AN EYE-TRACKING-BASED ADAPTIVE MULTIMEDIA STREAMING SCHEME

*Gheorghita Ghinea° and Gabriel-Miro Muntean**
°School of Information Systems, Computing and Mathematics, Brunel University, United Kingdom
*Performance Engineering Laboratory, School of Electronic Engineering, Dublin City University, Ireland

ABSTRACT

Traditionally, adaptive multimedia streaming schemes aim to adjust features such as bandwidth to respond to changing network conditions, in the hope that the user perceptual impact of any quality loss is, if not unnoticed, minimised. However, current solutions equally affect the whole viewing area of the multimedia frames, despite research showing that there are regions of the frame which are perceptually more relevant than others. This paper presents a novel eye-tracking-based adaptive scheme (ETAS) for multimedia streaming that performs transmission-related quality adjustments by selectively degrading the quality of those regions of the image the viewers are the least interested in, leaving perceptually relevant regions unchanged.

1. INTRODUCTION

In the quest to better utilize network bandwidth, adaptive multimedia streaming solutions [1] [2] have been proposed, which reduce, in a controlled manner, the amount of multimedia data to be streamed, in the hope that end-user perceived quality is minimally affected. By sending less data, the pressure on the delivery network is released and eventually the loss rate decreases, increasing the end-user perceived quality.

Existing solutions, though, affect equally the whole viewing area of multimedia frames when adjusting the multimedia stream. However research has shown [3] that there are regions of the multimedia display image on which the viewers are more interested in than on others.

This paper proposes a novel eye-tracking-based adaptive multimedia streaming scheme that when performing adaptive transmission-related quantity and consequently quality adjustments, selectively affects the quality of those regions of the image the viewers are the least interested in. As the quality of the regions the viewers are the most interested in will not change, it is expected that the proposed scheme will provide higher overall end-user perceived quality than any of the existing adaptive solutions.

2. RELATED WORK

The large majority of the proposed adaptive schemes for multimedia streaming are **sender-based**, giving a significant role to the server in taking adjustment decisions, such as the ***Loss-Delay based Adjustment algorithm (LDA)*** proposed in [4]. It relies on RTCP reports and on a packet-pair technique to estimate round trip delay, loss rate and the bottleneck link bandwidth. The scheme controls the transmission rate using these estimates and some user parameters. The ***enhanced Loss-Delay Adaptation algorithm (LDA+)*** [5] also makes use of RTCP reports to collect loss and delay statistics and to adjust the transmission rate like TCP connections subject to equal losses and delays. The ***Rate Adaptation Protocol (RAP)*** proposed in [6] uses TCP-like acknowledgement of the packets to estimate loss rates and delays. In case of no loss, the rate is additively increased function of round trip delay, whereas in case zero loss, the rate is halved as TCP does. ***Layered Quality Adaptation (LQA)*** [7] is one of the most significant schemes that make use of the properties of layered-encoding in supporting rate-controlled adaptations. It modifies the bitrate and consequently the quality of the transmitted multimedia by adding and removing a layer respectively. In [8] a ***TCP-Friendly Rate Control Protocol (TFRCP)*** is presented, based on a TCP model. In case of losses, the rate is limited to the equivalent TCP rate computed according to the TCP model otherwise the rate is doubled. One of the most significant sender-based adaptive schemes as it introduces an estimation of the end-user perceived quality in the adaptation loop is the ***Quality Oriented Adaptation Scheme (QOAS)*** [9].

The **receiver-based schemes** provide mechanisms that allow for the receivers to select the service quality and/or rate such as ***Receiver-driven Layered Multicast (RLM)*** [10] and ***Receiver-driven Layered Congestion Control (RLC)*** [11].

The ***TCP Emulation At Receivers (TEAR)*** scheme, described in details in [12] is a significant **hybrid adaptive mechanism** that involves both the sender and the receiver in the adaptation process. **Transcoder-based solutions** constitute another category. They focus on matching the available bandwidth of heterogeneous receivers through transcoding or filtering [13, 14].

All these schemes perform adjustments of the streamed multimedia data such as it affects equally all the frames' regions, regarding of viewer's interest. Unlike them the proposed eye-tracking-based adaptive scheme takes into account the differentiated importance these regions have for the end-user.

3. EYE-TRACKING-BASED ADAPTIVE MULTIMEDIA STREAMING SCHEME

The Eye-Tracking-based Adaptive Scheme (ETAS) for multimedia streaming is an unicast rate-based adaptive solution for delivering high quality multimedia. Its goal is to increase the end-user perceived quality when viewing remotely streamed multimedia sequences in highly loaded delivery conditions by taking into consideration viewer's aq interest in certain multimedia frame regions and consequently differentiating their tratment during adaptation process.

ETAS includes client and server-located components that are involved in the bi-directional exchange of video data and control packets through the delivery network. The client monitors some transmission-related parameters and regularly computes the quality of delivery scores, which are sent as feedback to the server. The server analyses these scores and proposes content-related adjustment decisions in order to increase user Quality of Experience in existing delivery conditions. Content adaptation is performed as research has shown that viewers of streamed multimedia content prefer controlled reduction in quality to the effect of random losses [15]. Therefore the transmitted quantity of multimedia data can vary during the streaming process.

When required to reduce the quantity and consequently the quality of transmitted multimedia-related information in order to meet the available bandwidth constraints, ETAS affects the streamed data in terms of some compression-related parameters such as resolution and frame-rate.

Existing adaptive multimedia streaming schemes involve content modifications that affect equally the whole viewing area of the multimedia frames being transmitted. However as eye-tracking research has shown [3], there are some regions within multimedia streams' frames the viewers are more interested in than in others. Consequently ETAS enhance the classic adaptive solution for streaming multimedia with an eye-tracking-feedback loop that determines the regions of interest for the viewers. Information related to the location of the region of maximum interest is collected by the ETAS's client side component and is sent regularly as feedback to the server, keeping it informed.

ETAS's server side component maintains a viewer region of interest model that is updated regularly by feedback. Based on the information from this model, ETAS selectively adjusts the quality of those regions the viewer is the least interested in when transmission-related quality adaptations are required to be performed. As the quality of the regions the viewers are the most interested in will not change, the proposed scheme will provide much higher overall end-user perceived quality than any of the existing adaptive solutions, with significantly reduced bandwidth requirements. This is illustrated in Figure 1, which shows that the coverage of eye-tracking based regions of interest to be far less than that of the original frame, and highlights the opportunity to selectively stream only those perceptually relevant regions at higher quality.

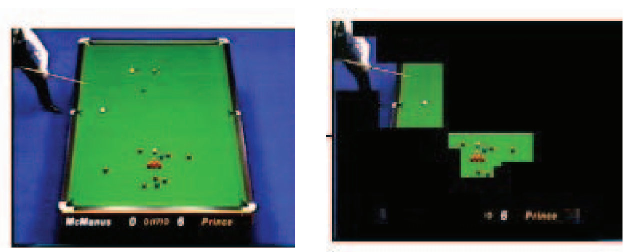

Figure 1. a) Original frame b) Eye-tracking based regions of interest

Figure 2 presents the schematic principle of the eye-tracking-based adaptive multimedia streaming scheme. It involves a server and a client that communicate via a bi-directional channel in order to exchange multimedia information and feedback. The server has associated a number of different server states that are each assigned to a different potential stream quality. For example Figure 2 presents a five-state model. The different stream quality versions are obtained in real time by adapting compression-related parameters such as resolution and frame rate of the areas the viewers are the least interested in. For example the highest quality stream could have the regions on non-interest played at 25 fps and at maximum resolution, the above-average quality stream will have them displayed at 20 fps and at a resolution degraded with 20% and so on. Consequently these streams will have different bandwidth requirements.

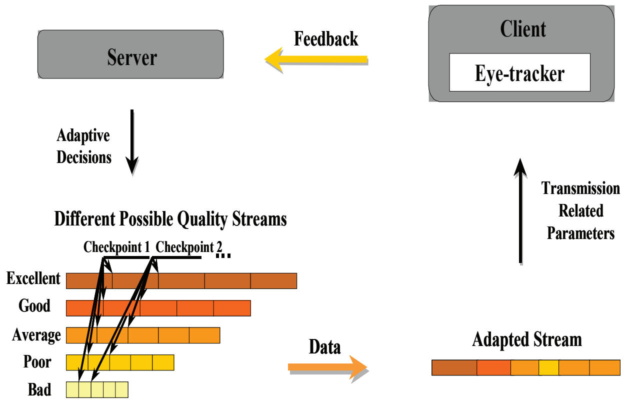

Figure 2. Eye-tracking-based adaptive principle

During transmission the server dynamically varies its state according to the client feedback. For example, when the client reports a decrease in end-user quality due to packet loss, the server switches to a lower quality state, which reduces the quantity of data sent. As a consequence to a lower loss rate, the end-user perceived quality increases, in spite of transmitting less information. In improved conditions, the server gradually increases the quality of the delivered stream and in the absence of loss, this determines an increase in the end-user perceived quality. Consequently the viewer will receive an “adapted” stream (see Figure 2).

4. PRELIMINARY TESTING

Preliminary tests involved simulations using Network Simulator version 2 (NS-2) [16]. The “Dumbbell” topology that assumes a single shared bottleneck link (A-B) presented in Figure 3 was used. The sources of traffic are located on one side of the bottleneck link, whereas the receivers are on the other side. ETAS’s components were deployed at both the server and client (ETAS Si, and ETAS Ci, respectively, where 1<=i<=N). No other traffic was involved. Buffering at the bottleneck link uses a drop-tail queue of length proportional with the product between the round trip time and the bottleneck link’s bandwidth. During simulations this bandwidth was set to 100 Mbps and the bottleneck link’s delay was set to 0.1 s. Apart from the bottleneck link, the other links are over-provisioned such as they will not influence the simulation results.

Simulations involved ETAS-based adaptive and non-adaptive multimedia streaming. *diehard1* – a multimedia sequence with very high motion content – was MPEG2 encoded at five different rates between 2 Mbps and 4 Mbps using the same frame rate (25 frames/sec) and the same IBBP frame pattern (9 frames/GOP). Traces were collected and used during NS-2 simulations. ETAS-based adaptation involved five server quality states, each having associated one of the encoded multimedia stream quality versions. The non-adaptive streaming used the maximum quality sequence with average bitrate 4 Mbps. Table 1 presents the properties of all multimedia versions used during simulations.

Table 1. Statistics related to the different quality encoded versions of the diehard1 multimedia clip

Version (0-4)	Avg. Encoding Rate (Mbps)	Peak/Mean Ratio
0	2.00	7.48
1	2.50	7.43
2	3.00	6.31
3	3.50	5.65
4	4.00	4.06

The simulations started with a number of clients randomly selecting the starting point from within the multimedia clip in order to allow for independence from the natural multimedia bitrate variations. The resulting streaming sessions lasted 500 sec from which 50 sec transitory periods at the beginning and at the end were not considered. Simulation results were assessed in terms of average throughput, average loss rate and estimated end-user perceived quality. The quality is assessed using the non-reference moving picture quality metric proposed in [17] and expressed on the 1-5 ITU-T R. P.910 subjective quality scale, where 1 represents “Bad” quality and 5 - “Excellent”. Table 2 presents comparative simulation results when ETAS and the non-adaptive scheme were used in turn with an increasing number of clients.

Table 2. Average throughput, loss and quality when increasing the number of ETAS and non-adaptive clients

	ETAS			Non-adaptive		
No. Clients	Through. (Mbps)	Loss (%)	Qual. (1-5)	Through. (Mbps)	Loss (%)	Qual. (1-5)
23	4.00	0	4.56	4.00	0	4.56
24	4.00	0	4.54	3.89	0.81	3.48
25	4.00	0.02	4.52	3.52	4.36	1
26	4.00	0.03	4.51	3.48	9.34	1
27	3.62	0.05	4.42	3.35	11.57	1
28	3.33	0.25	4.17	3.18	14.03	1
29	3.26	0.04	4.39	3.12	21.73	1
30	3.18	0.22	4.19	3.06	26.64	1
31	3.12	0.32	4.07	3.00	31.34	1
32	3.09	0.37	4.01	2.84	37.12	1

It could be clearly seen how when using ETAS the number of clients streaming multimedia at least at “Good” quality (level 4 on the ITU-T 1-5 scale) can be increased with 40% in comparison with the non-adaptive case, reaching 32. By using adaptation the loss rate was maintained below 1% in all the cases when ETAS was used, unlike when the non-adaptive streaming scheme was employed. In those situations the loss rate rapidly increased up to 37% with the increase in the number of simultaneous multimedia clients, severely affecting the end-user perceived quality.

5. CONCLUSIONS AND FURTHER WORK

This paper introduced ETAS, an Eye-Tracking-based Adaptive Scheme for multimedia streaming, in which adaptation is based on regions of interest obtained as a result of eye-tracking monitoring. ETAS selectively adjusts the quality of those regions from the multimedia frames the viewer is the least interested in when transmission-related quality adaptations to existing delivery network conditions are required to be performed. As the quality of the regions the viewers are the most interested in will not change, the proposed scheme provides high overall end-user perceived quality.

Simulation results involving multiple clients streaming multimedia simultaneously show how ETAS performed much better in terms of average client throughput, loss and estimated end-user perceived quality than when a non-adaptive streaming scheme was used.

Work in progress compares the performance of ETAS-based multimedia streaming with that when other adaptive schemes are employed. Future work envisages the deployment of ETAS on a prototype system for multimedia streaming. Consequently subjective testing will complement simulations in order to fully assess the benefit of ETAS in terms of end-user perceived quality.

6. REFERENCES

[1] S. Dravida, D. Gupta, S. Nanda, K. Rege, J. Strombosky, M. Tandon, “Broadband Access over Cable for Next-Generation Services: A Distributed Switch Architecture”, IEEE Communications Magazine, Vol. 40, No. 8, August, 2002, pp. 116–124

[2] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, J. M. Peha, “Streaming Video over the Internet: Approaches and Directions”, IEEE Trans. On Circ. and Sys. for Video Tech., vol. 11, no. 3, 2001, pp. 282–300

[3] S. R. Gulliver, G. Ghinea, “Stars in their eyes: what eye-tracking reveals about multimedia perceptual quality”, IEEE Transactions on Systems, Man and Cybernetics, Part A, vol. 34, No. 4, July 2004, pp. 472 – 482

[4] D. Sisalem, H. Schulzrinne, “The Loss-Delay Based Adjustment Algorithm: A TCP-Friendly Adaptation Scheme”, Proc. ACM NOSSDAV, 1998

[5] D. Sisalem, A. Wolisz, “LDA+ TCP-Friendly Adaptation: A Measurement and Comparison Study”, ACM NOSSDAV, 2000

[6] R. Rejaie, M. Handley, D. Estrin, “RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet”, Proc. IEEE INFOCOM, 1999

[7] R. Rejaie, M. Handley, D. Estrin, "Layered Quality Adaptation for Internet Video Streaming", IEEE Journal on Selected Areas of Communications, vol. 18, no. 12, Dec. 2000, pp. 2530-2543

[8] J. Padhye, J. Kurose, D. Towsley, R. Koodli, “A Model Based TCP Friendly Rate Control Protocol”, ACM NOSSDAV, 1999

[9] G.-M. Muntean, P. Perry and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks", IEEE Trans. Broadcasting, vol. 50, no. 1, March 2004, pp. 1-10

[10] S. McCanne, V. Jacobson, M. Vetterli, "Receiver-Driven Layered Multicast", Proc. ACM SIGCOMM, Stanford, California, USA, August 1996, pp. 117-130

[11] L. Vicisano, J. Crowcroft, L. Rizzo, “TCP-like Congestion Control for Layered Multicast Data Transfer”, Proc. IEEE INFOCOM, vol. 3, San Francisco, CA, USA, March 1998, pp. 996-1003

[12] I. Rhee, V. Ozdemir, Y. Yi, “TEAR: TCP Emulation at Receivers - Flow Control for Multimedia Streaming”, Technical Report, CS Department, NCSU, April 2000

[13] L. Wang, A. Luthra, B. Eifrig, “Rate Control for MPEG Transcoders”, IEEE Trans. On Circuits and Systems for Video Technology, vol. 11, no. 2, Feb. 2001 pp. 222–234

[14] N. Yeadon, F. Garcia, D. Hutchison, D. Shepherd, "Filters: QoS Support Mechanisms for Multipeer Communications", IEEE J. on Selected Areas in Comp., vol. 14, no. 7, Sept. 1996, pp 1245-1262

[15] G. Ghinea, J. P. Thomas, “QoS Impact on User Perception and Understanding of Multimedia Video Clips”, Proc. ACM Multimedia, Bristol, United Kingdom, 1998, pp. 49-54

[16] Network Simulator-2, http://www.isi.edu/nsnam/ns

[17] O. Verscheure, P. Frossard, M. Hamdi, “User-Oriented QoS Analysis in MPEG-2 Video Delivery”, Journal of Real-Time Imaging, vol. 5, no. 5, Oct. 1999, pp. 305-314