Application level Wireless Multi-level ECN for Video and Real-time Data

Ouldooz Baghban Karimi, Mahmood Fathy, Saleh yousefi
Computer Engineering Department, Iran University of Science and Technology
ouldoozbk@comp.iust.ac.ir, {mahfathy, syousefi}@iust.ac.ir

Abstract

Multi-level Explicit Congestion Notification (MECN) and Wireless MECN (WMECN) are the amendments proposed in recent research to have an improved feedback from network backlog. These methods are implemented in conjunction with TCP and modify the TCP in its response to congestion. The improvement of these methods on TCP is derived from the fact that they modify the window size variations in response to congestion. Quantitative benefits of ECN, MECN and similar proposed methods for TCP have always been an issue to debate. Adding ECN to UDP API and end-to-end congestion control is also one of the topics of interest. In this paper we present a new WMECN-Like mechanism which is modified in a way that could be used by applications to adapt transmission rate or video encoding quality. Low complexity of the proposed mechanism makes it more suitable than UDP which is recommended protocol for QoS over wireless networks.

Keywords*- adaptive end-to-end QoS, application level QoS, congestion, multi-level ECN, UDP, video, and wireless networks.*

1. Introduction

Multimedia communication over the current best effort Internet faces well-known challenges with respect to quality of service, congestion control and network friendliness. Among multimedia applications visualized and real-time applications are more sensitive to network state changes. Transmission of multimedia and specifically video data over wireless networks needs to go further to consider the problems of wireless and mobile networks as well.

The quality of the video transmitted over the Internet highly depends on the available bandwidth. The traffic load along the path changes throughout the transmission in an unpredictable way. Hence, the bandwidth requirements have to be modified accordingly. When the amount of bandwidth available is not sufficient, the video transmission rate is adjusted by lowering the quality of the video in a controlled way. The quality adjustment is done by excluding some frames from the transmission or by reducing frame sizes. All these depend on the information we obtain form the network which could be done in number of ways: packet dispersion, network state information, congestion and loss notifications, etc. [1]. Among the ways above, network state information is used for obtaining a feedback from network backlog. One of the important problems in transmission of multimedia data over Internet is network awareness [2]. Being aware about the network backlog will help the application to gain valuable information about the network state which is useful in the subsequent decisions. This is especially important in wireless networks in which the lossy link, temporary disconnection durations and congestion events mostly affect the multimedia quality. Network backlog information, user-perceived quality, playback buffer information and arrival rate altogether would help us obtain reliable information about the network to decide about the next action accurately.

Multi-level ECN for early congestion notification is proposed to gain an enhanced feedback from network [3]. WMECN mechanism is also proposed to provide this feedback in wireless networks [4]. The proposed schemes all aim at changes in TCP's reaction to different ECN levels. The reaction is altered by changes in ECN semantics and also by the coefficients used for optimizing the TCP (Transport Control Protocol) window shrink in response to different congestion levels.

The novelty of our approach lies in the fact that we do not use network feedback in transport layer, i.e. we do not shrink the window size in response to loss or congestion notifications. Instead we apply an algorithm that would respond to these events in application level with a rate-quality adaptation algorithm.

We assume that the application does not rely only on the end-point observations, but also uses a network interface to get feedback directly from the network. A network router or a Base Station (BS) can inform the user about the congestion based on the fact that its buffer occupancy has reached certain occupancy level. A BS or a user cell can inform the application layer about the possible disconnection duration. However, predicting the lack of resources and reacting to it without packet loss is crucial for video quality. Therefore, our effort concentrates on the prediction and adaptation in the case of the network congestion, loss and disconnection duration.

The Internet draft specifying alternate semantics for explicit congestion notification, [5], [6] has proposed that the alternate semantics for ECN should offer a good compatibility with previous definitions so that avoiding the old router mistaking the notation used with the previous definitions of ECN. Also this will support coexistence with routers without ECN support. Our proposed scheme is tested to have a coexistence possibility with previous versions of ECN.

In this paper we propose a way to overcome the video streaming problems in mobile wireless networks and provide an algorithm for network feedback that shows better QoS and multimedia delivery over wireless networks in application layer. The proposed mechanism aims at adaptive end-to-end QoS support so it is classified in pure end-to-end protocols category. In this paper, we first consider the ECN marking and its effects on video delivery over Internet, study the multi-level ECN marking in wireless networks and its effect on multimedia services. Then we will present a new WMECN-like mechanism to be used by adaptive end-to-end schemes in application layer to consider and control the network events as well as user perceived quality to support QoS. Then, we present the simulation experimental results that show the performance of our approach.

The rest of this paper is organized as follows: In second section we study the problems encountered in QoS support over wireless networks. In third section, transport layer protocol selection for QoS over wireless networks is discussed. Fourth section explains the network state feedback and how it could be interpreted in the best possible way. Fifth section describes the proposed scheme in detail. Section six includes the simulation settings and results and finally section seven includes the concluding remarks and ways to future works.

2. QoS over Wireless Networks

Providing QoS over wireless networks have been a problem over recent years. A number of amendments on present protocols have been proposed to solve the problem in different levels. Each algorithm tackles the problem from one point of view. The main problems encountered in QoS support in wireless networks are as follows:

2.1 Limited bandwidth

Wireless networks offer limited bit rates. For example GSM Evolution (EDGE) offers up to 384 kbps. On the other hand new IEEE LAN standards offer higher bandwidth. The proposed mechanism should come over this wide range of bandwidth to offer acceptable multimedia quality over the variable bandwidth available in wireless networks [7], [8], and [9].

2.2 Short Flows

The majority of data in transmission in wireless networks consists of short flows. This may cause problems to TCP mechanisms that are still in the state of slow start in the begging of the connection. The data transmission is already completed in most of the times before the window of the sender reaches the maximum size, so full use of the link is not possible [7], [8], and [10].

2.3 Long round trip times

In general wireless media show longer latency delays than wired ones. This would cause the increased needed times for obtaining feedback from network or long times for systems like TCP to shrink or increase their window size. So when mechanisms like TCP are being used, the throughput would be low due to moderate growth in additive increase phase in the congestion window. Offering a mechanism that has no need to additive changes in window size and sending rate would help gaining a better throughput [7], [8].

2.4 Random Losses

Wireless media are more prone to disconnections and transmission losses. In TCP when a packet loss is interpreted as congestion, this would cause in the decreased window size and the degraded throughput. Since the multimedia data is very sensitive to loss and delay variations caused afterwards, there is a need for a mechanism for notification of such states and durations in multimedia delivery applications [7], [8], and [9].

2.5 User Mobility

Wireless networks enable the user to move freely. The term mobility means that the user can roam freely and at the same time seamlessly enjoy network services without interruptions. This differs from portability which implies a situation that a user could carry a laptop and plug it into a network connector whenever needed in a fixed terminal. A number of issues related to user mobility led to the creation of the Mobile IP working group by the IETF.

Wireless networks are usually designed in a cellular fashion, where each cell covers a particular area. Users inside this area are serviced from a single BS (Base Station) or AP (Access Point), a host that is connected to the wired network and offers wireless connectivity to wireless hosts. During a handoff, all necessary information must be transferred between the two base stations so that the mobile host can continue to be connected. The handoff process usually takes about 300ms and sent packets may loss in this duration of time. There is now a consents within the research community that it is desirable for reliable transport protocols to be able to differentiate between congestion-related, transmission (random, interference, etc.), and motion-related (handoff, handover) losses [7], [8], [10] and [11].

2.6 Power consumption

For the devices that use battery for normal operation, power consumption is a very important aspect. Typically communication over a wireless links needs more energy than CPU processing. Energy efficiency does not only depend on the amount of avoidable extra data, but also on the total duration of connection. Prolonged communication times lead to excessive power consumption too [8].

3. Transport Layer Protocol Selection

For an end-to-end QoS support, there always has been a choice of TCP or UDP in underlying layers. An important problem in wireless computing is the use of TCP as the transport layer protocol. Normal TCP mistakes packet loss due to lossy nature of wireless links for congestion and call upon congestion control mechanisms to shrink the congestion window size and pull back the rate of packet flow. This leads to a significant reduction in throughput. Less research has been done with using UDP as the transport protocol for delay, jitter and loss sensitive streams. UDP is connection-less and does not maintain packet numbering and timing. It also does not provide any kind of information about the network state. The advantage of using UDP is that it does not employ any congestion control mechanisms like TCP. UDP does not take any preventive measures in the case of a loss of few packets which the user might not mind either. There have been a number of research proposing UDP in companion with RTP for time stamping and sequence numbering and using RTCP reports for obtaining network information to support QoS over wireless networks [7, 8, 9, and 10].

As the compression volume increase, the need to retain lost packets increase. TCP works effectively over traditional networks but it does not work effectively for multimedia services as its performance degrades due to window size decrease in the response to packet loss resulting from the congestion. TCP shows even worse performance over the wireless networks due to main problems discussed in wireless networks. There are number of ways to solve these problems including pure end-to-end protocols, link layer protocols, split connection protocols, soft-state transport-layer caching protocols, cross layer signaling protocols [7]. All of the proposed mechanisms show a defection point that avoids us from deploying them for ensuring QoS over wireless networks. In this paper we have chosen the UDP for our transport layer mechanisms and have coupled an application layer mechanism with it to ensure application level QoS, time stamping and sequence numbering.

4. Network State Feedback

Traditional network feedback is present in the form of Explicit Congestion Notification and Explicit Loss Notification in TCP. For multimedia delivery over the Internet, these notifications are not sufficient to provide QoS for multimedia data. Wireless networks even need more enhanced type of feedback as they need information about temporary disconnections and lossy link state. This information is needed in application level feedback.

In this section we first examine the MECN mechanism, and then we use an extension of it as WMECN for wireless networks which is used for data rate and video quality adaptation in application level. The WMECN method uses a kind of history for storing a memory of previous events so that it could use this memory in future events in response to MECN [4]. It gives a weight to each kind of network feedback with zero weight for wireless losses. Then the final summation of these values helps to decide about the next decision. We do not maintain the memory part of the algorithm and just use the marking scheme. In our algorithm, the rate adjustment coefficients are applied in application level.

As it is proposed in [3], MECN uses two bits that are being specified for the use of ECN in the IP header [6] to indicate four different levels of congestion instead of binary feedback provided by ECN. Four levels obtained by two bits are assumed for following feedbacks: "00" for non-ECN capable feedback, "01" for no congestion notification, "10" for mild congestion notification and "11" for severe congestion notification.

Three different thresholds have been proposed on the red queue, minimum threshold, maximum threshold and middle threshold. When the queue size is lower than minimum threshold no packet is marked. When the queue size is between minimum and middle threshold, packets are marked with "10" with maximum probability $P_{\max}$. The rest of packets remain unmarked. When the queue size is between middle and maximum thresholds, packets are marked with "11" with maximum probability $P_{\max}$ and the rest of packets are marked with "10". After the maximum threshold, all the packets are marked with "11".

Table 1: Different ECN values and how to interpret them

ECN Value	How to interpret
00	Non ECN capable
01	No Congestion
10	Mild Congestion
11	Severe Congestion

The proposed mechanism offers coexistence with competing traffics. It also offers coexistence possibility with old routers with different definitions of ECN fields in IP header, as proposed by Floyd in [5]. This will help using it with old routers as well as avoiding mistakes scenarios in routers.

5. Application Level WMECN (AWMECN)

Our proposed method, Application Level MECM, is using the WMECN marking mechanism in network layer in companion to decision making algorithm which works in application layer. For this purpose, an application layer header is added on UDP packets. This header contains the sequence number, timing fields and MECN field as well as an ACK field used for application level acknowledgement for received packets.

MECN is set in the router when the queue size reaches the defined thresholds. This is done using 7 and 8 bits in IP header. Then these values are set into application layer header and sent to the receiver. Arrival and transmission rates and their values in comparison is another factor in adaptation decision. The adaptation decision is made in the receiver and a rate adapt or quality adapt request is sent to the sender based on the MECN value and transmission and arrival rate values. The sender then adapts the quality and rate based on the received requests. Since we consider the frame sized to be fixed here, the quality adaptation is done with excluding some frames from transmission.

Using network backlog feedback in transport layer will only help us in rate adaptation by increasing or decreasing the window size. This will help to avoid congestion, but huge shrinks in window size have bad effects on video data which are very sensitive to rate variations. This leads to one of the important benefits of using the feedback in application level, which is the fact that it will enable us to consider more parameters than congestion in increasing or decreasing the sending rate as well as adapting the rate and quality in a way that have the least degrading effect on video stream. For example a feedback that adds disconnection duration, notified by the receiver to the sender, could be added. This feedback could help the sender to avoid sending packets in the short wireless disconnection periods in which all the packets are subject to loss. Another parameter that could be used in adaptation is the playback buffer occupancy which is beneficial in better understanding of video on demand situation. Adding these parameters would also cause in adding a quality adaptation scheme, which is a valuable character we gain in addition to the rate adaptation. Figure 1 shows the parameters that could be brought into account in the rate and quality adaptation schemes.

Since the decision making relies on different parameters rather than only a network backlog and loss, this kind of decision making avoids two mistakes scenarios for loss; identification of congestion loss as wireless link loss and identification of wireless link loss as congestion loss.

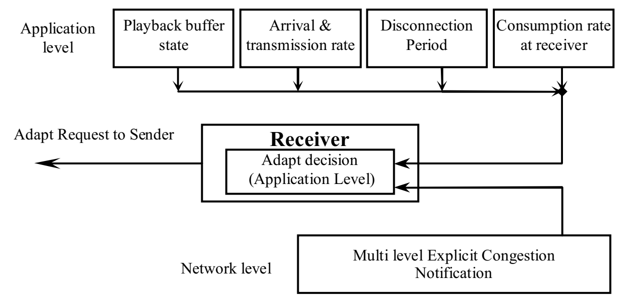

Figure 1: adaptation mechanism used

6. Simulation Results

We run the simulations using network simulator, NS-2 [13]. The general setting for the algorithm is assumed to be like the one in Figure 2. We tested the algorithm with different numbers of source and destination nodes.

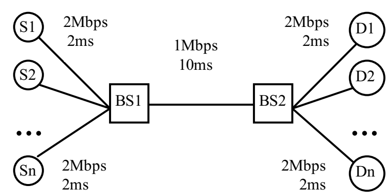

Figure 2: Simulation network setting

A loss generating module is used to generate loss function and disconnection durations with variable intervals of disconnection. This module is implemented in NS-2 and is used to simulate the lossy link state and disconnection durations caused by mobility and handoffs. The user can change the lossy link state simulation parameters by altering the loss probability and mean disconnection duration input to this function. The receiver has no additional information about the possibility, time and duration of disconnection.

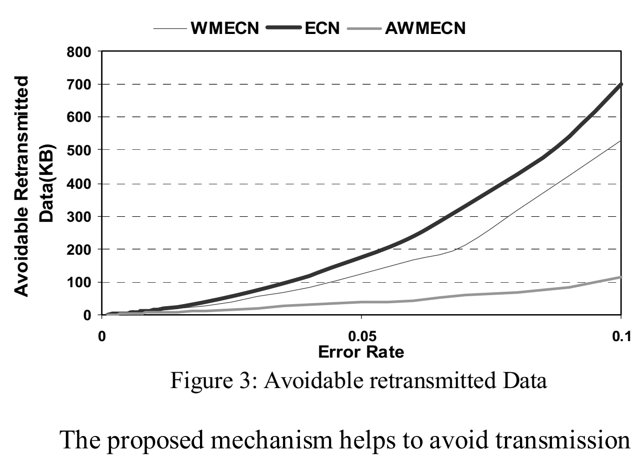

Figure 3: Avoidable retransmitted Data

The proposed mechanism helps to avoid transmission of extra data which is lost and retransmitted in the other similar algorithms. It also helps to reduce the transmission time by avoiding the effect of drastic shrinks of window size in

Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL'06)
0-7695-2552-0/06 $20.00 © 2006 IEEE

TCP-like mechanisms on video stream. This would help in better user perceived video quality as well as reduce transmission duration which is a factor in energy consumption in wireless devices.

Based on the facts discussed above avoidable retransmitted data and connection duration are used to be the main metrics used for evaluation of the mechanism. Evaluation of total avoidable data in ECN-using schemes is shown in comparison with the proposed method in Figure 3. It is obvious that the proposed mechanism shows a great enhancement in avoiding retransmissions.

Transmission duration is used for both evaluations of performance and average power consumption estimations. Different scenarios used for evaluation of the mechanism, show an average reduce of 1.9% in transmission duration compared with similar methods. Although the protocol is used in the application level and it may be sensible to think that application level adaptation will cause a prolonged communication, the proposed method shows a good communication duration and low extra data so leading to an efficient power consumption scheme.

As discussed before, the main advantage of this method is in the fact that it adds features that could not be supported in transport layer which result in short communication duration and less retransmitted data as well as improved quality achieved by the user.

7. Conclusion and Future Work

In this paper a modification of WMECN which is used in application level is proposed to provide an application level notification of loss, congestion, and lossy link state. Coupling this feedback with an efficient adaptation mechanism in application layer was used for obtaining a reliable multimedia transmission over wireless networks. All of the adaptive decisions are made in the receiver in application level. Then the results of these decisions making procedure is transmitted to sender via communication messages in the form of small requests in acknowledge packets send for each block of data. The sender then applies the requested changes in rate and quality. The transport layer protocol used for this communication is UDP which is the best option for multimedia transmission over wireless networks. Sequence numbering, frame acknowledge, time stamping and adaptation schemes are added to have a perfect protocol over UDP. The proposed scheme is simulated using NS-2 and has shown a better performance than other MECN mechanisms in transport layer. As this mechanism is implemented in application layer, other upcoming factors, such as multimedia characteristics can also be used to have a better decision in adaptation scheme.

8. References

[1] Ewa Kusmierek, David H.C. Du, "Streaming video delivery over Internet with adaptive end-to-end QoS", *The Journal of Systems and Software,* ELSEVIER, Vol.75, No.3, 2004, pp 237-252

[2] Veselin Rakocevic, "Congestion Control for multimedia applications in the wireless Internet", *International Journal of Communication Systems*, Vol.17 ,No.7 , 2004, pp. 723-734

[3] Arjan Durresi, Mukundan Sridharan, Chunlei Liu, Mukul Goyal, "Traffic Management using Multilevel Explicit Congestion Notification", Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics SCI'2001, ABR over the Internet, Orlando, FL, July 22-25, 2001, pp. 12-17

[4] Mukundan Sridharan, Arjan Durresi, Chunlei Liu, Raj Jain, " Wireless TCP Enhancements using Multi-level ECN", *Proceedings of SPIE Internet Quality of Service*, ITCOMM 2003 , Vol. 5245, Orlando Florida (2003) 180-187

[5] S.Floyd, "Specifying Alternate Semantics for the Explicit Congestion Notification (ECN) bit", Internet Draft (2005)

[6] K. Ramakrishnan, S. Floyd, D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168 (2001)

[7] Linying Xu, Manyun Liu, Lianfang Zhang, "TCP Performance Evaluation over Wireless Networks", *Canadian Conference on Electrical and Computer Engineering*, IEEE, Vol.2, May 2004, pp. 983-986

[8] Francis Jing Jyi Chai, Micro-Mobility Management and Freeze-TCP Implementation in a wireless network, MS Thesis, University of Canterbury, Christchurch, New Zealand

[9] Wenqing Ding, Abbas Jamalipour, "A New explicit loss notification with acknowledgement for wireless TCP", *12th IEEE Symposium on PIMRC*, IEEE, Vol.1 , 2001, pp. 65-69

[10] V. Venugopal Reddy, Vindo Sharama, M.B, Suma, "Providing QoS to TCP and Real Time Connections in the Internet", *Queuing Systems,* Kluwer Academic Publishers, No. 46, 2004, pp.461-480

[11] Ajay Kr. Singh, Sridhar Iyer, "ATCP: Improving TCP performance over mobile wireless environments", IEEE *4th international workshop on mobile and communication networks*, IEEE, 2002, pp.239-243

[12] Srisankar Kunniyur, R. Srikant, "End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks", *IEEE Transactions on Networking*, IEEE, Vol.11, No.5, October 2003, pp.689-702

[13] NS-2 distributed code web page at http://www.isi.edu/nsnam/ns/ns-contributed.html

0-7695-2552-0/06 $20.00 © 2006 IEEE