# How Network Coding Benefits Converge-Cast in Wireless Sensor Networks

**Zhenzhou Tang[1, 2], Hongyu Wang[1], Qian Hu[2] and Long Hai[1]**
[1] School of Information and Communication Engineering, Dalian University of Technology
Dalian, Liaoning - China
[e-mail: mr.tangzz@gmail.com]
[2] College of Physics and Electronic Information Engineering, Wenzhou University,
Wenzhou, Zhejiang - China
*Corresponding author: Zhenzhou Tang

*Received November 18, 2012; revised January 23, 2013; revised March 30, 2013; accepted April 25, 2013; published May 31, 2013*

## Abstract

Network coding is one of the most promising techniques to increase the reliability and reduce the energy consumption for wireless sensor networks (WSNs). However, most of the previous works mainly focus on the network coding for multicast or unicast in WSNs, in spite of the fact that the converge-cast is the most common communication style in WSNs. In this paper, we investigate, for the first time as far as we know, the feasibility of acquiring network coding benefits in converge-cast, and we present that with the ubiquitous convergent structures self-organized during converge-casting in the network, the reliability benefits can be obtained by applying linear network coding. We theoretically derive the network coding benefits obtained in a general convergent structure, and simulations are conducted to validate our theoretical analysis. The results reveal that the network coding can improve the network reliability considerably, and hence reduce number of retransmissions and improve energy-efficiency.

***Keywords:*** Network coding, wireless sensor networks, converge-cast

A preliminary version of this paper appeared in IEEE VTC 2012, Sept. 3-6, Quebec City, Quebec, Canada. This version includes a concrete analysis and supporting implementation results based on Network Simulator II. This work is supported by the Zhejiang Provincial Natural Science Foundation of China (LQ12F02009), the Nonprofit Technology Application Research Projects of Zhejiang Province (2011C310290029) and the Scientific Research Fund of Zhejiang Provincial Education Department (Y201121034).

**http://dx.doi.org/10.3837/tiis.2013.05.014**

## 1. Introduction

Explosive growth in embedded computing and rapid advances in low power wireless networking technologies are fueling the development of wireless sensor networks (WSNs). And WSNs have been attracting a great attention due to their wide range of potential applications. However, sensor nodes are generally powered by batteries which only provide a limited amount of power, and it is often difficult and costly to recharge or replace the batteries. Moreover, the lossy wireless links which often lead to packet delivery failures further accelerate the energy consumption due to frequent retransmissions, as we all known that data transmission dominates the energy consumption in wireless sensor networks. Therefore, reliability and energy-efficiency are essential for wireless sensor networks.

Network coding is a technique that can increase network throughput and reduce energy consumption by combining data from different input links in an intermediate node [1, 2]. Many applications have incorporated this idea since its initial proposal by Ahlswede [1]. Network coding is particularly well-suited for wireless networks due to the broadcast nature of their communications [3, 4]. It is precisely because of so many advantages of wireless network coding, that the applications of network coding in WSNs are intensively studied in recent years [5-12].

However, by reviewing the related literatures, we have found that the issue of network coding in wireless converge-cast still remains open. The consideration on network coding in converge-cast is stimulated by the special way of how a WSN typically works. Environmental data generated by the sensor nodes scattered in the monitoring regions are gathered towards the sink node hop by hop, as shown in **Fig. 1**. Since the network engages to collect data from the sensor nodes for most of the time, *the most common communication style in WSNs should be converge-cast*. However, to the best of our knowledge, few works have been done on it. To this end, we investigate how the network coding benefits the wireless sensor networks during converge-casting, and how to obtain the network coding benefits as many as possible in this paper.

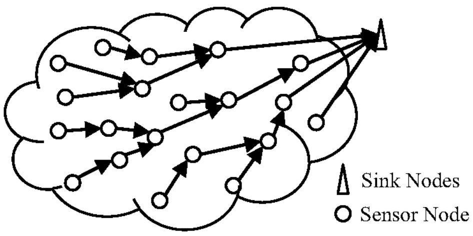



**Fig. 1.** Converge-cast in wireless sensor networks

The main contributes of our work are summarized as follows:

1) In this paper, we investigate, for the first time as far as we know, the feasibility of acquiring network coding benefits in converge-cast. We reveal that with the ubiquitous convergent structures self-organized during converge-casting in the network, the reliability benefits can be obtained by applying linear network coding.
2) We define the typical convergent structures where to perform network coding. And on the basis of that, we also derive the theoretical expressions to calculate the reliability benefits in a general convergent structure.

3) We perform different simulations to evaluate the network coding benefits. The results show that network coding is able to bring about reliability benefits. And with network coding, converge-cast in WSN costs much fewer retransmissions and can reduce energy consumption considerably.

The rest of this paper is organized as follows. In Section II, we discuss related work. Section III investigates the feasibility of acquiring network coding benefits in converge-cast. The suitable network coding style and the source of network coding benefits in converge-cast are discussed. In Section IV, the linear network coding model is proposed. And also in this section, the amount of reliability benefits is derived. The numerical and simulation results are presented in Section V. Further discussion is presented in Section VI, followed by Section VII giving the conclusions.

## 2. Related Work

The famous work in [1] introduced the concept of network coding with the objective of improving the throughput of a multicast traffic flow in a wired network. Instead of only forwarding the incoming packets, the essence of network coding is a paradigm shift to allow coding. Thus, network coding can achieve the maximum information flow of a multicast network. Koetter [13] and Li. [14] studied the Linear Network Coding (LNC), and they proved that a linear combination of packets can also achieve the maximum throughput of the multicast network. And subsequently in [15], Ho. et al extended the LNC and proposed the novel idea of Random Linear Network Coding (RLNC). They showed that with the coefficients randomly taken from a sufficient large finite field $\mathbb{F}_q$, the probability that all the receivers can decode the source processes is considerably high.

Due to the broadcast nature of wireless media, network coding was introduced to wireless network naturally. Katti et al. proposed COPE, which is a network coding aware unicast routing for wireless mesh networks [3]. It is the first work on the system architecture for wireless network coding, and is also the first deployment of network coding in a wireless network. And Chachulski et al. proposed a MAC-independent LNC-based opportunistic routing called MORE (MAC-independent Opportunistic Routing & Encoding) [4]. COPE and MORE are prominent representatives of two different types of network coding schemes respectively, i.e. *inter-flow* network coding (Inter-NC) and *intra-flow* network coding (Intra-NC) [16]. For Inter-NC, coding is done over packets from different flows and towards different next-hop nodes [3, 17, 18]. And Intra-NC schemes typically combine multiple packets heading to the same destination into a single coded-packet which should be decoded at that common destination [4, 19].

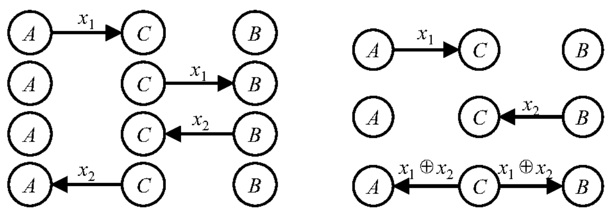
Wireless network coding is able to introduce benefits on throughput, reliability, and energy-efficiency [16], which greatly inspires the researches on applying network coding into WSNs. These works mainly involve network coding aided multi-cast and unicast [5-12], and the specific applications include online reprogramming and data dissemination [5-8], distributed data storage and acquisition [9, 10], and reliable data transmission [11, 12], etc. However, work on network-coding-aided converge-cast can seldom be found, although converge-cast is the most common communication style in WSNs. Current converge-cast in WSNs typically employs an inverted multicast tree [20-22], and packets are still simply forwarded to the parents without encoding.

# 3. Feasibility of Acquiring Network Coding Benefits in Converge-cast

In this section, we discuss about the feasibility of acquiring network coding benefits in converge-cast, that is, which kind of benefits can be produced in converge-cast.

## 3.1. Inter-flow Network Coding

In inter-flow network coding, the router, in which the packets from different flows are combined (coded), broadcasts the newly generated packet to different neighbors in a single transmission. Inter-flow network coding can help improve throughput because coding allows the routers to compress the transmitted information given what is known at various nodes. Consider the example in **Fig. 2(a)**, where node *A* and *B* attempt to exchange a pair of packets. The limited transmission range prevents them from communicating directly and thus they have to turn to node *C* for forwarding. In the traditional store-and-forward scheme, four transmissions are required. While in case of network coding, node *A* and *B* could transmit their respective packets to node *C*, which XORs (exclusive-ORs) the two packets and broadcasts the coded packet. Thus, node *A* can recover node *B*'s packet by reXORing with its own packet, and so does node *B*. In this case, the exchange completes only within three transmissions. Hence, the network throughput is improved.




(a) Traditional scheme needs four transmissions

(b) Inter-flow network coding scheme only needs three transmissions

**Fig. 2.** A simple example of how network coding can improve throughput

Let $O(C, f)$ denote the set of nodes who can overhear the flow *f* and are known by node *C*. And $D(C, f)$ denotes the set of all downstream nodes of node *C* on flow *f*. Generally, supposing the flow *f* intersects with $f_x$ at node *C*, *C* is a coding node if and only if both the following conditions are satisfied:

$$\begin{cases} \exists d_1 \in O(C, f) \cap D(C, f_x) \\ \exists d_2 \in O(C, f_x) \cap D(C, f) \end{cases} \tag{1}$$

where $d_1$ and $d_2$ are two nodes in the network. However, as shown in **Fig. 1**, so long as different flows arrive at the same node in case of converge-cast, their following paths come to the same. Therefore, neither of these two conditions in (1) can be satisfied. In other words, inter-flow network coding cannot benefit converge-cast unfortunately.

## 3.2. Intra-flow Network Coding

Intra-flow network coding allows the intermediate nodes to combine the packets heading to the same destination. The encode operation is merely a simple random linear combination of *G* original packets as

$$X_k = \sum_{i=1}^{G} g_k^i M_i \quad (2)$$

where $M_i$ ($i$ = 1, 2, ... ,$G$) are the original packets, $g_k^i$ are the coefficients randomly selected from a $q$-order Galois field $\mathbb{GF}(q)$, and $X_k$ is the coded packet. The destination node can recover all the original packets if it has received no less than $G$ linear independent encoded packets of $M_i$ [2].

In the presence of intra-flow network coding, each received packet contains some information about all the original packets. No packet is specific for both the sender and the receiver, and no specific packet is indispensable. As a result, unlike the traditional automatic repeat request (ARQ), where the sender needs to know exactly which packets the destination misses so that it can retransmit them, the sender does not need to learn which particular packets the destination misses, and it only needs to get an ACK from the destination once it has received enough packets to decode and recover all the original packets. By this way, intra-flow network coding is able to dramatically reduce the bandwidth consumed by those feedbacks.

We are delighted to discover that network coding in converge-cast shares certain characteristics with intra-flow schemes. Specifically, in intra-flow network coding, all the coded packets should finally be decoded in the same specific node. And similarly, packets sent by various nodes during converge-casting are tend to be joined up with each other in certain intermediate nodes or the sink node, as shown in **Fig. 1**. That means those converging nodes have the opportunities to recover all the original packets if those packets have been coded elsewhere. Hence, in converge-cast, benefits can be obtained by exploiting the ubiquitous convergent structures self-organized during converge-casting in the network.

### 3.3. The Source of Benefits

Before discussing about the source of network coding benefits in converge-cast, we first introduce some terminologies which will be used throughout the paper.

*Definition 1*: The *converge-tree* $\mathbb{T}$ is a directed tree which is a collection of nodes connected to the sink and paths from those nodes to the sink in the network. The $\mathbb{T}$ is organized by non-network-coding data collection protocols, such as Collection Tree Protocol (CTP) [20]. The root of $\mathbb{T}$ is just the sink node.

**Fig. 1** shows a converge-tree.

*Definition 2*: An $n$-order *converge-structure* $\mathbb{CS}_n$ is a three-layer subtree of the $\mathbb{T}$ composed of $n$ leaves $S_i$ ($i$ = 1, 2, …, $n$), $n$ interior nodes $C_i$ ($i$ = 1, 2, …, $n$), and a root $D$, as shown in **Fig. 3** (a). A $\mathbb{CS}_n$ possesses the following characteristics:

1) There is a one-to-one correspondence between the set of leaves and the set of interior nodes. Assume that $S_i$ is the only child of $C_i$.
2) $C_i$ ($i$ = 1, 2, …, $n$) can overhear the communications of $S_{i-1}$ and $S_{i+1}$. While $C_1$ and $C_n$ can only overhear the communications of $S_2$ and $S_{n-1}$ respectively.
3) A complete process of linear network coding can be performed in the $\mathbb{CS}_n$. The leaves are source nodes, the interior nodes are encoders and routers, and the root is the decoder. And the $\mathbb{CS}_2$ is the most basic one to perform network coding, as shown in **Fig. 3(b)**.

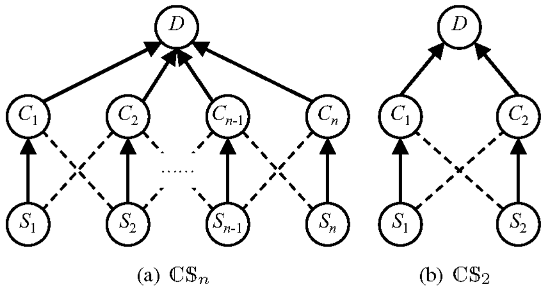



**Fig. 3.** Converge-structures. The nodes and solid lines with arrows form a subtree of converge-tree. The dashed lines represent the overhearing links.

During converge-casting in WSNs, the probability of forming $\mathbb{CS}_n$ is considerably high, especially for low order $\mathbb{CS}_n$. And the computational complexity of performing network coding in $\mathbb{CS}_n$ is low. A encoder only need to combine no more than three original packets.

Next, we explain how the network coding benefits converge-cast. Consider the simple example illustrated in **Fig. 4** (a). It is a $\mathbb{CS}_2$. The packets $x_1$ and $x_2$ are to be joined up with each other in $D$. However, the link from $S_2$ to $C_2$ ($\mathcal{L}_{S_2 \to C_2}$) fails. Evidently, in case of non-network-coding schemes, at least one retransmission is required to ensure the success of packet delivery. But it is different if the idea of network coding is employed. Since $C_1$ can overhear $x_2$, it would combine $x_1$ and $x_2$ by means of XOR or linear network coding [2] (in fact, XOR is a special kind of linear network coding scheme with $\mathbb{GF}(2)$). Assume the coded packet is $(x_1 \oplus x_2)$. And likewise, $C_2$ can overhear $x_1$ and just only forward it to $D$ for it only get one packet. Finally, $D$ is able to recover $x_2$ by reXOR the coded packet with $x_1$, as shown in **Fig. 4** (b). Hence, network coding in converge-cast is able to reduce the probability of retransmission. That is the benefit!

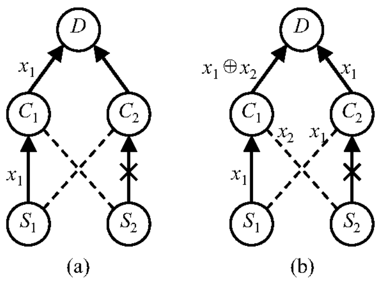



**Fig. 4.** The reliability benefits introduced by employing network coding in converge-structure.

## 4. Quantitative Analysis of Network Coding Benefits in Converge-cast

In this section, we study how many benefits can be obtained by applying network coding to converge-cast in WSNs. We first introduce the typical network coding model, and then we derive the theoretical benefits introduced by network coding.

### 4.1. The Network Coding Model

A $\mathbb{CS}_n$ can be represented by a matrix defined as:

*Definition 3*: A *n*-order *converge-matrix* $\mathbf{M}_n$ is mapped from $\mathbb{CS}_n$. The element of $\mathbf{M}_n$, denoted as $c_{i,j}$ ($i, j = 1, 2, ..., n$), represents the coefficient of the original packet sent by $S_j$ when performing linear network coding in $C_i$. The coefficients are chosen over all elements of a

Galois Field $\mathbb{GF}(q)$. $\mathbf{M}_n(i,j)=0$ if the corresponding $\mathcal{L}_{S_j \to C_i}$ fails. And if all the links work, the $\mathbf{M}_n$, which is called as the perfect $\mathbf{M}_n$, should be:

$$\mathbf{M}_n = \begin{pmatrix} c_{1,1} & c_{1,2} & 0 & \cdots & 0 & 0 & 0 \\ c_{2,1} & c_{2,2} & c_{2,3} & \cdots & 0 & 0 & 0 \\ 0 & c_{3,2} & c_{3,3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_{n-2,n-2} & c_{n-2,n-1} & 0 \\ 0 & 0 & 0 & \cdots & c_{n-1,n-2} & c_{n-1,n-1} & c_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & c_{n,n-1} & c_{n,n} \end{pmatrix}$$

The rank of $\mathbf{M}_n$ is denoted as $\mathbb{R}(\mathbf{M}_n)$.

Hence, let $\mathbf{X}_n=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$ be the vector of original packets. The vector of coded packets, which is denoted as $\mathbf{Y}_n=(y_1,y_2,\ldots,y_n)^{\mathrm{T}}$, can be calculated as $\mathbf{Y}_n=\mathbf{M}_n\cdot\mathbf{X}_n$. And *D* is able to recover all the original packets with $\mathbf{X}_n=\mathbf{M}_n^{-1}\cdot\mathbf{Y}_n$. Obviously, this equation has a unique solution if and only if $\mathbb{R}(\mathbf{M}_n)=n$.

Taking into consideration the limited computing capacity of a sensor node, it is reasonable to assign predetermined values to these coefficients. Since a coded packet includes no more than three original ones, for example, $\mathbb{GF}(2^3)$ is adopted. Hence, the coefficients should be chosen from (0, 1, $\alpha$, $\alpha^3$) where $\alpha$ is the primitive element of $\mathbb{GF}(2^3)$. And the binary and decimal notations are (000, 001, 010, 011) and (0, 1, 2, 3) respectively as follows.

$$\mathbf{M}_n = \begin{pmatrix} \alpha & \alpha^3 & 0 & \cdots & 0 & 0 & 0 \\ 1 & \alpha & \alpha^3 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \alpha & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha & \alpha^3 & 0 \\ 0 & 0 & 0 & \cdots & 1 & \alpha & \alpha^3 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \alpha \end{pmatrix} \xlongequal{\text{decimal notation}} \begin{pmatrix} 2 & 3 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 2 & 3 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & 3 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 2 & 3 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 2 \end{pmatrix} \tag{3}$$

It is simple to prove that the perfect $\mathbb{R}(\mathbf{M}_n)=n$.

### 4.2. Quantitative Analysis

*Theorem* 1: For a feasible converge-cast in a $\mathbb{CS}_n$ with lossy links and a network coding in which the coefficients are determined by the corresponding $\mathbf{M}_n$, assuming that all the links share the same delivery ratio *r*, the probability $p_n$ that the decoder *D* can recover all the original packets can be calculated as:

$$p_n = \begin{cases} r^4(2-r^2) & \text{if } n=2 \\ p_{n-1}(r+r^2-r^3)r+(p_{n-2}-p_{n-3}r^3)r^4(1-r) & \text{if } n \geqslant 3 \end{cases} \tag{4}$$

Particularly, $p_0 \triangleq 1$ and $p_1 \triangleq r^2$.

*Proof*: First, $p_2$ can be calculated easily with the exhaust algorithm.

$$p_2 = \left[2r^2(1-r)^2 + \binom{3}{4} r^3(1-r) + r^4\right] r^2 = r^4(2-r^2) \tag{5}$$

Then we calculate $p_n (n \geqslant 3)$. *D* can recover all the *n* original packets if and only if $\mathbb{R}(\mathbf{M}_n) = n$ and all the links from the encoders to the decoder are successful. In other words, $p_n = p[\mathbb{R}(\mathbf{M}_n)]r^n$ where $p[\mathbb{R}(\mathbf{M}_n)]$ is the probability of $\mathbb{R}(\mathbf{M}_n) = n$. According to the total probability formula, we have:

$$p[\mathbb{R}(\mathbf{M}_n)] = (p_n^1 + p_n^2) \tag{6}$$

$$p_n^1 = p[\mathbb{R}(\mathbf{M}_{n-1})] \cdot p[\mathbb{R}(\mathbf{M}_n) \mid \mathbb{R}(\mathbf{M}_{n-1})] \tag{7}$$

$$p_n^2 = p[\overline{\mathbb{R}(\mathbf{M}_{n-1})}] \cdot p[\mathbb{R}(\mathbf{M}_n) \mid \overline{\mathbb{R}(\mathbf{M}_{n-1})}] \tag{8}$$

where $p[\overline{\mathbb{R}(\mathbf{M}_n)}] \triangleq p[\mathbb{R}(\mathbf{M}_n) < n]$.

In case that $\mathbf{M}_{n-1}$ is a full-rank matrix, $\mathbf{M}_n$ is also full rank if either of the following conditions is satisfied:

1) $\mathcal{L}_{S_n \to C_n}$ works. Or
2) $\mathcal{L}_{S_n \to C_n}$ fails, but both $\mathcal{L}_{S_{n-1} \to C_n}$ and $\mathcal{L}_{S_n \to C_{n-1}}$ are successful. Moreover, $\mathbf{M}_{n-2}$ must be a full-rank matrix.

Hence, we have

$$\begin{aligned} p_n^1 &= p[\mathbb{R}(\mathbf{M}_{n-1})] \cdot p[\mathbb{R}(\mathbf{M}_n) \mid \mathbb{R}(\mathbf{M}_{n-1})] \\ &= p[\mathbb{R}(\mathbf{M}_{n-1})]r + p[\mathbb{R}(\mathbf{M}_{n-1})\mathbb{R}(\mathbf{M}_{n-2})]r^2(1-r) \end{aligned} \tag{9}$$

And because

$$\begin{aligned} & p[\mathbb{R}(\mathbf{M}_{n-1})\mathbb{R}(\mathbf{M}_{n-2})] = p[\mathbb{R}(\mathbf{M}_{n-1})(\Omega - \overline{\mathbb{R}(\mathbf{M}_{n-2})})] \\ & = p[\mathbb{R}(\mathbf{M}_{n-1})] - p[\mathbb{R}(\mathbf{M}_{n-1})\overline{\mathbb{R}(\mathbf{M}_{n-2})}] \end{aligned} \tag{10}$$

therefore

$$p_n^1 = p[\mathbb{R}(\mathbf{M}_{n-1})](r + r^2 - r^3) - p[\mathbb{R}(\mathbf{M}_{n-1}) \cdot \overline{\mathbb{R}(\mathbf{M}_{n-2})}]r^2(1-r) \tag{11}$$

Moreover, $\mathbf{M}_n$ must be a singular matrix if $\mathbb{R}(\mathbf{M}_{n-1}) < n-2$. Assuming that $\mathbb{R}(\mathbf{M}_{n-1}) < n-1$, $\mathbf{M}_n$ is full rank if both of the following conditions are satisfied:

1) $\mathbf{M}_{n-2}$ is a full-rank matrix. And
2) Both $\mathcal{L}_{S_{n-1} \to C_n}$ and are successful.

Therefore, we have

$$\begin{aligned} p_n^2 &= p[\overline{\mathbb{R}(\mathbf{M}_{n-1})}] \cdot p[\mathbb{R}(\mathbf{M}_n) \mid \overline{\mathbb{R}(\mathbf{M}_{n-1})}] \\ &= p[\mathbb{R}(\mathbf{M}_n)\overline{\mathbb{R}(\mathbf{M}_{n-1})}] = p[\overline{\mathbb{R}(\mathbf{M}_{n-1})}\mathbb{R}(\mathbf{M}_{n-2})]r^2 \end{aligned} \tag{12}$$

By utilizing (12), it can be inferred that

$$p[\mathbb{R}(\mathbf{M}_{n-1})\overline{\mathbb{R}(\mathbf{M}_{n-2})}] = p[\overline{\mathbb{R}(\mathbf{M}_{n-2})}\mathbb{R}(\mathbf{M}_{n-3})]r^2 \tag{13}$$

Substitute (11), (12) and (13) into (9), we have

$$\begin{aligned} p_n = p[\mathbb{R}(\mathbf{M}_{n-1})](r + r^2 - r^3) + p[\overline{\mathbb{R}(\mathbf{M}_{n-1})}\mathbb{R}(\mathbf{M}_{n-2})]r^2 \\ - p[\overline{\mathbb{R}(\mathbf{M}_{n-2})}\mathbb{R}(\mathbf{M}_{n-3})]r^4(1-r) \end{aligned} \tag{14}$$

Because

$$p[\overline{\mathbb{R}(\mathbf{M}_{n-1})}\mathbb{R}(\mathbf{M}_{n-2})] = p[\mathbb{R}(\mathbf{M}_{n-2})] - p[\mathbb{R}(\mathbf{M}_{n-1})\mathbb{R}(\mathbf{M}_{n-2})] \tag{15}$$

$$p[\overline{\mathbb{R}(\mathbf{M}_{n-2})}\mathbb{R}(\mathbf{M}_{n-3})] = p[\mathbb{R}(\mathbf{M}_{n-3})] - p[\mathbb{R}(\mathbf{M}_{n-2})\mathbb{R}(\mathbf{M}_{n-3})] \tag{16}$$

And according to (9), we have

$$p[\mathbb{R}(\mathbf{M}_{n-1})\mathbb{R}(\mathbf{M}_{n-2})] = p[\mathbb{R}(\mathbf{M}_{n-2})]r + p[\mathbb{R}(\mathbf{M}_{n-2})\mathbb{R}(\mathbf{M}_{n-3})]r^2(1-r) \tag{17}$$

Substitute (15), (16) and (17) into (14), and simplify it, we have

$$\begin{aligned} p[\mathbb{R}(\mathbf{M}_n)] =& p[\mathbb{R}(\mathbf{M}_{n-1})](r + r^2 - r^3) \\ &+ p[\mathbb{R}(\mathbf{M}_{n-2})]r^2(1-r) - p[\mathbb{R}(\mathbf{M}_{n-3})]r^4(1-r) \end{aligned} \tag{18}$$

That is:

$$p_n = p_{n-1}(r + r^2 - r^3)r + (p_{n-2} - p_{n-3}r^3)r^4(1-r) \tag{19}$$

The theorem is proved.

The probability that all the packets are successfully received by *D* without network coding, denoted as $p'_n$, can be calculated as

$$p'_n = r^{2n} \tag{20}$$

And the reliability benefit *B* introduced by network coding in $\mathbb{CS}_n$ is

$$B = \frac{p_n - p'_n}{p'_n} \tag{21}$$

### 4.3. Numerical Results

**Fig. 5** presents the comparisons between $p_n$ and $p'_n$ in $\mathbb{CS}_n (n = 2, 3, 4)$. $p_n$ and $p'_n$ are calculated from (19) and (20) accordingly, and they can be interpreted as packet collection rates with and without network coding. The link delivery ratio $r$ is taken from 0 to 1. It can be observed in **Fig. 5** that in case of employing network coding, the packet collection rates always surpass those of non-network-coding schemes under various link delivery ratios. Network coding improves the reliabilities dramatically. **Fig. 6** illustrates the reliability benefits (*B*) introduced by network coding in $\mathbb{CS}_n (n = 2, 3, 4, 5, 6)$. We can see that the benefits are proportional to the orders of converge structures for a given link delivery ratio, and meanwhile, are inversely proportional to the link delivery ratio for a given converge structure, which can be deduced from (21).

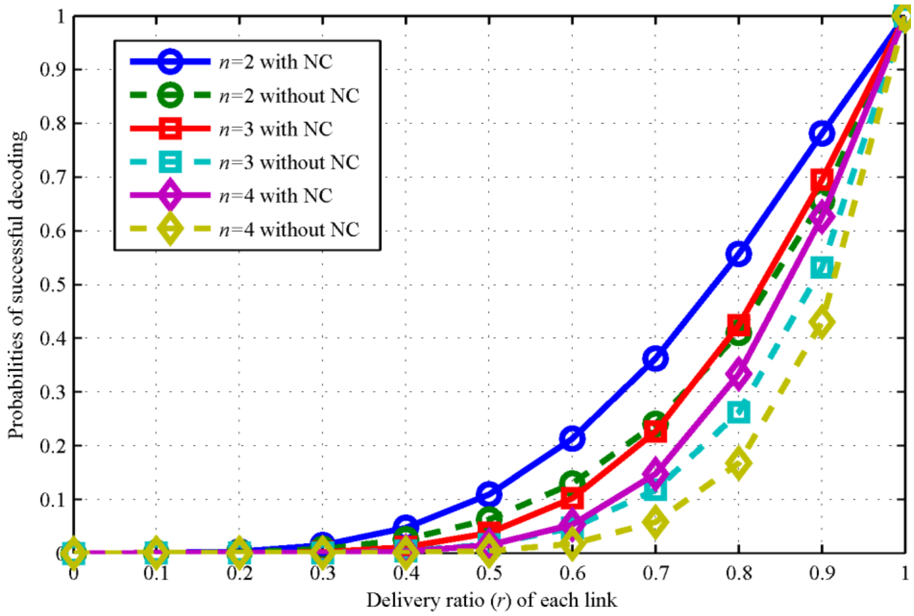



**Fig. 5.** Numerical results for $p_n$ and $p'_n$ in $\mathbb{CS}_n(n = 2, 3, 4)$.

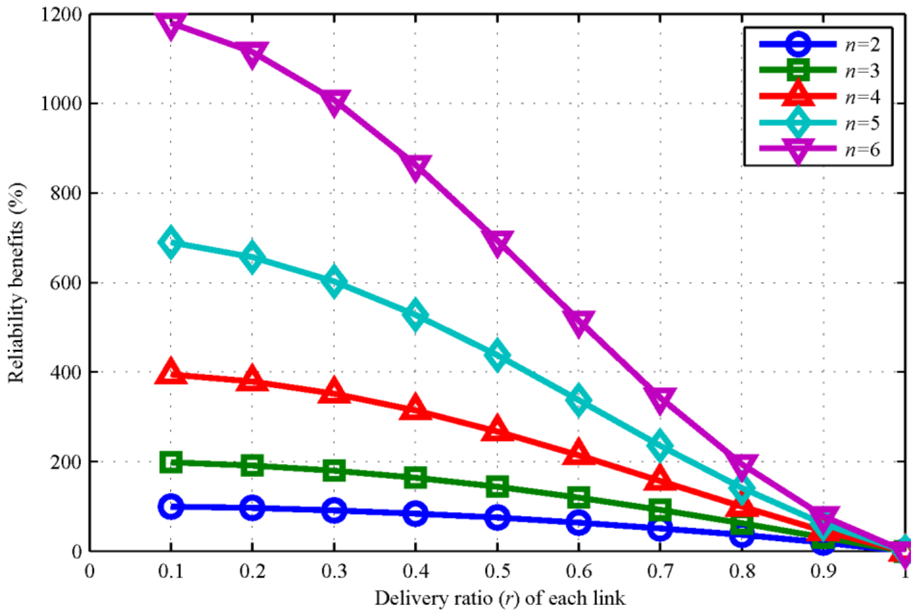



**Fig. 6.** Reliability benefits of employing network coding in $\mathbb{CS}_n(n = 2, ..., 6)$

## 5. Simulation Results

We have implemented the network-coding-aided converge-cast in a $\mathbb{CS}_n$ on Network Simulator II (NS2). The topologies are shown in **Fig. 7**. To ensure that the decoder is reachable for all coders, the distances between adjacent coders or source nodes in $\mathbb{CS}_4$ are closer than those in $\mathbb{CS}_2$ and $\mathbb{CS}_3$. Since our primary objective is to evaluate network coding benefits, we simply assign the converge-tree manually as shown in **Fig. 3 (b)**. The coefficients of the coded packets are predetermined according to (3). Each source node generates CBR packets destined to $D$. And the packet size is defined as 1000 bits. All the converge-cast schemes in our simulations are based on a simple Time Division Multiple Access (TDMA) Medium Access Control (MAC) protocol. For simplicity, TDMA schedules with global time synchronization are employed. In detail, the first $n$ slots of a TDMA frame are allocated to the source nodes, the next $n$ slots are allocated to the coders, and the last slot is assigned for the decoder. And in the simulations, no mobility is assumed.

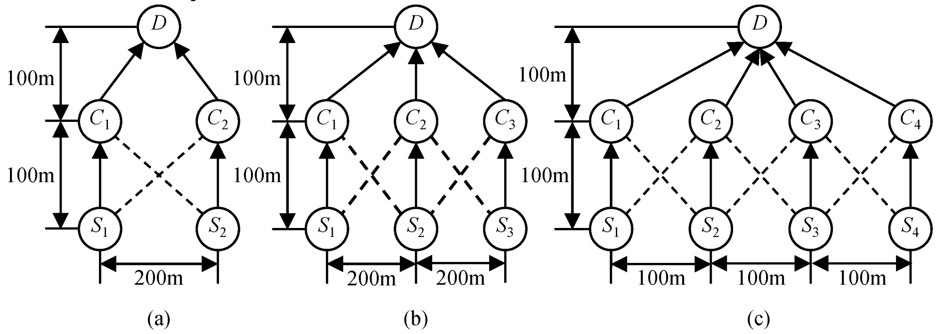



**Fig. 7.** network topologies for simulations, (a) for $\mathbb{CS}_2$, (b) for $\mathbb{CS}_3$, (c) for $\mathbb{CS}_4$,

The objective of the first set of simulations is to validate Theorem 1. The experiment is very simple: Each source node simply sends out 100 CBR packets; and the coders mix what they have received and forward them to the decoder. No ACK mechanism is applied, that is, a sender does not care about whether the packet is successfully received by the destination or not. Altogether 900 experiments are done to count the packet collection rates with and without network coding under different link delivery ratios. Moreover, the theoretical results are obtained according to (19) and (20). The comparisons between the theoretical and simulation results are shown in **Fig. 8**, where we can find that the theoretical and simulational results are almost perfectly matched, thus verifying our analysis.

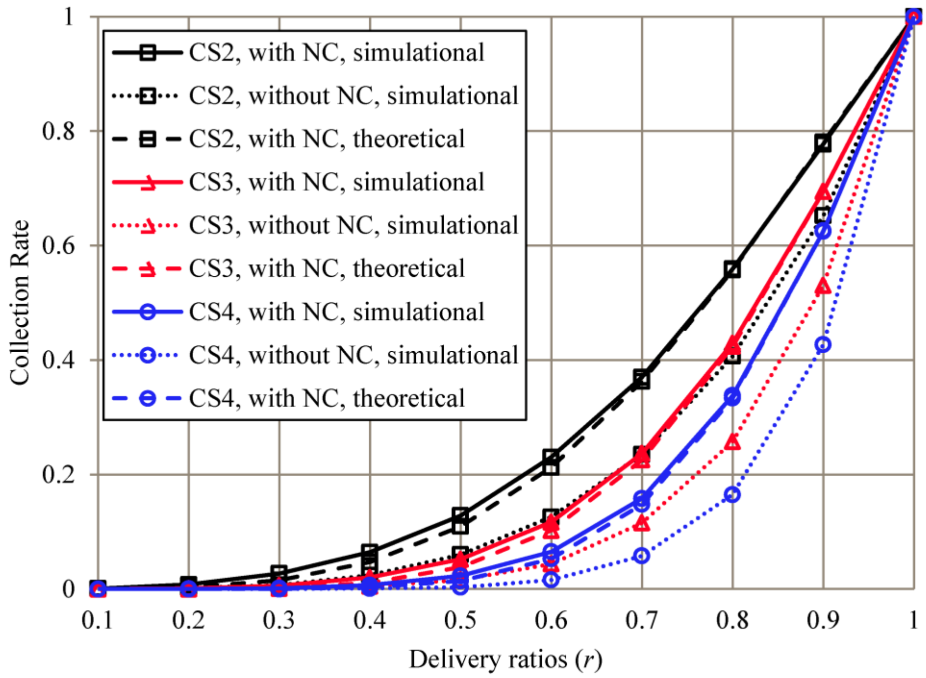



**Fig. 8.** Packet collection rates of $\mathbb{CS}_n(n = 2, 3, 4)$ under different delivery ratios

Considering that, in practical WSNs, ACK mechanism and retransmissions should be employed to ensure the packet delivery, another set of simulations are conducted. In this set of simulations, link-level ACK mechanism is adopted. A total of 100 CBR packets are sent out, and the numbers of transmissions of the source nodes are recorded when all the packets are successfully collected by the destination. Altogether 100 simulations are performed for each link delivery ratio and the average values are adopted.

**Fig. 9** shows the average numbers of transmissions per packet of the source nodes under various link delivery ratios with and without network coding. It can be observed that, firstly, much fewer transmissions are required if network coding is applied in all scenarios. The numbers of transmissions are reduced by 47.6%, 52.8% and 55.4% respectively in $\mathbb{CS}_2$, $\mathbb{CS}_3$ and $\mathbb{CS}_4$ as $r = 0.1$, and the gains monotonously drop to about 10% as $r = 0.9$. Secondly, numbers of transmissions required in $\mathbb{CS}_2$, $\mathbb{CS}_3$ and $\mathbb{CS}_4$ are totally identical without network coding, while are different if network coding is employed, that is, the higher the converge structure's order is, the fewer transmissions are required. The reason is that the source nodes locate at the edge of a $\mathbb{CS}_n$, such as $S_1$ and $S_3$ in $\mathbb{CS}_4$, can take advantage of only one overhearing link, while the others can exploit two. Hence, there are more average overhearing links available for per source node in a higher-order converge structure, which can bring about more reliability benefits and reduce more transmissions.

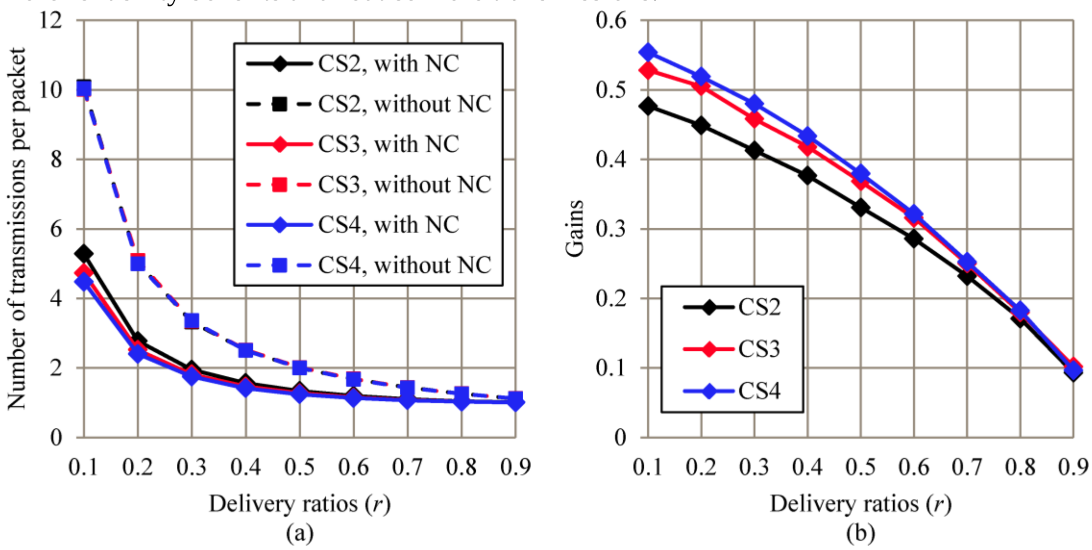



**Fig. 9.** (a) The average numbers of transmissions of the source nodes under different link delivery ratios. (b) The gains of network coding on number of transmissions under different link delivery ratios.

**Fig. 10** (a) illustrates the performance on end-to-end delay in the three scenarios described above for various link delivery ratios. First, we can find out that the end-to-end delay in a high-order converge structures is longer than that in a low-order one. And as it is expected, the delay is inversely proportion to link delivery ratio. **Fig. 10** (b) shows the results of $dl_{nc}/dl_{none\text{-}nc}$, where $dl_{nc}$ and $dl_{none\text{-}nc}$ are the average end-to-end delays of the CBR packets with and without network coding. Clearly, in case of poor wireless link quality, network coding is able to reduce the delay by means of its reliability benefits. In these cases, retransmission is the chief reason of the delay. However, when the link quality gets better, i.e., the delivery ratio is higher than $6\times10^{-4}$, none-network-coding schemes outperform network-coding ones. The reason is that, with the delivery ratio improves, the delay caused by retransmissions decreases rapidly and tends to play a minor role of the total delay. And meanwhile the inevitable waiting period, which is introduced by the decoder for collecting enough coded packets to recover the original ones, significantly increases the end-to-end delay, just as Voigt et al have also argued in [23].

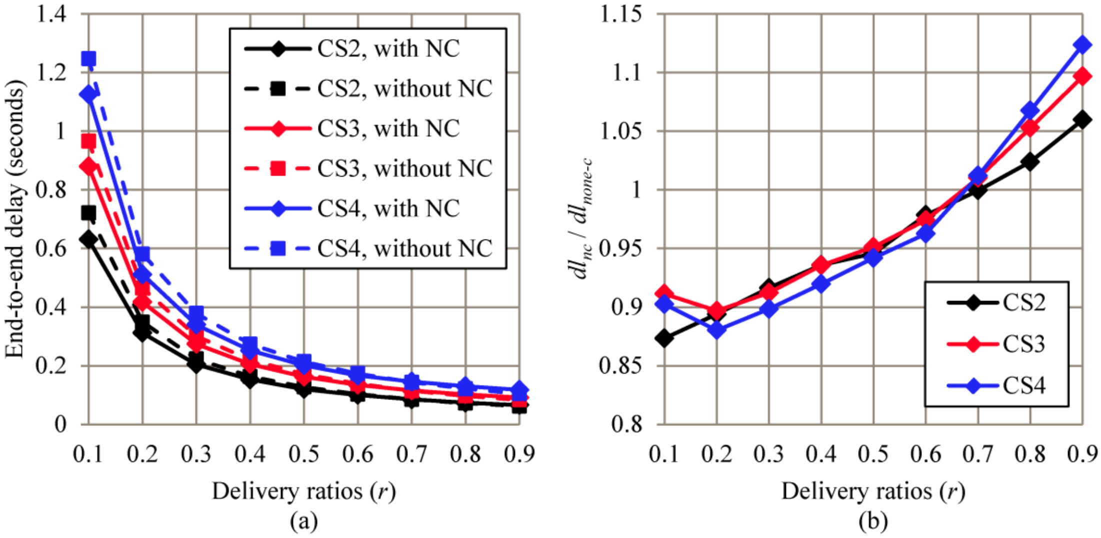



**Fig. 10.** (a) The end-to-end delay of the CBR packets with and without network coding under different delivery ratios. (b) The ratio of the average end-to-end delay of the CBR packets with and without network coding under different link delivery ratios.

We further study the performance on energy saving via this set of simulations. The detailed parameters for energy model and propagation model are listed in **Table 1** and **Table 2**. These parameters ensure that each node's signal is able to effectively cover a sphere with the radius of 250 meters. And at the edge of the sphere, the received power will be equal to RXThresh_, meaning that any packet will be discarded if the receiver is outside this area.

**Table 1**. Simulation parameters for energy model

| Parameters | Value | Description |
|---|---|---|
| Pt_Consumer | 0.386 (W) | Transmitting power consumption |
| Pr_Consumer | 0.3682 (W) | Receiving power consumption |
| P_Idle | 0.3442 (W) | Idle listening power consumption |
| P_Sleep | 0.00005 (W) | Sleeping power consumption |
| P_transition | 0.05 (W) | State transition power consumption |
| P_Init | 1000 (J) | Initial energy level |

**Table 2**. Simulation parameters for propagation model

| Parameters | Value | Description |
|---|---|---|
| propType | TwoRayGround | Propagation model. The TwoRagGroud model considers both the direct path and a ground reflection path. |
| RXThresh_ | $5.00256\times10^{-10}$(W) | RXThresh_ is the reception threshold. If the received signal strength is greater than this threshold, the packet can be successfully received. |
| CSThresh_ | $2.13551\times10^{-11}$(W) | CSThresh_ is the carrier sensing threshold. If the received signal strength is greater than CSThresh_, the packet transmission can be sensed. However, the packet cannot be decoded unless signal strength is greater than RXThresh_. |

**Fig. 11** shows the average energy consumed by the source nodes for transmitting and by the coders for receiving respectively under different link delivery ratios. In fact, **Fig. 11** (a) and (b) are completely compatible with **Fig. 9**, for the transmitting energy consumption is determined by the total number of transmissions if the transmitting power and the length of the CBR packets are given. And by **Fig. 11** (c) and (d), firstly, one can see that the energy consumed by the coders for receiving is much more than that consumed by the sources for transmitting. The reason is that the coders spend lots of energy overhearing data from other nodes. Secondly, due to the reliability benefits introduced by network coding, the network-coding-aided scheme saves considerable amount of energy spent by receiving and overhearing. However, as the delivery ratio becomes high, the superiority is not so obvious anymore. And we also observe that the gains of different converge structures are almost the same. It is because that all the coders behave in the same way and produce almost the same benefits.

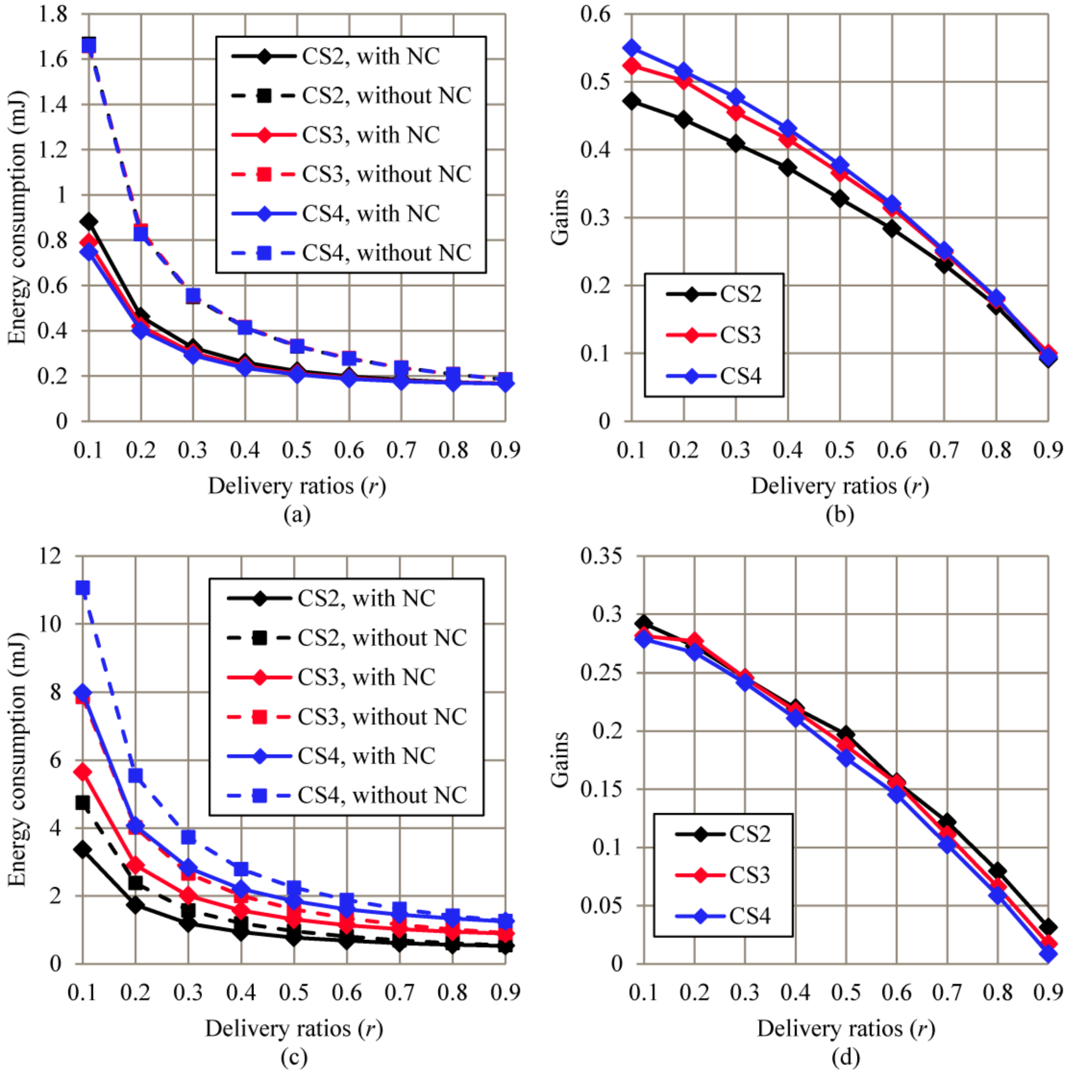



**Fig. 11.** (a) The average energy consumed by the source nodes for transmitting one packet under different link delivery ratios; (b) The gains of network coding on energy consumed by the source nodes for transmitting one packet under different link delivery ratios; (c) The average amount of energy consumed by the coders for receiving under different link delivery ratios. (d) The gains of network coding on energy consumed by the coders for receiving under different link delivery ratios.

# 6. Discussion

## 6.1. Prerequisites of Obtaining Network Coding Benefits

It can be found that: Firstly, without overhearing, none of $C_i$ ($i$ = 1, 2, …, $n$) can receive any packet sent by source node other than its own child, which makes network coding impossible. The existence of overhearing can be interpreted as the introduction of redundant links, which is certainly capable of improving transmission reliability. Secondly, if all the links were completely reliable, no retransmission would be needed, and the network coding would not introduce any benefit at all. Therefore, both the feasibility of overhearing and the unreliability of links are the prerequisites to obtain the benefits of network coding in $\mathbb{CS}_n$.

## 6.2. The need of Network Coding

There may be a puzzle about the necessity of network coding in a $\mathbb{CS}_n$. For example, as shown in **Fig. 4** (b), instead of sending the coded packet $(x_1 \oplus x_2)$, $C_1$ only need to forward $x_2$ to $D$, which can also achieve the aim. Moreover, there would be overhead on combining $x_1$ and $x_2$. But the question is how $C_1$ could learn about which packet to forward. In this case, information exchange between $C_1$ and $C_2$ is indispensable for them to coordinate the forwarding. And what's more, in a $\mathbb{CS}_n$, a $C_i$ ($i$ = 1, 2, …, $n$) has to negotiate with up to five adjacent interior nodes about the right packet to forward, which introduces much more overhead and latency than network coding.

## 6.3. How to Explore More Coding Opportunities

As mentioned above, there may be plenty of $\mathbb{CS}_n$ formed during converge-casting. Some of them could be overlapped to explore more network coding opportunities, as shown in **Fig. 8**. A node may act as a source node in a $\mathbb{CS}_n$, and meanwhile, it is the decoder or one of the encoder in another $\mathbb{CS}_n$.

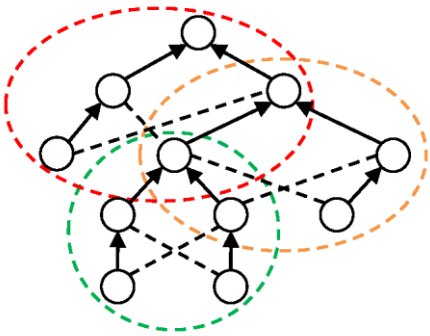


**Fig. 8.** The overlapped $\mathbb{CS}_n$. Each ellipse contains a $\mathbb{CS}_n$.

# 7. Conclusion

In this paper, we have investigated the feasibility of acquiring network coding benefits in converge-cast, and we have demonstrated that with the ubiquitous convergent structures self-organized during converge-casting in the network, the reliability benefits can be obtained by applying linear network coding. We have theoretically derived the network coding benefits obtained in a general convergent structure, and simulations have been conducted to validate our theoretical analysis. The results reveal that the network coding can improve the network reliability considerably, and hence reduce number of retransmissions and improve energy-efficiency.

## References


[1] R. Ahlswede, N. Cai, S.-Y. R. Li*, et al.*, "Network information flow," *IEEE Transactions on Information Theory,* vol.46, no.4, pp. 1204-1216, 2000. Article (CrossRef Link).

[2] H. Tracey, M. Muriel, S. Jun*, et al.*, "On randomized network coding," in *Proc. of 41st Annual Allerton Conference on Communication, Control, and Computing*, pp. 11-20, October 1-3, 2003. Article (CrossRef Link)

[3] S. Katti, H. Rahul, W. Hu*, et al.*, "XORs in the air: practical wireless network coding," in *Proc. of ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 243-254, September 11-15, 2006. Article (CrossRef Link)

[4] S. Chachulski, M. Jennings, S. Katti*, et al.*, "Trading structure for randomness in wireless opportunistic routing," in *Proc. of ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 169-180, August 27-31, 2007. Article (CrossRef Link).

[5] A. Hagedorn, D. Starobinski, A. Trachtenberg, "Rateless Deluge: Over-the-Air Programming of Wireless Sensor Networks Using Random Linear Codes," in *Proc. of International Conference on Information Processing in Sensor Networks*, pp. 457-466, April, 22-24, 2008, 2008. Article (CrossRef Link).

[6] I. H. Hou, T. Yu-En, T. F. Abdelzaher*, et al.*, "AdapCode: Adaptive Network Coding for Code Updates in Wireless Sensor Networks," in *Proc. of 27th IEEE Conference on Computer Communications*, pp. 1517-1525, April 13-18, 2008. Article (CrossRef Link).

[7] D. Wei, C. Chun, L. Xue*, et al.*, "A Lightweight and Density-Aware Reprogramming Protocol for Wireless Sensor Networks," *IEEE Transactions on Mobile Computing,* vol.10, no.10, pp. 1403-1415, 2011. Article (CrossRef Link).

[8] X. Wang, J. Wang, Y. Xu, "Data dissemination in wireless sensor networks with network coding," *Eurasip Journal on Wireless Communications and Networking,* vol.2010, 2010. Article (CrossRef Link).

[9] S. Acedański, S. Deb, M. Médard*, et al.*, "How good is random linear coding based distributed networked storage," in *Proc. of 1st Workshop on Network Coding, Theory, and Applications*, April 7, 2005. Article (CrossRef Link)

[10] D. Wang, Q. Zhang, J. Liu, "Partial network coding: Concept, performance, and application for continuous data collection in sensor networks," *ACM Transactions on Sensor Networks,* vol.4, no.3, 2008. Article (CrossRef Link)

[11] Y. Yang, C. Zhong, Y. Sun*, et al.*, "Network coding based reliable disjoint and braided multipath routing for sensor networks," *Journal of Network and Computer Applications,* vol.33, no.4, pp. 422-432, 2010. Article (CrossRef Link)

[12] L. Ting-Ge, H. Chih-Cheng, C. Cheng-Fu, "On Reliable Transmission by Adaptive Network Coding in Wireless Sensor Networks," in *Proc. of IEEE International Conference on Communications*, pp. 1-5, June 14-18, 2009. Article (CrossRef Link)

[13] R. Koetter, M. Medard, "An algebraic approach to network coding," in *Proc. of 2001 IEEE International Symposium on Information Theory*, p. 104, June 29, 2001. Article (CrossRef Link)

[14] S. Y. R. Li, R. W. Yeung, C. Ning, "Linear network coding," *IEEE Transactions on Information Theory,* vol.49, no.2, pp. 371-381, 2003. Article (CrossRef Link).

[15] T. Ho, R. Koetter, M. Medard*, et al.*, "The benefits of coding over routing in a randomized setting," in *Proc. of IEEE International Symposium on Information Theory*, p. 442, June 29 - July 4, 2003. Article (CrossRef Link).

[16] C. Fragouli, D. Katabi, A. Markopoulou*, et al.*, "Wireless Network Coding: Opportunities & Challenges," in *Proc. of IEEE Military Communications Conference*, pp. 1-8, October 29-31, 2007. Article (CrossRef Link).

[17] L. Jilin, J. C. S. Lui, C. Dah-Ming, "DCAR: Distributed coding-aware routing in wireless networks," *IEEE Transactions on Mobile Computing,* vol.9, no.4, pp. 596-608, 2010. Article (CrossRef Link).

[18] P. Pahlavani, V. Derhami, A. M. Z. Bidoki, "FENC: Fast and efficient opportunistic network coding in wireless networks," *KSII Transactions on Internet and Information Systems,* vol.5, no.1, pp. 52-67, 2011. Article (CrossRef Link)

[19] M. Halloush, H. Radha, "Network Coding with Multi-Generation Mixing: A Generalized Framework for Practical Network Coding," *IEEE Transactions on Wireless Communications,* vol.10, no.2, pp. 466-473, 2011. Article (CrossRef Link).

[20] O. Gnawali, R. Fonseca, K. Jamieson*, et al.*, "Collection tree protocol," in *Proc. of 7th ACM Conference on Embedded Networked Sensor Systems*, pp. 1-14, November 4-6, 2009. Article (CrossRef Link)

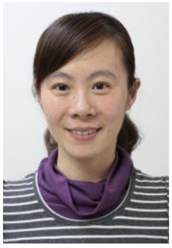
[21] L. Dijun, Z. Xiaojun, W. Xiaobing*, et al.*, "Maximizing lifetime for the shortest path aggregation tree in wireless sensor networks," in *Proc. of 30th IEEE Conference on Computer Communications*, pp. 1566-1574, April 10-15, 2011. Article (CrossRef Link).

[22] W. Bechkit, M. Koudil, Y. Challal*, et al.*, "A new weighted shortest path tree for convergecast traffic routing in WSN, " in *Proc. of IEEE Symposium on Computers and Communications*, pp. 187-192, July 1-4, 2012. Article (CrossRef Link)

[23] Thiemo Voigt, Utz Roedig, Olaf Landsiedel*, et al.*, "Practical Network Coding in Sensor Networks: Quo Vadis?," in *Proc. of 3rd International Workshop on Networks of Cooperating Objects*, April 16, 2012. Article (CrossRef Link)

**Zhenzhou Tang** received his B.S. degree in Electronic Engineering and M.S. degree in Communications and Information System from Zhejiang University, Hangzhou, China, in 2001 and 2004 respectively. He is currently an associate professor of the College of physics and electronic information engineering, Wenzhou University, Wenzhou, China, and is pursuing his Ph.D. in Wireless Communication at Dalian University of technology. His research interests include cooperative communications, network coding, and wireless sensor networks.

**Hongyu Wang** received his B.S. degree from Jilin University of Technology, Changchun, China, in 1990, and M.S. degrees from Graduate School of Chinese Academy of Sciences, Changchun, China, in 1993, both in Electronic Engineering. He received the Ph.D. in Precision Instrument and Optoelectronics Engineering from Tianjin University, Tianjin, China, in 1997. Currently, he is a Professor in the institute of Information Science and Communication Engineering, Dalian University of Technology, China. His research interests include algorithmic, optimization, and performance issues in wireless ad hoc, mesh and sensor networks.



**Qian Hu** received her B.S. degree in Communication Engineering from Nanjing Institute of Posts and Telecommunications, Nanjing, China, in 2001, and M.S. degree in Circuits and Systems from Zhejiang University, Hangzhou, China, in 2005. She is now a lecturer of the College of physics and electronic information engineering, Wenzhou University, Wenzhou, China. Her research interests are in the areas of wireless communications, including cooperative communication, network coding, and MAC and network layer protocols for wireless sensor networks.

**Long Hai** is pursuing his Ph.D. in Wireless Communication at Dalian University of technology. His major research areas are wireless communications and network coding.