Joint Opportunistic Network Coding and Opportunistic Routing for Correlated Data Gathering in Wireless Sensor Network

Chong Tan[1,2], Junni Zou[2], Min Wang[2]
[1]Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences
[2]Department of Communication Engineering, Shanghai University, Shanghai, 200072, P.R. China
Emails: {dudumaotc, zoujn, wangmin}@shu.edu.cn

***Abstract*—In this paper, we study the problem of correlated data gathering in wireless sensor networks. An opportunistic routing protocol joint with opportunistic network coding is proposed for correlated data gathering. In order to reduce the total transmission, we introduce the expected number of coded transmission (ECTX) as metric for the routing selection. Moreover, we also define the coding gain for the network nodes to count the coding opportunistic, with which the throughput of the wireless sensor network is further increased. Numerical simulations are conducted for performance analysis with the proposed routing scheme.**

***Index Terms*—Wireless sensor network, opportunistic network coding, opportunistic routing, correlated data gathering, distributed source coding**

I. Introduction

Due to typically battery-powered sensor nodes, effective data gathering from sensor nodes to end sites or sink nodes is crucial for energy-constrained wireless sensor networks. Since sensor nodes partially monitor the same spatial region and are often densely deployed, the data collected by nearby sensor nodes are either redundant or correlated. This data correlation can be exploited to reduce the amount of data transmitted in the network, resulting in energy savings. In this paper, we consider the data gathering scenario where data is sampled at a number of distributed correlated sources and needs to be routed to one or a few sink nodes.

The data gathering for correlated sources can be discussed in the prior works. Slepian-Wolf coding (S-W coding) is a distributed source coding that can remove data redundancy without requiring inter sensor communication. However, S-W coding is based on the assumption that each sensor node has a priori knowledge of the correlation structure, which depends on the distances between sensor nodes and the characteristics of the observed phenomenon. Unfortunately, S-W coding requires a perfect prior knowledge of the whole network which is not scalable in large networks. In this paper, we adapt a localized version of S-W coding from [1], such that only local correlation information is required at each sensor node.

Network coding has demonstrated a wide range of applications for improving the performance of wireless networks [2]. However, a packet gained by a relay node has to wait to be network-coded with others which results in large delay and packet-loss rate. A novel network coding approach, referred as opportunistic network coding (ONC), is presented in [3] to overcome these limitations. With the opportunistic network coding, the intermediate node can determine whether a packet is transmitted with or without network coding. As it is difficult and complicated to design dynamic scheduling and coding algorithms, a practical network coding protocol is proposed in [4] which is based on the opportunistic reception from the download nodes. Although different approaches to network coding are disused and illustrated in the previous works, they focus on how to apply network coding at the intermediate nodes. It is obvious that network coding can improve the network performance only when the bottleneck link occurs or the data flow is beyond the limited link capacity. Therefore, we consider joining the opportunistic network coding with the routing protocol to establish an effective data gathering framework for wireless networks.

Moreover, opportunistic routing emerges a promising technique that achieves higher throughput than traditional deterministic forwarding in the face of highly unpredictable and lossy wireless channels. Biswas et al. in [5] proposed an opportunistic protocol ExOR to send information through multiple relays and introduce delicate scheduling policies to reduce duplicate transmission. Chachulski et al. in [6] developed an improved ExOR protocol called MORE, in which no specific scheduler is required and network coding is applied to avoid the unnecessary forwarding. However, the metric they adopted for the relay set selection is very simple (e.g. ETX metric) which largely limits their capability in improving the network throughput [7]. In [8], coding-aware routing protocol is proposed which takes into account network coding at network nodes in route selection. Ni et al. in [9] defined the coding gain as the routing metric and proposed a coding-aware routing scheme with which the total number of coded transmissions can be reduced. However, the above coding-aware routing protocols only fix the route selection issue, and neglect how to apply the network coding at the network nodes. Hence, in the paper, we not only consider the opportunistic network coding to reduce the number of transmissions, but also take account into the opportunistic

978-1-4673-6187-3/13/$31.00 ©2013 IEEE

routing mechanism to improve the performance of the whole network.

Our main contributions in this paper are as follows. First, we join the opportunistic network coding and the opportunistic routing together to establish an efficient data gathering framework for correlated sources. Then, we consider both the number of coded transmissions and the network coding opportunistic to decrease the total number of transmissions.

The rest of this paper is organized as follows. Sec. II presents the network model, related constraints and formulates the problem. An opportunistic routing strategy joint with ONC is then proposed at Sec. III. Numerical results are presented in Sec.IV, and we conclude this article at the last section.

II. System Model

A wireless sensor network is represented by a directed graph $G = (V, E)$, where V is the set of nodes, E is the set of wireless links, $S \in V = \{s_1, s_2, ..., s_{N_s}\}$ and $T \in V = \{t_1, t_2, ..., t_{N_t}\}$ denote the set of sensor nodes and sink nodes, respectively. Each node $i \in V$ has a radio range d_i and can communicate with other nodes within its range. We denote a link either by a single index l or by the directed pair (i, j) of nodes it connects. And each link $(i, j) \in E$ has a fixed finite capacity C. u_{ij} and r_{ij} are the flow rate and the packet reception ratio of the link $(i, j) \in E$, respectively.

For a sensor node $s_i \in S$, T_i is the set of sink nodes which can gain data from node s_i; For a sink node $t_j \in T$, S_j is the set of source nodes with the same sink node t_j.

Each source s_i periodically measures a continuous random observation X_i. The values of the sources are drawn from some joint distribution $p(X_1 = x_1) = p(x_1)$ and can be either continuous or discrete. The joint source vector $X = \{X_1, ..., X_{N_S}\}$ is characterized by a joint probability distribution $p(X_1 = x_1, ..., X_{N_S} = x_{N_S}) = p(x_1, ..., x_{N_S})$.

A. Distributed Source Coding

To quantify the performance of the aggregation scheme, we need to quantify the amount of information generated by the sources. Let $H(X_i)$ denote the entropy of observation X_i collected by sensor s_i, $H(X_i) = -\int p(X_i) \log_2 p(X_i) dX_i$. Once, the source nodes transmit enough information to the terminals, then they can recover the original data, losslessly. Intuitively, each source s_i can code its data at a rate R_i which is greater than or equal to their respective entropy, $R_i \geq H(X_i)$.

With S-W coding, the correlated data from different sensor nodes can be coded independently without communication with each other, $R_i \geq H(X_i|X_i^c)$, where X_i^c is the correlated data with X_i and gained by other sensors. The S-W region specifies the minimum encoding rate that the sensor nodes must meet in order to transmit all independent information to the sink nodes. It is satisfied when any subset of sensor nodes encode their collected data at a total rate exceeding their joint entropy.

However, the rate admissibility constraints require each sensor node to have global correlation information, which is not scalable in large networks. Here, a localized version of S-W coding is adapted to relax the rate admissibility constraints, such that only local correlation information is required at each sensor node. Here, we describe the localized S-W coding as:

$$R_{s_i}^{t_j} \geq H(X_{s_i}|N^{s_i}) \quad t_j \in T_i \tag{1}$$

where N^{s_i} is a subset of the sensors within the neighborhood of sensor s_i. For each node $i \in N^{s_i}$, i is the one-hop neighbor of s_i and can connect with s_i directly. Moreover, each $i \in N^{s_i}$ has the same sink node t_j as s_i.

B. Minimum Cut Capacity Constraints

The capacity region of the terminal t_j with respect to S_j as:

$$\begin{aligned} C_{t_j} = \quad & \{(R_1^{t_j}, ..., R_{N_S}^{t_j}) : \\ & \forall B \subseteq S_j, \textstyle\sum_{i \in B} R_i^{t_j} \leq min - cut(B, t_j)\}. \end{aligned} \tag{2}$$

Thus, C_{t_j} consists of inequalities that define the maximum flow from each subset of S_j to the terminal t_j. A rate vector $(R_1^{t_j}, ..., R_{N_S}^{t_j}) \in C_{t_j}$ can be transmitted from the source nodes $s_1, ..., s_{N_S}$ to terminal t_j.

Definition 1 (Traffic matrix) We call the traffic matrix A of a graph G, the elements of the $N_S \times N_R$ square matrix A is given by:

$$A_{ij} = \begin{cases} 1, & \text{if source i is transmitted to sink j,} \\ 0, & \text{else.} \end{cases} \tag{3}$$

With this notation, $T_i = \{j : A_{ij} = 1\}$ and $S_j = \{i : A_{ij} = 1\}$.

C. Channel Interference Constraint

Suppose any link originating from node $k \in V$ will interfere with link (i, j) if $d_{ki} < (1+\Delta)d_{ij}$ or $d_{kj} < (1+\Delta)d_{ij}$. Here, $\Delta > 0$ specifies the interference range. Also define $\Psi_{i,j}$ for each link (i, j) as the cluster of links that cannot transmit as long as link (i, j) is active, then the wireless network channel interference constraint can be defined as:

$$u_{ij} + \sum_{(p,q) \in \Psi_{ij}} u_{pq} \leq C, \quad j \in V_{out}(i) \tag{4}$$

D. Network Coding Constraints

With network coding, flows from different sources to different sinks are allowed to share the network capacity by coding together at the intermediate nodes. For any intermediate node i, if network coding is applied, the actual physical flow sent by node i to its next hop j is only to be the maximum of the received information flows. That is

$$u_{ij} = \begin{cases} \max\limits_{s,t} \; B_{ij}^{st} \cdot R_s^t & \text{if network coding is applied,} \\ \sum\limits_{s,t} B_{ij}^{st} \cdot R_s^t & \text{otherwise.} \end{cases} \tag{5}$$

where,B_{ij}^{st} is defined as the path matrix for source s by:

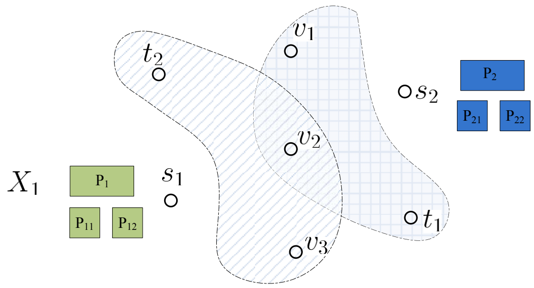

Fig. 1. Example of NC-aware OR for correlated sources

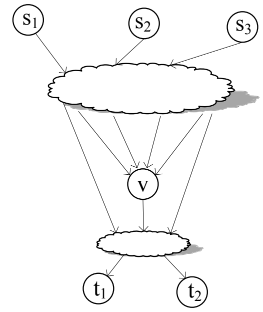

Fig. 2. Example of intermediate node gain data from different sources

$$B_{ij}^{st} = \begin{cases} 1, & \text{if source } s \text{ to sink } t \text{ occupied link } (i,j), \\ 0, & \text{else.} \end{cases} \tag{6}$$

III. Opportunistic Routing Scheme for Correlated Source

In this section, we consider the data gathering scenarios where an opportunistic routing mechanism joint with opportunistic network coding is applied to increase the network throughput and reduce the number of transmissions in wireless video sensor networks. According to the localized S-W region, each node can compress its data with the sensors belong to its one-hop neighbor set. During the transmission, sensor node attempts to broadcast its data packet. With opportunistic network coding, the packets can potentially be compressed again at the intermediate nodes, packetized, and forwarded to the corresponding sinks.

The proposed opportunistic routing scheme addresses two important issues: candidate selection and relay priority assignment. An example is present in Fig.1 to illustrate the opportunistic routing scheme. Both sources s_1 and s_2 want to send packets to its corresponding sink. The packets sent by node s_1 can be received by $\{v_2, v_3, t_2\}$, and the packets from s_2 can be received by $\{v_1, v_2, t_1\}$. With opportunistic network coding, the data from s_1 and s_2 can combined at node v_2. So, node v_2 is selected as the relay node for both the of sources. Then we calculate the expected coded transmission (ECTX) and select the node with a small ECTX opportunity as the forwarder in each packet transmission.

A. Forwarding Node Selection

When a node has packets to send, it first broadcasts a RTC to its neighbors which can be considered as the forwarding candidates. Each node in the candidate set calculates its ECTX and reports it to the sender with a CTS. With the information gained from its candidates, the sender can select a node as its relay node to transmit its packets.

In the previous routing schemes without network coding, the routing metric most commonly used is minimum hop count, such as expected transmission count (ETX). According to the ETX metric, for a single link (i, j) with the delivery probability $r_{ij}$the value of expected number of transmissions on the link is $1/r_{ij}$. With network coding, the expected number of coded transmissions is defined to capture the number of transmissions required to make the coded packet received successfully.

As shown in Fig.1, without network coding, node s_1 can gain two ETXs from its forwarding candidates, $\{\frac{1}{r_{s_1v_2}} + \frac{1}{r_{v_2t_1}}, \frac{1}{r_{s_1v_3}} + \frac{1}{r_{v_3t_1}}\}$. In order to reduce the number of transmission, the candidate with lower ETX can be chosen as the relay node by s_1. And s_2 can choose its relay node as the same as s_1. With network coding, node v_2 can gain the data from both s_1 and s_2. Then, the ECTX is presented as

$$(ECTX) = \frac{1}{\min_{l \in V_{in}(v)} r_{lv}} + \frac{1}{\min_{k \in V_{out}(v)} r_{vk}} \tag{7}$$

Comparing with the ETX, the number of transmissions for both sources reduced by network coding is denoted as $\min\{\frac{1}{r_{s_1v_2}r_{v_2t_1}}, \frac{1}{r_{s_1v_3}r_{v_3t_1}}\} + \min\{\frac{1}{r_{s_2v_2}r_{v_2t_2}}, \frac{1}{r_{s_2v_3}r_{v_3t_2}}\} - \frac{1}{\min\{r_{s_1v_2}r_{s_2v_2}\}} + \frac{1}{\min\{r_{v_2t_1}r_{v_2t_2}\}}$.

B. Opportunistic Network Coding

However, another problem we have to consider is which packets should be coded at the intermediate node. Once, there are several senders send their data simultaneously and the intermediate node can receive the data from different sources as shown in Fig.2. We assume that the intermediate node has several data buffers to sort the packets from different sources to different sink nodes. As shown in Fig.2, a relay node v has three buffers and can receive data from s_1, s_2, s_3.

We assume that sensor s_1 and s_2 delivery their data to sink t_1, s_2 and s_3 to sink t_2. Therefore, for sink node t_1, the set of source node $S_1 = \{s_1, s_2\}$, and $S_2 = \{s_2, s_3\}$ for sink t_2. And the traffic matrix is

$$A_{3\times 2} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

.

Random linear coding scheme can be applied at intermediate nodes. For each intermediate node, the packets on its outgoing links are random linear combinations of the packets on its incoming links. Therefore, node v has four probable combinations of data coding, as shown in table 1.

where ρ_v^i is defined as the network coding gain. Obviously, the relay node can choose the combination with the highest

TABLE I
THE CASES FOR INTERMEDIATE NODE V TO COMPRESS THE DATA

	$R^{t_1}_{s_1}$	$R^{t_1}_{s_2}$
$R^{t_2}_{s_2}$	ρ^1_v	ρ^2_v
$R^{t_2}_{s_3}$	ρ^3_v	ρ^4_v

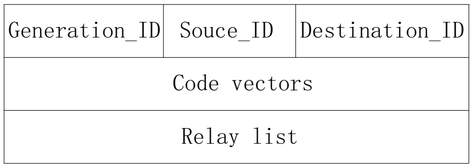

Fig. 3. The format of the packet header

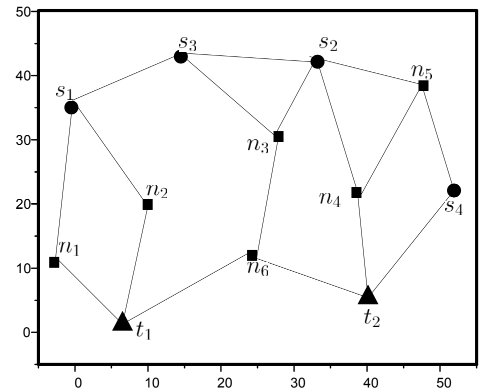

Fig. 4. The Grid topology for wireless sensor network

network coding gain, if the decoding is feasible at the sink node. Therefore, the coding gain of correlated sources can be defined as:

$$\rho_v = \max\{\rho^i_v\} = \max \quad \frac{\sum_{k\in V_{in}(v)} u_{kv} - \max u_{kv}}{\sum_{k\in V_{in}(v)} u_{kv}} \tag{8}$$

where $u_{kv} = \sum_{s_i\in S}\sum_{t_j\in T_i} R^{t_j}_{s_i} B^{s_i t_j}_{kv}$, for $k \in V_{in}(v)$.

With the coding policy, the relay node can combine the packets with high coding gain when several packets from different sources are received. The original source and destination IP address and the list of nodes which could participate in forwarding it, as depicted in Fig.3. The sender keeps transmitting the coded packets of the current generation until it is recovered at the destination, and then it moves to the next generation.

C. Priority of Forwarder

The priority for each candidate is to choose the node with low ECTX to reduce the network transmissions. The forwarding node can compress the data with a high coding gain to improve the end-to-end throughput. As defined in equation 8, the coding gain can be assigned to the intermediate node which predicts the throughput improvement of data transmission.

For each node v, a max-coding gain problem is formulated as following:

$$\textbf{P1:} \qquad \max_{s_i\in S, t_j\in T_i} \rho_v \tag{9}$$

$$\textbf{s.t.} \qquad (1), (2), (4), (5).$$

Each node s_i can choose a set of forwarding candidates with different priorities. The nodes within a source's forwarding candidate set agree on that the highest priority node keeps the packet and all the other nodes drop the packet to prevent unnecessary multiple forwarding of the same packet. If the packet is not received by any node in the source's candidate set, the source broadcasts the packet again until it is received by at least one node in the candidate set or the maximum number of trials is reached.

D. Complexity Analysis

Assume that a sensor node in the wireless network has an average degree D including both in-degree and out-degree, $(D > 1)$. The time for the node to decide which packets should be coded is $\mathbf{O}(D(D-1))$. Moreover, a data sender needs $\mathbf{O}(D^2)$ to select a relay node to forward its data to the sink node.

In the proposed routing mechanism, the time for constituting the optimal data gathering framework is $\mathbf{O}(D^2)$.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our routing scheme by simulation and compare the performance with the other exited routing protocols. As shown in Fig.4, we consider a wireless network with 10 sensor nodes and 2 sink nodes. The source set to sink t_1 is $S_1 = \{s_1, s_3\}$, and $S_2 = \{s_2, s_4\}$ takes t_2 as the sink node. As all the sources are identically distributed, we assume that $H(s_i) = 0.5Mbps$ and $H(s_1|N_{s_1}) = 0.3Mbps$, $H(s_2|N_{s_2}) = 0.3Mbps$, $H(s_3|N_{s_3}) = 0.2Mbps$. Also, all the links have a static capacity $1Mbps$.

First, suppose that all the packets can be received correctly. For SPT scheme, each source node transmits its data along the shortest path to it sink, although it has to compete the capacity of some link with other sources. As shown in Fig.5, link (n_3, n_6) is occupied by s_2 and s_3, simultaneously. Therefore, s_2 and s_3 have a low end-to-end throughput. With network coding, the data from s_2 and s_3 can be combined at node n_3 and the capacity of link (n_3, n_6) can be shared by both sources. The coding is applied at the bottleneck link when it is necessary and passive with NC-based scheme. With the opportunistic routing joint with ONC scheme, the data sender takes into account the availability of coding opportunities during route selection and increases the throughput positively. The data rate for sink node t_2 can gain its bound of min-cut.

To evaluate the performance of opportunistic routing in lossy wireless environments, we assume that each link has a packet reception rate randomly chosen in the range of $[0, 1]$. The end-to-end throughput for each source in lossy wireless network is illustrated in Fig.6. It is seen that the opportunistic routing joint with ONC scheme can significantly improve the

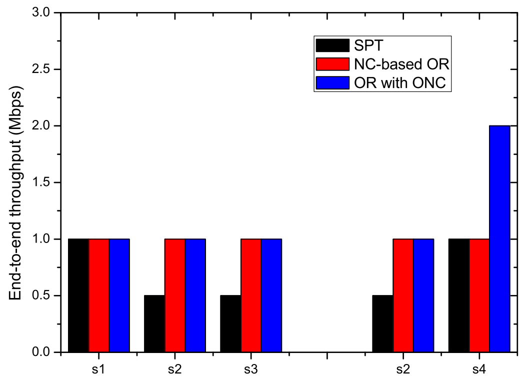

Fig. 5. The end-to-end throughput for each source with different routing scheme

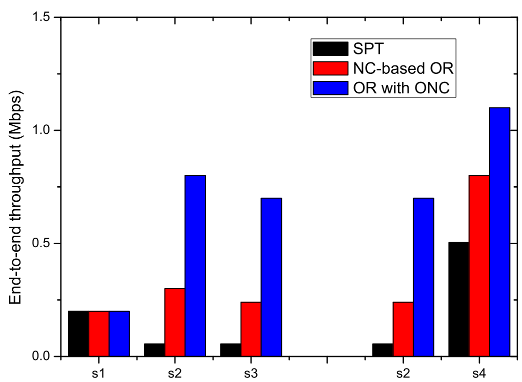

Fig. 6. The end-to-end throughput for each source within lossy wireless network

end-to-end throughput as compared with the other routing schemes.

We assume that the flow for each source has a packet rate ranging from $0.5Mbps$ to $1.5Mbps$. Clearly, in order to recover the original data losslessly, the average data rate must be no less than its entropy. Fig.7 shows the network throughput with SPT, NC-based OR and OR joint with ONC, respectively.

When the offered load is small, it is hard to create more coding opportunities. As a result, OR joint with ONC scheme performs similarly to SPT and NC-based OR. As the offered load increases, OR joint with ONC and NC-based OR achieve higher network throughput than SPT due to the introduction of network coding operations, which help reduce the number of transmissions. Furthermore, OR joint with ONC outperforms NC-based OR as it can create much more coding opportunities through opportunistic forwarding. With the offered load increasing further, the network throughputs with OR joint with ONC and NC-based OR become close again because in this case paths for different flows become more likely to overlap.

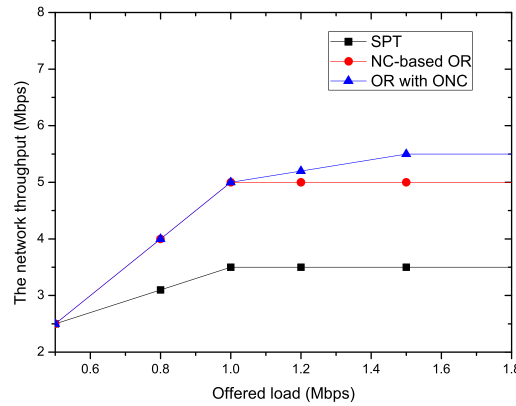

Fig. 7. The network throughput under different offered load

V. Conclusion

The problem of correlated data gathering in wireless sensor networks is considered in this paper. In order to increase the network throughput and face the highly unpredictable wireless links, an opportunistic routing protocol joint with opportunistic network coding is proposed. We consider both the number of coded transmissions and the network coding opportunistic to decrease the total number of transmissions. With numerical simulations performance analysis are illustrated and good results are obtained using the proposed scheme.

VI. Acknowledgment

The work has been partially supported by the NSFC grants No. 61271211, No. 60632040, No. 60802019, and No. 60736043.

References

[1] Y. He, I. Lee, L. Guan, "Network lifetime maximization in wireless visual sensor networks using a distributed algorithm," *In: Proceedings of the IEEE International Conference on Multimedia and Expo*, 2007, pp. 2174C2177.

[2] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, Muriel, J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," *IEEE/ACM Transactions on Networking,* June 2008, pp. 497-510.

[3] W. Chen, K. Letaief, Z. Cao, "Opportunistic network coding for wireless networks," *In: Proc. IEEE ICC'07,* June 2007, pp. 4634-4639.

[4] T. Cui, L. Chen, T. Ho, "Energy efficient opportunistic network coding for wireless networks," *In: Proc. INFOCOM'08,* Apr. 2008, pp. 1022-1030.

[5] S. Biswas, R. Morris, "Opportunistic routing in multi-hop wireless networks," *In: Proc. ACM SIGCOMM'05*, Aug. 2005.

[6] S. Chachulski, M. Jenning, S. Katti, D. Katabi, "Trading structure for randomness in wireless opportunistic routing," *In: Proc. ACM SIGCOMM'07*, pp.169-180, Aug. 2007.

[7] Y. Yan, B. Zhang, J. Zheng, J. Ma, "CORE: A coding-aware opportunistic routing mechanism for wireless mesh networks," *IEEE Wireless Communications,* vol.17, no.3, June 2010, pp.96-103.

[8] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: the case for coding-aware routing," *In: Proc. IEEE INFOCOM'07,* May 2007, pp. 1028-1036.

[9] B. Ni, N. Santhapuri, Z. Zhong, S. Nelakuditi, Routing with opportunistically coded exchanges in wireless mesh networks," *In: Proc. IEEE WiMesh'06,* Sept. 2006, pp. 157-59.