

# CONGESTION CONTROL FOR MULTIMEDIA APPLICATIONS IN THE WIRELESS INTERNET

**Veselin Rakocevic**

School of Engineering and Mathematical Sciences

City University

London EC1V 0HB, United Kingdom

E-mail: [V.Rakocevic@city.ac.uk](mailto:V.Rakocevic@city.ac.uk)

Phone: +44 (0)20 70408136; Fax: +44 (0)20 70408568

**Abstract:** This paper provides a parallel study of two important issues for the next-generation multimedia networking. Firstly, the emerging multimedia applications require a fresh approach to congestion control in the Internet. Currently, congestion control is performed by TCP; it is optimised for data traffic flows, which are inherently elastic. Audio and video traffic do not find the sudden rate fluctuations imposed by the TCP multiplicative-decrease control algorithm optimal. The second important issue is the mobility support for multimedia applications. Wireless networks are characterised by a substantial packet loss due to the imperfection of the radio medium. This increased packet loss disturbs the foundation of TCP's loss-based congestion control. This paper contributes to the ongoing discussion about the Internet congestion control by providing a parallel analysis of these two issues. The paper describes the main challenges, design guidelines, and existing proposals for the Internet congestion control, optimised for the multimedia traffic in the wireless network environment.

**Keywords:** Internet, Multimedia, Wireless Networks, Congestion Control, Quality of Service

## 1. Introduction

This paper analyses congestion control for unicast multimedia (video and audio) traffic in the wireless Internet. The current Internet congestion control is performed by TCP and is best suited for data traffic. Data traffic consists of non-prioritised packet flows that are insensitive to rate fluctuations. Congestion in the Internet is controlled by halving the traffic rate each time a single packet loss is experienced. TCP assumes that packet losses happen only due to the congestion in the network. Congestion control actions are performed at the network end-systems, with intermediate Internet routers only indicating the congestion by dropping or marking packets that arrive at a congested buffer. This mechanism has been very successful, producing a highly stable and scalable network.

The emergence of new, multimedia applications has brought the necessity for a change in many of the Internet design concepts. Real-time multimedia traffic flows require strictly bounded delay, delay variation and packet loss, and guaranteed throughput with limited fluctuations. The multimedia applications can broadly be classified into three classes: streaming of stored multimedia, streaming of live multimedia and real-time interactive multimedia. *Streaming* refers to the technique used when the client plays out multimedia content from one location in the file while receiving later parts of the file from the server. For stored multimedia, it is possible to pause, rewind, fast-forward or index through the content, while for streaming of live media the fast-forward option is not possible. Real-time interactive multimedia enables people to communicate with each other in real-time. Applications for this multimedia include voice-over-IP and video telephony.

Multimedia traffic in the current Internet can be transported over either TCP or UDP. UDP does not perform any congestion control. Multimedia traffic is sent over UDP at a constant rate equal to the drain rate at the receiver. TCP does perform congestion control, but this control creates large fluctuations in the fill rate in the receiver buffer. This is far from optimal for the

multimedia traffic, since a typical video traffic flow is highly sensitive to sudden and large rate changes. Multimedia traffic requires congestion control, but it requires a new approach to it, which would be more suitable for the end-to-end performance of multimedia applications.

This necessity for a new congestion control paradigm has been acknowledged by the IETF [3][4][26]. A number of studies [4][5][6][7][8] showed that streaming audio and video are better served by a congestion control mechanism which reacts slowly on packet losses, achieving smooth throughput changes. New congestion control schemes [8][9][10] are being developed and analysed and an increasing number of experimental [11] and simulation studies [5][12][13] compare the behaviour of these schemes in different network environments. Terms *TCP-friendliness* and *TCP-compatibility*, frequently used within the new schemes, suggest a dramatic change in the Internet design. These issues are discussed in more detail in section 3.

The second important trend in the multimedia networking today is the emergence of broadband wireless networks. It is possible to say that the wireless Internet has recently become a reality. Wireless networks - local networks (both structured and ad-hoc WLANs) and wide-area mobile networks - use radio medium to deliver high data rates to mobile users. Multimedia applications for wireless networks are emerging, and the integration of the wireless network domain with the fixed Internet is becoming increasingly important.

When it comes to congestion control, the two main problems with wireless networks come from the radio medium and from the pauses in the communication – handoffs. The radio medium generates very high level of bit error rates compared to the wired Internet. This produces large packet losses and confuses TCP congestion control, which understands these errors as a result of congestion. On the other hand, the handoffs disrupt the timeout-based retransmission deployed by TCP.

This paper contributes to the ongoing research on multimedia wireless networking by giving a parallel overview of the new congestion control paradigms for multimedia traffic and

the important design and research issues surrounding the deployment of both TCP and TCP-compatible control schemes in the wireless environment.

The paper is organised as follows. We first present briefly the features of the multimedia traffic and the current operation of congestion and flow control under the TCP/IP protocol suite in the fixed Internet. After that, the new schemes which enhance the congestion control performance in the multimedia environment are presented. This is followed by a survey of methods for deploying TCP over the wireless medium. Finally, the analysed issues are integrated and guidelines for the successful design are given.

## **2. Multimedia Traffic**

### **2.1. The Traffic**

Multimedia traffic comprises audio, video and data traffic. Even though numerous technical improvements are continuing to decrease the performance requirements multimedia applications put on the network, it is still vital to precisely define the expectations for throughput, data loss, and delay. The emerging approach for the multimedia traffic is adaptability – it is now possible to prepare audio and video bitstreams for the transmission on different bandwidth levels. This adaptability is absolutely crucial – the Internet is a packet-oriented connectionless network which is not designed to provide fixed circuit-like bandwidth connections. It is extremely difficult for the network to allocate the requested bandwidth to individual multimedia flows. To make that feasible, a complicated set of protocols would be needed to reserve bandwidth and enable special priority scheduling mechanisms to operate within the network. This is the main reason why adaptability became the most important feature of the video traffic when it comes to Internet transmission. The traffic itself must be flexible and adaptive.

When it comes to audio signals, we are concerned with two types of audio signal: speech signal as used in variety of interpersonal applications including telephony and video telephony, and music-quality audio as used in applications such as CD-on-demand and broadcast TV. Audio traffic requires relatively low rates, typically between 5Kbps and 10Kbps for the speech signal and up to 128Kbps for the CD-quality music signal.

Bandwidth requirements of video traffic are much greater than those of audio traffic. Video traffic uses the MPEG compression standards that generate video traffic flows with rates ranging from 1.5Mbps to 6Mbps. When it comes to delay and packet loss, multimedia applications in general are highly delay sensitive. It is possible to roughly estimate [11] that end-to-end delays under 100ms will be required to support interactive streaming multimedia.

On the other hand, data traffic has the only requirement a zero packet loss. End-to-end delay and jitter do not usually represent a performance requirement for a typical data transfer. Data traffic attempts to utilize the network bandwidth in full and has the objective of achieving the maximal throughput. TCP/IP protocol suite and associated TCP congestion control serve this type of traffic perfectly.

Unlike data transfers, streaming audio/video seeks to achieve smooth playback quality, rather than to simply transmit at the highest possible rate. Audio/video bitstreams should be prepared to expect (at least to a certain level) variable bandwidth levels and unexpected buffering, and to react to the information about the network availability. Naturally, the levels of adaptation are limited, and it is not possible to have a completely flexible real-time traffic flow. The next section presents several techniques that can be used to achieve this adaptability.

## **2.2. Quality Adaptation**

The prevailing conclusion in the Internet community is that the complexity in the network should be moved to the network *edges*. Multimedia traffic sources should be able to

change the sending rate based on the information about the state of the network. If this is achieved, the network could treat multimedia traffic flows as variable-rate flows, and the need for virtual-circuit reservation protocols will disappear, or at least would be reduced.

The set of mechanisms used to prepare the multimedia traffic for the Internet transmission is usually called *quality adaptation*. Quality adaptation actions include a number of distinctive techniques of compression, signal processing and encoding, developed to achieve better adaptation of audio/video traffic in the network. The design objective with these mechanisms is to prepare the traffic to be sent using a set of discrete transmission rates.

The audio/video bitstream should be structured in the following way. There should be a basic transmission rate, which will be used when the network congestion is high. This rate should accommodate essential audio/video information, producing the lowest acceptable quality level for the end-user. Depending on the congestion level in the network, additional audio/video data would be added to the bitstream, creating more complex (but more demanding) traffic flow and increasing the end-user quality of service. There is an increasing amount of research work being done on different quality adaptation schemes, especially for the MPEG-encoded video [5][14]. Since quality adaptation is not the main objective of this study, here we only briefly mention several solutions.

The most popular option is Hierarchical encoding [5] – the video stream is encoded at a base layer and one or more *enhancement* layers, which can be combined to render the stream at higher quality. There are many ways [14] to divide a video signal into two or more layers in the standard block-based hybrid video coder. A video can be temporally down-sampled, and the base layer can include the bit stream for the low-frame rate video, whereas the enhancement layer(s) can include the error between the original video and that up-sampled from the low frame-rate coded video. The same approach can be applied to the spatial resolution, so that the base layer contains a small frame-size video.

For hierarchically encoded data, a new layer is added only when there are assurances that there are enough resources in the network. Usually, this is done when the available bandwidth is greater than the consumption rate of the existing layers plus the new layer, and when there is sufficient total buffering at the receiver to survive an immediate backoff and continue playing all the existing layers plus the new layer.

Another method for quality adaptation is *simulcast* [5]. In simulcast, the bitstream is encoded at various target bitrates and it switches between the previously encoded layers as the available bandwidth changes.

### **3. Internet Congestion Control**

#### **3.1. TCP**

The dominant transport protocol in the Internet today is TCP. TCP provides transparent segmentation and reassembly of user data and handles flow and congestion control. TCP maintains a congestion window, which is an estimate of the number of packets that can be in transit without causing congestion. TCP packets are cumulatively acknowledged as they arrive in sequence. A TCP receiver sends two types of acknowledgements – it sends positive ACKs for segments received correctly and in-order, and duplicate ACKs for segments received correctly but out-of-order. The TCP sender retransmits the packet either when three successive duplicate ACKs for the same sequence number arrive (fast retransmit), or - when the session becomes idle or acknowledgements are lost - after waiting a certain amount of time (timeout-based retransmission).

Since TCP was designed with reliable fixed network in mind, the loss of a TCP segment indicates network congestion. TCP congestion control uses *Additive Increase, Multiplicative Decrease* (AIMD) linear control algorithm to control the sending rates. A

sending window  $W$  controls the sending rate.  $W$  increases from  $W$  to  $W + a$  following a receipt of an acknowledgement and decreases from  $W$  to  $W(1-b)$  following a congestion indication. The usual values for  $a$  and  $b$  are  $a = 1$ ;  $b = 0.5$ . In other words, the congestion window is halved for every window of data containing a packet drop, and increased by roughly one packet per window of data otherwise. The AIMD-based congestion control is optimal for rate-hungry data traffic, providing *fairness* and maximal resource utilisation [18], as long as all competing flows use the TCP congestion control. This means that TCP flows with similar Round Trip Times (RTT) control their rates so that the available bandwidth is distributed equally among them.

### 3.2. TCP-Friendliness

In the today's Internet TCP congestion control does not seem to be optimal for everyone. The multiplicative decrease of the rates in the case of congestion causes large oscillations in transmission rates. As we have seen before, in the case of a video stream, this degrades the smoothness, resulting in low end-to-end quality of service for the end-users. In order to avoid the congestion collapse, non-TCP (audio/video) traffic should make slower changes in the sending rates.

There are a number of proposed mechanisms for a congestion control that would be more suitable for multimedia traffic than the AIMD control. These include TCP-Friendly Rate Control (TFRC) [8], TCP Emulation at Receivers (TEAR) [10], Binomial controls [9], AIMD schemes with different values for the increase and decrease parameters, etc. All of these schemes are based on the notion of *TCP-friendliness* [3][4][8][9][19]. A congestion control mechanism is said to be TCP-friendly if its bandwidth usage, in the presence of a constant loss rate, is the same as TCP. The TCP-friendliness paradigm has extended the requirement for universality of the congestion control mechanism to the notion that all competing traffic flows



should be TCP-friendly. A good survey of TCP-friendly congestion control mechanisms is given in [20]. We will briefly present the most important schemes in the following.

TCP-Friendly Rate Control (TFRC) proposed by Floyd et. al. [8] is a rate-control method, which means that the transmission rate is explicitly calculated based on the information on the current packet loss rate in the network. This method is also called equation-based congestion control. The cornerstone of this approach is the observation that the bandwidth usage of a TCP flow in the presence of a constant packet loss rate  $p$  is proportional to  $1/\sqrt{p}$  [12].

Naturally, the algorithm for calculating the loss event rate is a key design issue in the equation-based congestion control. The appropriate control equation is the TCP response function (1). This formula gives an upper bound of the acceptable sending rate  $T$ , in the function of the packet size  $s$ , round-trip time  $R$ , steady-state loss event rate  $p$ , and the TCP retransmit timeout value  $t_{RTO}$ :

$$T = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RTO}\left(3\sqrt{\frac{3p}{8}}\right)p(1+32p^2)} \quad (1)$$

TCP Emulation at Receivers (TEAR) [10] is a mechanism in which the receiver emulates the congestion window evolution of a TCP sender. The receiver is aware of the packet loss rate, and it uses this information to maintain an exponentially weighted moving average of the congestion window. It then obtains a TCP-friendly sending rate by dividing the size of the congestion window by the estimated RTT.

Binomial controls [9] generalize TCP increase/decrease rules to derive a family of TCP-friendly congestion control algorithms with a varying degree of oscillations. The general functions for increase/decrease in binomial congestion control are:

$$\begin{aligned} \text{I: } & W_{t+\delta t} \leftarrow W_t + \alpha/W_t^k, & \alpha > 0 \\ \text{D: } & W_{t+\delta t} \leftarrow W_t - \beta W_t^l, & 0 < \beta < 1 \end{aligned} \quad (2)$$

$k$  and  $l$  are the parameters of binomial controls and  $W_t$  is the instantaneous window value, which governs the transmission rate. It has been shown [9] that a binomial congestion control satisfying the  $(k, l)$  rule, i.e.  $k + l = 1$ , is TCP-friendly. One member of this family, the so-called SQR T case ( $k = l = 0.5$ ) is especially attractive for streaming multimedia delivery due to its smaller magnitude of oscillations.

As we have seen, all of the new congestion control mechanisms have in common the need for slower response to the congestion indication in addition to the regular need for avoiding high network congestion.

The need for a new direction in the Internet transport layer has been recognized at the standardization level, with a number of proposals emerging [4][26]. The example of this is the proposal for the *Datagram Congestion Control Protocol* (DCCP) [4].

DCCP is intended for applications that ‘require the flow-based semantics of TCP, but which do not want TCP’s in-order delivery and reliability semantics, or which would like different congestion control dynamics than TCP’ [4]. DCCP provides a negotiation of a suitable congestion control mechanism, where the mechanism is defined using an 8-bit *Congestion Control Identifier* (CCID). The congestion control mechanisms currently under the standardization include TCP-like congestion control, which halves the congestion window in response to a packet drop, and TCP-Friendly Rate Control, which was explained briefly earlier in this section.

The next section analyses the challenges of the wireless Internet, for which TCP congestion control is not optimal even for the data traffic. Section 5 will integrate the conclusions into guidelines for research direction that will lead us to a scalable, efficient and transparent video service over both wired and wireless Internet domains.

## **4. Wireless Internet**

### **4.1. Mobility and Radio medium**

Introduction of wireless local area networks and the emergence (however slow) of multimedia applications for the 3<sup>rd</sup> generation mobile network architectures made the support for mobility an important issue in the QoS-oriented network design. Furthermore, the necessity for a comprehensive end-to-end QoS network design drives the research for a full interoperability between the wireless network and the fixed TCP/IP-based Internet.

High packet loss rate and limited bandwidth are the two features of the radio medium that present the biggest challenges for the successful design of multimedia wireless networks. For example, in a performance study of wireless networks [16][21] a wireless LAN suffered from a frame error rate (FER) of 1.55% when transmitting 1400 byte frames over an 85ft distance, with clustered losses. Such error rates are unmatched in the fixed networks. Furthermore, the throughput of the radio channel is substantially reduced due to factors such as multipath fading, co-channel interference and noise disturbances. The reduced throughput and high error rates in the wireless networks have a massive impact on the performance of loss-triggered TCP congestion control. This impact is analysed in detail in the next section.

Another important problem TCP has with the mobile systems are the long communication pauses whenever mobile devices move between adjacent cells – the handoffs. These pauses greatly effect the timeout-based TCP retransmission. Research [22] has shown that there are long pauses in transmission (up to 800msec) when handoffs occur if the TCP congestion control is used. An active TCP connection can also lose up to a full transmission window's worth of packets and related acknowledgements during each handoff. Recovery from losses due to handoffs in wireless systems can be achieved by TCP receiving appropriate signals from lower layers. Also, wireless link endpoints can choke TCP senders during

handoffs by transparently closing the receiver's advertised window [16]. Finally, the fast retransmission scheme proposed in [22] can significantly reduce the transmission pause caused by handoffs.

#### 4.2. Congestion Control in Wireless Networks

When it comes to congestion control and deployment of TCP over wireless links, a common conclusion is that the degraded performance of TCP over wireless links is mostly due to mistaking wireless losses for congestion.

File transfer tests [16][21] over a WLAN with a nominal bandwidth of 1.6Mbps achieved a throughput of only 1.25 Mbps. This 22% throughput reduction due to a FER of only 1.55% is caused by the frequent invocations of congestion control mechanisms that repeatedly reduce the transmission rate.

What can be done with TCP in order to enable it to be used over the wireless links? Fig. 1 shows the main components in the end-to-end TCP connection. One option is for the TCP sender to solve the problem itself, by distinguishing between random and congestion losses. The other option is to use the Access Point link layer and the Mobile User link layer to correct the errors using the link-layer retransmissions and thus hide the losses from TCP. Finally, it is possible to split the TCP connection into two connections, one 'fixed', from the TCP sender to the Access Point, and one 'wireless', from the Access Point to the Mobile User. In other words, TCP over wireless schemes can be classified into three broad categories: *link-layer protocols*, which provide local reliability; *split-connection protocols*, which break the end-to-end connection into two connections at the base station; and *end-to-end protocols*, where the TCP sender is aware of the wireless link.

### 1) *Link-layer protocols*

In general, to implement a high-quality video communication over wireless networks, intelligent adaptive transceivers are needed, and the choice of compression and error control schemes is absolutely critical. Even single bit errors in the compressed information can result in enormous errors in the decoded video. Due to predictive coding which is usually implemented, errors in the video bitstream will propagate from one frame to another until uncorrupted reference is re-established. A good overview of the link-layer mechanisms needed for wireless video communications is given in [14].

From the point of view of congestion control, the role of the link-layer error control is to hide the losses due to the wireless medium from the transport layer. If this is successful, TCP will see only congestion losses and the throughput reduction explained earlier will be avoided. The main link-layer mechanisms for dealing with errors are error correction techniques such as Forward Error Control (FEC) and retransmission of lost packets in response to Automatic Response Requests (ARQ) messages. Link-layer protocols can provide good results. A detailed performance comparison of congestion control mechanisms given in [17] shows that a reliable link-layer protocol can improve the throughput 10-30%.

Novel link-layer mechanisms include *Snoop TCP* [23]. Snoop is a link-layer protocol that uses information from TCP. It monitors and caches packets that pass through the TCP connection in both ways. Using the loss indications conveyed by duplicate acknowledgements or local timeouts, the link layer transparently retransmits lost data. The information about lost packets never reaches the TCP layer, thereby avoiding unnecessary congestion control invocations. However, like other link-layer solutions, Snoop suffers from not being able to completely shield the sender from wireless losses. This can happen either because of timer interactions between the two layers, or because duplicate acknowledgements themselves can be lost.

## 2) *Split-connection protocols*

Since end-to-end retransmissions are slow, TCP connections may be split using Access Points (see Fig.1). End-to-end connections are thus decomposed into separate TCP sessions for the wired and wireless parts of the path. Another protocol, optimised for error recovery, may be substituted over the wireless links.

The main problem with split schemes is that the splitting violates end-to-end TCP semantics. Acknowledgements may reach the sender before data packets reach their destination. To preserve TCP semantics, acknowledgements must be delayed, thus reducing throughput. Furthermore, Access Points face significant overhead, since packets undergo TCP processing twice, and considerable per connection state is maintained there.

The typical split-connection scheme is *Indirect TCP (I-TCP)* [24]. I-TCP uses regular TCP connection over the wireless link. This causes problems, since TCP sender on the wireless link often timeouts, causing the original sender to stop transmitting, which reduces throughput and increases delay. Further problem with this scheme is with handoff procedures, which tend to be slow due to the large amount of per-flow state maintained at the base station.

Another interesting scheme is WTCP [33]. This scheme also uses the base station of the wireless network as an Access Point where packets are buffered and retransmitted in the case of a loss due to the wireless medium. WTCP does maintain end-to-end TCP semantics by effectively hiding the time packets spend at the base station. This way, the round trip time estimation at the source is not affected and the source is able to detect the congestion.

## 3) *End-to-end protocols*

The end-to-end protocols are used when the sender is aware of the wireless link, for example when it is not possible to incorporate TCP-awareness in the base station. One approach is to enable the sender to somehow distinguish between the random and congestion losses. An interesting scheme for this is given in [34], where the losses are identified using inter-arrival

times at the receiver point. Although the scheme works only under heavy assumptions of high network load and having only a single wireless link in the end-to-end path, it presents an interesting approach to the problem.

The second mechanism that does allow the sender to distinguish between congestion and random losses is Explicit Loss Notification (ELN) mechanism [17]. ELN mechanism operates in the following way: when a packet is dropped on the wireless link, future cumulative ACKs corresponding to the lost packet are marked to identify that a non-congestion related loss has occurred. Upon receiving this information with duplicate ACKs, the sender may perform retransmissions without invoking the associated congestion control procedures.

Losses in wireless networks usually happen in clusters, resulting in losing multiple packets from a single TCP sender window. When this happens, TCP performs poorly, unnecessarily retransmitting packets and thus reducing the throughput. The use of selective acknowledgements (SACKs) [26] was standardised as a prevention mechanism in the case of these multiple losses. SACK packets are sent by the receiver to inform the sender about the data that has been successfully received. SACKs do not carry only the information about the next expected frame, but also a bitmask of packets seen so far. This information helps the sender to retransmit only the frames that have actually been lost. The main problem with this approach is the overhead. In the case of large windows, selective ACKs carry lot of data even if there are no losses.

There are several schemes that can reduce the overhead when using selective ACKs. SMART scheme [27] builds the bitmask of correctly received packets at the sender, instead of carrying it in every selective ACKs. The idea is that each acknowledgement should carry the cumulative acknowledgement and the sequence number of the packet that initiated the acknowledgement. This informs the sender which packets have been correctly received. SMART scheme improves the selective acknowledgements method, while retaining the drawback of not being able to fully distinguish between congestion and random losses.

A very important problem for implementing congestion control on wireless networks is the problem of *spurious timeouts* [16][25]. Spurious timeouts are timeouts that would not have occurred had the sender waited long enough. These timeouts are not typical for the wired network, but in the wireless network they are very important due to frequent disconnections – handoffs – that may last on the order of seconds.

Ludwig and Katz noticed [25] that spurious timeouts could be avoided if the TCP sender was able to distinguish between the acknowledgements for the original transmission of a segment and the acknowledgement for the retransmission. They propose *Eifel* scheme that uses TCP timestamp option [35] to distinguish between the two acknowledgements. Timestamps are echoed in acknowledgements, thus allowing false timeouts to be easily avoided. This introduces certain overhead to the system but does not affect TCP semantics and improves the performance in the environments of heavy load and frequent handoffs.

We have seen in this section that main challenges for the congestion control in wireless networks come from frequent and bursty random losses, fluctuating bandwidth levels and disconnections (handoffs). There are a number of schemes that attempt, partially or in full, to solve the problem, either by introducing additional intelligence at base stations, or by modifying the end-to-end TCP operation.

## **5. Integrated Solutions**

So far, this paper presented the most important congestion control issues for the multimedia Internet and for the wireless Internet. The objective of our study, however, is to analyse the challenges surrounding the integration of the mechanisms presented above. The question we want to answer is – what do we need to do to provide high-quality, reliable



streaming multimedia communication on the wireless Internet, without producing the danger of over-congesting the fixed Internet?

There are several studies and proposals for the integrated solution. Rejaie et al [28] propose a scheme for streaming video over TCP/IP using receiver buffering scheme for transmission and playback of hierarchically-encoded video. In their scheme, long-term adaptation is performed by adding and dropping layers of the video stream, and AIMD linear control of the sending rate is used to react to congestion. Buffer space is allocated between layers so as to place a greater importance on lower layers, thereby protecting these layers from increased transmission delays.

Lee et. al. [26] provide another integrated scheme for video streaming, consisting of an efficient video coding algorithm and a TCP-friendly rate-based congestion control scheme. The congestion control they propose is an AIMD control with 'history', where they add a 'history' parameter in the decrease algorithm to account for random or non-persistent losses. In this way, the rate reduction is done more gently which is suitable for multimedia applications. The coordination between the video transcoder and the congestion control mechanism is presented in [29]. This coordination is achieved through an *interaction* layer, which transmits the rate information to the application layer that performs the encoding.

Zhang et. al. [7] propose a cross-layer architecture for video delivery over the wireless network. In the transport layer, they introduce an intermediate node (gateway) on the edge of the wireless network domain. The sender performs application-level error control and a TCP-friendly congestion control based on congestion feedback from the mobile host and the gateway. The mobile host informs the gateway about packet losses by sending negative acknowledgements. The gateway retransmits the lost packets if they are still in the buffer. Another document [30] explains how to discriminate between the losses due to the congestion and the random losses due to the radio medium.

One of the increasingly important research issues is the nature of the congestion indication. Currently, packet loss is the indication for network congestion. Packet marking using Explicit Congestion Indication ECN [15] is used for congestion indication. A number of researchers [6] argue that in the future multiservice Internet, congestion should not be a binary value anymore. Congestion signalling should reflect the degree of congestion. The network should explicitly tell the receiver the state of congestion and how to react to it. This would decrease the possibility of invoking congestion avoidance in the case of a single, non-persistent loss. A typical example for this is the high packet loss in the wireless networks. A typical implementation of complex congestion indication would be in the case of a structured video bitstream. Complex congestion information will naturally produce larger overhead, but will at the same time allow the senders to increase/decrease their sending windows in a way that both reduces the congestion and does not have a negative effect on the end-to-end application performance.

As we have seen, the deployment of streaming multimedia over the wireless Internet brings multiple challenges. A cross-layer architecture is needed, providing these three main functionalities:

- Adaptive video encoding at the application layer
- TCP-friendly congestion control at the transport layer
- Efficient error control mechanism at the link layer

Interaction between these three functionalities remains the crucial design and research issue. An optimal scheme for multimedia transport on wireless networks should avoid duplicating of functionality between the three layers while satisfying the rigid requirements of the multimedia traffic.

## 6. Conclusion and Future Directions

This paper analysed the issues surrounding the design of a congestion control algorithm optimised for the multimedia traffic over the wireless links. The design problems have been identified as twofold: the first problem is to make the congestion control suitable for the streaming multimedia traffic; secondly, there needs to be a mechanism to distinguish between the losses which are result of congestion and the losses due to the wireless medium. This paper has analysed both issues extensively.

Main research issues for the future include the communication between the three layers included in the congestion control process – application, transport, and link layers; the behaviour of the TCP-friendly schemes in dynamic environments; and, increasingly, the performance of multicast applications in wireless environments. Finally, the lack of comprehensive simulation comparison of the integrated solutions presents a big obstacle and an interesting direction for the future research.

## 7. References

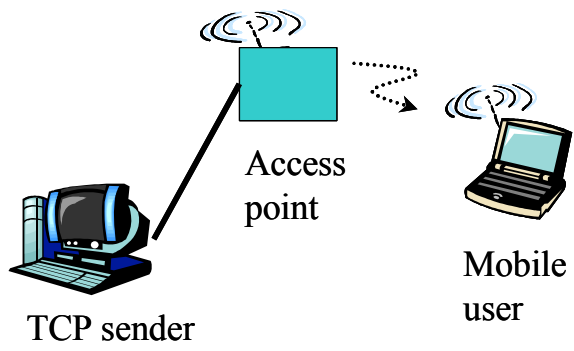
- [1] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RTP: a Transport Protocol for Real-Time Applications, IETF RFC 1889, January 1996
- [2] H. Schulzrinne, A. Rao, R. Lanphier, Real-Time Streaming Protocol (RTSP), IETF RFC 2326, April 1998
- [3] S. Floyd, Congestion Control Principles, IETF RFC 2914, September 2000
- [4] E.Kohler, M.Handley, S.Floyd, J.Pahdye, Datagram Congestion Control Protocol (DCCP), *IETF Draft (work in progress), draft-ietf-dccp-spec-01.txt*, March 2003

- [5] N. Feamster, D. Bansal, H. Balakrishnan, On the Interactions between Layered Quality Adaptation and Congestion Control for Streaming Video, *11<sup>th</sup> International Packet Video Workshop, Kyongju, Korea*, May 2001
- [6] D. Katabi, M. Handley, C. Rohrs, Internet Congestion Control for Future High Bandwidth-Delay Product Environments, *ACM SIGCOMM 2002*
- [7] Q. Zhang, W. Zhu, Y-Q. Zhang, A Cross-Layer QoS-Supporting Framework for Multimedia Delivery over Wireless Internet, *International Packet Video Workshop, PV2002, Pittsburgh, USA, 2002*
- [8] S. Floyd, M. Handley, J. Padhye, J. Widmer, Equation-based Congestion Control for Unicast Applications, *ACM SIGCOMM 2000*
- [9] D. Bansal, H. Balakrishnan, Binomial Congestion Control Algorithms, *IEEE INFOCOM 2001*
- [10] I. Rhee, V. Ozdemir, Y. Yi, TEAR: TCP Emulation at Receivers – Flow Control for Multimedia Streaming, NCSU Technical Report, April 2000
- [11] D. Loguinov, H. Radha, Large-scale Experimental Study of Internet Performance Using Video Traffic, *ACM Computer Communications Review*, 32(1), 2002
- [12] D. Bansal, H. Balakrishnan, S. Floyd, S. Shenker, Dynamic Behaviour of Slowly-Responsive Congestion Control Algorithms, *SIGCOMM'01, San Diego, CA, USA*
- [13] D. Wu, Y.T. Hou, W. Zhu, Y-Q. Zhang, J.M. Peha, Streaming Video over the Internet: Approaches and Directions, *IEEE Trans. On Circuits and Systems for Video Tech.*, 11(3), 2001
- [14] C. Adsumilli, Y.H. Hu, Adaptive Wireless Video Communications: Challenges and Approaches, *International Packet Video Workshop, PV2002, Pittsburgh, USA, 2002*
- [15] S. Floyd, TCP and Explicit Congestion Notification, *ACM Comp. Comm. Review*, 24:10-23, October 1994
- [16] G. Xylomenos, G. C. Polyzos, P. Mahonen, M. Saarinen, TCP Performance Issues over Wireless Links, *IEEE Communications Magazine*, 39(4), 2001

- [17] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, R. H. Katz, A Comparison of Mechanisms for Improving TCP Performance over Wireless Links, *ACM SIGCOMM '96*
- [18] D-M. Chiu, R. Jain, Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks, *Computer Networks and ISDN Systems 17* (1989), 1-14
- [19] S. Floyd, K. Fall, Promoting the Use of End-to-end Congestion Control in the Internet, *IEEE/ACM Transactions on Networking*, 7(4), August 1999
- [20] J. Widmer, R. Denda, M. Mauve, A Survey on TCP-Friendly Congestion Control, *IEEE Network*, 15(3), May-June 2001
- [21] G. T. Nguyen, R. H. Katz, B. Noble, M. Satyanarayanan, A Trace-based Approach for Modelling Wireless Channel Behaviour, *Winter Simulation Conference*, Dec. 1996
- [22] R. Caceres, L. Iftode, Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments, *IEEE JSAC*, 13(5), June 1995
- [23] H. Balakrishnan, S. Seshan, R. H. Katz, Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks, *ACM Wireless Networks*, 1(4), December 1995
- [24] A. Bakre, B. R. Badrinath, I-TCP: Indirect TCP for Mobile Hosts, *15<sup>th</sup> International Conference on Distributed Computing Systems (ICDCS)*, May 1995
- [25] R. Ludwig, R. H. Katz, The Eifel Algorithm: Making TCP Robust Against Spurious Timeouts, *Computer Communications Review*, 30(1), Jan 2000, pp. 30-36
- [26] M. Mathis, J. Madhavi, S. Floyd, A. Romanow, TCP Selective Acknowledgements Options, IETF RFC 2018, October 1996
- [27] S. Keshav, S. Morgan, Smart Retransmission: Performance with Overload and Random Losses, *INFOCOM'97*
- [28] R. Rejaie, M. Handley, D. Estrin, Layered Quality Adaptation for Internet Video Streaming, *IEEE JSAC, special issue on Internet QoS*, winter 2000

- [29] K-W. Lee, R. Puri, T-E. Kim, K. Ramchandran, V. Bharghavan, An Integrated Source Coding and Congestion Control Framework for Video Streaming in the Internet, *IEEE INFOCOM2000*
- [30] F. Yang, Q. Zhang, W. Zhu, Y-Q. Zhang, An Efficient Transport Scheme for Multimedia over Wireless Internet, *IEEE International Conference on 3G Wireless and Beyond (3Gwireless '01)*, June 2001, San Francisco, CA, USA
- [31] J. Kurose, K. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*, Addison-Wesley, 2001
- [32] D. Loguinov, H. Radha, Increase-Decrease Congestion Control for Real-Time Streaming: Scalability, *IEEE INFOCOM 2002*
- [33] K. Ratnam, I. Matta, WTCP: An Efficient Mechanism for Improving TCP Performance over Wireless Links, *3<sup>rd</sup> ISCC98*, Athens, Greece
- [34] S. Biaz, N. H. Vaidya, Discriminating Congestion Losses from Wireless Losses using Inter-Arrival Times at the Receiver, *IEEE Symposium ASSETT'99*, March 1999
- [35] V. Jacobson, R. Braden, D. Borman, TCP Extensions for High Performance, *IETF RFC 1323*, May 1992

*Veselin Rakocevic* obtained a PhD degree in Electronic Engineering from University of London in 2002 and a Diploma in Electronic Engineering and Telecommunications from University of Belgrade, Yugoslavia in 1998. He is currently a Lecturer at City University, London, where he teaches Computer Networks and Wireless Communications. His main research interests include multimedia wireless networking, evaluation and design of QoS-aware networking technologies, and network optimization. Dr. Rakocevic is an Associate Member of IEEE, member of IEEE Communication and Computer Societies, and a member of IASTED Technical Committee on Telecommunications.



V. Rakocevic; Fig. 1. A Typical End-to-end wired-wireless connection