Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Quality incentive based congestion control for multimedia communication over IP networks*

JOHANSEN Stian, KIM Anna N., PERKIS Andrew
(*Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology, Trondheim N-7034, Norway*)
E-mail: {stianjo; annak; andrew}@q2s.ntnu.no
Received Dec. 10, 2005; revision accepted Feb. 23, 2006

Abstract: In this paper we introduce a framework for using quality as an incentive to promote proper application level congestion control. Through integrating a joint-source channel coder and feedback-based congestion control scheme, we are able to construct accurate and efficient quality incentives. The framework is applicable in all network architectures where end-to-end congestion control may be used, and is as such not specific to either best-effort or traffic class-based architectures. The concept is presented along with preliminary simulations that highlight the resulting rate control accuracy. We also discuss how to implement some well-known congestion control schemes within our framework.

Key words: Congestion control, Incentives, Video communication, Joint source/channel coding, Unequal error protection
doi:10.1631/jzus.2006.AS0007 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

Heterogeneous and delay sensitive multimedia communication applications make up an increasing fraction of Internet traffic. These applications generally have lower requirements in terms of errors/losses than traditional applications of file-transfer nature. Meanwhile, when the number of users and/or amount of data communicated in the shared network is large, proper regulation of the network traffic must be in place to avoid potential congestion collapse. It is well understood that congestion control for multimedia applications should be tailored for their intrinsic characteristics, such that a certain level of media quality is retained at the receiver.

In general, a congestion control algorithm should take the following aspects into consideration:

(1) Fairness: the steady-state operation of the global system should give a "fair" allocation of bandwidth to each user.

(2) Link utilization: when the network is operating in equilibrium, in the router in question, the incoming traffic has an aggregate rate that is equal to the capacity of the outgoing link.

(3) Responsiveness: the congestion control scheme should react quickly and accurately when congestion occurs.

Additional requirements may be appropriate depending on the network architecture in question. For example, in a best-effort network, it is important that the rate of a multimedia source is "TCP friendly". Roughly speaking, this means that the source in question should not use more resources than a TCP connection would under the same network conditions.

In this paper, we introduce an end-to-end congestion control framework based on a quality incentive for the point-to-point unicast transmission scenario. When congestion occurs in the network, the users/end systems are encouraged to reduce their transmission rates by receiving rewards in terms of optimal quality. Quality can in this context be defined as, but certainly not limited to, for example, a distor-

* Project supported by the Research Council of Norway, Norwegian University of Science and Technology (NTNU) and the Norwegian Resarch Network (UNINETT)

tion metric or the tolerable delay. Each user is considered non-cooperative and only acts upon improvement of his or her own quality of transmission. The optimal transmission rate is determined by the congestion control scheme (CCS) in use.

Most congestion control literature targets the best-effort case where multimedia data needs to coexist with TCP traffic. In more recently proposed (and gradually deployed) Quality of Service (QoS) providing architectures such as DiffServ, separate traffic classes may be available for multimedia communications. Obviously, "TCP friendliness" is not an issue in this case. Furthermore, congestion may be avoided in such architectures through using deterministic traffic service guarantees. However, the perhaps more likely to be deployed probabilistic traffic guarantees imply the possibility of congestion. As such, intra-class congestion control in these priority-enabled architectures is not fundamentally different from the best-effort case. In any case, our framework is equally applicable in both the priority-enabled and best-effort cases.

Another advantage of our proposed framework is that it can theoretically function with any source-based CCS. The appropriate transmission rate is determined through effectively utilizing the network information that is available to the sender through a feedback channel. This is in turn implemented in the joint source-channel coder to align the rate given by the CCS and the transmission rate that gives the best end-to-end quality.

CCSs can be classified into increase-decrease or model/equation based. Two examples of CCSs in the former class are AIMD and AIPD. In the AIMD (Additive Increase, Multiplicative Decrease) algorithm, the transmission rate is to be reduced by a factor of the current rate upon detection of packet loss and increased linearly per round-trip-time in the absence of packet losses. The AIPD (Additive Increase, loss-Proportional Decrease) algorithm differs from AIMD in that rate reduction is made proportional to the packet-loss rate. It was shown in (Lee *et al*., 2001) that both AIMD and AIPD converge to fair rate allocation, while AIPD competes more aggressively for bandwith than AIMD (and is thereby less TCP friendly). Thus, AIPD may be more suitable for intra-class congestion control in DiffServ or other traffic-class based architectures. An example of model/equation-based CCSs is TFRC (TCP-Friendly Rate Control) (Floyd *et al*., 2000) which ensures TCP friendliness through modelling its throughput explicitly as a TCP connection.

Rate control protocols using these CCSs include e.g. (Rejaie *et al*., 1999) where the sender probes the network for available bandwidth and implements increase-decrease schemes depending on the network conditions. In (Puri *et al*., 2001), an Increase-Decrease algorithm based on packet loss history was proposed. This CCS is then further integrated with source coding by using a transcoding mechanism that effectively utilizes the obtained rate information.

The main difference between the above cited approaches and our proposed framework is that the above rely on the collaboration of the end users. Such an assumption is naturally no longer valid when there are selfish users aggressively occupying bandwidth and being unresponsive to rate control. It was stressed in (Floyd and Fall, 1999) that non-cooperative users can result in extreme unfairness or the potential of congestion collapse. Hence it is important that some form of incentive is in place to promote the use of end-to-end congestion control mechanisms. The introduced quality incentive also ensures that the end users' perceived quality is an integrated part of the congestion control mechanism.

In the following sections, we describe the quality incentive based congestion control and briefly review some CCSs as examples of how to determine the appropriate transmission rate, with the potential of the framework being illustrated by simulations. We summarize the paper in the concluding remarks and give possible extensions for future work.

QUALITY INCENTIVES FRAMEWORK

The main idea proposed in this paper is the introduction of an incentive in terms of quality for encouraging proper rate control. We propose to implement this by designing the joint source-channel coder (JSCC) in end systems such that, when packet losses increase, the user (or, realistically, the transmission rate controller, see Fig.1) should be encouraged to reduce the rate. The rate yielding lowest distortion (or equivalently, the best quality) should be according to the one determined by the congestion control scheme in question. To clarify the logical structure it can be useful to think of this as two rate control systems operating in a hierarchical relationship at the sender. The inner rate controller is the one implementing congestion control. This is located in the JSCC while the outer rate controller is the simpler (presumably opportunistic) user-level control which merely aims to choose the distortion-optimizing rate at every time instant. This logical structure is shown in Fig.1. Since the inner rate controller (the one performing congestion control) in practice dictates the rate behavior of the flow (assuming that the incentive is indeed followed), this rate control scheme needs to be in accordance with the network configuration in question. We stress that the actual rate control algorithm implemented can theoretically be any end-to-end congestion control scheme. Development of new rate control algorithms for congestion control and avoidance is outside the scope of this work, we focus on the application of existing schemes. Examples and implications of some well-known congestion control algorithms are discussed in Section 3.

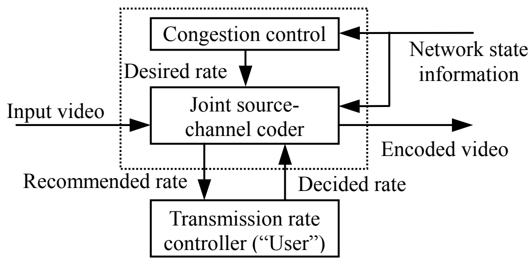

Fig.1 Logical structure of inner/outer rate controllers

In the following we formulate the distortion-rate requirements of a JSCC providing congestion control through quality (distortion) incentives. To unify the analysis, we introduce the end-to-end distortion D as a function of total transmission rate R and a second parameter γ. The γ parameter can simply be the packet loss rate, but it may also be a vector of network state parameters. The interpretation of $D(R,\gamma)$ is that it gives the expected end-to-end transmission rate R under channel conditions given by γ. In the following, the index i in R_i can be thought of as a time index.

Using an arbitrary congestion control scheme, we assume that the appropriate transmission rate R_1 is given as a function of the parameter set γ (and possibly the current transmission rate R_0):

$$R_1 = CC_R(R_0, \gamma), \tag{1}$$

where CC_R is the rate control function in the CCS. We wish to find the distortion-rate behavior of our JSCC such that the user-level (outer) rate control can simply be stated as the following minimization problem:

$$\hat{R}_1 = \min_R D(R,\gamma)|_{\gamma=\gamma'}, \tag{2}$$

where $\hat{R}_1$ is the new transmission rate found as the result of this minimization. The distortion-incentive requirements are then given by

$$D(R,\gamma)|_{\gamma=\gamma'} > D(R',\gamma)|_{\gamma=\gamma'},\ \forall R \neq R' = R_1. \tag{3}$$

That is, given a network state as represented by the parameter set γ', the distortion-minimizing transmission rate should be as given by Eq.(1). It is useful to consider the following two cases separately.

Case A (Congestion $R_1<R_0$) In the congested phase, the non-ideal (lossy) network state will inevitably degrade end-to-end quality. The error-resilience properties of the joint source-channel coder can then be adjusted in such a way that the optimal transmission rate (in a distortion sense) is equal to the rate given by Eq.(1). This is the foundation of the distortion-incentive scheme presented here.

Case B (No congestion $R_1 \geq R_0$) In the congestion-free case, the application should be allowed to increase its rate in a controlled manner. Generally, all congestion control schemes give an upper bound for how much the transmission rate should be allowed to increase (in order to prevent "unfair" resource usage). Assuming that our congestion control scheme allows a rate increase of R_Δ, Eq.(3) becomes

$$D(R)|_{\gamma=\gamma_0} > D(R')|_{\gamma=\gamma_0},\ \forall R \neq R' = R_0 + R_\Delta, \tag{4}$$

where γ_0 is the parameter set when there is no congestion. A somewhat counterintuitive implication of Eq.(4) is that the distortion-rate function must be strictly increasing for rates greater than R_0+R_Δ in the congestion-free case. Clearly, this does not harmonize well with the properties of distortion-rate functions as known from information theory. It is however necessary in order to constrain the rate increase. In the unconstrained case (with D being a strictly monotonically decreasing function of R), a user performing rate control according to Eq.(2) would increase its rate ad infinitum.

For a visualization and example of the incentives scheme (specifically, Case A above), consider Fig.3a (a more thorough explanation of how the plots are generated is given in Section 4). Given that the current transmission rate is 600 kbps and there is no congestion, we are observing the maximum possible PSNR. Now consider the onset of congestion with a packet loss rate increase to, say, 5%. Our CCS (in this case AIPD) will indicate that the appropriate transmission rate in this congested phase should be 525 kbps. The user/end system may or may not choose to adhere to this rate control regime. However, considering the incentives framework as exemplified by the plot, the PSNR-maximizing transmission rate given that the packet loss is 5% is approximately the recommended rate, 525 kbps. All other transmission rates will give an inferior PSNR performance under these network conditions.

EXAMPLE CONGESTION CONTROL SCHEMES

To clarify the properties of the quality-incentive scheme, we give a short summary of two important congestion control schemes and their possible implementation in the proposed framework.

AIMD: Additive Increase, Multiplicative Decrease

This scheme is the foundation of TCP. Upon detection of packet loss, the rate is reduced by a fraction of the current transmission rate. Rate increase is, as the name suggests, done linearly through incrementing the sending rate by a fixed amount per round-trip time. In this case Eq.(1) becomes

$$R_1 = R_0 + \alpha, \text{ when } \gamma=0, \tag{5}$$

$$R_1 = R_0(1-\beta), \text{ when } \gamma>0, \tag{6}$$

where γ in this case is simply the packet loss rate. The rate control of AIMD can be approximated by the distortion-incentives scheme, as seen in Section 4.

AIPD: Additive Increase, loss-Proportional Decrease

Making the rate decrease proportional to the experienced packet loss rate gives a less dramatic rate decrease upon packet loss detection and generally a less oscillating rate evolution than that of AIMD. This is the main principle of AIPD, where the rate control relations are as follows:

$$R_1 = R_0 + \alpha, \text{ when } \gamma=0, \tag{7}$$

$$R_1 = R_0(1-\beta\gamma), \text{ when } \gamma>0, \tag{8}$$

where γ is once again the packet loss rate. AIPD is well suited for implementation through the incentives scheme, due to the linear dependency on packet loss rate. It is certainly possible to make Eq.(3) promote a small rate decrease at low and larger rate reductions in the case of higher γ. Simulation results for AIPD are presented in Section 4.

SIMULATIONS

We provide a set of model-based simulations to validate the performance of the proposed scheme. The subband-based video coder 3D-SPIHT (Kim and Pearlman, 1997) is used as the source model in this work. The PSNR performance of the coder is approximated through averaging the coding performance (frame-by-frame) for the standard test sequences Foreman, Akiyo and Stefan at a number of bit rates. We model the PSNR-rate-function of this coder through least-squares curve fitting to a Weibull parametric model as described in (Charfi, 2004). Since we use PSNR as the quality measure rather than distortion, the inequality in Eq.(3) must be turned with $PSNR(R,\gamma)$ in place of $D(R,\gamma)$.

The channel model used in the simulations shown here is the simplified Gilbert model (Yee and Weldon, 1995). This two-state markov model is defined by the two parameters p (avg. packet loss probability) and ρ (correlation between consecutive packet losses). The random (binomial) packet loss model was also tested, yielding results similar to that of the Gilbert model with low ρ. In the simulations shown here we use a packet size of 512 bytes.

We use Unequal Error Protection (UEP) to implement the distortion incentives (Mohr *et al*., 2000; Stankovic *et al*., 2002). In this context it is sufficient to say that this error protection scheme turns an embedded (progressive) bitstream into a packetized representation where all packets are equally important for reconstruction quality. Error protection (FEC) is allocated according to the relative importance of the data in question. Considering the rate decrease case, the distortion incentive is implemented through using UEP to make a less error resilient allocation for the higher rates compared to the lower rates. This concept is illustrated in Fig.2. For high rates we naturally have a better PSNR performance initially, but as packet losses increase this allocation gives a rapidly decreasing performance. For lower rates we enforce a more robust error protection. This is implemented through optimizing the UEP allocation for the specific packet losses and at the transmission rates that are appropriate at this packet loss rate [according to Eqs.(6) and (8), etc].

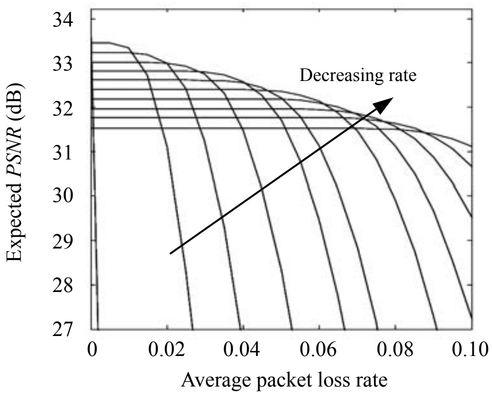

Fig.2 Varying error resilience performance at different rates. Lower transmission rates are more error resilient

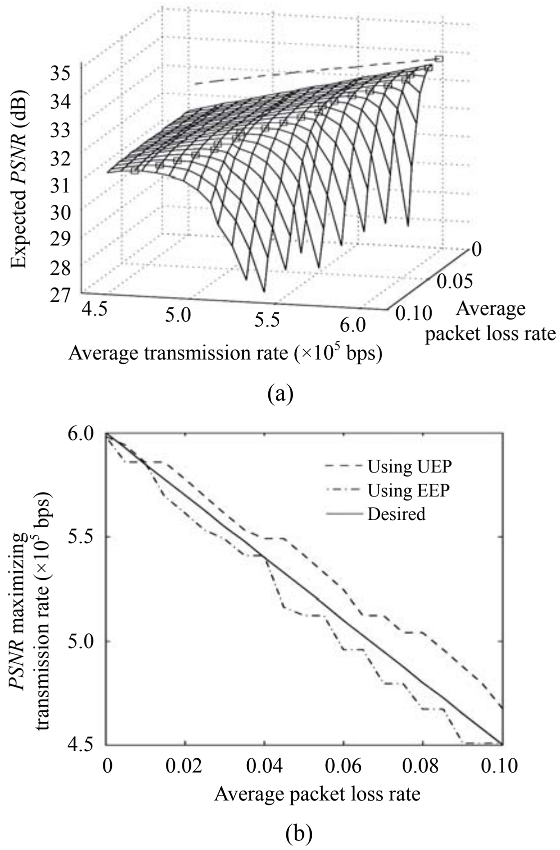

Fig.3 Simulation results, AIPD with β=2.5 and Gilbert channel with ρ=0.15. (a) *PSNR* performance as a function of packet loss and transmission rate when using UEP; (b) The resulting ability of UEP (dashed line) and EEP (dashed-dotted line) to track the rate as given by the AIPD congestion control scheme (solid line)

1. AIPD

Fig.3 gives simulation results for AIPD rate control with an initial transmission rate of 600 kbps, β=2.5 and Gilbert channel parameters ρ=0.15, respectively. Fig.3a shows *PSNR* as a function of packet loss and total transmission rate. Fig.3b shows the ability of this scheme to track the rate that should maximize *PSNR* (note that EEP, Equal Error Protection, is also simulated). That is, the plot shows, for a given packet loss rate, what transmission rate actually maximizes *PSNR*. These rates should ideally coincide with the rate given by the congestion control relation [that is $\hat{R}_1 = R_1$, see Eq.(7)]. In Fig.3b, R_1 is shown as a solid line.

Common to Fig.3a and Fig.4 is that the dashed line along the zero-packetloss axis shows the performance of the video coder when no redundancy is used at the corresponding rates.

2. AIMD

Fig.4 shows the corresponding results for the AIMD rate control scheme (Eq.(6)) with β=0.4. The plot shows *PSNR* as a function of both total transmission rate and packet loss rate. As is evident from the figure, the performance is as desired with a strong incentive for reducing the rate by a factor of β for all nonzero packet losses.

3. Rate increase

As formulated in Eq.(4), it is (in the noncongested case) necessary to force an increasing $D(R)$ curve at rates higher than that indicated by the rate controller. The actual shape of the curve can theoretically be arbitrary as long as Eq.(4) is satisfied, but the practical effect should also be considered. Specifically, the slope of the $D(R)$ curve at rates $R>R_1$ will determine the incurred penalty when transmitting at these rates. A simple way of implementing this is to, for $R>R_1$, use an increasing proportion of the total transmission rate as stuffing (dummy) data. As an example, consider a "mirroring" of the $D(R)$ curve around the maximum allowed rate. This would give us an actual transmission rate of

$$R_{\text{actual}} = R_1 - (R - R_1), \text{ when } R > R_1. \quad (9)$$

We also mention that, at the end of a congested phase, all redundancy (FEC) that may have been added by the source-channel coder during congestion will be removed. In this way, the actual information rate communicated will increase, while the transmission rate conforms to the limitations as imposed by the CCS in question.

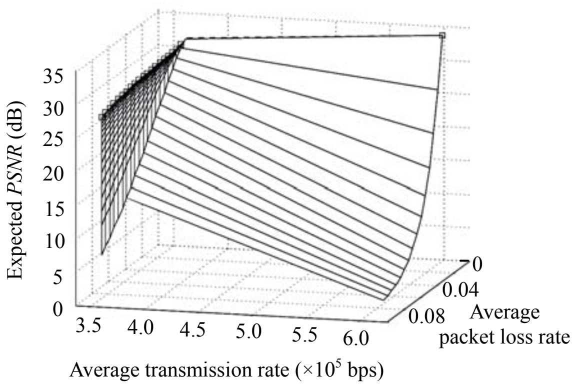

Fig.4 Simulation results, AIMD with β=0.4 and Gilbert channel with ρ=0.65

CONCLUDING REMARKS AND FUTURE WORK

This paper has presented a framework for promoting rate control through quality-based incentives. The framework is applicable in all scenarios where congestion may occur and it is possible to feed back network state information to the sender. We have also presented initial simulation results that highlight the intended operation of the system. Results showed that the incentives can be accurately matched to different congestion control schemes using familiar channel coding techniques like UEP and EEP.

The proposed framework can be extended in a range of directions. We are currently looking at the following areas as future work.

From an application layer perspective, the notion of quality can be extended beyond the distortion-centric view considered in this paper. In wireless communication, rate control can relate directly to power utilization, which is an important design parameter for handheld devices. JSCC can then be combined with transmission power control. Similar arguments can be raised for considering delay constraints.

Regulating intra-class traffic flows in DiffServ architecture using the framework defined in this paper is being looked into.

Finally, it is necessary to develop JSCC schemes that can result in better accuracy in the tracking of the desired rates (Fig.3b).

ACKNOWLEDGEMENT

We sincerely thank P. J. Emstad, A. Lie and Ø. Kure, NTNU, for valuable comments to this work.

References

Charfi, Y., 2004. Efficient Error Protection for Image and Video Transmission over Noisy Channels. Ph.D Thesis, University Konstanz, Germany.

Floyd, S., Fall, K., 1999. Promoting the use of end-to-end congestion control in the Internet. *IEEE/ACM Trans. on networking*, **7**(4):458-472. [doi:10.1109/90.793002]

Floyd, S., Handley, M., Padhye, J., Widmer, J., 2000. Equation-Based Congestion Control for Unicast Applications. ACM SIGCOMM Communication Review, p.43-56.

Kim, B.J., Pearlman, W., 1997. An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierachical Trees. Proceedings of Data Compression Conference. Snowbird, USA.

Lee, K.W., Kim, T., Bharghavan, V., 2001. A Comparison of Two Popular End-to-End Congestion Control Algorithms: the Case of AIMD and AIPD. Proceedings of GLOBECOMM.

Mohr, A.E., Riskin, E.A., Ladner, R.E., 2000. Unequal loss protection: graceful degradation over packet erasure channels through forward error correction. *Journal on Selected Areas in Communication*, **18**(6):819-828. [doi:10.1109/49.848236]

Puri, R., Lee, K.W., Ramchandran, K., Bharghavan, V., 2001. An integrated source tarnscoding and congestion control paradigm for video streaming in the Internet. *IEEE Trans. on Multimedia*, **3**(1):18-32. [doi:10.1109/6046.909591]

Rejaie, R., Handley, M., Estrin, D., 1999. RAP an End-to-End Rate-Based Congestion Control Mechanism for Realtime Streams in the Internet. Proceedings of INFOCOMM.

Stankovic, V., Hamzaoui, R., Xiong, Z., 2002. Packet Loss Protection of Embedded Data with Fast Local Search. Proceedings of IEEE International Conference on Image Processing. Rochester, NY, USA.

Yee, J.R., Weldon, E.J., 1995. Evaluation of the performance of error-correcting codes on a Gilbert channel. *IEEE Trans. on Communications*, **43**(8):2316-2323. [doi:10.1109/26.403764]