Computer Networks 54 (2010) 1418–1425

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

A transmission control SCTP for real-time multimedia streaming

Kyung-Hoe Kim [a], Kwang-Min Jeong [b], Chul-Hee Kang [a], Seung-Joon Seok [c,*]

[a] Dept. of Electronics and Computer Engineering, Korea Univ., Anam-dong, Sungbuk-gu, Seoul 136-701, Republic of Korea
[b] Samsung Electronics, Co. Ltd., 416 Maetan-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-742, Republic of Korea
[c] Dept. of Computer Engineering, Kyungnam Univ., 449 Wolyong-dong, Masan-si, Kyungnam 631-701, Republic of Korea

ARTICLE INFO

Article history:
Received 31 October 2007
Received in revised form 3 September 2009
Accepted 30 November 2009
Available online 11 December 2009
Responsible Editor: Prof. Qian Zhang

Keywords:
SCTP
Multimedia streaming
Partial reliability

ABSTRACT

Multimedia streams over the Internet have a strict playback delay time, i.e., multimedia data arriving after the playback time cannot be played by the receiver and they are discarded. In this paper, we introduce a transmission control SCTP (TC-SCTP) which has a transmission control sub-layer (TCSL) in which the multimedia streaming server determines whether data can be played in the receiver before sending the data and decides whether to send data messages or not. In addition, TCSL employs differentiated retransmission policy depending on the type of multimedia. We evaluate the performances of SCTP, partial reliability (PR)-SCTP, and TC-SCTP via ns-2 simulator. The simulation results demonstrate that the TC-SCTP protocol can decrease the amount of transmissions and increase the video decodable ratio, compared with other protocols.

1. Introduction

Transmission control protocol (TCP) is a transport layer protocol that is widely used for various Internet applications, e.g., World Wide Web (WWW), email, and file transfer. However, the retransmission scheme in TCP is not appropriate for multimedia streaming applications because it can increase the end-to-end delivery latency. Therefore, user datagram protocol (UDP) is a commonly used transport layer protocol for multimedia streaming applications. UDP does not employ any flow control schemes in response to network congestion, and therefore it can burden other users on the network and, ultimately, lower its service quality [1].

To overcome this limitation, real-time transport protocol (RTP) and real-time control protocol (RTCP) can be adopted on the top of UDP in multimedia stream applications [2,3]. That is, the RTP/RTCP layer supplements the functions of UDP by correcting out of order data and controlling the volume of data transmitted by senders for congestion control. However, these actions rely on periodic reports between the sender's and receiver's RTCP, they cannot control by packets nor respond actively to network conditions.

A new transport layer protocol, stream control transmission protocol (SCTP) has been proposed by Internet Engineering Task Force (IETF SIGTRAN Working Group. Although it was first developed for telephone signaling, it is gradually expanded into a general-purpose transmission layer, and it has been standardized as RFC 2960 in 2000 [4]. Like TCP, SCTP provide reliable service and flow control mechanisms. In addition, similar to UDP, it can support unreliable transmission [5]. SCTP can provide multi-stream and multi-homing services for a single connection. In particular, it can differentiate the level of reliability provided to messages, which is called SCTP partial reliability (PR-SCTP) [6]. PR-SCTP has the function of setting the reliability level for specific messages. The preset reliability level is used to determine the timing when the retransmission of specific data message is stopped. The function can be effectively applied to traffic containing different types of data, such as I, P, B frames in MPEG streaming applications. However, PR-SCTP attempts transmission at least once even for messages that do not require any retransmission

* Corresponding author. Tel.: +82 11 255 7693; fax: +82 55 248 2554.
E-mail addresses: kyunghoe@widecomm.korea.ac.kr (K.-H. Kim), sjseok@kyungnam.ac.kr (S.-J. Seok).

1389-1286/$ - see front matter © 2009 Elsevier B.V. All rights reserved.

due to the stringent delay constraint. In addition, if retransmission is given up, it has to send a Forward-TSN Chunk to the receiver.

Recently, multimedia streaming protocols are required to control its sending rate in response to the congestion condition of network [9–12,17–19]. It is because non-responsive streaming to network congestion, such as UDP, starves TCP flows under congestion condition. In this paper, we propose to use SCTP's congestion control for multimedia streaming.

Real-time data must reach the receiver within the tolerable playing time. Thus, a protocol like TCP, aiming for full reliability, cannot be used because a retransmission delay can exist. However, its alternatives, such as UDP and the combined UDP + RTP, do not attempt retransmission for lost packets. In this case, even if there is a chance for retransmission, based on the maximum playing time, they do not retransmit lost data. Considering this problem, the present study explains a multimedia data transmission protocol which is a modification of PR-SCTP.

In this paper, we introduce a sub-layer, called as transmission control sub-layer (TCSL), between PR-SCTP and the application layer. The TCSL sub-layer maintains as many buffers as the number of data types, and stores messages from the transmission server depending on the type. After that, TCSL of TC-SCTP evaluates the necessity of transmission before transmission. If the message does not need to be sent, it is simply removed. Compared with PR-SCTP with the fixed reliability level, TC-SCTP can dynamically consider network condition in each packet transmission.

The rest of this paper is organized as follows. Section 2 reviews existing video data transmission protocols. Section 3 describes a proposed scheme for streaming multimedia data, called TC-SCTP. Section 4 evaluates the performance of TC-SCTP and Section 5 concludes this paper.

2. Existing protocols for streaming multimedia data

As mentioned above, PR-SCTP can differentiate its retransmission service for each message. Differentiated retransmission can be achieved by applying different retransmission limits depending on the data type, e.g., delay-sensitive data and delay-tolerable data. The differentiated retransmission policy in PR-SCTP can be applied to MPEG-type video transmissions. In order to differentiate the reliability of each frame in MPEG video transmissions, [7,8] restrict the number of transmissions according to the type of frame to be transmitted. For example, only one retransmission is allowed for I frames and P frames with stringent end-to-end delay bound because additional retransmissions are not necessary due to excess playing time. On the other hand, no retransmission is provided for B frames because losses of B frames do not have any significant effects on the quality of video streaming. In this mechanism, each frame data is retransmitted only once or not retransmitted at all. However, because all data have to be transmitted at least once, if the network is congested, data may be queued for a long period of time, which may in turn lower the overall performance of the video transmission. Moreover, when the retransmission is given up, the information on the corresponding packet should be delivered to the receiver using forward TSN, which takes a long one-way trip time, and may cause faster retransmission than required, resulting in a reduction of the size of cwnd (congestion window). Furthermore, if the volume of network traffic is small and delay characteristics are favorable, the quality of video transmission is lower than the case where the number of retransmissions is not restricted.

Like [7,8], Media-SCTP [5] also restricts the number of retransmissions according to the frame reliability. Media-SCTP assumes that MPEG frames belonging to the same GOP have the same playout deadline. Because each GOP is composed of the same number of frames, the playout deadline appears at a regular interval. The receiver's frame dropping filter compares the current time with the playout deadline for each retransmitted frame. The difference between the two points of time is the time allowed for retransmission. Because it takes around 0.5 round trip time (RTT) in average for the retransmitted frame to be delivered from the sender to the receiver, the frame dropping filter compares the time allowed for retransmission with 0.5 RTT. If the allowed time is longer than 0.5 RTT, the corresponding data is retransmitted. If not, the receiver sends the Forward-TSN chunk to the sender so that the corresponding data chunk should not be retransmitted. On receiving the Forward-TSN, the sender takes an action to prevent the time-out of the corresponding data and the consequent reduction of the congestion window. That is, Media-SCTP's receiver side decides data retransmission only when the playback time of the data remains more than 0.5 RTT regardless of the type of the data frame. In the proposed TC-SCTP, however, the differentiated retransmission restricts the maximum number of retransmission in the consideration of remaining time of data frame until its playback and of data frame type.

Ahmed and co-workers [13,14] propose a sub-layer above RTP which is referred to as MP-RTP. Like Media-SCTP, MP-RTP's receiver side has responsible for retransmission request and also sender side considers frame's lifetime before sending low priority frame. However, high priority frame is retransmitted through all available path simultaneously. In this paper, proposed TC-SCTP control sending multimedia data according to data types. TC-SCTP does not request any receiver's message for retransmission and based on SCTP's multi-streaming properties, not multi-homing's.

Wu et al. [15] introduce six special issues about Internet streaming video. Application-layer QoS control issue of these issues is classified into congestion control and error control. In particular, this paper introduces a Delay-Constrained Retransmission as an application layer error control method. This Delay-Constrained Retransmission is similar mechanism to Media-SCTP [5,13,14]. In this scheme, receiver's application requests packet retransmission and sender's application retransmits the lost packet when arrival in time for display is expected.

The RaDiO in [16] determines which packets to select, when to transmit and how to transmit to minimize decoded video distortion at the receiver. This algorithm is a general scheme for data transmission and is mainly based on loss probability. Also RaDiO considers playback deadline on scheduling decision and decoding dependency of individual packets with relative priority. This algorithm treats the rate distortion at encoding level but the TCSL that we proposed in this paper is controls the (re) transmission at transport layer level.

3. Proposed Transmission-Control SCTP (TC-SCTP)

In real-time video transmission techniques using the existing PR-SCTP, the retransmission function is differentiated depending on the importance of data chunk. That is, by restricting the number of retransmissions for each data chunk, it prevents unnecessary retransmissions. However, PR-SCTP is regardless of the condition of network congestion, so its performance is usually low in dynamically changing network environments. Moreover, PR-SCTP forces to transmit each data chunk at least once even under network congestion.

The goals of this work are to minimize unnecessary transmission and to adapt network. In general, it is not expected for all transmitted data to meet the restrict delay constraint in video transmission. Data that have been transmitted but failed to meet the delay time constraint can be discarded without affecting the receiver's video playing. Although the occurrence of such unnecessary transmissions may not worsen the network situation, with it we cannot expect any active improvement in the network situation. Thus, in order to prevent unnecessary transmissions and to improve network condition, we need a technique where the sender omits data with short marginal time, for transmission in advance, and sends the following data promptly. Consequently, we introduce a Transmission Control-SCTP (TC-SCTP) as transport layer protocol at sender side.

3.1. TC-SCTP

TC-SCTP is located at transport layer between an application layer and a network layer and composed of Transmission Control sub-layer (TCSL) and PR-SCTP sub-layer. Fig. 1 shows the position and architecture of the TC-SCTP. The TCSL intercepts video data incoming form the upper layer toward PR-SCTP sub-layer and stores the data at buffers according to its type. The TCSL sub-layer checks if each frame stored in the buffers has sufficient time for arriving at receiver side or not, and determines whether to deliver each frame to the PR-SCTP sub-layer. Time constrainer in the TCSL sub-layer checks an expected margin of time until each data should be played back at receiver. If the time constrainer decides that a message cannot reach the receiver within the playback time, the message is removed by the TCSL's sender. This process in the TCSL sub-layer prevents unnecessary transmission. (Also,) TC-SCTP provides differentiated transmission service for real-time video data. For example, MPEG video stream is composed of voice data, image data (I, P and B frames), and control data. To provide differentiated transmission service to data coming from the video server of the upper layer, TCSL of TC-SCTP has four buffers for voice, I-frame, P-frame, B-frame, and control messages. a classifier of TCSL receives the MPEG frames directly from the application layer, discriminates them, and puts them in the separated queues according to the types of frame. Also, the classifier marks the frames with its arrival time which is necessary to calculate each frame's playback time. The Time Constrainer checks the remaining playback time for each message de-queued from buffers. Then, the sender selects a PR-SCTP stream according to the remaining playback time.

TC-SCTP adopts existing PR-SCTP which supports various types of reliability and make up for the weak points for multimedia service. PR-SCTP sub-layer of TC-SCTP opens multiple streams for differentiated retransmission when establishing a session, and sets the maximum number of retransmissions differently for each stream. Fig. 2 shows an illustrative example, which opens four streams and sets the maximum number of retransmissions at 4, 2, 1 and full, respectively. The sender module of TCSL sub-layer sends each frame data to one of the four PR-SCTP streams (stream 1, 2, 3, 4) depending on its data type (I frame, P frame, B frame, Voice/Control) and remaining playback time. Because I frame has more important information to play MPEG video back than P and B frames, stream 1 attempt more retransmissions (four times) than stream 2 (two times) and stream 3 (one time). So, B frame is only retransmitted once. The last stream 4 provides control information and voice data with fully reliability. It is because control and voice data are tightly related to receiver's playback and user's service satisfaction, respectively.

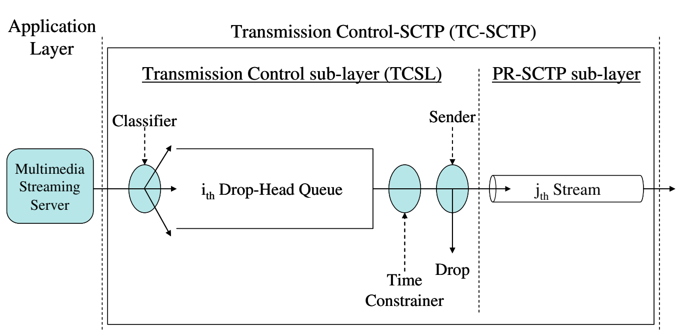


Fig. 1. TC-SCTP layer architecture.

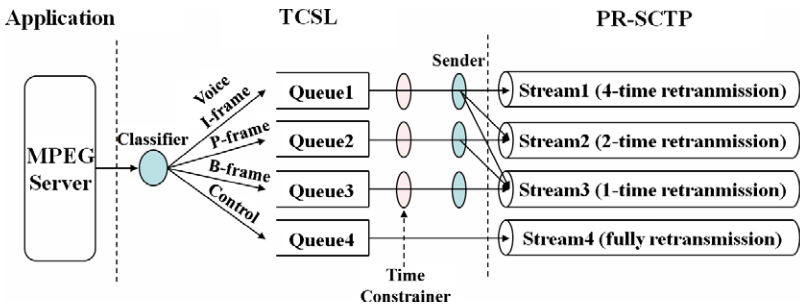


Fig. 2. Differentiated transmission according to data type.

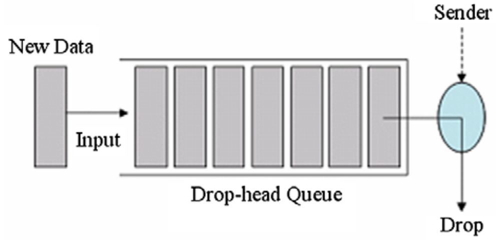
3.2. Drop-head queue

The proposed TCSL sub-layer has multiple buffers, and the number of buffers is determined by the number of types of data incoming from the application layer. For example, MPEG video data are usually divided into five types – I-frame, B-frame, P-frame, voice data and control data - according to the degree of classification of **OCI** information. For differentiated transmission of these messages, the TCSL sub-layer maintains a buffer for each type of messages. Voice and I-frame share the same buffer because these data have similar properties.

A typical buffering method is drop-tail queuing. If the queue is full when a new frame arrives, the frame is dropped. In this case, en-queued data are served first. TCSL sub-layer's buffers, however, use drop-head queuing property. That is, if a drop-head queue has no room for newly arriving frames, it removes data in the forefront of the buffer and then stores the new frames. It is because video streaming data is sensitive to delay. The longer they stay in the buffer, the shorter their marginal time for transmission. The buffers become full of frames when the network situation is deteriorated and frames are not transmitted smoothly in PR-SCTP sub-layer. In a buffer, the foremost frame is oldest and is most likely to have shorter marginal time for transmission than the time constraint for dropping. Thus, when a new frame arrives to a buffer occupied already, the foremost frame is dropped first and the new frame is added. Fig. 3 shows the drop-head queuing process.

3.3. Time constraint

Time constraint is used by the Time Constrainer of buffers in the TCSL sub-layer, to determine whether to send data via PR-SCTP. TCSL sub-layer assigns different time constraints according to the data type. In order to assign time constraints, it is required to estimate the maximum time allowed for each data in the sender's TCSL sub-layer to be played in the receiver's application program. This section defines the parameters involved in assigning time constraints.



The offset delay (T_{OS}) is the basic delay given when the sender and the receiver's application layers begin to send and receive video streaming data. The delay means the time taken for multimedia data transmitted by the sender's server to be played in the receiver's application program. Data that arrives at the receiver after exceeding the time cannot be played, and is discarded. In this paper, offset delay was defined as the maximum marginal time for a data message received by TCSL sub-layer from the upper layer, to be delivered to the receiver's application. As mentioned above, for the Time Constrainer to determine whether data in the buffer can be played in the receiver's program, it should know how much time has passed since the data were created.

In assuming that the sender's server sends data to TCSL of TC-SCTP as soon as it creates the data, TCSL can use the time of receiving data from the upper layer (T_{AR}) as the time of creation for the data. The playback time (T_{PB}) is the point of time when each frame data should be played in the receiver's upper layer. So, each data should arrive at the receiver before the playback time, because data arriving after the time cannot be played and removed. TCSL defines the estimated time of data playing by adding the offset delay time (T_{OS}) to the time when the data arrives at the sender's TCSL sub-layer. The offset delay time is determined by sender's application and notified to TCSL. This paper premises that the sender's application should negotiate for service parameters, such as receiver's buffer size, service rate, and so on, after service request from receiver's application. The sender's application derives the offset delay time from values of receiver's buffer size and transmission delay.

$$T_{PB} = T_{AR} + T_{OS}. \quad (1)$$

For data to be sent via PR-SCTP sub-layer, the time constrainer of TCSL sub-layer estimates the time remaining until the playback time. In this paper, the remaining time is defined as the marginal time to playback (T_{MG}) at receithe corresponding data will be necessary. If the marginal time for transmission is long, it is more likely that the data is transmitted to the receiver successfully. However, if it is short, it is more possible that even if the data is transmitted to the receiver successfully, the transmission would be useless. Thus, to prevent unnecessary transmission, the time constrainer in the TCSL sub-layer buffers checks the marginal time for transmission of each data before transmission. If the marginal time for transmission is shorter than a specific value, the TCSL's sender discontinues transmission and removes the data from the buffer. The marginal time for transmission is the difference between the estimated time of playing (T_{PB}) of each data and the current time of the time constrainer (T) for the data as below (see Fig. 4):

$$T_{MG} = T_{PB} - T. \tag{2}$$

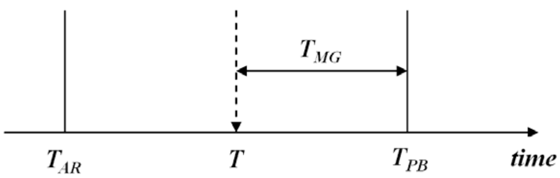


Fig. 4. Marginal time for transmission.

3.4. TCSL data transmission

TC-SCTP is composed of TCSL sub-layer described in this section and PR-SCTP of [6,8]. A TCSL's sender forwards messages to multiple streams pre-allocated by PR-SCTP according to message type and marginal time length (T_{MG}) which a TCSL's time constrainer calculates for messages from its local queue. This TCSL data transmission algorithm is illustrated in Fig. 5.

TCSL's sender classifies the marginal time lengths into three cases according to message type: $T_{MG} \geqslant Th_1$, $T_{MG} \geqslant Th_2$, $T_{MG} \geqslant Th_3$ ($Th_1 > Th_2 > Th_3$). Also, TCSL sub-layer uses multiple streams of PR-SCTP with different reliability levels, such as maximum number of retransmission [8] or available time limit for retransmission [6]. For example, the stream 1, 2, and 3 of Fig. 2 allow four, two, and one times of retransmission. And, stream 4 provides full reliability i.e., it attempts retransmissions until transmission is complete.

As mentioned before, the proposed TC-SCTP handles five types of MPEG data – I-frame, P-frame, B-frame, voice data and control data. Among these data, voice and I-frame image data are critical information in terms of user satisfaction. Thus, TC-SCTP attempts retransmission as many times as permitted by marginal time for transmission. For this, if the marginal time for transmission (T_{MG}) of an incoming message is longer than Th_1, the TCSL's sender sends it to Stream 1 of PR-SCTP sub-layer. Stream 1 attempts retransmission of the data up to four times. If the

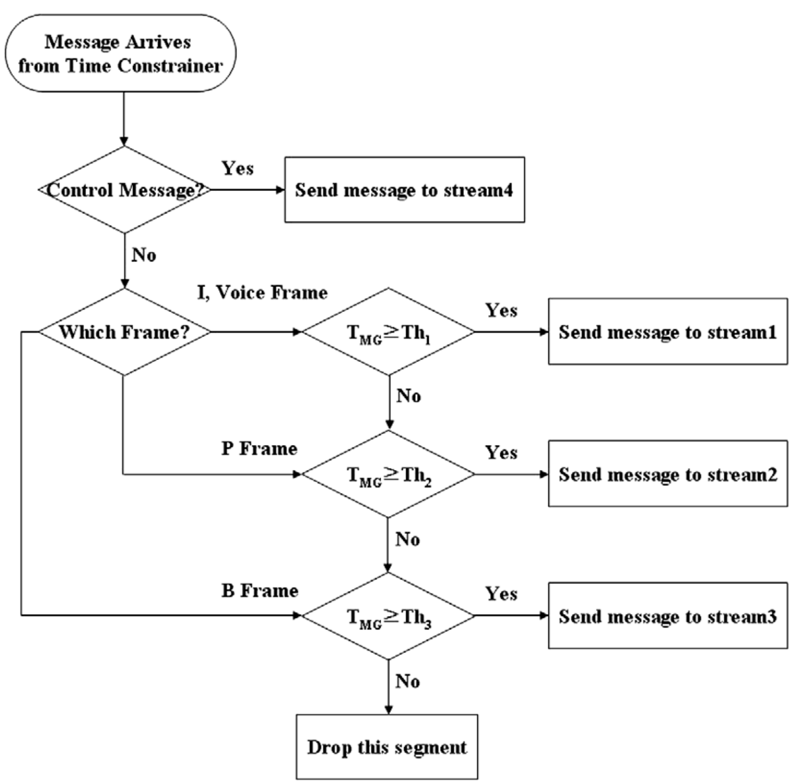


Fig. 5. TCSL data transmission algorithm.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.