# The Data Generation and Distribution using the Streaming Network Coding in WSN

Li Chenchao (IIS), Iwai Masayuki (IIS), Sezaki Kaoru (CSIS & IIS)

**Abstract---Network coding (NC) is a novel mechanism proposed to improve the throughput utilization of a given network topology [1]. Compared to other traditional approaches network coding makes optimal use of the available network resources and, moreover, computing a scheduling scheme that achieves such rate is computationally easy. Nowadays, network coding is applied in many fields such as wireless Ad-hoc networks, wireless sensor network, wireless mesh network and so on. In this paper, a new data generation and distribution using the streaming network coding in WSN has been proposed. In detail, we will demonstrate its principle and algorithm when utilizing it into WSNs.**

**Keyword: network coding, wireless sensor network, data generation and distribution.**

## 1. Introduction

Network coding (NC) was a new coding conception first proposed in 2000 [1], which has attracted much attention in coding field. The basic principle of network coding is illustrated in Fig. 1. Network coding breaks the traditional routing principle in current communication network that information can only be stored and forwarded separately without superposition, and allows network nodes to code the incoming data with appropriate coding methods, such as exclusive or (XOR) operation, linear operation [2] and so on, so as to achieve the maximum transmission capacity defined by the "Max-flow Min-cut" theorem of Shannon.

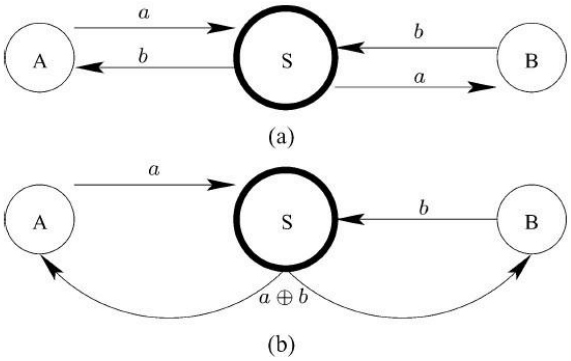


Figure.2 A typical wireless network coding example.

To exchange messages a and b, nodes A and B must route their packets through node S. Clearly, the traditional scheme in Fig.(a) would require four transmissions. However, if S is allowed to perform network coding with simple XOR operations, a _ b can be sent, as shown in Fig. (b), in one single broadcast transmission (instead of one transmission with b followed by another one with a). By combining the received data with the stored message, A, which possesses a, can recover b and B can recover a using b. Thus, network coding saves one transmission.

With lots of benefits for communication network, NC can be used in many network structures. In terms of the link unreliability and physical layer radio characteristic in wireless networks, NC is more suitable for wireless applications, such as wireless Ad-hoc networks, wireless mesh networks [3], and wireless sensor networks (WSN) [4].

Wireless sensor network (WSN) consists of many small, inexpensive distributed sensor nodes that organize themselves into a multi-hop wireless network as illustrated in Fig. 2. Compared with others, WSN has a high node density, little mobility and usually runs on the remote unattended environment; especially each node has limited energy. Considered as data-centric networks, in WSN, information processing and exchange from node to node using multi-hop transmission are the main tasks because it is required by many basis operations, such as cluster head election and data aggregation. Such an information exchange using all-to-all broadcast causes heavy traffic and consumes enormous energy. However, owing to the location and cubage, each node has limited energy. Once the energy is used up, this node cannot work anymore, which will then bring in the change of network topology.

There are security risks in WSNs. For example, in a houses sensor network, we use nodes to collect statistics about water, humidity and electricity consumption within a large neighborhood in order to get resources planning purposes and usage advice. But try to imagine that a bad guy attacks your nodes and your private information is leaked. The bad guy will know your personal activities just like when you take a shower, when all family members are gone. Hence privacy protection is extremely important for WSNs.

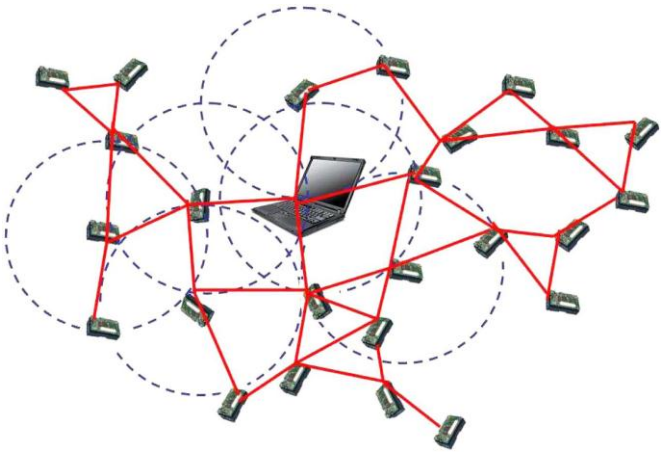

Fig.2. Wireless sensor network is a collection of small devices that, once deployed on a target area, organize themselves in an ad-hoc network, collect measurements of a physical process, and transmit the data over the wireless medium to a data-fusion center for further processing. If a mobile node (represented here by a laptop computer) is available, then it can be used to establish secure links between sensor nodes and, thus, bootstrap the network.

In this paper, we firstly give a survey of NC and scan applications in WSN. Then, by given the basic principle and algorithm, we will demonstrate our proposed resolution briefly. Last, we do the conclusion and make a plan about the future work.

## 2. Related work

WSNs have been studied extensively recently and lots of surveys can be found in [5][6]. In our application, we are interested in distributing data from the sensor node into its neighbors and generate the original data unrestrainedly.

Network coding is a powerful tool for data storage. A typical coding scheme is the *erasure codes* [7], in which a centralized server gathers all $N$ data segments and builds $C$ coded segments, $C \gg N$. If any $N$ out of $C$ coded segments are collected, the original data segments can be decoded [8].But unfortunately, these centralized operations are not suitable for wireless sensors networks which are distributed system.

Network coding was first introduced in [1] to improve the performance of multicast routing. In the proposal, a router in the network can not only relay the data packets, but also code the data packets. The throughput of the network can thus be significantly improved. To guarantee decoding of the coded data packets by the receivers, a centralized management of the co-efficient matrix is needed. *Random linear coding* [9] was proposed thereafter to relax this central control feature. The co-efficient of the codes are picked randomly from a finite field by each individual node. If the finite field is chosen appropriately (large enough), then the linear dependency of the co-efficient is negligible and the probability of successful decoding of the code words is high. Our work also incorporates this feature.

## 3. The data distribution and generation system

When doing the data distribution, the system works as Figure.3. A node encodes the file which needs to be stored distributed. Then it transmits the encoded data blocks and stores them into severe reliant nodes, at the same time, submits the location file which included the address of blocks.

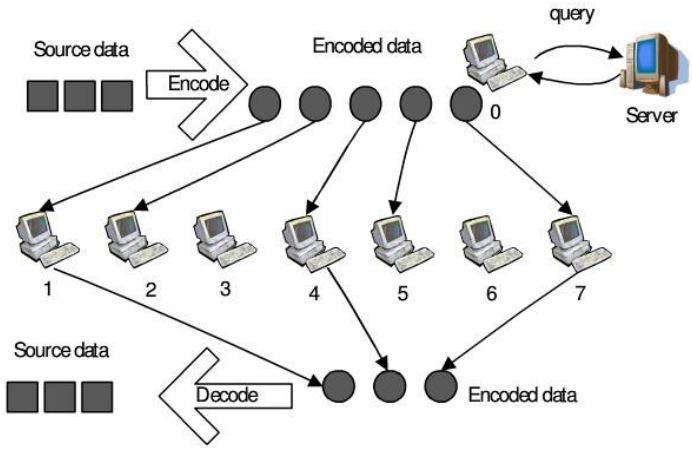


Figure.3 how the sensor network works when doing the data distribution

And when we need to generate the original data, the client node requests the location files and according to that file, takes back the data blocks. By decoding algorithm which will be explained in detail in the next section, the original data will be generated. This process is demonstrated in Figure.4.

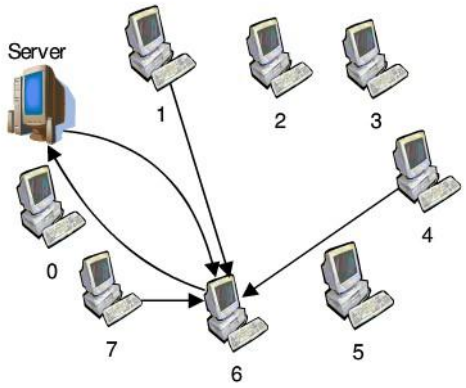


Figure.4 how the sensor network works when doing the data generation

## 4. Proposed encoding and decoding principle

### 4.1 Linear network coding (LNC)

If the encoding and decoding functions in NC are all linear operations, then we call this process linear network coding

2003[2]. In particular, all the nodes except source and sink nodes in the network, do the linear operation such as addition and multiplication for all the received packets, when receiving enough independent packets, nodes can decoding the original information. Besides, if the length of two combinations is not the same, LNC will put a couple of 0 behind the shorter one. This code is very useful, since encoding and decoding are computationally inexpensive. This is especially attractive in resource-constrained Wireless Sensor Networks.

### 4.2 Galois field

Field is any set of elements which satisfies the field axioms for both addition and multiplication and is commutative division algebra. A field with a finite number of members is known as a finite field or Galois field. For a given Galois field of size q, if q − 1 powers of an element x (x1, x2, . . . , xq−1) produce all non-zero elements, that element x is called a generator of the given Galois field. And as shown in Table 1, the probability of linear independency is over 99.6% for $q = 2^8$, and this is almost independent of $N$.

| $q$ | Probability | $q$ | Probability | $q$ | Probability |
|---|---|---|---|---|---|
| $2^1$ | 0.288788 | $2^5$ | 0.967773 | $2^9$ | 0.998043 |
| $2^2$ | 0.688538 | $2^6$ | 0.984131 | $2^{10}$ | 0.999022 |
| $2^3$ | 0.859406 | $2^7$ | 0.992126 | $2^{11}$ | 0.999511 |
| $2^4$ | 0.933595 | $2^8$ | 0.996078 | $2^{12}$ | 0.999756 |

Table1 Probability of Linear Independency as a Function of Finite Field Size ($q$).

### 4.3 Reed-Solomon Coding

In Reed–Solomon coding, source symbols are viewed as coefficients of a polynomial $p(x)$ over a finite field. The original idea was to create $n$ code symbols from $k$ source symbols by oversampling $p(x)$ at $n > k$ distinct points, transmit the sampled points, and use interpolation techniques at the receiver to recover the original message.

For a given data, let us break down it into m messages w0,w1,w2, . . . ,wm−1. And construct P(X) using these messages as coefficients such that

$$P(X) = \sum_{i=0}^{m-1} w_i x^i$$

And evaluate this polynomial P(X) at n different pointsx1, x2 …, $x_n$. Then P(x1), P(x2)…P( $x_n$ ) can be represented as multiplication of matrix and vector as shown in Figure 3. Here we can see that matrix A is Vandermonde matrix, W is a vector of messages, and code words are contained in a vector AW. If we have any m rows of A and their corresponding P(X) values, we can obtain vector W which contains coefficients of polynomial, which is again the original messages. Figure.4 shows how Reed-Solomon code works in our data distribution system.

$$\begin{pmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ 1 & x_2 & \cdots & x_2^{m-1} \\ 1 & x_3 & \cdots & x_3^{m-1} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & x_{n-1} & \cdots & x_{n-1}^{m-1} \\ 1 & x_n & \cdots & x_n^{m-1} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \end{pmatrix} = \begin{pmatrix} p(x_1) \\ p(x_2) \\ p(x_3) \\ \vdots \\ \vdots \\ p(x_{n-1}) \\ p(x_n) \end{pmatrix}$$

Figure.4 how Reed-Solomon code works

Besides, if some part of erasure codes are original messages, when every packet in this part arrives we can get original messages without decoding. At the encoding side, we don`t need any computation for the portion of code words containing original messages. This reduces encoding overhead. At the decoding side, when we have all code words containing original messages, we get messages without any further computation. In this case, we can achieve huge save in decoding computation. The mechanism is demonstrated in Figure.5.

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 1 & x_{m+1} & \cdots & x_{m+1}^{m-1} \\ 1 & x_{m+2} & \cdots & x_{m+2}^{m-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \end{pmatrix} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \\ p(x_{m+1}) \\ p(x_{m+2}) \\ \vdots \\ p(x_n) \end{pmatrix}$$

Figure.5 How Reed-Solomon code works when server gets original messages without decoding.

At the decoding side, the more original message part we have, the closer decoding matrix would be to identity matrix, and the quicker decoding process becomes. Therefore, if we have most of code words containing original messages, and some messages constructed by matrix, decoding will be very fast. Reed-Solomon code can give benefit even when we lose some packets.

## 5. Encoding and decoding algorithms

### 5.1 Data distribution algorithm

In order to increase the flexible of encoding, the client node can choose the coding rate by itself. When storing the encoded data block, the client node chooses the nodes which existed in reliance neighbor list and then transmit the blocks. When done, it submits the location file which including the destination of blocks to the server and backup the data in local storage. The process is shown as follow in detail:

**Algorithm Data distribution**

getFile(); *//get the data file from local storage*

setMN(M,N); *//set the coding ratio where M means the number of encoded blocks and N means the essential number of decoding blocks*

checkRatio(); *//check the coding rate whether it can be applied or not*

buildField(); *// Construct Vandermonde Matrix under Galois Field*

EncodeFile(); *//Encode the file data and divide it into M pieces*

MD5(file_block); *//Do the MD5 algorithm(MD5 Message-Digest Algorithm) and store the information into blocks*

sendFileBlock(); *//Complete the encoding process and store the blocks into neighbors nodes according to the reliance list*

sendSavedInfo(); *// Submit the location file which including the destination of blocks to the server*

5.2 Data generation algorithm

**Algorithm Data generation**

f_info=getFileInfoFromServer(); *//get the storage destination file from the server*

for i=0 to f_info.n getFileBlock(client); *//request the blocks from the neighbors nodes*

wait(); *//wait until receiving the enough number of blocks*

if(receiveCount>=N) *//start the verify process when received enough blocks*
For each i=0 to receiveCount
if(!checkBlockOk(file_block))
deleteDirtyFile(file_block) *// verify the received blocks and delete the defiled ones.*

erasureDecodeProcedure() *//do the decoding process*

When trying to generate the data, the client firstly should check whether the backup data information is in the local storage. If negative, the client node requests the location files and according to destination file, takes back the data blocks. When the node receive the enough number of blocks (in our case, the number is N), it should verify them and delete the defiled ones. When the number of verified blocks reaches N, it begins to generate the data, otherwise it will request for the remainder blocks until the generation complete. The process is shown as above.

## 6. Conclusion and future work

A data generation and distribution using the streaming network coding in WSN has been presented. The main contribution of our proposal was to add a new functionality of data storage to the sensor nodes. As those nodes in many WSNs scenarios are exposed to attackers, a research effort should be dedicated not only the storage but also the safe and simple encryption and decryption algorithm when considering the buttery consumptions of the nodes. From the mathematics, our algorithm is safe. For future work, it is intended to create a buttery balance function into our algorithm and implement the proposed scheme on reality world.

## References

[1] R.Ahlswede, N. Cai, S. Y. R. Li and R. W. Yeung, "Network information flow," IEEE. Trans. Inform. Theory, 46(4), pp. 1204-1216, July 2000.
[2] S. Y. R. Li, R. W. Yeung and N. Cai, "Linear Network Coding," IEEE. Trans. Inform. Theory, 49(2), pp. 371-381, Feb. 2003.
[3] A. Al Hamra, C. Barakat and T. Turletti, "Network coding for wireless mesh networks: A case study," 2006 Int. Symp. on a World of Wireless, Mobile and Multimedia Networks, Buffalo-Niagara Falls, NY, United states, pp. 103-111, June 26, 2006 - June 29, 2006.
[4] Sachin Katti, HariharanRahul, WenjunHu, Dina Katabi, Muriel Medard, Jon Crowcroft , " XORs in the Air: Practical Wireless Network Coding" *SIGCOMM'06,* September 11–15, 2006, Pisa, Italy.
[5] Akyildiz, I., Su, W., Sankarasubramaniam, Y., and Cayirci, E. 2002. A survey on sensor networks. *IEEE Communications Magazine 40*, 8 (Aug.), 102-114.
[6] Al-Karaki, J. and Kamal, A. 2004. Routing techniques in wireless sensor networks: A survey. *IEEE Wireless Communications 11*, 6 (Dec.), 6-28.
[7] Blahut, R. 1983. *Theory and Practice of Error Control Codes*. Addison-Wesley, Reading, MA.
[8] Luby, M. 2002. Lt codes. In *Proc. IEEE FOCS'02*. Vancouver, Canada.
[9] Medard, M., Acedanski, S., Deb, S., and Koetter, R. 2005. How good is random linear coding based distributed networked storage? In *Proc. NETCOD'05*. Riva del Garda, Italy.
[10] Deb, S. and Medard, M. 2004. Algebraic gossip: A network coding approach to optimal multiple rumor mongering. In *Proc. Allerton Conference on Communication, Control and Computing 2004*. Urbana, IL.