Cross-Layer Quality Assessment of Scalable Video Services on Mobile Embedded Systems

Kyungtae Kang, *Member*, *IEEE*, Won J. Jeon, *Member*, *IEEE*, Kyung-Joon Park, *Member*, *IEEE*, Roy H. Campbell, *Fellow*, *IEEE*, and Klara Nahrstedt, *Fellow*, *IEEE*

Abstract—The recent development of high-speed data transmission over wireless cellular networks has enabled the delivery of multimedia broadcasting services to mobile users. These services involve a range of interactions among different system components, including the wireless channel, the network, and mobile devices, making it crucial for the service provider to verify the model, design, and behavior of a new service before it is deployed. However, previous studies have largely relied on network simulations or scaled experiments, and there has been little work on the sort of unified framework for quality-of-service (QoS) assessment, which considers the interactions between components, that we propose in this paper. Accurate models of the wireless channel, the network, and the data processing that takes place on an embedded system of a mobile client, are integrated within our framework, and allow us to predict several key system metrics and the quality of the video stream as it is perceived by users. Furthermore, different models of system components can be easily plugged in to extend this framework. As an example application, we analyze the performance of the process of decoding scalable videos on ARM-based mobile embedded systems in CDMA2000 wireless cellular networks.

Index Terms—Cross-layer quality assessment, scalable video, mobile device.

1 Introduction

The recent development of high-speed data transmission over wireless cellular networks has enabled a range of video broadcasting services, but these require a higher data rate than earlier services such as web browsing and email. As the network data rate increases, the timely processing of video streams by mobile client devices becomes more important than preserving the full integrity of the data streams.

Multimedia broadcasting services over wireless cellular networks involve many interactions among different system components, including the wireless channel, the network, and client devices. Therefore, it is crucial for the service provider to verify the system model, the design, and its behavior before a new type of service is deployed. However, it is not easy to predict how all the system components will interact. Different wireless operators and device manufacturers have their own specifications for production and maintenance, and it is common to observe unexpected malfunctions and loss of performance. Due to the complexity and size of these systems, the prototyping or temporary deployment of a real system is not usually considered to be feasible.

- *K. Kang, R.H. Campbell, and K. Nahrstedt are with the Thomas M. Siebel Center for Computer Science, Department of Computer Science, University of Illinois at Urbana-Champaign, 201 North Goodwin Avenue, Urbana, IL 61801-2302. E-mail: {ktkang, rhc, klara}@illinois.edu.*
- *W.J. Jeon is with the Computer Science Laboratory, Samsung Information Systems America (SISA), 75 West Plumeria Drive, San Jose, CA 95134. E-mail: won.jeon@samsung.com.*
- *K.-J. Park is with the School of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea. E-mail: kjp@snu.ac.kr.*

Manuscript received 24 Sept. 2009; accepted 17 Mar. 2010; published online 25 June 2010.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2009-09-0384. Digital Object Identifier no. 10.1109/TMC.2010.119.

The difficulties of ensuring that system components are fully compatible before deployment has been addressed in various contexts. The use of network simulation tools such as OPNET [1] and ns-2 [2], simulated network models [3], [4], or mimicking the network system with scaled emulations [5], can provide an approximate assessment of a particular network. However, these approaches are not the same as an analysis of the entire system, including data processing and the quality delivered by applications running on mobile embedded systems.

In this paper, we propose a holistic cross-layer quality-of-service (QoS) assessment framework for video broadcasting services over wireless cellular networks. Taking this approach is of fundamental importance since it should not only improve the accuracy of performance evaluation, but may also provide valuable insights leading to the design of new protocols. Our framework involves a precise model of a real system architecture and can be used to verify models. By simulating the physical wireless channel, the network, and the data processing that takes place on a an embedded system of a mobile client, the system throughput and so-called *goodput* (which is a user-oriented metric especially for video streams) can be predicted.

Our main contributions are as follows: First, we provide a flexible QoS assessment framework for video broadcast services in wireless networks, which can easily be extended to other scenarios with different network and device models. This framework covers the whole system stack, including the wireless physical channel and network, the processing power of the device, and the quality assessment of applications. Second, we analyze the system and provide a guideline for the performance required from a client device in terms of error correction (i.e., the error-recovery rate, the delay jitter, and the perceived quality of video streams). This analysis relates specifically to CDMA2000 networks, and is based on a range of channel parameters,

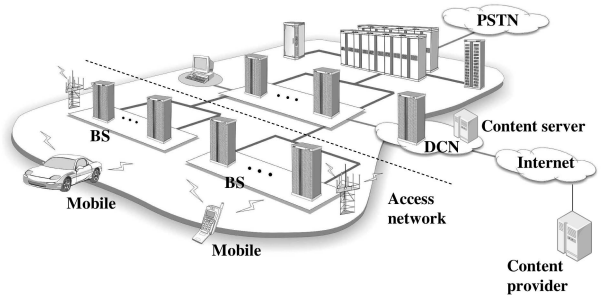


Fig. 1. Network architecture for a video broadcasting service.

such as the signal-to-noise ratio (SNR) and the mobile speed, as well as on the service parameters, such as the type of error-correction scheme.

The rest of our paper is organized as follows: In Section 2, we abstract the requisite network architecture and protocols, and introduce our framework for the assessment of quality in video broadcasting services. In Section 3, we describe the different network and system modules in the framework in detail. In Section 4, we analyze key QoS parameters and their simulation. In Section 5, we show how this framework can be used to analyze scalable video decoding on mobile devices in CDMA2000 wireless networks. We discuss other aspects of our framework in Section 6, and finally, conclude this paper in Section 7.

2 Overview

We will review the network architecture and protocols used to provide a video broadcasting service over wireless cellular networks. We will also introduce the structure of our simulation and analysis framework.

2.1 Video Transmission over Mobile Broadcast Networks

Fig. 1 shows a generic cellular network architecture to support video broadcast services, which is composed of a radio access network (RAN) and a data core network (DCN). The broadcast content provider, who can be located in the serving or home network, or anywhere in an IP network such as the Internet, makes broadcast content available within an IP stream. The content server may store and forward content from a provider, or merge content from multiple content providers. Finally, the base station (BS) transmits a video stream to a number of mobile stations over the wireless channel. We will focus on this forward link and assume the forward traffic channel is time-multiplexed between the different access mobiles.

A wireless mobile device is equipped with a radio modem and a data interface, allowing it to access time-division multiplexed channels. Fig. 2 shows the generic protocol layers and the interactions between them that are required for reliable data communication. The broadcast security protocol encrypts contents to cope with the threat of a user obtaining unauthorized access to a particular content stream. The broadcast data-link protocol encapsulates the physical layer to create a link responsible for node-to-node communications, and provides error control and medium access control (MAC) for reliable communications. In general,

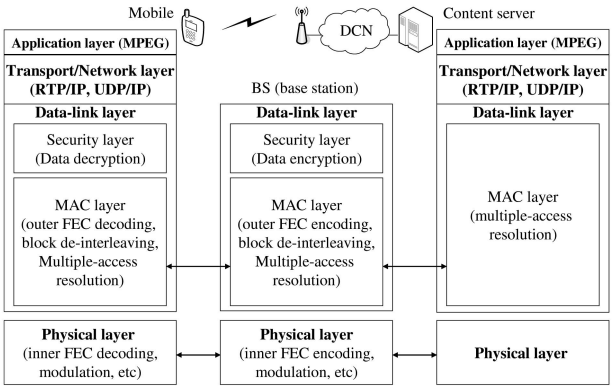


Fig. 2. Generic network protocol layers used in a video broadcasting service.

automatic repeat request (ARQ) cannot be used to correct damaged packets in a broadcasting service, because 1) there is no reverse channel for ARQ, 2) a storm of negative acknowledgments (NAKs) will occur when a channel condition is bad, and 3) it does not meet the real-time constraints of video applications. Reliability is increased by using the outer forward error-correction (FEC) code and the block interleaver to rearrange code symbols so as to spread bursts of errors over multiple codewords to make error-correction codes more effective. The broadcast physical layer provides the structure of the broadcast channel.

2.2 Simulation and Analysis Framework

As shown in Fig. 3, each module in the simulation framework has a corresponding model which allows it conceptually to mimic the operations of the appropriate components of a wireless broadcasting system. The lower level system modules feed information to the upper level modules, which allows the individual analysis of each module, as well as an analysis of the whole system. The following modules are incorporated in the simulation framework:

- Wireless channel module: The wireless channel model proposed by Zorzi et al. [6], [7], [8] uses Markov approximations for the block error process that occurs in fading channels, with a range of modulation schemes. This represents the behavior of block errors at different mobile speeds more precisely than other additive white Gaussian noise (AWGN) models [9]. It also provides a good approximation to the relationship between the probability of successful transmission, the steady-state packet error-rate (PER), and the mobile speed.

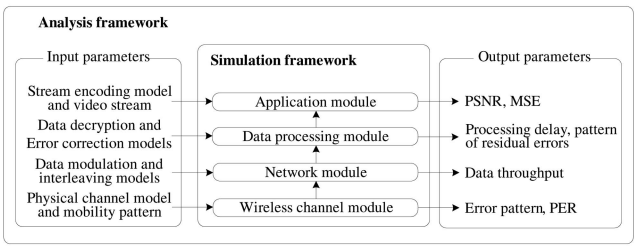


Fig. 3. The simulation and analysis framework.

The PER is strongly affected by the SNR of the channel and the implementation of inner FEC coding in the physical layer.

- Network module: The wireless network model is based on time-division multiplexing (TDM) technology, which enables the deterministic delivery of video data to multiple mobile devices, by careful packet scheduling at the access network.
- Data processing module: Data processing at the mobile device mainly takes place in the data-link layer, and involves correction of the errors that occur during the transmission of data across a fading channel, and content protection to restrict the reception of broadcast services to authorized users. The error-correction scheme that we model is the Reed-Solomon (RS) FEC scheme, and the encoded data are aligned by block interleaving [10]. Data encryption and decryption are performed by an Advanced Encryption Standard (AES, [11]) block cipher. (If link-layer encryption is used, the content-encryption function is typically located in the RAN. If higher layer encryption is used, content encryption takes place in the content server.) This module simulates RS decoding for different levels of block interleaving and coding parameters, and AES block decryption.
- Application module: A model of scalable video decoding is used to determine the quality of video streams perceived by users, for a given system throughput and pattern of packet loss in the underlying network layer.

While the whole system is being monitored, probes can also be inserted into individual network and service components. This analysis framework allows detailed monitoring throughout the simulation period. The system also checks the compatibility of parameters across simulation modules, and translates them if necessary.

3 Proposed System Model for Video Broadcasting

We will now provide more details about the system models introduced in the previous section, representing the wireless physical channel, the network, data processing operations, and the decoding of scalable video streams on mobile devices.

3.1 Modeling Wireless Channels

Zorzi et al. investigated the behavior of the block errors which arise in data transmission over fading channels, and showed that a Markov chain is a good approximation to a range of modulation schemes, block lengths, and error-correction capabilities.

They also showed that the transition probabilities in Markov models are largely insensitive to parameters such as the rate of inner FEC coding and the modulation format, and depend only on the PER and the Doppler frequency normalized to the block transmission rate, V, which can be derived as follows:

$$V(f_c, v, L_{\text{phy}}, \mu_p) = f_D L_{\text{phy}}/\mu_p, \tag{1}$$

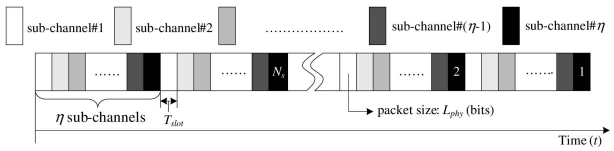


Fig. 4. Abstraction of the TDM structure.

where L_{phy} is a size of a physical layer packet, μ_p is the reference data rate of the physical channel, and f_D is the Doppler frequency of the system, which equals $f_c v/c$, where v is the mobile speed, c is the speed of the electromagnetic wave, and f_c is the carrier frequency [12]. This relationship is, in fact, almost the same as that in the simple threshold model, for which closed-form analytical expressions are available [6], [8] for the Rayleigh fading case. This observation allows each transition probability to be obtained as a function of PER and V.

3.2 Wireless Network Model

Fig. 4 is a high-level abstraction of a forward link channel based on the TDM structure with η subchannels, which we have generalized from competing wireless technologies. This multiple-access structure enables the timely delivery of video data by assigning a dedicated subchannel to each mobile. If the length of each time slot is T_{slot} and that the transmission of one encoded physical layer packet takes exactly one slot, the size L_{phy} of each physical layer packet of a particular subchannel k can be determined as follows:

$$L_{\text{phy}} = \mu_{p_k} T_{\text{slot}}, \tag{2}$$

where μ_{p_k} is the data rate of that subchannel. When there are η subchannels, the total data rate of the forward data channel μ_p is given as follows:

$$\mu_p = \sum_{k=1}^{\eta} \mu_{p_k}. \tag{3}$$

This data rate is determined by the modulation, the coding rate of the inner FEC code in the physical layer, and the spreading used in each subchannel. For example, a CDMA2000 [13], [14] 1xEV-DO system supports a range of channel data rates between 38.4 and 2,457.6 Kb/s. The rate depends on: the modulation (which may be quadrature phase-shift keying (QPSK), 8-phase-shift keying (PSK), or 16-quadrature amplitude modulation (QAM)), the code-rate (1/5 or 1/3), and the number of slots scheduled for each physical layer packet.

3.3 The Model of Data Processing on the Device

3.3.1 Physical Layer

Channel coding and modulation in the physical layer are performed on the information-bearing digital signal, with the aim of optimizing the performance of the communications system. While modulation is concerned with packing bits in a chip which is the fundamental unit of transmission in a CDMA system, channel coding is the packing of data in a slot. Channel coding can cope with bits that are flipped due to noise and interference by employing *smart redundancy*. The most significant coding techniques for implementing redundancy are error-control coding, using inner

FEC codes such as convolutional and turbo codes, and interleaving. Typically, turbo coding is more powerful than convolutional coding when the physical layer packets are large (i.e., several hundred bits or more). In this case, turbo coding significantly improves performance by allowing the use of a lower radio frequency (RF) power while still achieving the same error rate. This is so effective that the communication channel is able to perform close to the Shannon limit [13], [15], [16].

Additionally, concatenated error-correcting codes, in which an inner code is combined with an outer FEC code, improve the performance of error correction. To optimize the performance of this form of coding, we need to allow for the characteristics of the channel through which transmission takes place, so as to select the most suitable channel coding and modulation techniques for a given application.

3.3.2 MAC Layer

The MAC protocol provides error control, as well as a method of medium access, by adding an outer FEC code that forms a concatenated code in conjunction with the physical layer inner FEC code. This outer code provides additional data protection, and RS codes are a good choice for the outer code because of their superior performance at lower error rates [17], [18].

An RS code is specified as a tuple (N, K, R) with $\hat{s}$-bit symbols. This means that the encoder takes K data symbols of $\hat{s}$ bits each and adds R parity symbols of $\hat{s}$ bits each to make an N-symbol codeword. Reed-Solomon is therefore known as a systematic code because the appended parity symbols leave the data unchanged. The algebraic RS decoder can correct up to $t = R$ *erasures* when the positions of the erroneous symbols are known or up to $t = R/2$ *errors* when these positions are unknown. Erasure coding outperforms error-correction coding because more information is available. Large values of R mean that more erasures or errors can be corrected, but more processing power will be required because the amount of computation needed to encode and decode RS codes increases with the number of parity symbols in each codeword.

The RS decoding algorithm, which we have analyzed in detail elsewhere [19], consists of three computational components, C_1, C_2, and C_3. C_1 operates on every codeword received, C_2 is called once for each codeword which contains any errors, and C_3 is invoked to deal with each error in that codeword. The total execution time T_{cw} required for the decoding of an RS codeword, encoded by an (N, K, R) code with $\hat{s}$-bit symbols and containing ν erroneous symbols, can be estimated as follows:

$$T_{cw}^{(\hat{s},N,K)}(\nu) = \begin{cases} T_{C_1}^{(\hat{s},N,K)} + T_{C_2}^{(\hat{s},N,K)} + \nu T_{C_3}^{(\hat{s},N,K)} & (\text{if } 0 < \nu \le t) \\ T_{C_1}^{(\hat{s},N,K)} & (\text{otherwise}), \end{cases} \quad (4)$$

where $T_{C_i}^{(\hat{s},N,K)}$ is the execution time of component C_i.

This RS coding scheme is usually combined with block interleaving. In a wireless mobile channel environment, error bursts occur frequently, and interleaving allows an error-correcting code to function more effectively, although interleaving itself has no intrinsic error-correcting capability. The basic idea is that data block is entered into a 2D array from the top to the bottom, with each row written from left to right. The block is then read out in columns from left to right and from top to bottom, as shown in Fig. 5.

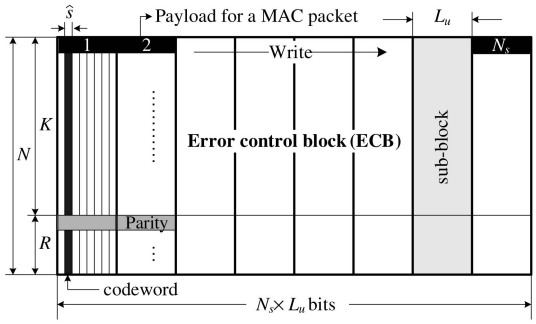


Fig. 5. The block interleaving scheme used in an ECB.

In order for block interleaving to work properly, a special region of array memory (transmit buffer), which is called an error-control block (ECB), is required to buffer data at both the sender and the receiver. Fig. 5 shows the structure of the ECB, which is a table made up of N rows and N_s columns of entries, each entry being L_u bits in length, where N_s is the number of MAC layer packets in each ECB row, and L_u is the size of the payload in each MAC layer packet. A sequence of RS codewords, each $\hat{s}$ bits wide and encoded with the outer RS code (N, K, R), spans the columns of N_s subblocks, each of which is L_u bits wide. An ECB subblock therefore contains $N_{cw} = (L_u/\hat{s})$ codewords. The value of N_s determines the level of block interleaving. As it increases, the error bursts which inevitably occur during data transmission are interleaved more effectively, so as to favor RS decoding and reliable communication.

Each of the N rows is then individually encoded by the outer code and sent over the air to multiple mobiles. The access network concatenates trailer bits to the payload of size L_u, to form a complete MAC layer packet of size L_{mac}. The resulting packet is then forwarded to the physical layer protocol for transmission. A single physical layer packet of size L_{phy} contains $z = \lfloor L_{\text{phy}}/L_{mac} \rfloor$ MAC layer packets ($\lfloor\rfloor$ is a floor function).

3.3.3 Security Layer

Encryption of the broadcast content and distribution of the decryption key to authorized subscribers to the channel are used to counter the threat of pirate users obtaining free access to the content. We focus on the use of AES encryption and decryption procedures, operating in the security layer within the link layer. The AES algorithm is a symmetric block cipher that can both encrypt and decrypt information. We are only concerned with deciphering, because it is the operation that needs to be performed in a mobile that is receiving a broadcast service.

The AES algorithm operates on a 2D array of bytes called the *state*. This consists of four rows, each containing N_b bytes, where N_b is the block length divided by 32. The length of the cipher key can be 128, 192, or 256 bits. The key length N_k is 4, 6, or 8, which corresponds to the number of 32-bit words (the number of columns) in the cipher key. The number of rounds N_r that have to be performed

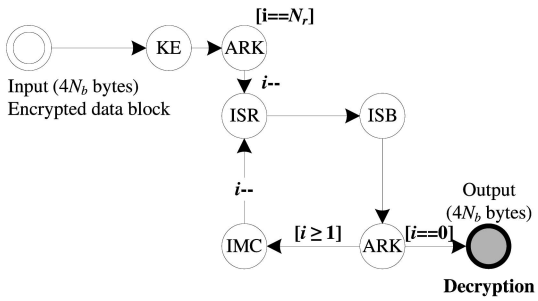


Fig. 6. Flow diagram for the AES decryption algorithm.

during the execution of the AES algorithm depends on the key size, so that $N_r = N_k + 6$.

At the start of the decipher, the input is copied to the state array. After an initial round-key (RK) addition, this array is transformed by performing the round function 10, 12, or 14 times, depending on the key length, with the final round differing slightly from the first $N_r - 1$ rounds, as shown in Fig. 6 (i.e., the final round does not include the IMC transformation [11]). A round function is composed of four different byte-oriented transformations, as follows:

- Inverse Shift-rows (ISR): The inverse of the Shift-rows (SR) transformation, where the bytes in the last three rows of the state are cyclically shifted over different numbers of bytes (offsets).
- Inverse Subbytes (ISB): The inverse of the byte substitution transformation, in which the inverse S-box is applied to each byte of the state.
- Inverse Mix-columns (IMC): The inverse of the Mix-columns (MC), where data is exchanged between columns of the state array.
- Add-round-key (ARK): a round key is added to the state by a bitwise XOR operation.

The AES decryption algorithm also includes a key expansion (KE) routine which generates a key schedule for each new value of the cipher key. Thus, the execution time $T_{IC}^{N_b,N_k}$ required to encrypt a data block containing $4 \times N_b$ bytes with a cipher key of $4 \times N_k$ bytes can be determined as follows:

$$T_{IC}^{N_b,N_k} = T_{KE}^{N_k} + N_r\left(T_{ISR}^{N_b} + T_{ISB}^{N_b} + T_{ARK}^{N_b}\right) + (N_r - 1)T_{IMC}^{N_b} + T_{ARK}^{N_b}, \quad (5)$$

where T_A^B is the execution time of transformation A with the parameters B.

3.4 Application Model for Scalable Videos

Typically, a video stream is encoded for efficiency in such a way that successive frames share data and are therefore mutually dependent. This means that the quality of a video perceived by a user is degraded more than one might expect by the loss of a single frame during transmission over a wireless network, because dependent frames will also be damaged. Several different methods of error concealment have been suggested to deal with this so-called *cascading effect*. The dependencies among video frames and error concealment methods in use must be considered in evaluating the perceived quality of video streams for given wireless channel and network conditions.

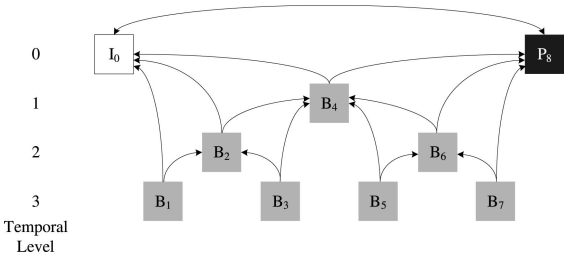


Fig. 7. Hierarchical structure of H.264/SVC frames in a video stream.

H.264/SVC [20] was suggested and developed as an extension of H.264/AVC, and is currently recognized to be the most efficient video encoding scheme. SVC streams have a hierarchical B-frame structure which enables higher coding efficiency than nonhierarchical coding schemes such as AVC. H.264/SVC allows temporal scalability, because B-frames at lower levels can be discarded to generate lower resolution video streams. Fig. 7 shows a typical configuration of group of pictures (GoPs) in the base layer of H.264/SVC, with a GoP size (G) of 8. The quality of decoded frames at the receiver is naturally affected by this hierarchical structure and also by the picture copy concealment technique, which is used to hide errors.

Mansour et al. [21] analyzed the hierarchical structure and error concealment scheme in H.264/SVC video streams and suggested the following model[1] to relate quality degradation (D) to frame losses:

$$D(p_{net}) = p_{net} \times \frac{E[D_{I/P}] + (G-1)E[D_B]}{G}, \quad (6)$$

where

$$E[D_{I/P}] = [(1-p)\,\triangle_G + 2p\triangle_G] \times G + \frac{1}{\gamma}\sum_{j=0}^{M-1}\sum_{i=1}^{M-j-1}\triangle_G\left(1 - \frac{i}{M-j}\right) \times G \quad (7)$$

and

$$E[D_B] = \sum_{k=1}^{\log_2(G)} \frac{2^{k-1}}{G-1} D_{B_k}. \quad (8)$$

Here, p_{net} is the overall probability of frame loss, p is the probability that the previous key frame (i.e., the I or P frame) is lost, $E[D_{I/P}]$ and $E[D_B]$, respectively, are the expected quality degradation, expressed as a mean-square error (MSE), when key frames and B-frames are lost. M is the intraframe distance between key frames (i.e., between I and I, or I and P frames), G is the size of a GoP, $\triangle_G$ is the quality degradation due to concealment mismatch between two frames that are G pictures apart, γ is the discount in the signal error (which is $\frac{1}{M}$ for I/P frames and $\frac{1}{M-j}$ for B frames in (7)), and D_{B_k} is the distortion due to the loss of frame B_k, where k is the temporal level of that frame.

1. Note that we consider packet losses to occur in the base layer only. In reality, there could be a drift effect due to losses in the enhancement layers. This could be modeled in a similar manner to the base layer losses, by scaled quality compensation.

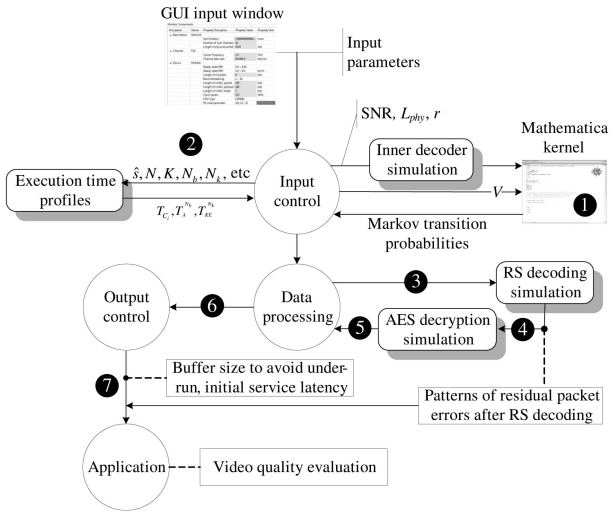


Fig. 8. Structure of the simulation analysis.

4 Cross-Layer Simulation and Performance Analysis of Scalable Video Transmission

We will now describe how our framework is implemented and used to analyze the QoS parameters of wireless video transmission.

4.1 Simulation Framework

Our analysis tool is written in Java, and its structure is shown in Fig. 8. The tool runs with parameters that describe the physical channel, the network, and the video broadcasting service. Once this input has been received, the input control module passes the SNR measure of the forward data channel, the code-rate r, and the physical layer packet size to a simulator of the inner decoding process, which derives the corresponding PER for the channel. (But we will not consider the timing aspect of inner FEC coding because it is implemented typically in hardware and introduces little overhead.) This PER, together with the correlation properties of the fading process, representing the dependency between the transmission of consecutive blocks, is passed to the Mathematica [22] kernel through J/Link [23], which is a toolkit that connects Mathematica and Java. Mathematica code is used to calculate the Markov state transition probabilities of consecutive packets by applying model of the wireless fading channel, which is described in Section 3.1. These Markov state transition probabilities can then be used to approximate the behavior of the packet errors in the physical layer which arise in data transmission over fading channels. Thus, the input control module describes the functionalities of the wireless channel module and the network module, which were introduced in Section 2.2.

The analysis tool then obtains profile data for the execution time $T_{C_i}^{(\hat{s},N,K)}$ required by the three computational components that are involved in decoding a codeword. This execution time depends on the symbol length $\hat{s}$ of the RS codeword and the RS code in use. Then, it obtains profile data for the execution time $T_A^{N_b}$ required by the transformation A performed by the AES decryption algorithm, and the time $T_{KE}^{N_k}$ required by the key expansion routine of the example software implementation of AES block decryption. $T_A^{N_b}$ depends on the length of the AES cipher block, while $T_{KE}^{N_k}$ depends on the length of the cipher key. This profile is also dependent on the computational power of the hardware platform that is executing the RS decoder and the AES decryption code. We created several profiles by measuring the time required by each computational component of the RS decoder for different RS codes, and the time required by all the transformations involved in AES decryption for different values of N_b and N_k, in execution environments based on different processors.

Next, the data processing module uses Monte Carlo simulation to model the consecutive stages of MAC layer RS decoding and security layer AES decryption, which interact with each other across the layers. This simulation uses the timing profile data described above and a stochastically generated pattern of errors in the physical layer packets which matches the expected occurrence of errors during transmission over a fading channel, which is obtained using the state transition probabilities. A Monte Carlo approach using repeated random sampling can obtain reliable value for the decoding time of each ECB, as well as the pattern and number of the residual errors in each ECB after RS decoding, for a given channel condition, provided that the number of iterations is sufficiently large. The time required for AES decryption can then also be determined.

The results of the simulation decided above are forwarded to the output control module, which estimates the size of buffer needed to avoid service dropout due to buffer underrun, which would cause significant interruption in video streaming. The buffer must be large enough to contain the amount of data needed to accommodate the worst case execution time (WCET) for the consecutive operations of RS decoding and AES decryption in processing a single ECB. The resulting initial service latency can now also be determined, and finally, the application module is invoked to determine how the perceived quality of the final video is affected by the end-to-end delivery of data from the service provider to the application running on the mobile device.

Physical layer packets are processed by a mobile device, and the resulting payload is fed into the application buffer for display. If some packets do not arrive at the application buffer in time, so that they do not meet the timing constraint imposed by the display rate, they are regarded as missed. An error concealment algorithm within the application minimizes the effect of such misses on the video as it appears on the screen. The quality perceived by a user must be determined in the context of an appropriate distortion model (6) for scalable videos.

4.2 Estimation of the Buffer Size Required to Avoid Service Interruption

Estimating the WCET required by the module that performs RS decoding and AES decryption at the mobile is critical in determining an adequate size of buffer to absorb jitter and provide a seamless video service. Recall that our timing model (5) predicts that the execution time required for RS decoding will increase with the number of erasures ν in an RS codeword, up to R, because C_3 must run more frequently. But the time required for AES decryption is not greatly affected by the number of erasures, provided that the RS decoder is able to correct them. This means that the WCET for processing an ECB, also defined as T_{WC},

increases as the channel condition deteriorates. We have also shown [24] that the WCET is relatively long in a slow fading environment.

Let us assume that the value of SNR within the service area covered by a particular BS, with a code-rate of r, is larger than or equal to $\xi_{\min}$, and that the mobile speed v is in the range $v_{\min}$ to $v_{\max}$. In this case, the WCET occurs when the channel condition, which can be expressed as a pair made up of SNR and V, is Φ_{WC}, which corresponds to the tuple $(\xi_{\min}, V(f_c, v_{\min}, L_{\text{phy}}, \mu_p))$. We can determine the WCET for a particular level of block interleaving that satisfies a Υ percent confidence limit by Monte Carlo simulation.

Providing a seamless service is critically important for video applications. We therefore need to absorb the jitter caused by the variation in data processing required at the mobile by buffering the data for the period equal to the WCET. The total size of buffer required at the data-link layer L_{buf}, and the service latency T_L, when the level of block interleaving is N_s, can then be determined as follows:

$$\begin{aligned} L^N_{buf}(\eta, N_s|\Phi_{WC}, \Upsilon) &= \overline{\mu_u} T^N_L(\eta, N_s|\Phi_{WC}, \Upsilon), \\ T^N_L(\eta, N_s|\Phi_{WC}, \Upsilon) &= T^N_p(\eta, N_s) + T_{WC}(N_s|\Phi_{WC}, \Upsilon), \end{aligned} \tag{9}$$

where $\overline{\mu_u}$ is the data rate of the MAC payload in a particular subchannel (the effective data rate of the MAC payload excluding parity information $\overline{\mu_e}$ is $\overline{\mu_u} \times \frac{K}{N}$), T_p is the period required to fill each ECB with its MAC payload when there are η sub-channels, and Υ is the confidence limit in determining the value of T_{WC}. The values of $\overline{\mu_u}$ and T_p can also be determined as follows:

$$\overline{\mu_u} = \frac{zL_u}{\eta T_{\text{slot}}}, \tag{10}$$

$$T^N_p(\eta, N_s) = \frac{NN_sL_u}{\overline{\mu_u}} = \frac{\eta N N_s T_{\text{slot}}}{z}. \tag{11}$$

5 Example QoS Analysis of Scalable Video Broadcasting over a CDMA2000 Network

We now present our experimental analysis of scalable video broadcasting over a CDMA2000 network. The 3GPP2 recently baselined the specification for a CDMA2000 high-rate broadcast packet data air interface [25], [26], which allows the high-speed delivery of packets to multiple-access terminals [27]. The air interface of this packet data system comprises a group of protocols collectively called the broadcast protocol suite [26]. The form of a transmission begins with framing protocol, which is used to fragment higher layer packets at the access network, and specifies how the access terminal determine higher layer packet boundaries. Next, the broadcast security protocol specifies how framing packets are encrypted using the AES block cipher. The broadcast MAC protocol defines the procedures used to transmit over the broadcast channel. It also provides FEC using RS coding, and multiplexing to reduce the radio-link error rate seen by the higher layers. Finally, the broadcast physical layer provides the channel structure for the broadcast channel. We will now go on to the detailed implementation of each layer specified in the standard, in the order in which they are processed by a mobile device.

TABLE 1
Time Required by Each Component of the RS Algorithm to Decode One Codeword

	$T^{(8,N,K)}_{C_i}$		
RS code	C_1	C_2	C_3
(32,24,8)	20.3μs	237.5μs	13.2μs
(32,26,6)	15.7μs	175.1μs	12.1μs
(32,28,4)	10.2μs	118.9μs	11.3μs

5.1 Simulation of the Inner Turbo Coding

A turbo code is known to be one of the best channel coding schemes for coping with bit flipping in communication systems. In the CDMA2000 forward channel, a coding-rate r of either $1/3$ or $1/5$ is used for the turbo encoder, which provides increased redundancy to help ensure that the packets are decoded by the mobile even when the channel conditions are poor. A turbo code of rate $1/3$ is a punctured version of the $1/5$ code. Details of the turbo code specified in CDMA2000 are given elsewhere [13], [14].

To evaluate the performance of turbo coding, it is necessary to know the size of a physical layer packet and to determine the number of decoding chips that will be required when a particular code rate is used. When the rate of the turbo code is r and QPSK modulation is used, $L_{\text{phy}}/2r$ chips are required for a L_{phy}-bit packet because one chip can transport two bits using QPSK. A further constraint is that the number of chips for processing each physical layer packet must be a multiple of 1,536. As a result, 1,536 are required when the code rate is $1/3$ and 3,072 are needed for a code rate of $1/5$. The performance of the CDMA2000 turbo code for a given number of chips can be determined using the Coded Modulation Library from Iterative Solutions [28], which runs on MATLAB.

5.2 Implementation of the RS Decoder

We used a software implementation of the RS *erasure* decoder, which is highly optimized for speed and memory, especially as regards the Berlekamp algorithm (i.e., the modified Berlekamp-Massey algorithm [31], [32]). This algorithm is used in Minsky's version of the RS decoder [29].

This version of the RS decoder runs in an ARM11-based mobile platform,[2] operating at 400 MHz in pipeline mode. Table 1 shows the time required to decode an 8-bit-wide RS codeword $(\hat{s} = 8)$, which is measured on the IAR [30] embedded workbench for the ARM processor. The IAR embedded workbench is a cycle-accurate simulator that supports instruction-level or functional-level execution time profiling. In the simulations, the erasures were randomly located in one of the 8 bits of the codeword, and the results presented in Table 1 are average values; but the position of an erasure within a codeword appears to have negligible effect.

5.3 Implementation of AES Block Decryption in the Security Layer

The broadcast security framework in CDMA2000 is described in detail elsewhere [33]. The broadcast security

2. The ARM processor family is widely used in contemporary 3G phones for the GSM and CDMA2000 networks. For example, QUALCOMM's MSM7500 convergence platform for CDMA2000 1xEV-DO high-speed wireless multimedia has an integrated 400 MHz ARM11 processor.

TABLE 2
Time Required by Each Transformation to Decrypt a 128-Bit Data Block ($N_b = 4$) and by the Key Expansion Routine for Different Values of N_k

A	ARK	ISR	ISB	IMC
T_A^A	0.727μs	0.097μs	7.373μs	20.633μs
KE				
N_k		4	6	8
$T_{KE}^{N_k}$		30.273μs	28.271μs	41.074μs

TABLE 3
System Parameters Used in Our Analysis

Symbol	Value(s)	Description
Physical channel parameters		
f_c	1.8GHz	Carrier frequency
μ_p	614.4kb/s	Reference channel data-rate
$\overline{\mu_u}$	120kb/s	Data-rate of the MAC payload
$\overline{\mu_e}$	90kb/s	Effective data-rate of the MAC payload
$\xi_{\min}$	-1.414dB	Minimum SNR within the service area
$\upsilon_{\min}$	4km/h	$V = 0.0111$
$\upsilon_{\max}$	40km/h	$V = 0.111$
Network parameters		
T_{slot}	1.67ms	Slot duration
η	5	Number of sub-channels
r	1/3	Inner turbo-code code-rate
L_{phy}	1024 bits	Length of a physical-layer packet
L_{mac}	1002 bits	Length of a MAC-layer packet
L_u	1000 bits	Length of the payload in a MAC-layer packet
Data processing parameters for the mobile device		
$\hat{s}$	8 bits	Length of a symbol
(N, K, R)	(32,24,8)	Example RS code
N_s	$1 \sim 16$	Number of MAC-layer packets in an ECB row
N_b	4	Number of 32-bit words comprising the state
N_k	4	Number of 32-bit words comprising the cipher key

protocol in the access network encrypts framing packets to form broadcast security packets using the AES encryption procedures [26], [34], [35], and the AES decryption code deciphers those security packets at the mobile which is receiving a video broadcast service. The AES algorithm can be implemented in any combination of software, firmware, and hardware. The appropriate choice depends on several factors such as the application, the environment, and the technology used. There are a number of examples of efficient implementations of this AES algorithm [36], [37] on a variety of platforms.

In this analysis, we used Gladman's [38] reference implementation of the AES algorithm, which is heavily optimized. This code allows blocks of 128, 196, and 258 bits, and this size is set during compilation. Both the code and data for AES decryption are cache-locked, preventing this time-critical process being ejected from cache. This version of AES block decryption will also run in a mobile device equipped with an ARM11 processor, and the execution time of the AES block decipher is also measured using the IAR embedded workbench.

The times required by each transformation routine of the AES algorithm to decrypt a 128-bit ($N_b = 4$) data block are presented in Table 2, which also illustrates how the time required by the key expansion routine depends on the length of the cipher key. The pseudocode for key expansion [11] shows how the first N_k words of the expanded key are filled with the cipher key. Every subsequent word is the result of an exclusive OR operation between the previous word and the word N_k positions earlier. When a word occurs in a position that is a multiple of N_k, an additional transformation, consisting of a cyclic shift of the four bytes of the word, followed by the application of a table lookup to all four bytes, is applied to the preceding word and the result is subject to a further exclusive OR with the round constant. This transformation is therefore executed more frequently when the value of N_k is small, so that more time is required for key expansion when $N_k = 4$ than when $N_k = 6$, as shown in Table 2. However, as N_k increases from six to eight, a slightly different key expansion routine is applied and the time required for key expansion increases again because the transformation is applied even more frequently.

5.4 Parameter Settings in the Proposed Service Architecture

The parameters used in our case study are summarized in Table 3. The value of each parameter is extracted from the CDMA2000 1xEV-DO [13], [14], [26] standard. We assume the use of QPSK modulation with a reference data rate of 614.4 Kb/s for a forward data channel, which is divided into 1.67 ms time slots, because this is known [26] to provide sufficient coverage even with an RS code of (16,12,4). Additionally, we assume that the forward channel contains five broadcast subchannels.

In CDMA2000 1xEV-DO, the modulation parameter and data rate used for the forward data channel [13], [14] determine the size of each physical layer packet, which is 1,024 bits under the assumptions that we have made and thus, the payload in every physical layer packet is a single 1,002-bit MAC layer packet. Each MAC layer packet has a 1,000-bit security layer packet as its payload. The CDMA2000 1xEV-DO standard also specifies that the physical layer packets transmitted over the forward data channel are encoded with a code rate of 1/3 by the turbo encoder when the data rate of that channel is 614.4 kb/s.

For our QoS analysis of scalable video in the CDMA2000 broadcasting environment, we used an example RS code of (32,24,8). Therefore, each broadcast subchannel has a data rate for its MAC payload of 120 kb/s, because we are assuming that there are five subchannels, which means that the effective data rate is 90 kb/s.

Regarding the physical channel and data processing parameters, we used values of V between 0.0111 and 0.111, which correspond to mobile speeds between 4 km/h and 40 km/h, with a reference channel data rate of 614.4 kb/s, a carrier frequency of 1.8 GHz, and a physical layer packet length of 1,024 bits. We also set the minimum value of SNR experienced by a mobile within a region to -1.414 dB. Then, we analyzed the service reliability and execution delay achieved by the RS decoder and AES decryption for different amounts of interleaving (i.e., $1 \leq N_s \leq 16$). The unit size of an AES cipher block and the size of the cipher key are 128 bits, and thus, the values of N_b and N_k are both 4.

5.5 Characteristics of the Sample Video and Video Broadcasting Applications

The sample video streams used for this analysis were *Star Wars IV* and *Silence of the Lambs*. The characteristics of these

TABLE 4
Statistics of the Sample Video Used in Our Analysis

Description	Value(s)
Title	*Star Wars IV, Silence of the Lambs*
Frames per second	30
Frame size	CIF (352×288)
Scalability scheme	Temporal
Stream type	VBR (variable bit-rate)
GoP size (G)	8
Number of B frames in a GoP	4 or 7
Quantization parameter	10 or 24

TABLE 5
WCET, Latency, and Buffer Size for Different Block Interleaving Parameters

N_s	$\overline{\text{PER}}_r$	WCET (ms)	Buffer size (bits)	Latency (ms)
1	2.918E-2	67	39720	331
2	1.178E-2	133	80040	667
4	3.294E-3	263	159600	1330
8	1.201E-3	517	318000	2650
16	9.589E-4	1009	633720	5281

video streams [39] are summarized in Table 4. The frames are packetized into 1,000 bits, which is the maximum payload size of physical layer packet in CDMA2000 1xEV-DO network.

A specified packet loss behavior by both the underlying network module and the data processing module on the mobile is translated by the application module into frame losses in the sample video streams. The quality degradation perceived by a user can then be estimated from (6).

5.6 Results of QoS Analysis

We now present the simulation results obtained from our cross-layer framework. Table 5 shows the upper bound on the residual rate of packet errors after RS decoding, which is $\overline{\text{PER}}_r$, when the worst SNR is -1.414 dB, and the speed of the mobile device ranges from 4 to 40 km/h. Additionally, the table shows the WCET required for RS decoding and AES decryption of the data in an ECB, the buffer size required for the initial buffering of a video stream arriving at a mobile device, and the corresponding service latency, for a given block interleaving parameter N_s. These numbers are simulation results with a 99.9 percent confidence level ($\Upsilon = 99.9$).

Fig. 9 shows traces of the number of residual packet errors in each ECB subblock for different levels of block interleaving, when the instantaneous SNR experienced by a mobile is -1.329 dB (equivalent to a PER of about 0.05 when $r = 1/3$). As shown in Fig. 9a, more interleaving reduces the number of residual packet errors, because a lower density of errors in the ECB subblocks makes it more predictable that those errors will be recovered by the RS decoder. However, at the higher mobile speed, error bursts are almost perfectly randomized by the block interleaving, as shown in Fig. 9b, so the interleaving has no effect on the error-correction performance of RS decoding.

Fig. 10 shows time traces for the RS decoding of consecutive ECB subblocks with different levels of block interleaving, measured at mobiles moving at two different speeds, when the instantaneous SNR experienced by a

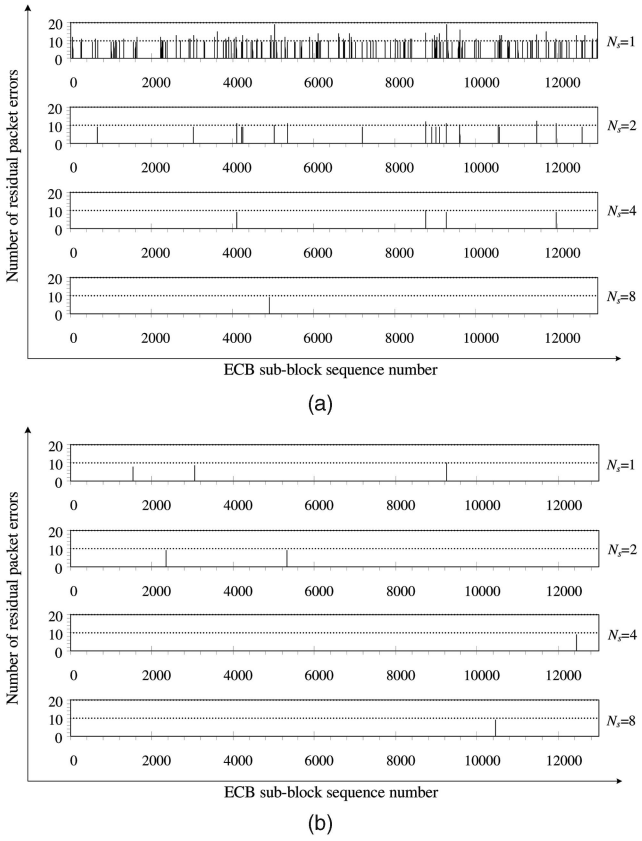


Fig. 9. Fluctuations in the number of residual packet errors in each ECB subblock as its sequence number increases, at different interleaving levels, with an SNR of -1.329 dB and at mobile speeds of 4 and 40 km/h. (a) 4 km/h. (b) 40 km/h.

mobile is -1.414 dB. With higher levels of block interleaving, error bursts in the physical layer are distributed across more ECB subblocks, reducing the average number of erased packets in a particular subblock but few subblocks contain no erased symbols, when error burst occurs. This explains the lower and more consistent RS decoding times. Conversely, low levels of interleaving concentrate errors into a smaller number of ECB subblocks and the number of error-free subblocks increases. This raises the density of errors in each subblock when the error burst occurs and the WCET increases, resulting in longer, and also more variable, RS decoding times, as shown in Fig. 10. In the faster mobile, these fluctuations are so pronounced that the results almost look random. In this case, the interleaving has no influence on the performance of RS decoding because the pattern of packet errors is already so highly irregular.

Fig. 11 shows traces of AES decryption time for successive ECBs, with different levels of block interleaving at a mobile speed of 4 km/h, when the instantaneous SNR experienced by a mobile is -1.329 dB. As the level of block interleaving increases, the variability of the decryption time is reduced because more errors are recovered, and thus, fewer errors remain after RS decoding, which is very clear from Fig. 9a. We only present the result for the mobile moving at 4 km/h because the time required to decrypt an ECB is almost constant for the faster mobile moving at around 40 km/h. This is because few residual packet errors

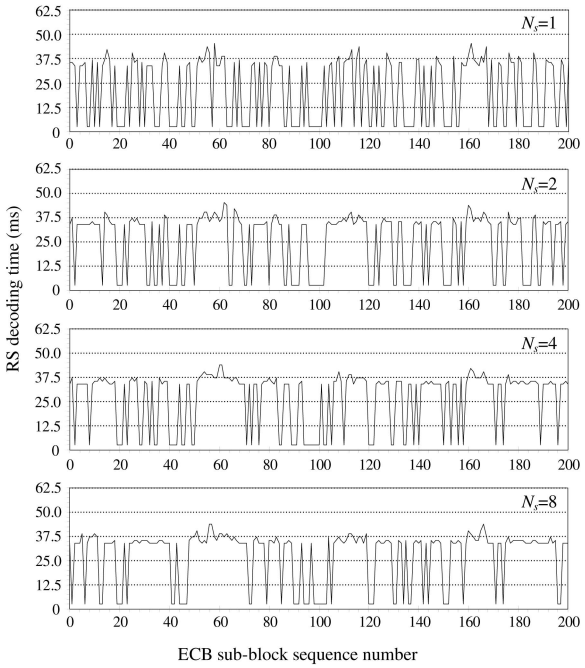


Fig. 10. Fluctuations of the time required for RS decoding of an ECB subblock as its sequence number increases, with an SNR of -1.414 dB and at mobile speeds of 4 and 40 km/h. (a) 4 km/h. (b) 40 km/h.

remain after RS error correction when the value of SNR is -1.329 dB, as shown in Fig. 9b.

Finally, the perceived quality degradation of the sample video streams, as mean-square errors, is shown in Figs. 12, 13, and 14. At both 4 and 40 km/h, distortion of the stream quality increases as the SNR decreases. At higher levels of block interleaving, the distortion decreases dramatically at both speeds. However, the distortion shows no clear relationship with the level of interleaving for the faster moving mobile. That is because the bursty errors caused by the fading channel are being randomized almost perfectly by the block interleaving, which can have no further effect on the perceived quality of the video stream.

Fig. 12 shows the perceived quality of the two video streams, *Star Wars IV* and *Silence of the Lambs*, with the same number of B frames and the same quantization parameter. We see that the way in which quality declines is similar for both streams. Fig. 13 shows the perceived quality of *Star Wars IV* with different numbers of B frames in a GoP. If there are fewer B frames in a GoP, then there must be more than P frames, which allow increased error-resilience. Finally, Fig. 14 shows the perceived quality of the *Star Wars IV* video stream with different quantization parameters. The use of AVC video compression with higher quantization parameters results in more compression and smaller frames, which are more tolerant of packet errors, and this reduces the perceived distortion of video streams over lossy wireless networks.

6 Discussion

Our framework could be useful in other scenarios if it were improved in a few ways:

- First, we have assumed that network operations are executed on a separate hardware chipset from the main processor (e.g., a modem processor), and this separate chipset is mainly present in order to correct errors in data packets. If some network operations have to be performed by the main processor, then an accurate assessment of computation time must allow for the scheduling of that processor.
- Second, our current approach to the validation and translation of input parameters is primitive. In order to provide more reliable guidelines for different system configurations, we need to improve the verification and translation of system configurations using a formal language such as the Architecture Analysis and Design Language (AADL) [40]. Using the AADL to describe a wireless system for broadcast services would allow an unfamiliar designer to check the performance of a configuration with different parameters. The AADL detects the use of invalid configurations, and the model itself constitutes a record of a set of acceptable system parameters. Many parameters are required for an analysis, so this arrangement would significantly reduce the possibility of parameters being misinterpreted and the scope for communication problems among system designers and engineers.
- Finally, it is desirable that feedback from users regarding perceived quality should be forwarded to the components in the underlying layers on the mobile device, and also sent back to the service provider, so that they can adjust either the application or network parameters accordingly. This is part of our program of work to increase the extent to which the user's perspective is considered in the cross-layer design of multimedia broadcasting services.

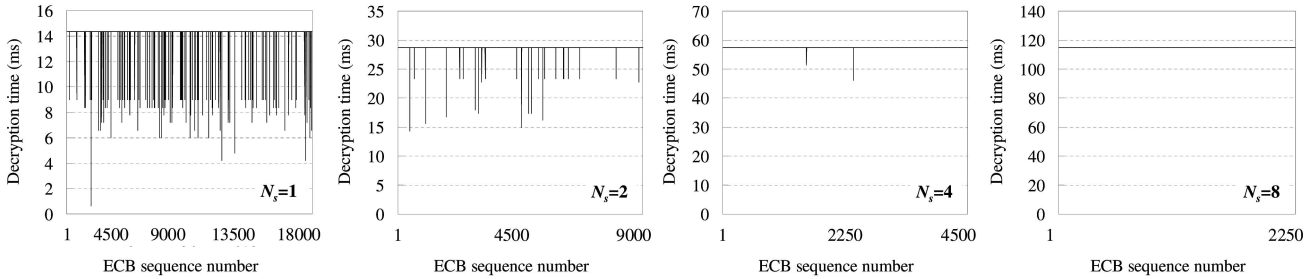


Fig. 11. Fluctuations of the time required for the AES decryption of an ECB as its sequence number increases, with an SNR of -1.329 dB and at a mobile speed of 4 km/h.

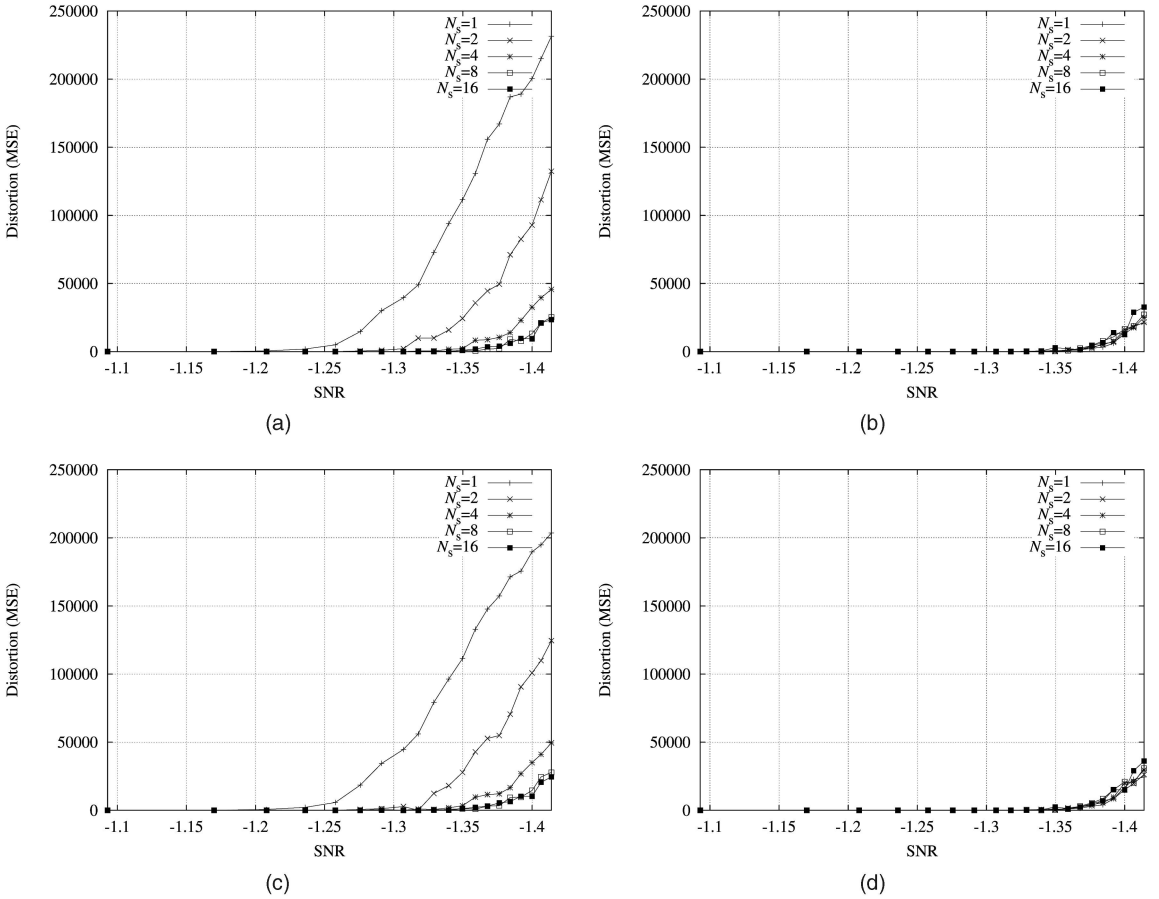


Fig. 12. Perceived quality of a scalable video stream for different levels of interleaving and mobile speeds (number of B frames = 7, quantization parameter = 10). (a) 4 km/h (*Star Wars IV*). (b) 40 km/h (*Star Wars IV*). (c) 4 km/h (*Silence of the Lambs*). (d) 40 km/h (*Silence of the Lambs*).

7 Conclusion

In this paper, we have proposed a QoS assessment framework for video broadcasting services in wireless cellular networks. The proposed framework is flexible in nature and can easily be extended to different system and network scenarios. By using a layered architecture of simulation modules, our analysis framework enables the systematic validation and evaluation of each module and the entire system before the deployment of a new type of system component or service. Using this framework, we analyzed the system and user performance of a mobile device in scalable video broadcasting in CDMA2000 1xEV-DO networks. By looking at different parameters for the system configurations such as wireless physical channel, network, and video streams, we were able to provide a baseline for client performance in terms of video quality in wireless broadcast networks.

Acknowledgments

A preliminary version of this paper appeared in the Proceedings of the 29th International Conference on Distributed Computing Systems (ICDCS '09). The corresponding author is W.J. Jeon.

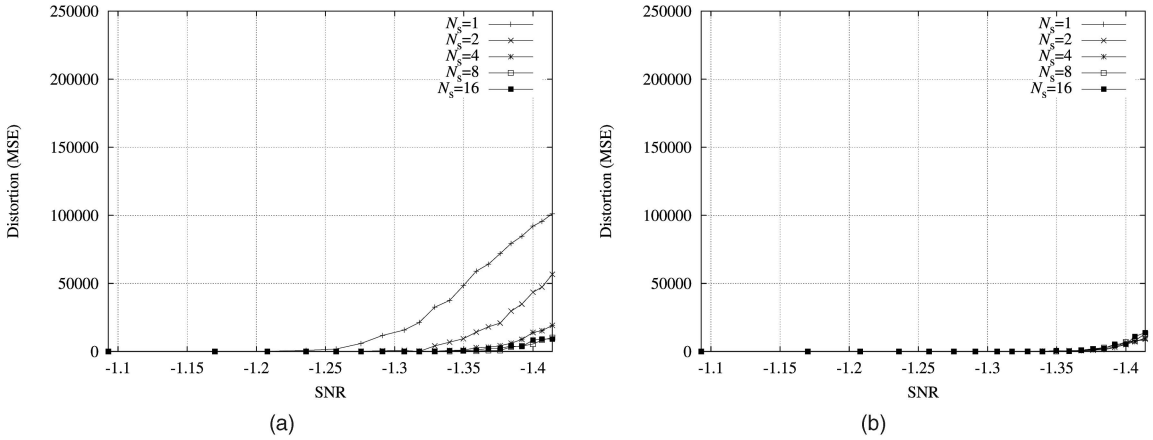


Fig. 13. Perceived quality of a scalable video stream for different levels of interleaving and mobile speeds (*Star Wars IV*, number of B frames = 4, quantization parameter = 10). (a) 4 km/h. (b) 40 km/h.

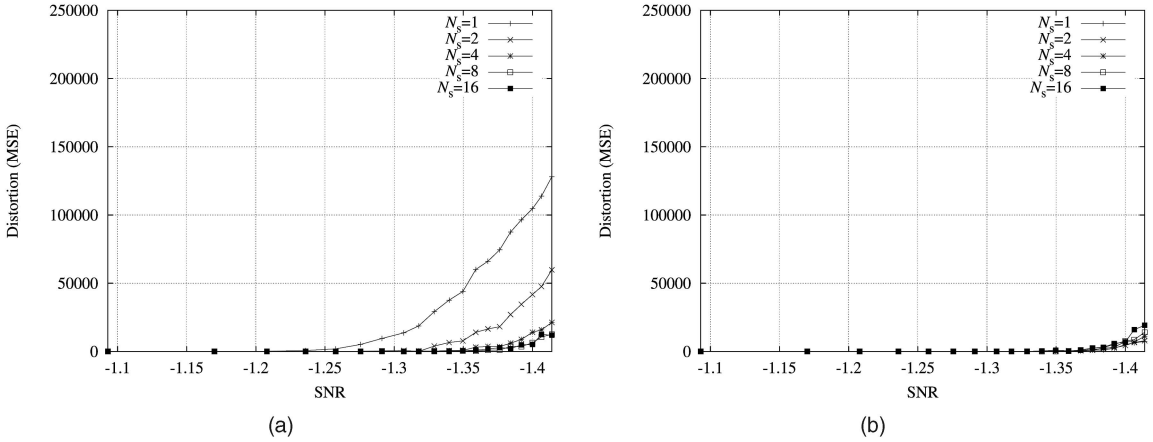


Fig. 14. Perceived quality of a scalable video stream for different levels of interleaving and mobile speeds (*Star Wars IV*, number of B frames = 7, quantization parameter = 24). (a) 4 km/h. (b) 40 km/h.

REFERENCES

[1] Opnet Technologies, http://www.opnet.com, 2010.

[2] The Network Simulator—(NS-2), http://www.isi.edu/nsnam/ns, 2010.

[3] C. Atici and M.O. Sunay, "High Data-Rate Video Broadcasting over 3G Wireless Systems," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 212-223, Mar. 2007.

[4] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall, "Delivery of Broadcast Services in 3G Networks," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 188-199, Mar. 2007.

[5] N.H. Vaidya, J. Bernhard, V.V. Veeravalli, P.R. Kumar, and R.K. Iyer, "Illinois Wireless Wind Tunnel: A Testbed for Experimental Evaluation of Wireless Networks" *Proc. ACM SIGCOMM Workshop Experimental Approaches to Wireless Network Design and Analysis*, pp. 64-69, Aug. 2005.

[6] M. Zorzi, R.R. Rao, and L.B. Milstein, "ARQ Error Control on Fading Mobile Radio Channels," *IEEE Trans. Vehicular Technology*, vol. 46, no. 2, pp. 445-455, May 1997.

[7] M. Zorzi and R.R. Rao, "On the Statistics of Block Errors in Bursty Channels," *IEEE Trans. Comm.*, vol. 45, no. 6, pp. 660-667, June 1997.

[8] M. Zorzi, R.R. Rao, and L.B. Milstein, "Error Statistics in Data Transmission over Fading Channels," *IEEE Trans. Comm.*, vol. 46, no. 11, pp. 1468-1477, Nov. 1998.

[9] D.R. Pauluzzi and N.C. Beaulieu, "A Comparison of SNR Estimation Techniques for the AWGN Channel," *IEEE Trans. Comm.*, vol. 48, no. 10, pp. 1681-1691, Oct. 2000.

[10] Y.Q. Shi, X.M. Zhang, Z.-C. Ni, and M. Ansari, "Interleaving for Combating Bursts of Errors," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 29-42, Aug. 2004.

[11] J. Daemen and V. Rijmen, *The Design of Rijndael: AES—The Advanced Encryption Standard*. Springer-Verlag, 2002.

[12] W.C. Jakes, *Microwave Mobile Communications*. Wiley-IEEE Press, May 1994.

[13] R. Parry, "CDMA2000 1xEV-DO [for 3G Communications]," *IEEE Potentials*, vol. 21, no. 4, pp. 10-13, Oct./Nov. 2002.

[14] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users," *IEEE Comm. Magazine*, vol. 38, no. 7, pp. 70-77, July 2000.

[15] C. Berrou and A. Glavieu, "Near Optimum Error Correcting Coding and Decoding: Turbo-Codes," *IEEE Trans. Comm.*, vol. 44, no. 10, pp. 1261-1271, Oct. 1996.

[16] S. Benedetto and G. Montorsi, "Unveiling Turbo-Codes: Some Results on Parallel Concatenated Coding Schemes," *IEEE Trans. Information Theory*, vol. 42, no. 2, pp. 409-429, Mar. 1996.

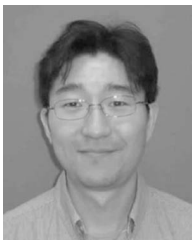
[17] W.J. Ebel and W.H. Tranter, "The Performance of Reed-Solomon Codes on a Bursty-Noise Channel," *IEEE Trans. Comm.*, vol. 43, nos. 2-4, pp. 298-306, Feb.-Apr. 1995.

[18] K. Kang, "Probabilistic Analysis of Data Interleaving for Reed-Solomon Coding in BCMCS," *IEEE Trans. Wireless Comm.*, vol. 7, no. 10, pp. 3878-3888, Oct. 2008.

[19] K. Kang, Y. Cho, and H. Shin, "Energy-Efficient MAC-Layer Error Recovery for Mobile Multimedia Applications in 3GPP2 BCMCS," *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 338-349, Mar. 2007.

[20] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 17, no. 9, pp. 1103-1120, Sept. 2007.
[21] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Modelling of Loss Distortion in Hierarchical Prediction Codecs," *Proc. IEEE Int'l Symp. Signal Processing and Information Technology,* pp. 536-540, Aug. 2006.
[22] P. Wellin, R. Gaylord, and S. Kamin, *Introduction to Programming with Mathematica.* Cambridge Univ. Press, 2005.
[23] T. Gayley, "Building User Interfaces Using J/Link," *The Mathematica J.,* vol. 9, no. 1, pp. 189-215, 2003.
[24] K. Kang and L. Sha, "An Interleaving Structure for Guaranteed QoS in Real-Time Broadcasting Systems," *IEEE Trans. Computers,* vol. 59, no. 5, pp. 666-678, May 2010.
[25] *CDMA2000 High Rate Broadcast-Multicast Packet Data Air Interface Specification, 3GPP2 Std. C.S0054-A Rev. 1.0,* Mar. 2006.
[26] P. Agashe, R. Rezaiifar, and P. Bender, "CDMA2000 High Rate Broadcast Packet Data Air Interface Design," *IEEE Comm. Magazine,* vol. 42, no. 2, pp. 83-89, Feb. 2004.
[27] J. Wang, R. Sinnarajaj, T. Chen, Y. Wei, and E. Tiedemann, "Broadcast and Multicast Services in CDMA2000," *IEEE Comm. Magazine,* vol. 42, no. 2, pp. 76-82, Feb. 2004.
[28] Iterative Solutions, http://www.iterativesolutions.com, 2010.
[29] A Reed-Solomon Error-Correcting Encoder/Decoder Library Written in C, http://sourceforge.net/projects/rscode, 2010.
[30] Embedded Development Tools, http://www.iar.com, 2010.
[31] J.L. Massey, "Shift Register Synthesis and BCH Decoding," *IEEE Trans. Information Theory,* vol. 15, no. 1, pp. 122-127, Jan. 1969.
[32] N.B. Atti, G.M. Diaz-Toca, and H. Lombardi, "The Berlekamp-Massey Algorithm Revisited," *Applicable Algebra in Eng., Comm. and Computing,* vol. 17, no. 1, pp. 75-82, Apr. 2006.
[33] *Broadcast-Multicast Service Security Framework, 3GPP2 Std. S.R0083 Rev. 1.0,* Oct. 2003.
[34] G. Rose and G.M. Koien, "Access Security in CDMA2000, Including a Comparison with UMTS Access Security," *IEEE Wireless Comm.,* vol. 11, no. 1, pp. 19-25, Feb. 2004.
[35] *Enhanced Cryptographic Algorithms, 3GPP2 Std. S.S0055 Rev. 2.0,* Jan. 2005.
[36] X. Zhang and K.K. Parhi, "Implementation Approaches for the Advanced Encryption Standard Algorithm," *IEEE Circuits and Systems Magazine,* vol. 2, no. 4, pp. 24-46, Oct.-Dec. 2002.
[37] G. Bertoni, L. Breveglieri, P. Fragneto, M. Macchetti, and S. Marchesin, "Efficient Software Implementation of AES on 32-Bit Platforms," *Lecture Notes in Computer Science,* pp. 159-171, Springer, Aug. 2002.
[38] Code for AES and Combined Encryption/Authentication Modes, http://www.gladman.me.uk, 2010.
[39] Video Traces for Network Performance Evaluation, http://trace.eas.asu.edu/tracemain.html, 2010.
[40] P.H. Feiler, D.P. Gluch, and J.J. Hudak, "The Architecture Analysis & Design Language (AADL): An Introduction," Technical Report CMU/SEI-2006-TN-011, The Software Eng. Inst., Carnegie Mellon Univ., Feb. 2006.

Kyungtae Kang received the BS degree in computer science and engineering and the MS and PhD degrees in electrical engineering and computer science from Seoul National University, Korea, in 1999, 2001, and 2007, respectively. Since 2008, he has been a postdoctoral research associate in the Department of Computer Science, University of Illinois at Urbana-Champaign. His research interests include QoS provisioning and energy minimization for real-time embedded applications in mobile systems. His recent research addresses problems in the interdisciplinary area of cyber-physical systems. He is a member of the IEEE.



Won J. Jeon received the PhD degree from the Department of Computer Science, University of Illinois at Urbana-Champaign. He is a senior engineer in the Computer Science Laboratory, Samsung Information Systems America, San Jose, California. Prior to that, he was a postdoctoral research associate at the University of Illinois at Urbana-Champaign. His research interests include distributed systems, secure mobile systems, resource management and adaptive middleware, multimedia systems, and information processing. He was the recipient of the Kodak Fellowship during his graduate study. He is a member of the IEEE.

Kyung-Joon Park graduated from the Seoul Science High School and received the BS, MS, and PhD degrees all from the School of Electrical Engineering and Computer Science (EECS), Seoul National University (SNU), Korea, in 1998, 2000, and 2005, respectively. He is currently a BK research assistant professor in the School of Electrical Engineering and Computer Science at SNU. He was a postdoctoral research associate in the Department of Computer Science, University of Illinois at Urbana-Champaign (UIUC), Illinois, from 2006 to 2010. He worked for Samsung Electronics, Suwon, Korea, as a senior engineer from 2005 to 2006, and was a visiting graduate student in the Department of Electrical and Computer Engineering at UIUC during 2001-2002. His current research interests include the design of medical-grade protocols for wireless healthcare systems, analysis of malicious and selfish behavior of wireless networks, design and analysis of self-adjusting protocols for wireless environments, and modeling and analysis of cyber-physical systems. He is a member of the IEEE.

Roy H. Campbell received the PhD degree in computer science from the University of Newcastle upon Tyne. He is the Sohaib and Sara Abbasi professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. His research interests include the problems of engineering and construction techniques of complex system software, security, continuous media, and real-time control, which pose a challenge, especially to operating system designers. His current research projects include security assessment of SCADA networks, operating system dependability and security, active spaces for ubiquitous computing, and the design of peer-to-peer distributed operating systems. He is a fellow of the IEEE.

Klara Nahrstedt received the PhD degree from the Department of Computer and Information Science, University of Pennsylvania. She is the Ralph and Catherine Fisher professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. Currently, she is the chair of the ACM Special Interest Group on Multimedia (SIGMM). Her research interests are directed toward multimedia middleware systems, QoS-aware resource management in distributed multimedia systems, multimedia security, and teleimmersive systems. She is a fellow of the IEEE.