Network-adaptive Scalable Video Streaming Over 3G Wireless Network

Qian Zhang, Wenwu Zhu, and Ya-Qin Zhang
Microsoft Research, China
5F, Bejing Sigma Center, No. 49, Zhichun Road
Haidian District, Beijing 100080, P. R. China

Abstract

Scalable video streaming over wireless link with Quality of Service (QoS) is a very challenging task due to the time-varying characteristics of wireless channel and limited battery resource in the handheld devices. This paper proposes an end-to-end network-adaptive architecture for video streaming over 3G wireless network. The proposed architecture not only dynamically estimates the varying network status on the fly, but also simultaneously performs application-level error control and transmission-level power control to protect the random and fading error occurred across the 3G network. Distortion/power-minimized rate allocation is presented to achieve the minimum end-to-end distortion or minimum total power consumption based on user's requirement. Simulation results demonstrate effectiveness of our proposed scheme.

I. Introduction

Wireless video applications and services have undergone enormous development recently due to the continuing growth of wireless communication, especially the emergence of 3G wireless network. However, wireless video streaming poses many challenges. It is known that there is limited bandwidth in wireless network and limited battery resource in the mobile host. Meanwhile, the capacity of a wireless channel is fluctuated due to the changing distance between the base station and the mobile host. Thus, from the system point of view, it is important to estimate the available wireless network condition dynamically. On the other hand, from the video compression point of view, it is necessary to employ scalable video coding in wireless environment. It has been shown that scalable video is suited to handle the variability of network condition gracefully [1].

Compared to the wired links, the wireless channels are typically much more noisy and have a higher bit error rate. Meanwhile, multipath and shadowing fading occur frequently in wireless channel. As a result, random and burst errors can have devastating effect on video streaming quality. Error control and power control are two very effective approaches for supporting quality of service (QoS) in robust video transmission. Error control is performed from individual user point of view by introducing redundancy to combat the transmission errors [2]. Power control is performed from group point of view by controlling the transmission power and transmission rate for a group of users [3]. Therefore, error control and power control techniques are necessary to ensure high-quality video delivery from application level and transmission level, respectively.

The existing error control schemes, e.g., ARQ (Automatic Retransmission reQuest), FEC (Forward Error Correction), hybrid FEC/ARQ, and UEP (Unequal Error Protection), are designed to minimize the distortion [4]. Most power control schemes, e.g., closed-loop and open-loop power control, are designed to achieve goals like guaranteeing the low bound of signal to noise ratio, balancing received power levels, or maximizing cell capacities. To date the techniques of error control and power control have been studied separately, and few of them consider the battery lifetime of mobile hosts in their designs. Although a strategy may not consider battery lifetime alone, it can balance the need for high battery lifetime with the need of providing reasonable cell capacity and quality of service.

In this paper, we propose an end-to-end architecture that simultaneously controls error correction and transmission power to maintain the desired QoS for scalable video streaming over 3G network, based on the measurements of throughput and error rate for 3G network. Note that our scheme is capable of dynamically adapting to the varying wireless network condition and minimizing the end-to-end distortion or the total power consumption in mobile host by performing optimal rate allocation.

II. End-to-end Architecture for Scalable Video Streaming over 3G Network

To efficiently streaming video over 3G wireless, we develop a network-adaptive architecture which consists of scalable video representation, wireless network measurement, network-adaptive application-level error control, and network-adaptive transmission-level power control. In this architecture, as the wireless network conditions change, sender can scale the video streams, protection data as well as transmitted signal power, and transport the scaled streams to the receiver with a good perceptual quality.

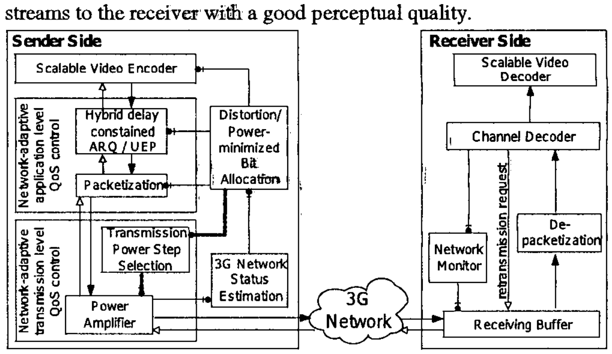


Figure 1. The end-to-end architecture for scalable video streaming over 3G network.

Figure 1 depicts our proposed end-to-end framework for scalable video streaming over 3G network. In this work, we use PFGS (Progressive Fine Granular Scalable) [1] as an example of scalable video. On the sender side, PFGS encodes raw video into several layers. Taking the video characteristics into account, these layers are packetized and protected against channel errors according to their importance and channel conditions using hybrid delay-constrained ARQ and UEP scheme. Then, video packets are transmitted over wireless channel through a power amplifier module. On the receiver side, after the process of channel decoding, the reconstructed packets are directed to the source decoder. In the mean time, the end system can sense the network status and feed back the related parameters to the sender. Having got the estimated 3G network status, the *distortion/power-minimized bit allocation* module on the server side re-allocates resource for achieving the minimal objective function under the required QoS and resource constraints. To achieve the minimum video distortion, the available bits are distributed among source, FEC, and ARQ to minimize the

0-7803-6725-1/01/$10.00 ©2001 IEEE

end-to-end distortion. Meanwhile, to achieve the minimum power consumption, the transmission power control step and channel protection levels are determined to minimize the total power consumption in the mobile host.

The key components in this architecture consists of *3G network status estimation, network-adaptive application-level error control, network-adaptive transmission-level power control* and *distortion/power-minimized bit allocation*. We will present details for those modules in the following sections.

III. 3G Network Status Estimation

Early work in 3G wireless channel estimation focused on the stochastic modeling of channel dynamics at physical layer, measured by received signal strength or bit error rate [5]. Such physical layer models cannot be directed used to evaluate the application-layer network performance, such as available bandwidth and packet loss ratio. This section will develop packet-level 3G network model and address how to estimate error rate and throughput using this model.

Transmitted over 3G network, video bitstream is first packetized and transported using UDP protocol. Then the corresponding packets are transferred to the RLC (radio link control) layer, where they are segmented into smaller RLC frames. Different ARQ schemes can be used to achieve the desired reliability. Consequently, the RLC frame is delivered through physical layer. One of the main services provided by the physical layer is the measurement of various quantities, such as physical-channel bit error rate, transport-channel block error rate (P_{BL}), transport-channel bit rate (B_{TRANS}), etc.

After obtaining the average block error rate, P_{BL}, we use a first-order Markov process to model the transmission on correlated Rayleigh fading channel and analyze the system performance at the data link layer. More specifically, a sequence of RLC frames' successes and/or failures can be approximated by a two-state Markov chain, which is defined by the transition matrix

$$M(x) = \begin{pmatrix} p(x) & r(x) \\ s(x) & q(x) \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}^x, \quad (1)$$

where p and $1\text{-}q$ are the probabilities that the j^{th} RLC frame transmission is successful, given that the $(j\text{-}1)^{th}$ RLC frame was successful or unsuccessful, respectively. Using this model, the steady-state transport block error rate, P_{BL}, is given by

$$P_{BL} = \frac{1-p}{2-p-q}. \quad (2)$$

As discussed in [6], for a Rayleigh fading channel with fading margin F, the average transport block error rate (P_{BL}) and the Markov parameter (q) can be expressed as

$$P_{BL} = 1 - e^{-1/F} \quad (3)$$

and

$$q = 1 - \frac{(1-P_{BL}) \times (Q(\theta, \rho\theta) - Q(\rho\theta, \theta))}{P_{BL}}, \quad (4)$$

where the definition for related parameters had been found in [6].

To measure the end-to-end throughput that applications are interested in, we need to consider the interaction of RLC and the UDP layer. As mentioned above, the total bandwidth used for each transport channel (B_{TRANS}) can be calculated in the physical layer. To accurately estimate the available throughput in the application layer, we further analyze the status of successive UDP packets. Let $P_{u,ss}$ and $P_{u,fs}$ be the probabilities that the current UDP packet is successful given that the previous UDP packet was successful or not. Further, let $P_{u,sf} = 1 - P_{u,ss}$ and $P_{u,ff} = 1 - P_{u,fs}$. Then, the available UDP throughput can be defined as

$$AUB = B_{trans} \times \frac{1-P_{u,ff}}{2-P_{u,ss}-P_{u,ff}} = B_{trans} \times \frac{1-P_{u,fs}}{1+P_{u,fs}-P_{u,ss}}. \quad (5)$$

The relation between packet transition probabilities and link layer frame transition probabilities had been derived in our previous work [6].

Based on the above analysis, the network performance, such as available UDP throughput, frame/packet transition probabilities, BER/BLER, etc., are reported to the application layer periodically. The application-level error control and transmission-level power control for video streaming then can be performed adapting to the varying network condition.

IV. Network-adaptive Application-Level Error Control

We propose a hybrid UEP and delay-constrained ARQ scheme for scalable video delivery. PFGS encodes raw video into base layer (BL), which carries the most important information, and enhancement layers (ELs), which carry less importance information. Taking the video characteristics into account, BL and ELs are protected differently. More specifically, we adopt hybrid delay-constrained ARQ and FEC for BL error protection, and adopt UEP to protect different ELs.

On the sender side, based on the delay constraint that is limited by video frame rate, current roundtrip time, and the estimated time consumed by processing procedure, we can calculate the maximum number of transmissions for current packet, $N_{\max}$. Upon receiving the retransmission request for the corrupted packet, the source side will only transmit the necessary part of higher protection for the packet. The sender determines the level of protection for each transmission such that the required residual error rate is within the desired range and the overhead is minimized. Similarly, to efficiently add error protection to ELs, sender determines the degree of protection for each layer to achieve the minimal objective function under the required QoS and resource constraints.

We will further analyze the expected end-to-end distortion and corresponding power consumption after employing the hybrid UEP and delay-constrained ARQ scheme. The bit rate of the source side is composed of bit rate in both BL and ELs. Mathematically,

$$R_s = R_{s\_base} + \sum_{i=1}^{L} R_{s\_enh}(i), \quad (6)$$

where L is the number of layers in ELs, R_{s_base} and R_{s_enh} represent the source rates of BL and of ELs, respectively. Then, source distortion can be described as

$$D_s(R_s) = D_s(R_{s\_base}) + \sum_{i=1}^{L} D_s(R_{s\_enh}(i)). \quad (7)$$

The consumed processing power for the source part is related to the source coding rate, which is denoted as

$$PC_s(R_s) = \sum_{i=1}^{bn} \rho s(R_{s\_base}(i)) + \sum_{j=1}^{L} \{ \sum_{i=1}^{bn_j} \rho s(R_{s\_enh}(i,j)) \}, \quad (8)$$

where $\rho s(.)$ is the rate-power relation for PFGS source coding, which can be obtained from [6].

Next, we will discuss the channel distortion. In this work, we use *RS (n, k)* code for forward error correction. The protection rate needed for the BL delivery is calculated as follows:

$$R_{ARQ} = \sum_{i=1}^{bn} R_{prot}(t_1, R_{s\_base}(i)) + \sum_{j=2}^{N_{\max}-1} \{ \sum_{i=1}^{bn} [P_{fail}(i, j-1) \times R_{prot}(t_j, R_{s\_base}(i))] \}, \quad (9)$$

where b_n is the number of source packets needed to be transmitted, $R_{prot}(t, R_{ss})$ is the bit rate needed for protecting R_{ss} at level t, and $P_{fail}(i, j)$ is the probability of the i^{th} packet failed in the past j times retransmission, which are calculated as in [6].

The channel distortion of BL after protection can be described as

$$D(R_{ARQ}) = \sum_{i=0}^{bn}[P_{fail}(i, N_{\max}-1) \times D_c(i)], \tag{10}$$

where $D_c(i)$ is the channel distortion caused by the loss of packet i.

The processing power consumption for the BL protection is represented as

$$\begin{aligned} PC_{ARQ}(R_{ARQ}) = & \sum_{i=1}^{bn} \rho c(R_{s\_base}(i), t_1) \\ & + \sum_{j=2}^{N_{\max}-1} \{\sum_{i=1}^{bn}[P_{fail}(i, j-1) \times \rho c(R_{s\_base}(i), t_j)]\}, \end{aligned} \tag{11}$$

where $\rho c(.)$ is the rate-power relation for RS coding, which can be obtained from [6].

Now we analyze the channel distortion in ELs. The protection rate needed for the ELs delivery is then represented as follows:

$$R_{FEC} = \sum_{i=1}^{L} R_{prot}(t_i, R_{s\_enh}(i)). \tag{12}$$

The channel distortion of ELs after UEP can be denoted as

$$D(R_{FEC}) = \sum_{i=1}^{L}[\sum_{j=1}^{bn_i}(P_{fail,layer}(i, j) \times D_c(j)], \tag{13}$$

where b_{ni} is the number of source packets needed to be transmitted in the i^{th} layer, $P_{fail,layer}(i, j)$ is the probability of which the j^{th} packet in the i^{th} layer is corrupted while the corresponding packets in the previous layers are correct, which can be calculated as in [6]. Similarly, the processing power consumption for ELs protections is represented as

$$PC_{FEC}(R_{FEC}) = \sum_{j=1}^{L}\{\sum_{i=1}^{bn_j}[\rho c(R_{s\_enh}(i, j), t_j)]\}. \tag{14}$$

V. Network-adaptive Transmission-Level Power Control

It is known that power control scheme can be used to achieve target quality of service in multimedia environment. In this work, we proposed an outer-loop power control and an inner-loop power control that operate in parallel. The outer-loop power control operates within the receiver, and is responsible for setting a target for the received signal-to-interference ratio (SIR). This target is set according to the video quality requirements and channel protection level used in the sender. Consequently, the receiver compares the received SIR from the sender with the target in every time-slot and sends the TPC (transmission power control) command correspondingly. Meanwhile, the inner-loop power control mechanism controls the transmitted power of the sender in order to counteract the fading of the radio channel based on the TPC command by the out-loop. To respond to TPC command, sender may adjust the power control step size to minimize the total power consumed in source coding, channel coding, and transmission.

In cellular radio system, reducing transmission power of a user will increase its battery lifetime, lower bit error rate for neighbors, and heighten bit error rate for its own transmission.

In Rayleigh fading environment with binary phase shift key (BPSK) modulation, the bit error rate is evaluated as

$$b_{cur} = 0.5 \times [1 - \sqrt{\Gamma/(1+\Gamma)}], \tag{15}$$

where $\Gamma = \varepsilon_b / N_0 \times \overline{\alpha^2}$ is the average value of the signal-to-noise ratio, and α has a Rayleigh distribution.

After adjusting the transmitted bit energy, i.e., ε_b, the channel status will be modified accordingly. Specifically, the BER of this channel will be changed, thereby affecting the channel distortion and corresponding error protection codes.

It is known that the transmission power is the output power delivered by the power amplifier. Energy consumption of the power amplifier is characterized by its power-aided-efficiency η. For the digitally programmable power amplifier [7] used in this work, the efficiency $\eta(P_t)$ can be approximated as follows:

$$\eta(P_t) = \eta_{\max}\sqrt{P_t / P_{t,\max}}, \tag{16}$$

where η_{max} is the maximum efficiency that is set to 50% in our work, and $P_{t,max}$ is the maximum transmit power. Thus the power consumed for transmitting a bitstream with bitrate R_T can be represented as

$$\varepsilon_t(P_t) = (\varepsilon_b \times R_T) / \eta(\varepsilon_b \times R_T). \tag{17}$$

VI. Distortion/Power-minimized Rate Allocation for Scalable Video Streaming

To efficiently utilize the system resources, such as limited available bandwidth and limited battery, we jointly consider the application-level error control as well as transmission-level power control, and allocate the available resource between source and channel so as to minimize the end-to-end distortion or total power consumption based on different requirements.

If one wants to obtain the minimum distortion without control the transmission power, we will allocate the bits among the source, the FEC, and the ARQ for a given fixed bandwidth capacity so as to achieve the minimal expected end-to-end distortion.

Suppose $R(t)$ is the available bit rate at time t, $R_S(t)$, $R_{ARQ}(t)$, and $R_{FEC}(t)$ are the bit rates used for the source, the FEC, the ARQ at time t, respectively. Then the distortion-minimized rate allocation can be formulated as

$$\begin{aligned} & \min_{\{R_s, R_{ARQ}, R_{FEC}\}} D_{end-to-end} = D_s(R_s) + D(R_{ARQ}) + D(R_{FEC}) \\ & \text{s.t.} \quad R_S(t) + R_{ARQ}(t) + R_{FEC}(t) \le R(t). \end{aligned} \tag{18}$$

If one wants to obtain the minimum total power consumption while satisfying the QoS requirements, such as end-to-end distortion and total bit rate, we can determine the source rate, channel rate, and transmission bit energy to achieve the minimal total power consumption in the mobile host. Suppose D_0 is the desired video distortion, the power-minimized rate allocation problem can be formulated as

$$\begin{aligned} \min_{\{R_s, R_{ARQ}, R_{FEC}\}} PC = & PC_s(R_s) + PC_{ARQ}(R_{ARQ}) + PC_{FEC}(R_{FEC}) \\ & + \varepsilon_t(P_t(\varepsilon_b, R_s + R_{ARQ} + R_{FEC})) \end{aligned}$$

$$s.t. \; D_{end-to-end} \le D_0 \text{ and } R_s(t) + R_{ARQ}(t) + R_{FEC}(t) \le R(t). \tag{19}$$

Note that optimization approaches, such as Lagrange multiplier and penalty function methods, can be used to solve the constrained non-linear optimization problem.

VII. Simulations

These simulations are to demonstrate effectiveness of our proposed network-adaptive scalable video streaming scheme. The testing video sequence *Foreman* is coded in CIF at a temporal resolution of 15 fps. The radio transmission parameters and propagation models for outdoor to indoor and pedestrian environment are adopted and can be obtained from the IMT-2000 evaluation methodology [8]. The large-scale path loss and small-scale fading are considered simultaneously in the proposed system. The other simulation parameters are tabulated in Table 1.

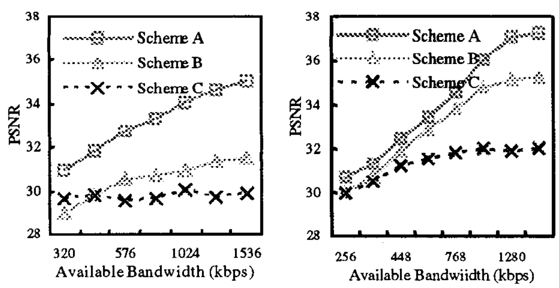
Table 1. Simulation parameters

Chip rate	11.0593 Mchips/sec
Power amplifier	$P_{t,\max} = 20$ mW, $\eta_{\max} = 50\%$, Noise figure = 5dB
Transmit power adjust step size	0.25dB – 1.5dB
Distance	20m – 200m
Moving speed	3km/h
Source and channel codec	0.18μ, 2.5V CMOS Technology
UDP packet size	576 Bytes
RLC frame size	320 bits, 640 bits
CRC bits	16
Maximal RLC retransmission times	3
Channel coding	RS (245, 200), (225, 180) Convolutional coding: 1/3, 2/3

7.1 Performance of distortion-minimized video streaming

In this simulation, we tested (1) Scheme A: our network-adaptive distortion-minimized rate allocation scheme for hybrid UEP and delay-constrained ARQ; (2) Scheme B: PFGS with fixed UEP (25% protection for base layer and 10% protection for enhancement layers); (3) Scheme C: PFGS with fixed channel protection in base layer (25% protection).

We conducted simulations under the channel bandwidth varying from 256kbps to 1.5Mbps. The simulations are performed in both high error and low error cases, which has the BLER from 8e-4 to 5e-3 and from 5e-3 to 2e-2, respectively. Note that in all these simulations the total rates including source and channel are the same for all the cases.



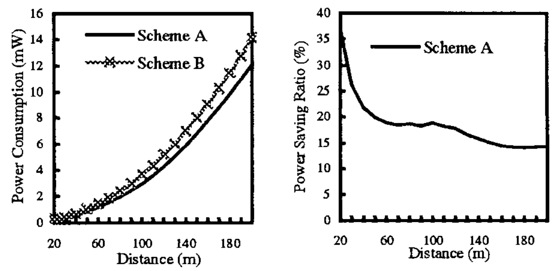
(a) High error channel. (b) Low error channel.
Figure 2. Average PSNR comparisons using three tested schemes under different bit rates.

Figure 2 shows the average PSNR comparison results at different available bandwidth under high-error and low-error channels. It can be seen that our proposed scheme achieves the best performance under different channel conditions and various channel rates. Notice that the higher channel rate, the larger difference between our scheme and the other two fixed UEP schemes. It can also be seen from Figure 2 that PSNR obtained using the other two schemes increases slower than ours. This is because the target bits are allocated according to the quality impact of each layer in our approach.

7.2 Performance of power-minimized video streaming

In this simulation, we tested (1) Scheme A: our network-adaptive power-minimized rate allocation scheme with adaptive source complexity and transmission power selection; (2) Scheme B: a power-controlled system with fixed source complexity and fixed ε_b / N_0 requirement.

Figure 3(a) illustrates the comparison results for the total power consumption of the above schemes when the desired PSNR is 30dB. It can be seen that our proposed scheme saves much power compared to the fixed power-controlled system. Figure 3(b) shows the corresponding power saving ratio using our scheme. As Figure 3(b) illustrates, the power saving ratio will decrease smoothly with the increasing distance. This is because when the distance is increased, the transmission power consumption must be enlarged to deal with the path loss and fading. Therefore, the optimization of source and channel processing power by adaptive rate allocation will introduce smaller power saving gain.



(a) Power consumption comparisons. (b) Power saving ratio.
Figure 3. Performance comparisons using Scheme A and B.

VIII. Conclusions

In this paper, we propose a network-adaptive architecture for scalable video streaming over 3G wireless network. Dynamically wireless network estimation, application-level error control, and transmission-level power control are integrated in this architecture. Distortion/power-minimized rate allocation is presented to achieve the minimum end-to-end distortion or minimum total power consumption. Significant video quality incremental ratio and power saving ratio were obtained using our scheme.

References

[1] S. P. Li, F. Wu, and Y.-Q. Zhang, "Study of a new approach to improve FGS video coding efficiency," ISO/IEC JTC1/SC29/WG11, MPEG99/m5583, December 1999, Maui.

[2] G. J. Wang, Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Channel-adaptive error control for scalable video over wireless channel", IEEE MoMuc 2000, Oct. 2000.

[3] M. P. J. Baker and T. J. Moulsley, "Power control in UMTS release'99", IEE 3G Mobile Communication Technologies, 2000.

[4] Q. Zhang and S. A. Kassam, "Hybrid ARQ with selective combining for fading channels", *IEEE Journal on Selected Areas in Communications*, pp. 867-880, vol. 17, no. 5, May 1997.

[5] H. W. Wang and C. H. Wei, "Pilot-symbol assisted channel estimation for coherent DS-CDMA communication", IEEE ISCAS'2000, May, 2000.

[6] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "QoS-Adaptive Resource Allocation for Video Transmission over 3G Wireless Network", submitted to IEEE Trans. on CSVT, 2001.

[7] M. Goel, S. Appadwedula, N.R. Shanbhag, K. Ramchandran, and D.L. Jones, "A low-power multimedia communication system for indoor wireless applications", *IEEE Workshop on Signal Processing Systems*, 1999.

[8] UMTS 30.03, Selection procedures for the choice of radio transmission technologies of the UMTS, 1998.