Wireless Multi-party Video Conferencing with Network Coding

Hui Wang, Ronald Y. Chang, and C.-C. Jay Kuo
Ming Hsieh Department of Electrical Engineering
Signal and Image Processing Institute
University of Southern California, Los Angeles, CA 90089-2564, USA
E-mails: wanghui@usc.edu, yjrchang@gmail.com, cckuo@sipi.usc.edu

***Abstract*— In this paper, we propose a cost-effective scheme for robust wireless multi-party video conferencing based on network coding (NC). The main idea is the adoption of a NC scheme to enhance robust transmission, to simplify the erasure protection procedure, and to reduce the downlink bandwidth by leveraging the properties of opportunistic NC and wireless broadcasting. We design a pipelining schedule to meet the delay requirement for real-time video conferencing. The proposed NC method outperforms the opportunistic network coding method by a significant margin and reduces the downlink bandwidth of the overall video bit rate.**

I. Introduction

Real-time multi-party video conferencing in wireless networks is a challenging task. Good video quality, large wireless bandwidth, and stringent delay requirements all pose formidable challenges in real-time video conferencing through error-prone wireless networks. To improve inefficient solutions with traditional methods, we propose a novel network coding scheme for the multi-party real-time video conferencing application.

Video transmission is subject to errors in wireless networks. With the traditional store-and-forward (S&F) method, all participants have to protect data against erasure transmission in each channel for graceful video quality degradation, which is complicated and bandwidth inefficient. Furthermore, transmission of multiple video bit-streams requires a large amount of bandwidth consumption when all video bit-streams are exchanged at the base station (BS). It limits the number of video conferencing participants since the bandwidth of downlink/broadcast from the base station to all users is a precious resource in a wireless network. Finally, real-time conferencing poses a stringent delay requirement. For example, it only allows $100 \sim 200ms$ delay in a commercial video conference system [1]. Most research on wireless network coding [2], [3] has primarily focused on the video streaming application, which allows longer delay.

In this work, we propose a cost-effective approach for robust wireless multi-party video conferencing based on network coding (NC). First, we propose a novel NC scheme for advanced erasure protection. The NC scheme not only enhances robust transmission but also simplifies the erasure protection mechanism, which only involves the base-station(BS)/access-point(AP) in erasure protection. It is shown by computer simulation that the NC scheme leads to a higher decoding probability to recover source video packets at the receiver end. Consequently, each user can receive better video quality. Second, we aim to reduce the bandwidth of the downlink channel by leveraging opportunistic NC and wireless broadcasting. Simulation results show that the downlink bandwidth can be greatly reduced by the proposed NC scheme. Third, we design a pipelining schedule and a buffering policy to facilitate NC on all participants so as to meet the delay requirement in real-time video conferencing. The maximum decoding delay is around $100ms$, which is within the delay requirement.

The rest of this paper is organized as follows. First, we offer a concise review on wireless NC in Sec. II. Then, we propose the NC scheme for wireless multi-party video conferencing in Sec. III. The performance of the proposed NC scheme is analyzed in Sec. IV. Simulation results are shown in Sec. V. Finally, concluding remarks are given in Sec. VI.

II. Review of Previous Work

Wireless NC has been studied by several researchers. Some of the main results are briefly reviewed in this section.

Wu *et al.* [4] studied the benefit of performing NC in a simple wireless ad hoc network. They proposed a bandwidth-reduction and power-saving method based on the XOR operation on intermediate nodes for wireless information exchange. The proposed scheme reduces the bandwidth consumption by broadcasting NC packets over the wireless channel. When NC is adopted in an error prone network, errors propagate. However, the issue of performance degradation due to error propagation was not discussed for NC-based message exchange in [4]. To address the error propagation effect, Karande *et al.* [5] proposed a cross-layer wireless NC method and studied the optimality condition, which can be achieved by selecting NC or S&F dynamically on intermediate nodes based on the error rate.

Katti *et al.* [6] proposed the use of opportunistic scheduling in the scenario of multiple unicasts to improve throughput. This method performs optimal scheduling based on the state information of neighboring nodes. With the overheard information from neighbors, optimally decodable NC codes can be generated for neighbors. As a result, the throughput is maximized.

Nguyen *et al.* [2] studied the application of NC in wireless networks for video broadcasting. They proposed an optimal scheme to generate erasure protection codes by performing NC for re-transmitting lost packets. By gathering ARQ (Automatic Repeat Request) messages from receivers, the broadcasting source generates optimal NC packets. Seferoglu *et al.* [3] proposed an extension of opportunistic scheduling with NC for multiple video unicasts. The quality of video streaming, instead of throughput, is optimized with a video time constraint. The NC codes are generated based on the priority and emergency of these packets.

III. Wireless Video Conferencing with NC

We focus on multi-party video conferencing in WiMAX networks. To facilitate our analysis and discussion, we have made the following assumptions on the system model.

- All channels are i.i.d. stationary, duplex and symmetric channels. Ignore the mutual interference among users and collisions in wireless channels.
- All channels are line of sight (LOS) channels, where the multi-path effect is ignored. Thus, the main channel behavior is signal

propagation (with a decaying magnitude and certain delay) corrupted by noise.

- The multi-carrier technique is adopted by all nodes in the network so that each node can transmit and receive data simultaneously.

A. System Overview

s A representative system diagram is shown in Fig. 1. Suppose U users join a video conference session, and every user expects to receive video streams from all other participants. Source video sequences are encoded at all user nodes, organized into packets of equal length. Meanwhile, every user overhears transmission from his/her neighbors. Each user encodes his/her own source packets and overheard packets from neighbors into a generation with random linear NC (RLNC) [7], [8]. Then, the user sends a generation of NC packets to the base station (BS).

After finishing sending a generation, users send requests to BS for more packets to successfully decode an NC generation. The BS receives packets from all users, and generates NC-coded packets for broadcasting based on user's requests. Then, users reconstruct the Global Coding Matrix (GCM) from the received and overheard packets, and decode the source data of other users. Finally, users can display these video contents.

NC packets are stored in buffers S and R of user nodes as well as buffer Q of the BS. Channels are classified as

- Uplink channels from users to the BS;
- Downlink/broadcast channels from the BS to all users;
- Overhearing channels between neighboring users when they work at the promiscuous mode.

The packet loss rate of both uplink and downlink channels of user i is α_i while the packet loss rate of the overhearing channel between user nodes i and j is $\beta_{i,j}$. The overhearing channel is symmetric in the sense $\beta_{i,j} = \beta_{j,i}$. In the following, we discuss the system, which consists of two parts: 1) an NC process, and 2) real-time scheduling.

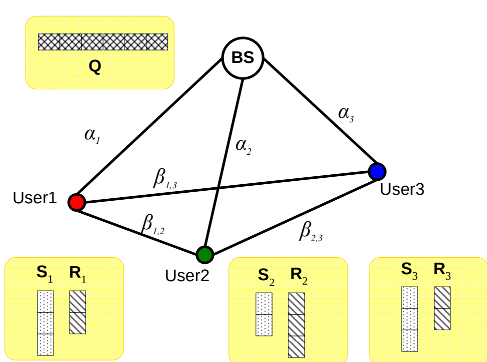

Fig. 1. Illustration of the NC wireless multi-party video conferencing system.

B. Proposed NC Process

In the proposed system, a generation of NC packets contains data blocks from all users in a period of time of length T. For example, if all users begin with the video conferencing session simultaneously, packets with the same frame index may be encoded into one generation. The NC process is performed on both user nodes and the BS as described below.

- In the 1st step, NC encoding is performed at each user node by mixing source packets from itself and overheard packets from neighboring nodes. For user i, buffer S_i stores source packets. The number of source packets for the current generation is $|S_i|$. A packet in this buffer is denoted by s_j with $0 \le j \le |S_i|$. Buffer R_i, whose size is denoted by $|R_i|$, stores overheard packets from other users. A packet in this buffer is denoted by r_j with $0 \le j \le |R_i|$. The kth NC-encoded packet at user i can be written as

$$Y'_k = \sum_{j}^{|S_i|} a_{j,k} s_j + \sum_{j}^{|R_i|} b_{j,k} r_j, \tag{1}$$

where $a_{j,k}$ and $b_{j,k}$ are randomly selected coefficients over a finite field. User i generates $|S_i|$ NC packets for each NC generation. The coefficients for one generation compose the local coding matrix on users (LCMU) which is a matrix of size $|S_i| \times (|S_i| + |R_i|)$. These NC packets are sent while packets of neighbors are received via overhearing channels. It takes a time interval of length T to send all packets for one generation. After finishing sending a generation, if user i finds that overheard packets are not sufficient to decode other users' packets for this generation, it requests m_i additional packets from the BS. To calculate m_i, user i counts two numbers: N' and $|R_i|$. N' is the total number of packets from all other users in this generation, and $N' = \sum_{j \ne i}^{U} |S_j|$ where $|S_j|$ is the number of source packets on user j. $|R_i|$ is the number of overheard packets on user i. It is clear that $m_i = N' - |R_i|$.

- In the 2nd step, BS generates NC packets by mixing all received packets after receiving users' requests. For every generation, BS generates M packets and broadcasts them to all participants, where $M = \max(m_i, \forall i \in [1 \sim U])$. Each packet in the BS buffer Q is denoted by q_j with $0 \le j \le |Q|$, where $|Q|$ is the number of packets stored in buffer Q. The BS does not send any NC packet immediately but wait for T until all users finish transmitting packets for a generation. Since more packets are received on the BS, a space of higher dimension is spanned by these NC packets. Thus, by waiting for T, the BS can generate, with a high probability, innovative NC packets which increases the chance to achieve full rank of GCM at the user end. The kth NC-encoded packet at the BS can be written in form of

$$Y_k = \sum_{j}^{|Q|} c_{j,k} q_j, \tag{2}$$

where $c_{j,k}$ is a randomly selected coefficient over a finite field. The coefficients of one generation compose the $M \times (|Q|)$ local coding matrix on the BS (LCMB). Those coefficients are transmitted along with NC packets [8] from the BS to each user node.

- In the 3rd step, each user performs NC decoding. For user i, the source packets in buffer S_i are of dimension $|S_i|$ in one generation. If they participate in the decoding process together with received NC packets in buffer R_i, the need of additional packets from the BS for successful decoding is eased and thus the downlink bandwidth saved. If user i wants to decode data from all other users $(S_j, \forall j \in [1 \sim U], j \ne i)$, it has to reconstruct an $N \times N$ GCM_i where $N = \sum_{j=1}^{U} |S_j|$. The upper part of GCM is an $|S_i| \times N$ submatrix. Columns corresponding to $[\sum_{k}^{i-1}$ to $\sum_{k}^{i}]$ are an identity matrix of dimension $|S_i| \times |S_i|$. The remaining submatrices are filled with zero. The lower part is an $N' \times N$ submatrix, which is composed of global coding coefficients from LCMU and LCMB. Then, the decoding equation becomes

$$\begin{bmatrix} S_i \\ R_i \end{bmatrix} = GCM_i \times \begin{bmatrix} S_1 \\ \vdots \\ S_U \end{bmatrix} \tag{3}$$

Source video packets from other users are decoded by applying Gaussian elimination to (3).

C. Real-Time Video Scheduling

We design a scheduler to achieve the real-time requirement for video conferencing as shown in Fig. 2. Based on the three NC steps stated in the last subsection, the following three operations can be pipelined: 1) user sending, 2) BS sending, and 3) user decoding/display. Each operation consists of two stages so as to process the odd and the even generations, respectively. As shown in the figure, t_0 is the start time for video conferencing and users start to send data. The BS waits for a time interval of length T and starts to send at t_1. The NC method proposed in [9] generates NC packets immediately after receiving any packet. In the video conferencing scenario, this may waste the downlink bandwidth since fewer packets are NC-encoded on the BS and the probability of generating innovative packets to users becomes lower. In our scheme, the BS gathers a sufficient number of packets to generate new NC packets by waiting for an interval of length T and, as a result, the chance of having innovative NC packets becomes higher. To decode the NC generation, users wait another interval of length T until a sufficient number of packets of the same generation are received for decoding at t_2 and then start to display.

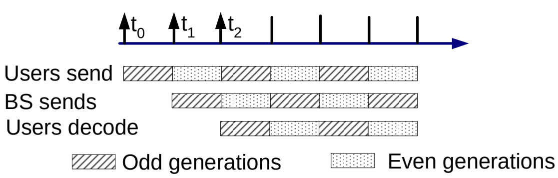

Fig. 2. Real-time scheduling with even and odd generations of video packets.

To enable the pipelining process for real-time video conferencing, we develop an alternative storage structure for buffers in Fig. 1. That is, we store even and odd generations of video packets alternatively. The life cycle to complete one even and one odd generations is $2T$. Those buffers are refreshed at the end of each life cycle. Within one life cycle, we have the following buffer management processes.

- Management of buffer S. In the first period of length T, S provides packets for the 1st-step NC encoding. In the end of the second T period, S provides packets for NC decoding.
- Management of buffer R. In the first period of length T, R stores packets received from overhearing channels. In the second period of length T, R stores packets received from broadcasting channels. At the end of the second period, R provides packets for NC decoding.
- Management of buffer Q. In the first period of length T, Q stores packets received from uplink channels. In the second period of length T, Q provides packets for the 2nd-step NC encoding.

With the above buffer management scheme, the decoding delay of the proposed NC process is equal to $t_2 - t_0 = 2T$ and the overall delay is $d = 2T + 2 \times \rho$ where ρ is the propagation delay in wireless media between users and the BS. Usually, ρ is small and negligible. If T is sufficiently small, we can meet the real-time transmission constraint. For example, if $T = 1/30s$, the NC decoding delay is $2 \times \frac{1}{30} \approx 66.7ms$.

IV. Performance Analysis

In this section, we analyze the proposed NC method and show that it offers a simple and robust erasure protection mechanism as compared with the opportunistic wireless NC method in [3]. In a wireless environment, video conferencing packets are subject to erasure over uplink/downlink/overhearing channels. The opportunistic wireless NC method in [3] has a problem in erasure protection. That is, when a packet sent from user i may be lost in both the overhearing and the uplink channels with probability $\alpha_j \times \beta_{j,i}$. In this case, BS is not able to send any useful packet to user j to decode this packet. When user i does not achieve the full rank of S_j from the other users, he/she cannot have the full rank of GCM_i in (3) to decode the current generation of packets regardless of the number of NC packets sent from the BS. Then, all data in this generation would be lost, which results in severer video quality degradation. An approach to alleviate this problem is to add redundancy for both users and the BS via channel estimation and optimized erasure protection.

In the analysis outlined above, the key is the rank of S_j of both the BS and user i, which composes a subspace as illustrated in Fig. 3(a), where the subspace, $I_{i,j}$, consists of a set of nodes where user i collects innovative packets after user j sends NC packets of a generation. For GCM_i to be of full rank, $S_j, (\forall j \in [1 \sim U], j \neq i)$ on user i have to be of full rank. Thus, it is essential to study ranks of all $I_{i,j}, (\forall j \in [1 \sim U], j \neq i)$. With the opportunistic NC method [3], the probability of achieving full rank over $I_{i,j}$ is

$$P_o = (1 - \alpha_i \beta_{i,j})^{|S_j|}. \tag{4}$$

Next, we study the rank of subspace $I_{i,j}$ of the proposed NC method. Besides the BS, more user nodes can contribute innovative packets to user j as shown in Fig. 3(b), The subspace $I_{i,j}$ of the proposed NC method includes nodes set $\{BS, user_k, \forall k \in [1 \sim U], k \neq i\}$. The probability of achieving full rank of S_i for the proposed NC method becomes

$$P_m = (1 - \alpha_i \prod_{k \neq j} \beta_{j,k})^{|S_j|}. \tag{5}$$

It is clear that $P_m > P_o$ if $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. If more users join the video conference, P_m is close to 1 whereas P_o does not change with U. The proposed method has a higher probability to obtain innovative packets in $I_{i,j}$. Therefore, it is more probable that (3) is decodable. Furthermore, we are free from erasure protection at user nodes, which eliminates the additional bandwidth required for redundancy packets in uplink channels. The proposed system needs only to protect the transmission in the downlink channel. We use a dynamic request to obtain more NC packets from the BS based on the sequence number of BS NC packets. That is, if a BS packet is lost in the downlink channel to user i, user i will request one more packet on top of m_i.

V. Simulation Results

We simulated the proposed system and NC algorithm with ns2. According to the analysis given in Sec. IV, when more users join the video conferencing session, the probability of achieving the full rank of GCM so as to have better video quality increases. For simplicity, we use the case $U = 3$ to demonstrate the advantage of the proposed method. We set $\beta_{i,j} = 0.3, \forall \{i, j\} \in [1 \sim U], j \neq i$. The 4CIF Crew sequence (848kbps) coded by H.264/AVC is placed in the buffer of

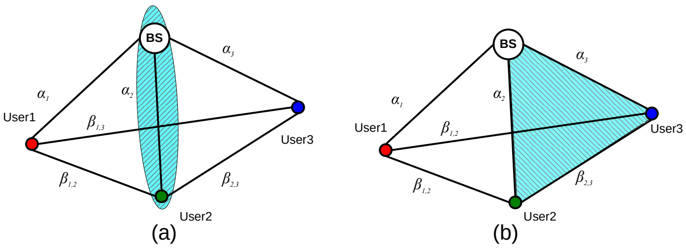

Fig. 3. The subspace for user node 1 with (a) the opportunistic NC method and (b) the proposed NC method.

TABLE I
THE PSNR VALUE COMPARISON OF EACH INDIVIDUAL USER WHEN $\alpha = 0.05$.

	Soccer1	Foreman1	Crew2	Foreman2	Crew3	Soccer3
[6]	27.8	28.2	30.1	28.9	29.8	26.4
Prop.	31.4	33.5	32.7	32.3	31.2	27.4

user 1. Similarly, coded Soccer (655kbps) and Foreman (553kbps) sequences are placed in the buffers of users 2 and 3, respectively. The frame rate is 30 fps, the QP parameter is 35, the GOP size is 8, and the intra period is 8. The finite field size for NC coefficients is 2^8, and the length of a generation period is $T = 1/30s$.

We use the averaged PSNR of all six received sequences to evaluate the overall system performance in Fig. 4. We see that the proposed NC method outperforms the traditional wireless NC method in terms of the averaged PSNR value. This is because the proposed NC method has a more efficient erasure protection capability. We also show the throughput of the downlink channel as a function of the average packet loss rate α. The throughput is about $1/3$ of the summing bit rates of three bitstreams. This reduction comes from effective overhearing of innovative NC packets from neighbors. Since additional NC packets are sent from the BS to compensate erasure packets via dynamic request, the throughput increases slightly with α.

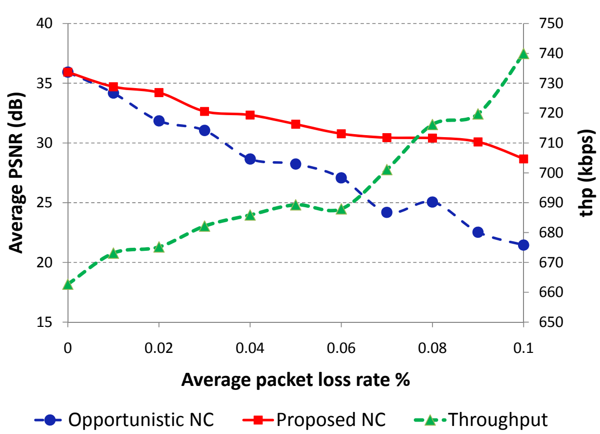

Fig. 4. The averaged PSNR performance and the throughput as functions of the average packet loss rate, α, of the downlink channel.

Table I shows the PSNR value of received sequences on three users. Again, the proposed NC method outperforms the wireless NC method for every sequence. Furthermore, we show one reconstructed

(a) Opportunistic NC (27.8dB) (b) Proposed NC (31.4dB)

Fig. 5. Comparison of the decoded 275th Soccer frame at user node 1 with $\alpha = 0.05$.

(a) Opportunistic NC (28.2dB) (b) Proposed NC (33.5dB)

Fig. 6. Comparison of the decoded 22nd Foreman frame at user node 1 with $\alpha = 0.05$.

frame of two received video sequences at user node 1 in Figs. 5 and 6. The visual quality of the proposed NC method is significantly better than that of the opportunistic wireless NC method.

VI. Conclusion and Future Work

A novel network coding method with a simple yet robust erasure protection mechanism was proposed for multi-party video conferencing. We also designed a scheduling scheme to achieve the low delay requirement for real-time applications. Simulation results show that the proposed NC method outperforms the opportunistic NC method in terms of received video quality and throughput by a significant margin. We will apply the proposed NC scheme to scalable video coding (SVC) bitstreams with unequal erasure protection in the near future.

References

[1] Cisco, "Telepresence network, http://www.cisco.com."
[2] D. Nguyen, T. Nguyen, and X. Yang, "Multimedia wireless transmission with network coding," *Proc. of Packet Video*, Nov 2007.
[3] H.Seferoglu and A.Markopoulou, "Opportunistic network coding for video streaming over wireless," *Proc. of Packet Video*, Nov 2007.
[4] Y. Wu, P. A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," *Proc. 39th Annual Conference on Information Sciences and Systems (CISS)*, Mar 2005.
[5] S. Karande., K. Misra, and H. Radha, "Clix: Network coding and cross layer information exchange of wireless video," *Image Processing, 2006 IEEE International Conference on*, Oct 2006.
[6] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, June 2008.
[7] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger, "On randomized network coding," Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing, Oct 2003.
[8] P. Chou, Y. Wu, and K. Jain, "Practical network coding," 51st Allerton Conf. Communication, Control and Computing, Oct 2003.
[9] D. Lun, M. Mdard, and M. Effros, "On coding for reliable communication over packet networks," *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, September-October 2004.