ENABLING USER COOPERATION THROUGH EMBEDDED CODING AND AWARENESS OF SOURCE CONTENT DYNAMICS

Andres Kwasinski
Department of Computer Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA.

ABSTRACT
A challenge when enabling cooperation for conversational multimedia is how to obtain enough resources from the network. One approach is to use the source dynamics and allocate freed up resources from sources in low activity levels. This introduces a tradeoff between cooperation effectiveness and the networks ability to admit more users. This paper presents a scheme that incorporates the use of embedded source codecs, allowing for a tradeoff between source compression and cooperation effectiveness, without affecting the ability to admit more users. This increases the number of users that can be admitted to the network for a target received quality. A scheme is presented to decide when to trade off compression ratio for cooperation.

Index Terms— Embedded codec, cooperative systems.

1. INTRODUCTION
Wireless communications are characterized by very severe challenges, such as the presence of time-varying, error-prone channels and the scarcity of radio resources. Cooperative communications have become a major research area as a technology that can help in overcoming these challenges. With user cooperation, multiple users collaborate by creating multiple signal paths to relay information for each other. These signals are combined at a destination to improve signal quality. User cooperation were introduced in [1] by presenting the idea of “decode and forward” processing at a relay node. In [2] the authors introduced the idea of cooperation through “amplify and forward” (AF) processing.

User cooperation presents a tradeoff between received signal quality and bandwidth efficiency. This presents a challenge when transmitting multimedia sources under strict delay constraints because efficient cooperation must be implemented using the least amount of system resources. Fortunately, most multimedia sources, such as speech and video, exhibit patterns of variation in the coding bit rate required to achieve a target quality [3]. In conversational video with constant quality, the bit rate used to represent the signal changes with the amount of motion (activity) in the video sequence because strict delay constraints prevent the use of large bit rate smoothing buffers. Speech shows a similar behavior, following patterns of talk spurs and silence activity. These patterns had been exploited in statistical multiplexing-like schemes [4] where users who are silent, release their resources so as to admit more users to the network. Both these sources can be modeled using a Markov chain, with states representing levels of source activity such that higher levels require larger bit rates [3]. In [5], we introduced the idea of a content-aware multiple access protocol that enables cooperation by using some of the resources released by those users in a low level of source activity. This idea builds upon [6], where the authors proposed a multiple access protocol that uses for cooperation reduced traffic periods arising from burstiness in data traffic that is not delay sensitive.

The protocol in [5] showed that enabling cooperation by using the different encoding rates allowed by the source dynamics results in significant improvement in the received quality. In [5], the resources freed up by users in low activity levels were divided into resources reserved for users to contend for channel access and resources for cooperation. This established a tradeoff between the networks ability to admit more users and the effectiveness of user cooperation. As the number of admitted users increases, the resources for access contention become insufficient and many users are blocked during channel access due to collision with other contending users. Also, as the number of users increases, cooperation becomes less effective due to insufficient resources. In this paper we address this problem and we present a scheme that allows to keep the effectiveness of user cooperation and meet a target channel access blocking probability, while trading off a smooth reduction in received quality. The result is a marked improvement in the received quality from that in [5], which is achieved by relying in the use of embedded source encoders. Embedded source encoders had been shown to allow for a tradeoff between a smooth and controlled reduction in reconstructed source quality and an increase in network capacity [7]. The scheme presented here maintains these same features while enabling and gaining from user cooperation.

2. SYSTEM MODEL
We consider a network carrying conversational media, such as a voice conversation or a video conference, from source nodes to a base station. In the sequel we will focus on voice traffic. Yet, the scheme we are presenting is extensible to video conferencing because this traffic and speech can be modeled in similar fashion [3]. During a conversation, the speech activity patterns can be modeled using a two-state Markov chain, with one state representing a silence state and the other a talk spur. During communication, the speech source is first compressed at a source encoder, where an activity detector evaluates the
state of the source. During a talk spurt, the source encoder generates a coded source block every \( T_b \) seconds, which will constitute the main system timing unit. The probabilities that a talk spurt with mean duration \( t_s \) and that a silence period with mean duration \( t_s \) end at the end of a coding period of \( T_b \) sec. are \( \gamma = 1 - e^{-T_b/t_s} \) and \( \sigma = 1 - e^{-T_s/t_s} \), respectively.

Key in the presented scheme is that the source encoder is an embedded encoder. This means that a coarser version of the source can be reconstructed from a truncated version of the encoded stream. We assume that the encoder outputs \( \Psi \) equal-size source packets, which are ordered from the packet that provides a basic, coarsest representation of the source to the one that adds the finest details.

The wireless medium is divided into elementary orthogonal channels. To simplify the presentation, we assume in the sequel the adoption of TDMA for channel partitioning. Then, the wireless medium is divided into time frames of duration \( T_b \) sec. which are divided into elementary time slots. Each elementary time slot is apt for transmission of a channel-coded source packet. To transmit the source block, elementary time slots from each user are grouped into time slots, of which we assume there are \( N \) in one frame.

Time slots are divided into three types: those used to transmit voice traffic from active users, those use for contention access to the medium and those used for cooperation. Following the order in which they appear within a time frame, the first \( N_T \) slots are reserved for the talking users, i.e. if there are \( M_T \) talking users that succeeded in accessing the wireless medium, \( N_T = M_T \). Next, of the remaining \( N - N_T \) free slots, the portion \( N_R = \text{round}(p_R(N - N_T)) \), controlled by the parameter \( p_R \), are allocated to enable cooperation. Lastly, the remaining \( N - N_T - N_R \) slots are used for contention. Since \( M_T \) changes from frame to frame, the slots are rearranged in order to maintain the frame structure.

Users in the network may be in one of three states: talk, silence or contention. Users ending a talk spurt transition from the talk to the silence state. Users starting a talk spurt transition from the silence to the contention state. Users in a contention state attempt to reserve a time slot by each trying to transmit a source block over a contention slot with probability \( p_c \). If the transmission succeeds, the user transitions to the talk state and reserves a slot for talking users until the end of its talk spurt. If the transmission fails, the user is said to have been blocked and remains in the contention state until the contention succeeds or the end of its talk spurt. The value of \( p_R \) controls the tradeoff between the resources allocated to contention access and to user cooperation. A larger value of \( p_R \) reserves more slots for improved transmission through cooperation but reduces the number of slots for contention, which translates into a higher probability of user colliding during access attempts. The use of an embedded source encoder introduces an extra element of control because talking user may be commanded to reduce their source encoding rate by transmitting a number \( \psi < \Psi \) of packets. In this case, users are said to operate at embedded level \( \psi \), for which each of the \( N_T \) time slots used by talking users are composed of \( \psi \) elementary time slots, instead of the \( \Psi \) used at maximum embedded level. By reducing the embedded level, the sources compression ratio is increased, requiring \( M_T(\Psi - \psi) \) less elementary time slots for communication. These freed-up elementary time slots are then reassigned for cooperation for a total of \( \Psi N_R + M_T(\Psi - \psi) \) elementary time slots.

The successful reception of a source coded block is acknowledged with a single ACK signal, regardless of the embedded level used. Acknowledging each source packet, although it would allow to fully exploit the embedded property, would also entail excessive control complexity and overhead. A relay node helps those failed transmissions, for which no ACK was sent, by implementing cooperation through Amplify and Forward (AF) signal processing. Consistent with the strict delay constraint of conversational media, the relay helps failed transmissions within the same frame they were transmitted. For \( M_T \) failed transmissions, the relay amplify the \( \nu_F = (\Psi N_R + M_T(\Psi - \psi))/M_T \) most important source packets of each failed transmission and resend them using the elementary time slots allocated for cooperation.

Transmission quality is characterized through outage probabilities, which are the probability that the receiver Signal to Noise Ratio (SNR) is less than a threshold \( \beta \). For a direct transmission with no cooperation this probability is
\[
P_{O_D} = \Pr\left\{ \frac{\text{SNR}_D = H_{sd}r_{sd}^{-\alpha} P_s}{N_0} < \beta \right\} = 1 - e^{-\frac{2\nu_{rd}r_{rd}^{-\alpha} P_r}{N_0}}, \tag{1}
\]
where \( H_{sd} \) is the random source-destination channel gain magnitude squared, which has a unit-mean exponential distribution, \( P_s \) is the source transmit power, \( r_{sd} \) is the source-destination distance, \( \alpha \) is the path loss exponent, and \( N_0 \) is the AWGN background noise variance. The equality in (1) follows from the exponential distribution of the received SNR. Also, the outage probability when using AF cooperation is
\[
P_{O_C} = \Pr\left\{ \frac{\text{SNR}_C = \frac{H_{sd}r_{sd}^{-\alpha} P_s}{N_0}}{H_{sr}r_{sr}^{-\alpha} P_s + H_{rd}r_{rd}^{-\alpha} P_r + N_0} < \beta \right\},
\]
where \( H_{sr} \) and \( H_{rd} \) are, respectively, the source-relay and relay-destination channel gains magnitude squared and \( P_r \) is the relay transmit power, which we assume \( P_r = P_s \). We will only consider the case of a symmetric network, where all the inter-users channels are statistically identical. Due to the harmonic mean in \( \text{SNR}_C \), there is no known closed form expression for \( P_{O_C} \). Although approximate expression exists, we will compute this probability numerically.

3. PERFORMANCE ANALYSIS

The performance can be studied by modeling the network as a two-dimensional Markov chain and calculating the steady state probabilities of the states \((M_C, M_T)\), where \( M_C \) and \( M_T \) are random variables denoting the number of users in the
contention and talk states, respectively. These probabilities are the elements of the vector, π, that is the left eigenvector of the minimum eigenvalue of the state transition matrix P. The elements of this matrix are the probabilities P(S1, S2) of state transition from S1 = (MC1, MT1) to S2 = (MC2, MT2). The state transition is determined by the equations MC2 = MC1 + mSC + mTC - mCT, and MT2 = MT1 + mCT - mTS - mTC, where mij are random variables denoting the number of users that transition from state i to state j and where i or j are S, C or T to denote the silence, contention or talk state, respectively. Since, after succeeding with access contention, a user stays in the talk state until the end of the talk spurt (when it transitions to the silence state), mTS = 0. The probability distribution of mCT can be computed by following the recurrence model in [5]. The recurrence is necessary because the number of contending users vary from slot to slot. From the Markov chain that models the speech activity level, and assuming independence between users, the distribution of mTS = 0, ..., MT, mCS = 0, ..., MC - mCT conditioned on mCT and mSC = 0, ..., MS are all binomial random variable with parameters γ, γ, and σ, respectively.

\[
\begin{align*}
Pr(m_{TS} = i) &= \binom{MT}{i} \gamma^i (1-\gamma)^{MT-i}, \\
Pr(m_{CS} = i|m_{CT}) &= \binom{MC-m_{CT}}{i} \gamma^i (1-\gamma)^{MC-m_{CT}-i}, \\
Pr(m_{SC} = i) &= \binom{MS}{i} \sigma^i (1-\sigma)^{MS-i},
\end{align*}
\]

where MS is the number of users in the silence state which, for M_u total users in the network, is MS = M_u - MC - MT.

The above distributions, the elements of \( \pi \) are calculated as \( P(S_1, S_2) = \prod_{i=0}^{M_T} \sum_{y=0}^{MC} \sum_{z=0}^{MC} Pr(m_{CS} = x|m_{CT} = y, S_1) \times \sum_{z=0}^{MC} Pr(m_{SC} = x+z) Pr(m_{CT} = y|S_1) \)

where \( M' = \min(MC_1 - x, N - MT_1 - MR_1), Pr(m_{TC} = MT_1 - x + y - z|S_1) = 0 \) if \( MT_1 - x + y - z > MT_1 - z \) and \( m_{SC} = MC_2 - MC_1 + x + y - z|S_1 = 0 \) if \( MC_2 - MC_1 + x + y - z > MS_1 \) (5)]. Note that for M_u users in the network, the number of silent users is M_u - MT - MC.

The proposed scheme performance can now be analyzed by deriving the blocking probability and the mean received quality. The blocking probability, PB, is the probability that at least one user will fail during the one frame contention cycle. Denoting \( \pi(S) \) the steady state probability of state S, \( PB = \sum_{S} (1 - R_{N-M_T-N_R}(0)) \pi(S) \)

where \( R_{N-M_T-N_R}(0) \) is the probability that there are no contending users at the end of a frame, which is calculated as part of the recursion that calculates the distribution of mCT.

The mean quality is given by the combination of expected quality from users needing and not needing the help of a relay:

\[
Q = Q(\psi)(1 - PO_D)\psi + Q_R[1 - (1 - PO_D)\psi],
\]

where \( Q(\psi) \) is the source encoding expected quality for embedded level \( \psi \). The expected quality when using a relay, \( Q_R \), can be obtained by conditioning on the state and number of initial failed transmissions:

\[
Q_R = \sum_{S} \left( \sum_{i=0}^{MT} Q_R(\nu_F) \sum_{i=0}^{MT} Pr(M_F = i) \pi(S) \right) = \sum_{S} \left( \sum_{i=0}^{MT} Q_R(\nu_F) \sum_{i=0}^{MT} Pr(M_F = i) \pi(S) \right) = \sum_{S} \left( \sum_{i=0}^{MT} Q_R(\nu_F) \sum_{i=0}^{MT} Pr(M_F = i) \pi(S) \right)
\]

where \( Pr(M_F = i) \) is the probability of i failed transmissions, conditioned on the state, as is given by

\[
Pr(M_F = i) = \left( \begin{array}{c} MT \\ i \end{array} \right) \left[ 1 - (1 - F_{OD})^\psi \right] \left[ (1 - F_{OD})^\psi \right]^{MT-i},
\]

for transmission at embedded level \( \psi \). The expected quality, conditioned on the state and on the number of failed transmissions, when the relay transmits at embedded level \( \nu_F \), \( Q_R(\nu_F) \), needs to consider the possible outcomes for the reception of the source packets; it is given by:

\[
Q_R(\nu_F) = Q(\psi)(1 - F_{OD})^\psi \left[ (1 - F_{OD})^\psi \right]^{MT-i} + \sum_{i=1}^{\psi} Q(i) \prod_{k=1}^{MT-i} 1 - F_{OD} \prod_{k=1}^{MT-i} 1 - F_{OD}^\psi
\]

where \( \nu_N = \max(0, \psi - \nu_F) \) is the number of source packets that cannot be helped by the relay and \( F_{OD}^\psi(i) = F_{OD} \) if \( \nu_F > i \) and \( F_{OD}^\psi(i) = F_{OD} \) otherwise.

### 4. RESULTS AND ANALYSIS

We implement a simulation scenario to evaluate the proposed scheme. For voice codec we used a two-state variation of an embedded QCELP codec from the University of Maryland. At maximum embedded level, \( \psi = 8 \), the source encoding rate is 9600 bps and the coding period \( T_b \) is 20 ms. We measured the mean source encoding quality at different embedded levels, \( Q(\psi) \), using the ITU-T recommendation P.862.1 (PESQ) on six varied speech sequences of male and female subjects. The PESQ algorithm provides a quality score highly correlated with that of the Mean Opinion Score (MOS) subjective test. The speech source model had parameters \( t_s = 1 \) sec. and \( t_s = 1.35 \) sec. Each frame has \( N = 50 \) time slots and \( p_e = 0.45 \) (chosen to minimize blocking probability). The radio transmission parameters where \( \beta = 15 \) dB, \( \alpha = 3.5 \), \( P_r = P_s = 15 \) mW, \( N_0 = 10^{-12} \), \( r_{sd} = 200 \) m and \( r_{sr} = 100 \) m. These settings reflect a challenging highly faded channel that results in a mean received speech quality equal to 1.93 with no cooperation, which is an unacceptable poor quality.

We first explore the choice of parameter \( p_B \). The criteria for this choice is to meet a reasonable low blocking probability. In order to maintain acceptable quality this probability is usually taken to be smaller than 5% because repeatedly blocked users are not able to transmit, resulting in bad communication quality. Fig. 1 shows the blocking probability as a function of the number of users in the network for different values of \( p_B \). These results lead to a choice of \( p_B = 0.1 \), which limited \( P_B \leq 0.04 \) and the number of users to 25.

Fig. 2 shows for different embedded levels the mean quality as a function of the number of users in the network, M_u.
The curve for $\psi = 8$ (which is equivalent to the scheme in [5]) confirms that the challenging radio environment is overcome by enabling cooperation, resulting in acceptable quality. Nevertheless, the effectiveness of cooperation quickly decreases with the increase in $M_u$ because meeting the blocking probability goal implied that only 10% of free resources were allocated for cooperation. As $M_u$ increases, the number of talking users increases and the amount of free resources decreases. Then, the quality decreases because cooperation is unable to protect all the source packets in the embedded level. It can be seen that the use of the embedded codec allows to increase the resources for cooperation, maintaining its effectiveness for a larger $M_u$ and resulting in a much smoother decrease in quality. With the use of an embedded codec, and taking $2.5$ as the smallest acceptable mean quality value, it is possible to accept up to 22 users. This is an increase of 50% with respect to the scheme with no embedded codec.

One question remains open, and that is how the base station decides when to command users to change the embedded level. This is, how to control the tradeoff between source compression and user cooperation. The observations from Fig. 2 points to a scheme where the embedded level should not be calculated but rather be estimated as a moving average.

5. CONCLUSIONS

We have studied how to enable user cooperation in a network carrying conversational multimedia traffic. We presented a cross-layer scheme that enables cooperation by exploiting the different needs in source encoding rate following the source activity levels and that introduces the use of a embedded source codecs. The use of embedded codecs allow for a flexible mechanism to assign resources for cooperation that avoids sacrificing resources needed for network access contention. This introduces a tradeoff between source compression ratio and cooperation effectiveness. The result is a notable increase in the number of users that can be admitted to the network while reducing quality in a smooth way. A simple, yet effective scheme is presented to decide when to trade off compression ratio for amount of cooperation in the network.

6. REFERENCES