 1-\sum_{i=1}^{N-1} P\{send \quad i \quad times\}$ since a packet can be maximally transmitted N times where $N = \lceil \frac{DTS}{RTO} \rceil$.

After the derivation for ϵ and ρ, we can easily compute the Actual Throughput for any value of RTO. The optimal RTO value can be obtained by numerical methods such as using exhaustive search or using bisection search method in the range of all possible RTO values.

4. PRELIMINARY RESULTS

In this section, we conduct some experiments to compare the performance of the traditional RTO selection and our proposed RTO selection. The backward path is chosen to be the same as the forward path. The parameters are set as : $DTS = 300ms$, $n_F = n_B = 2$, $\kappa_F = \kappa_B = 25ms$ and $\alpha_F = \alpha_B$. Basing on the transmission model, for a given average RTT and a packet loss rate, we attempt to find out the optimal RTO within the range from $\kappa_F + \kappa_B$ to DTS. In particular, we sample different RTO values with a constant interval of 25 ms. We compute the theoretical Actual Throughput for each RTO sample and choose the one with the largest Actual Throughput. On the other hand, for purpose of comparison , the Actual Throughput results by using the traditional RTO selection, i.e., $RTO = 2RTT$, are also calculated.

Fig. 1 and Fig. 2 show the performance comparison between the traditional RTO selection and our proposed scheme under different average RTT and packet loss rates. From these figures, one observation is that the Actual Throughput performance of our proposed RTO selection scheme is always better than that of the traditional approach, up to 5% gain. The Actual Throughput gain is more significant in the cases of larger packet loss rates, or smaller average RTT or larger average RTT.

5. SUMMARY

In this paper, we have proposed the algorithm of optimal RTO selection for end-to-end delay-constrained multimedia communications. Experimental results have shown our proposed scheme significantly outperforms the traditional RTO selection in the cases of larger packet loss rates, smaller average RTT or larger average RTT.

At the current stage, we have only conducted the experiments based on the transmission model, where we assume the sender knows the characteristics of the transmission paths such as RTT. In practice, the transmission model needs to be dynamically updated based on the feedback packets. Under such a circumstance, the estimated RTT may not be accurate which in consequence will affect the RTO selection. We will investigate this issue and also apply our RTO selection scheme to practical delay-constrained applications such as video streaming in our future work.

6. REFERENCES

[1] V. Jacobson, “Congestion avoidance and control,” in *ACM SIGCOMM*, 1988.

[2] M. Allman and V. Paxson, “On estimating end-to-end network path priorities,” in *ACM SIGCOMM,*, Sept. 1999.

[3] D. Loguinov and H. Radha, “On retransmission schemes for real-time streaming in the Internet,” in *IEEE INFOCOM 2001*, Apr. 2001, vol. 3, pp. 1310–1319.

[4] B. Mukherjee and T. Brecht, “Time-lined TCP for the TCP-friendly delivery of streaming media,” in *International Conference on Network Protocols*, Nov. 2000, pp. 165–176.

[5] M. Podolsky, S. mccanne, and M. Vetterli, “Soft ARQ for layered streaming media,” in *Tech. Rep. UCB/CSB-98-1024*, Nov. 1998.

[6] P. A. Chou and Z. Miao, “Rate-distortion optimized streaming of packetized media,” in *Microsoft Research Technical Report*, Feb. 2001.

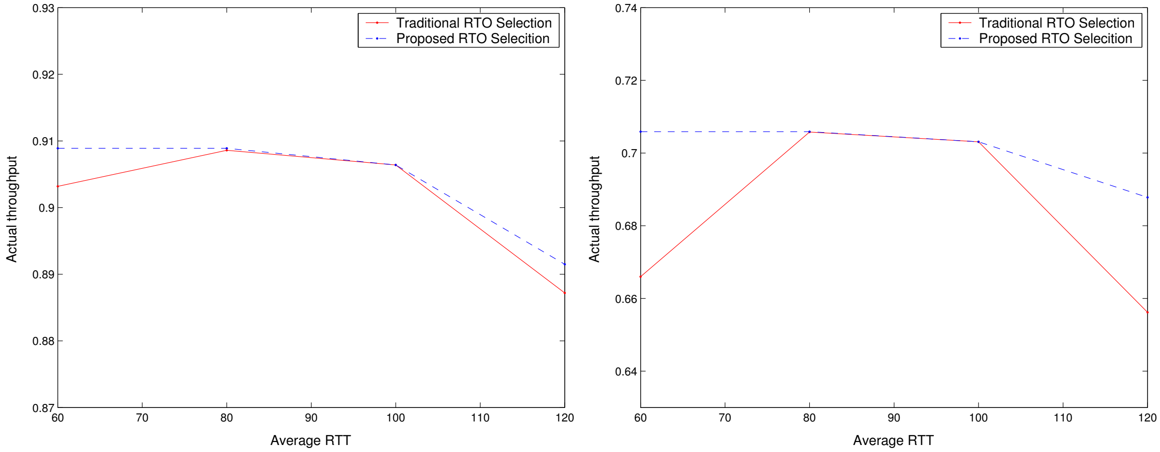

Fig. 1. The performance comparison of the Actual Throughput between the traditional RTO selection and the proposed RTO selection under different average RTT. Left: with 5% packet loss rate. Right: with 20% packet loss rate.

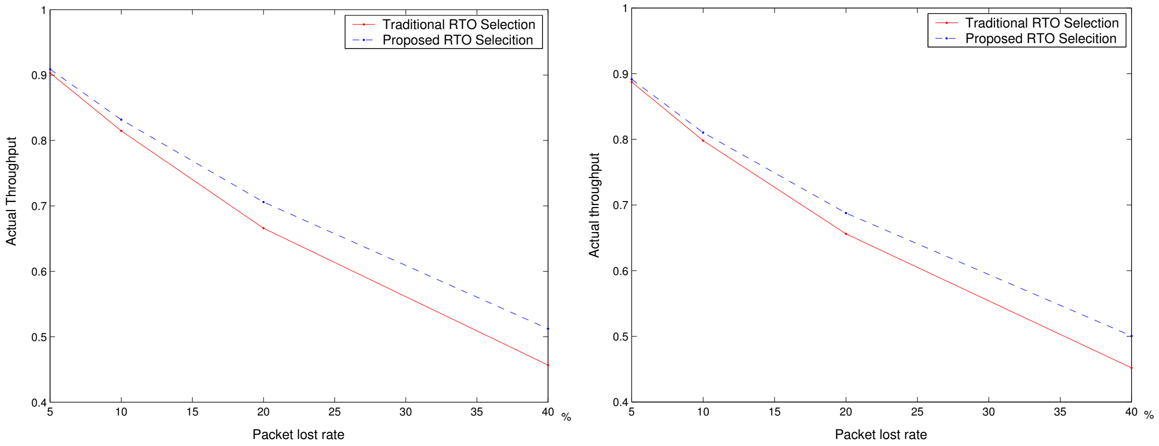

Fig. 2. The performance comparison of the Actual Throughput between the traditional RTO selection and the proposed RTO selection under different packet loss rates. Left: with the average RTT of $60ms$. Right: with the average RTT of $120ms$.