Data Recovery in Wireless Sensor Networks using Network Coding

A. A. Shahidan*, N. Fisal, Nor-Syahidatul N. Ismail, Farizah Yunus, Sharifah H. S. Ariffin

UTM-MIMOS Center of Excellent, Faculty of Electrical Engineering Universiti Teknologi Malaysia, 81310 UTM Johor Bahru Johor, Malaysia

*Corresponding author: ashahidan@utm.my

Graphical abstract

Abstract

Data transportation over resources constraint and noisy channel of wireless sensor network (WSN) is very challenging in term of guaranteeing the data survival along the transmission. However, with the convergence of different research areas such as routing, source and channel coding techniques, the WSN technology has successfully been tremendously developed. This paper proposes an on-the-fly data recovery (ODR) scheme using network coding in order to enhance the robustness of the network against packet loss. Along with the ODR scheme, the packet loss formulation is presented while a network model for a network coding designed is also introduced namely for erasure channel. The data generated by the sources are transferred to the destination through relay nodes via three transmission paths. In ODR process, the lost packet is recovered by a relay node by listening to the transmission of two adjacent nodes and performing the XOR operation on the listened packets. We provide the analytical study on network coding performance and conducting the simulation experiment to verify it. In the simulation studies, we have also compared the performance of the network using network coding with and without packet recovery. The result shows that the number of packet loss has been reduced significantly using the proposed scheme compared to the network with normal network coding.

Keywords: Network coding; wireless sensor networks; data recovery; multipath data transmission

1.0 INTRODUCTION

The WSN technology consists of many self-organized sensing nodes [1] which is currently becoming a platform for various applications such as environmental monitoring, hazardous environment exploration, and military tracking [2]. The development of WSN is very challenging in order to satisfy several constraints such as low power consumption, small buffer size, slow transmission rate and limited processing speed [3]. If the size of the data handled is large, the data should be compressed or coded in such a way that the transmission speed can be compromised. Network coding concept that was first introduced in [4] has the capability of enhancing the reliability of transmission in WSN by reducing the number of packet loss [5]. Based on network coding concept, instead of just forward the received packets, the intermediate node will combine them and create one or several output packets [6].

Network coding is known with its ability to improve the throughput [7] and reduce the energy consumption of the network [8]. Network coding has been used for various applications. As reviewed in [9], network coding has been adapted in multimedia applications for error control. Besides, network coding is realized based on fountain codes approach in order to achieve low-complexity characteristic of the network as described in [10].

Regarding the protection and recovery schemes, the authors in [11] have proposed mathematical programming model for network protection against single link error using network coding. However, the formulation focus on reducing cost while the packet loss formulation is not provided. Besides, the data protection is also demonstrated using network coding coupled with multipath routing where redundancy of transmission in multipath routing is exploited in order to realize the data recovery in underwater sensor networks [12] and in wireless body area networks [13]. Though, the proposed algorithm is limited for specific network topology. Furthermore, the loss recovery has also been implemented in vehicular safety communication where multiple packets transmitted by different vehicles is combined into a single transmission [14]. However, this scheme is limited to only one hop.

In particular, we consider the problem of reducing the energy consumption and increasing the throughput of the network in WSN using network coding as explained previously. In this paper, we introduce data recovery scheme using network coding couple with multipath transmission and investigate the benefit of the scheme in reducing the number of packet loss in WSN. The proposed scheme is deployed in flexible grid network arrangement where the topology can be extended depending on the application. Performance of the proposed mechanisms is analyzed using simulation. The remainder of this paper is organized as follows. Network coding concept are reviewed and summarized in Section

2.0. System model, analytical study of network coding and data recovery scheme are described in Section 3.0. The performance analysis, result and discussion are described in Section 4.0. Lastly, Section 5.0 presents the conclusion.

■2.0 NETWORK CODING CONCEPT

Network coding which has become one of the popular research areas in information theory since few years ago is applicable in practical networking system [15]. Through the following example, we are going to review the concept of network coding. Firstly, let *X* and *Y* are two nodes that are going to exchange information. However, they are not in transmission range among each other. The only communication can be made is via a relay node located between both nodes. Therefore, let *S* be an intermediate node between *X* and *Y* as illustrated in Figure 1.

  Let *a* be the data from *X* to *Y* while *b* be the data from *Y* to *X*. When conventional transmission method is used, four transmissions are required to complete the data exchanging procedure. In the first transmission, *T*1, *X* sends data *a* to *S* while in the second transmission, *T*2, data *b* is sent from *Y* to *S*. Then, *S* sends data *a* to *Y* in the third transmission, *T*3, followed by transmission of data *b* from *S* to *X* in the fourth transmission, *T*4. The process is illustrated in Figure 2.

  The total number of transmissions of the conventional method can be reduced from four to three transmissions if network coding is used. Based on this method, first transmission, *T*1, and second transmission, *T*2, are remain the same as conventional method. However, in the third transmission, *T*3, instead of transmitting data *a* to *Y*, *S* broadcast the combination of both data *a* and *b* to both *X* and *Y*. The combination process involves XOR operation on data *a* and *b*. The combined data received by both *X* and *Y* can be decoded by re-XOR the received data with the original data possessed by both nodes at the beginning of transmission procedure. The process is shown in Figure 3.

■3.0 SYSTEM MODEL

3.1 Network Model

A network model is designed and being used with the proposed algorithm while the performance of the system is analyzed through simulation studies. The network model consists of three parts called source, relay and destination. There are two nodes in source part that provide data streams in the network. The data from source part is passed to the relay part. Network coding is applied at the first node of a path in relay part where only one path will be transporting the coded data. In order to realize the proposed mechanism, a certain network must be able to provide three transmission paths without any overlapping nodes. In other word, one particular node can only belong to one path. Relay part of the proposed network is arranged in a grid form of three column with *h* nodes per column where *h* is the number of hops. Each column of the network model establishes one transmission path. In the last part, three receivers will collect the data of each path before passing to the Data Processing Unit (DPU). Figure 4 shows the proposed network model.

3.2 Network Coding Design

Packet loss of the data transmission in noisy network cannot be avoided. However, it is known that network coding is capable to reduce the number of transmissions as described in Section 2.0. Therefore, if the number of transmission in the network is fixed, the

number of packet loss can be reduced. The data encoding process is handled by a certain node using a simple XOR operation on two input packets. For example, let us refer to Figure 4 and let *A* be a data from source *S*1 and *B* is a data from source *S*2. The XOR process will be done by node $R_{1,2}$ since the node has two input packets from both sources. The encoding process will produce a combined packet or encoded packet, *A*-XOR-*B* and the packet will be passed through the same path to the destination using normal transmission. Besides the transmission of the encoded data, both original data *A* and *B* are also transported to the destination through the other two paths respectively.
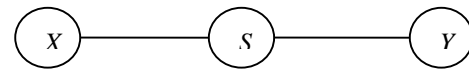

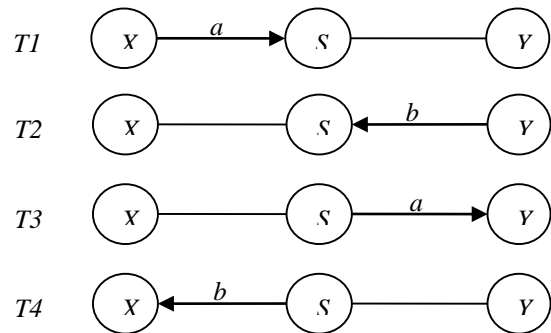**Figure 1** Node *S* as an intermediate node for node *X* and *Y*

Figure 2 Four transmissions are required for exchanging data between *X* and *Y*
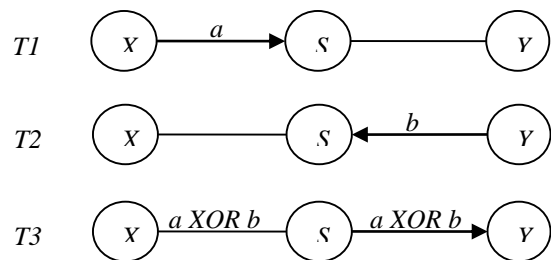

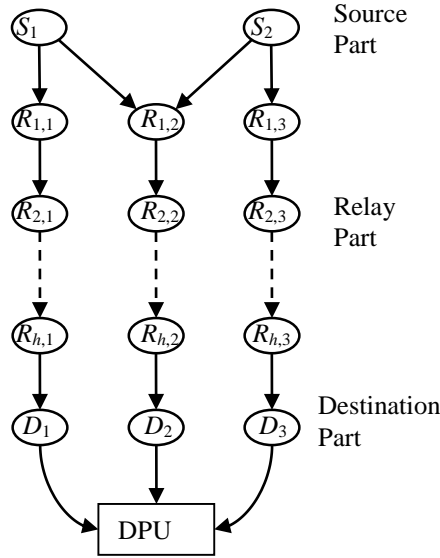**Figure 3** Three transmissions are sufficient for exchanging data between *X* and *Y*

Figure 4 System model used in the simulation

At the destination, the DPU will first identify the original data *A* and *B*. If one of those data is missing, the DPU will then identify the third received data which suppose to be the encoded data. The encoded data will be decoded using the original data received in order to recover the loss data. If the encoded data is also loss, then the DPU will only send the first original data received to the upper layer while the other one is considered loss.

The total number of packet loss in the transmission of a group of packets is directly proportional to the probability of packet loss of the network. Hence, the increment or decrement of total number of packet loss will follow the trend of the total value of packet loss probability of the network as shown in (1) where k_E is the number of packet loss, K is the total number of packets and P_E is the loss probability of a packet when it is transported in the network.

$$k_E = K P_E \qquad (1)$$

Meanwhile, the probability of a packet loss in the network, P_E, is closely related to the packet loss probability of every single link in the network, p_i where $i = 1, 2, 3, \ldots, h$, h is the total number of links (or number of hops) in one particular path. The probability of a packet loss of *l*-th path with *h* hops is shown in (2).

$$P_{E,l} = 1 - \prod_{i=1}^{h}(1 - p_i) \qquad (2)$$

If the probabilities of packet loss of all the links in the path are assumed to be the same which is *p*, equation (2) can be reduced to

$$P_E = 1 - (1 - p)^h \qquad (3)$$

The loss probability of a transmitted packet also affected when the number of paths that transported the same packet is increases. The relation between P_E and the number of paths, *L* is given as

$$P_{E,L} = \prod_{l=1}^{L} P_{E,l} \qquad (4)$$

The packet loss probability, P_E, in the case where the packet loss probabilities of all links are the same which is *p*, is expressed by

$$P_{E,L} = P_E^{\ L} \qquad (5)$$

It is known that network coding can only be applied when at least two packets are transported in a certain transmission session. In this paper, we will only consider the transmission of two packets per transmission session with network coding. Based on Fig. 4, the encoding process will be done at node $R_{1,2}$ in the second path and the output packet will be sent to the destination through the path. Meanwhile, the two original packets will also being transmitted to the destination through the other two paths respectively.

Now, let us review the example in section 3.2. Note that, the survival of data *A* in the transmission depending on the survival of transmission in the first path or the survival of both second and third paths. If this condition is void, data *A* will be loss. Let us see the probability of survival of data *A* in the transmission session. Firstly, the survival probability of data *A* in the first path, $P_{A,1}$ is given by

$$P_{A,1} = 1 - P_{E,1} \qquad (6)$$

Next, the probability of survival of data *A*-XOR-*B* and data *B* in second and third paths while data *A* in first path is loss is as follows

$$P_{A,2,3} = P_{E,1}(1 - P_{E,2})(1 - P_{E,3}) \qquad (7)$$

Therefore, the probability of data A is loss in the transmission, P_{EA}, is expressed by

$$\begin{aligned} P_{EA} &= 1 - (P_{A,1} + P_{A,2,3}) \\ &= 1 - \big((1 - P_{E,1}) + P_{E,1}(1 - P_{E,2})(1 - P_{E,3})\big) \\ &= P_{E,1}(P_{E,2} + P_{E,3} - P_{E,2}P_{E,3}) \end{aligned} \qquad (8)$$

Let the loss rate of every links be the same and the number of link in every path is assumed to be similar. In the same transmission session, the loss probability of any data in the network with network coding is shown as follows

$$P = 2P_E^{\ 2} - P_E^{\ 3} \qquad (9)$$

In (9), P_E is obtain from (3) and if the number of hops, *h*, is equal to one, then P_E is equal to *p*. The mathematical model here will be analyzed and compared to performance of the proposed algorithm through simulation experiment in Section 4.0

3.3 On-the-fly Data Recovery Process

The work is extended with the aim of improving the reliability of data transmission. To achieve the aim, instead of applying the network coding through end-to-end basis, we exploit the capability of network coding in recovering the lost packet and use it per hop basis. Hence, the On-the-fly Data Recovery (ODR) process has been introduced. ODR process is able to recover the lost packet as soon as after packet loss is detected in the middle of transmission. The on-the-fly terms is referred to the ability of the process to recover the lost packet at the intermediate nodes instead of at the destination node as in conventional method. It is known that as a packet travel through several hops in the network, the probability of the packet to be lost will increases. As a result, the probability of successful recovery process will reduces when the probability of packets reach the destination decreases. On the other hand, the ODR process conducted at the intermediate node allows the lost packet to be recovered immediately with higher possibility compared to the recovery process at the destination. Therefore, the throughput of the network will be improved subsequently.

The details of the ODR process are described as follow. The ODR process will start by a certain node when it fails to receive a data from its previous node after a certain waiting period. We are going to call the node as ODR node and the path that the node belongs to is called ODR path. The waiting period is obtained based on Time Division Multiple Access (TDMA) method where the

node will be given a time slot to transmit their data. Hence, a node can predict when the data should be transmitted by the previous node. If there is no data received at the predicted time slot, the node will begins the ODR process.

During the process, ODR node will listen to two transmissions from the other two paths. Both obtained data will be XORed to recover the lost data. Instant recovery can be done if there are two nodes, each from different path, which are in transmission range of the ODR node. However, if only one node of the non-ODR path is available, then the ODR node will listen to that node first. The transmission is continued until there is one node of ODR path and one node of the non-ODR path that is unavailable earlier that are in transmission range of each other. However, the node of non-ODR path is required to have the transmission time slot before the time slot hold by the current ODR node so that the node will be able to listen to the transmission. Then, the recovering process will be done by the ODR node

4.0 RESULTS AND DISCUSSION

Let us now analyze the result of the simulation experiment. For this simulation, the first hop is counted starting at the transmission of the first relay node to the next one. In short, the transmission from the source to the first relay nodes is assumed to be error free. In the first experiment, the performance of the network with three transmission paths is determined. The average numbers of packet loss for the transmission of 6144 packets using the network with and without network coding (NC) are compared as illustrated in Figure 5. Based on the figure, the number of packet loss is low in the transmission with NC at probability of packet loss less than 0.5. When the loss rate exceeds 0.5, the transmission with NC shows more packet loss than without NC. This is due to the insufficient number of received packets that can be used to reconstruct the loss packet as more than half of the total packet s is lost during the transmission. However, since 90% of wireless links of WSN have less than 0.005 packet loss probability [16], this occurrence will not significantly affect the performance of the WSN system.

Furthermore, the effect of the different number of paths used in the network is also examined. The simulation is carried out by varying the number of path from one to six. Besides, the simulation is repeated with the increment of the number of hop, h, from one to three hops. The result is shown in Figure 6. According to the figure, the increment of the number of path has reduced the number of packet loss. The reason is that the uses of many transmission paths will ensure higher availability of connection between the source and destination. When a path is failed for example, there are more paths that still available to transport the lost packet of the failure path. On the other hand, as the number of hop is increasing, the number of packet loss will also increase as proven by the expressions (2) and (3).

Besides, we have also conducted two simulation experiments in order to determine the effect of applying ODR process on the performance of the system. The simulation is carried on using the three transmission paths network based on Figure 4 with network coding applied. The simulation is repeated 20 times and the average value is taken. For the first experiment, the packet loss probability of every link is set to 0.2 while the number of hops is varied from two to eleven. The result of this experiment is illustrated in Figure 7. For the second experiment on ODR process, we fix the number of hops to two while the packet loss probability of every link is varied from 0.02 to 1 and the result of the experiment is shown in Figure 8.

Based on Figure 7, it is clearly shows that the number of packet loss is reduced significantly using the proposed recovery algorithm when the number of hops of the network is increased. The reason is

that the ODR process is done per hop basis, which allow the lost packet to be recovered as soon as possible. Meanwhile, Figure 8 shows that the number of packet loss is reduced considerably especially at the loss rate lower than 0.7. The recovery process per hop basis improves the packets survivability in every hop and hence the probability of recovery process to be successful will be higher.
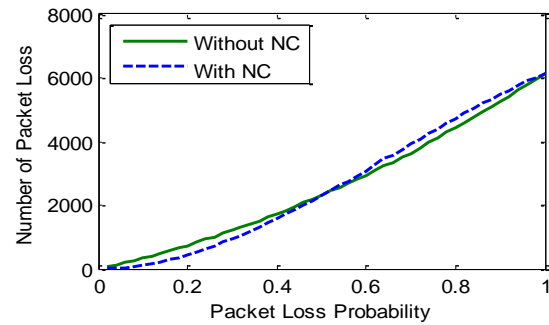
Figure 5 Comparison of the average number of packet loss in the network of three transmission paths with NC and without NC
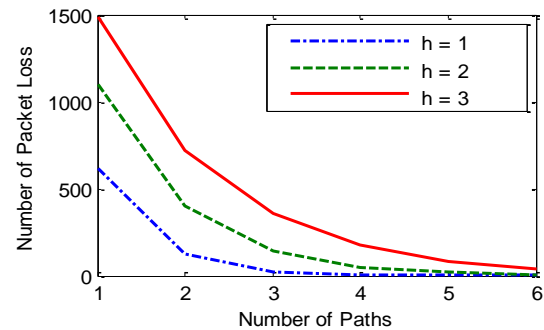
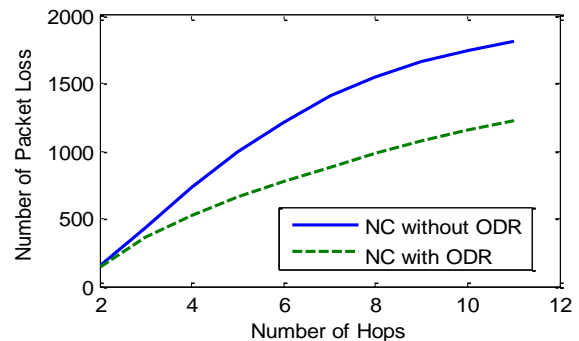**Figure 6** Effect on the performance of the system when different number of paths and hops are used

Figure 7 Comparison of the performance of the system using ODR and normal NC for different number of hops
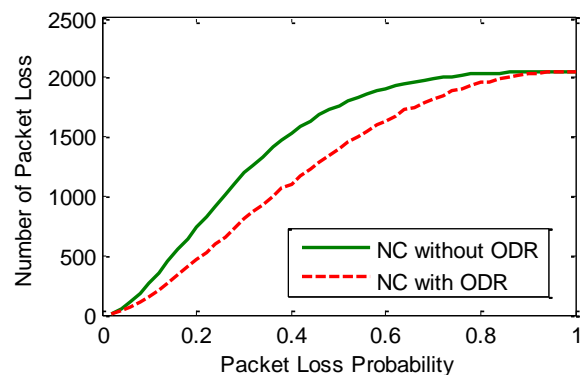
Figure 8 Comparison of the performance of the system using ODR and normal NC for different packet loss probability

■5.0 CONCLUSION

With the ability to enhance the throughput of the data transmission in wireless network, the NC has recently caught the eye of many researchers and becoming one of the subjects in their research works. In this paper, an on-the-fly data recovery scheme is proposed that couples NC and multipath data transmission in WSN. We first introduced a network model that has been used to study the performance of the proposed scheme. Then, we provided the analytical studies on the performance of network coding and followed by the simulation experiment in order to determine the effect of network coding on the performance of the system. Next, the performance of the system is examined through simulation in order to find out the benefit of the proposed recovery scheme compared to the normal NC. The result has confirmed that the proposed algorithm is able to reduce the number of packet loss in the transmission significantly.

References

[1]   Y. Wang. 2008. Study on Model and Architecture of Self-Organization Wireless Sensor Network, *4th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM '08*. 1–4.

[2]   I. F. Akyildiz, T. Melodia, and K. R. Chowdhury. 2007. A Survey on Wireless Multimedia Sensor Networks. *Computer Networks*. 51: 921–960.

[3]   C. F. García-Hernández, P. H. Ibargüengoytia-González, J. GarcíaHernández, and J. A. Pérez-Díaz. 2007. Wireless Sensor Networks and Applications: A Survey. *IJCSNS International Journal of Computer Science and Network Security*. 7: 264–273.

[4]   R. Ahlswede, C. Ning, S. Y. R. Li, and R. W. Yeung. 2000. Network Information Flow. *IEEE Transactions on Information Theory*. 46: 1204–1216.

[5]   A. A. Shahidan, N. Fisal, N.-S. N. Ismail, and F. Yunus. 2011. Proposed Network Coding for Wireless Multimedia Sensor Network (WMSN). vol. 107, J. J. Park, H. Arabnia, H.-B. Chang, and T. Shon, Eds., ed: Springer Netherlands, 387–395.

[6]   S. Y. R. Li, R. W. Yeung, and C. Ning. 2003. Linear Network Coding. *IEEE Transactions on Information Theory*. 49: 371–381.

[7]   S. Chieochan and E. Hossain. 2013. Channel Assignment for Throughput Optimization in Multichannel Multiradio Wireless Mesh Networks Using Network Coding, *Mobile Computing, IEEE Transactions on*. 12: 118–135.

[8]   A. Akhtar, M. Nakhai, and A. Aghvami. 2013. On the Use of Cooperative Physical Layer Network Coding for Energy Efficient Routing, *Communications, IEEE Transactions on*. PP: 1–12.

[9]   E. Magli, M. Wang, P. Frossard, and A. Markopoulou. 2013. Network Coding Meets Multimedia: a Review. *IEEE Transactions on Multimedia*. PP: 1–1.

[10] D. Vukobratovic, C. Stefanovic, and V. Cmojevic. 2008. On low-complexity Network Coding Coding for Data Collection in Wireless Sensors Network. *16th Telecommunications Forum*. Belgrade, Serbia.

[11] A. H. A. Muktadir, A. A. Jose, and E. Oki. 2012. An Optimum Mathematical Programming Model for Network-Coding Based Routing with 1+1 Path Protection, *World Telecommunications Congress*. 1–5.

[12] Z. Guo, B. Wang, and J.-H. Cui. 2007. Efficient Error Recovery Using Network Coding In Underwater Sensor Networks. *6th International IFIP-TC6 Conference on Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet*. Atlanta, GA, USA. 227–238.

[13] S. Marinkovic and E. Popovici. 2009. Network Coding for Efficient Error Recovery in Wireless Sensor Networks for Medical Applications. *First International Conference on Emerging Network Intelligence*. 15–20.

[14] Z. Wang, M. Hassan, and T. Moors. 2010. Efficient Loss Recovery Using Network Coding in Vehicular Safety Communication. *IEEE Wireless Communications and Networking Conference*. 1–6.

[15] C. Fragouli, J.-Y. L. Boudec, and J. Widmer. 2006. Network Coding: An Instant Primer. *SIGCOMM Computer Communication Review*. 36: 63–68.

[16] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. 2002. Wireless Sensor Networks: A Survey. *Computer Networks*. 38: 393–422.