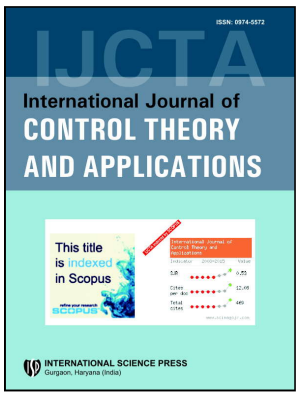

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 15 • 2017

Defense against Pollution Attack for Robust Linear Network Coding

Sonia Sharma[1]

[1] *Research Scholar; Department of CSE, IKGPTU, Jalandhar, Punjab, India, Cum Associate Professor; Department of Information Technology; Malout Institute of Management and Information Technology, Malout, Punjab, India, Email: sonia.mimit@gmail.com*

Abstract: Network coding is having the capability to improve throughput and reliability of the network. These benefits are vulnerable to various malicious attacks like corruption of message blocks and trade off among nodes. In this way, to conquer these issues, detection and correction strategies for pollution are proposed, but these strategies produces effective outputs in some particular cases only. In this paper, a new protocol for limiting the pollution attack (LPA-NC) is proposed in which the enhancement of existing error control methods is done. This will result in the improvement of throughput and robustness of the network. The proposed scheme will identify the possible attacks exists in a network by checking whether the syndromes are all zeros or not. After that, all the syndromes that are non-zero and that are inside the scope of the error decoding capacity of the linear network coding are disposed of and the genuine message can be recovered again. Hypothetical investigations and the simulation consequences of proposed protocol LPA-NC demonstrates that the performance of the overall network increments significantly alongside the increment of the computational overhead.

Keywords: network coding, error detection, error correction, null keys, throughput efficiency.

1. INTRODUCTION

Network coding is another way of communication utilized as a part of network atmosphere to upgrade the throughput, reliability & robustness of a network. The basic idea of network coding is to encode the messages coming from various nodes in a network and then forwards that coded packet to further destination nodes. A destination node successfully decodes the received packets. Network coding is mostly grouped into two structures: Inter session and Intra session network coding. The framework of network coding deals with topology changes, random packet losses and delays [13]. In a directed network, there is a tradeoff between greatest multicast stream rate and computational intricacy given by the network.

The idea of network coding was initially proposed by Ahlswede et al. [2]. Benefits of network coding can be achieved only when there is no any kind of network pollution and all nodes are free of any malicious attacks. The research work on both combating the malicious attacks and network pollution is now centered on linear network coding, presented in Li et al. [5]. It was appeared by Koetter and Medard [4] that to accomplish multicast limit by coding, linear codes are sufficient. A more practical approach to outline linear codes is the utilization of

random linear network codes proposed by Ho et al. [6]. Gkantsidis and Rodriguez [11] demonstrated that how the random network coding standards can be connected to the settings of point –to-point content distribution and by doing this, time taken for downloading a file can be diminished.

Various methods based on linear network coding are categorized into two schemes that are error detection and error- correction. The difference between both these schemas is that in one technique, various errors are identified at the intermediate nodes of a network and in other schema, correction of various errors are made at the sink node. Indeed even, the complicated nature for performing encoding and decoding are unpredictable, hence the error correction based schema appears to be of more interest.

However, there are few constraints of schema based on error-correction, which make error-detection schemes more attractive. To ensure the correctness of network flow, Krohn et al. [7] used homomorphic hash functions. Any packet in the network will be disposed of in the event that it does not pass the check at various intermediate nodes between the source and sink nodes. This scheme is used in random network coding because of reduction of communication overhead. However, the complexity of computations is very high [15]. There occurs high delay because of the computations of too many hash values for a large scale network. Kehdi and Li [12] proposed an error-detection schema depends on the idea of Null keys. The fundamental idea of this schema is to isolate the n-dimensional linear space over GF (q^n) into two orthogonal subspaces that are symbol subspace of measurement k and invalid or null key space of measurement n-k. The null keys schema is more effective as compared to the homomorphic hash function and also has no message delay. Unfortunately, the weakness of this scheme is that every single corrupted packet will be disposed of. A large sized - packet is partitioned into little pieces in a packetized network. If any malicious node corrupts one fragment of the packet, that fragment will be discarded instead of the whole message. By along these, the network transmission proficiency can be shut to zero.

In our work, another schema is proposed (LPA-NC) that consolidates both error-detection and error-correction schema redress to diminish network contamination attacks. Based on the quantity of mistakes per symbol, a coding parameter (n, k) for message encoding is chosen. Whenever any error is detected at the intermediate node, the packet will not be discarded; it will further forwards the message along with the non-zero syndromes. It is possible to recover the corrupted packet by the subsequent nodes, as the length of the errors is within the decoding limit. Our proposed scheme LPA-NC can handle small pollution attacks successfully, where as scheme given by Kehdi and Li cannot handle it.

The real commitments given by this paper are: -

1) In our proposed scheme LPA-NC, Throughput and robustness factors are enhanced when contrasted with the Null keys schema.

2) In LPA-NC, the overhead happens because of calculations is extremely direct.

3) Theoretical investigations and simulation consequences of the LPA-NC are also given to contrast this schema with the Null-keys scheme.

The rest of the paper is composed as: Section II represents work related with this paper. Section III provides the essential ideas utilized as a part of error-detection & error-correction. Section IV represents the proposed scheme. Section V gives theoretical analysis and simulation results. Section VI concludes our work.

2. RELATED WORK

The related work of this paper is about the usage of network coding concept for reduction of the pollution attack within a network. Two categories of Network pollution are error-detection and error-correction schema. Cai and Yeung [3] demonstrates a schema for the correction of errors at the destination nodes. Some coding bounds such as Hamming bound and Gilbert- Varshamov bound are also derived in error-correction schema. Jaggi et al. [9] has given linear codes in the form of two parts of the rate region that further depends on the channel codes provided by

BEC. There is requirement of only sink node for the correction of error; however the pollution spreads throughout the whole network. Krohn et al. [7] provided a scheme that uses homomorphic hash functions for the verification of the messages at various intermediately nodes. Charles et al. [10], follows cryptographic method for discarding the packets. Null keys scheme is provided by the Kehdi et al. [12], in which two orthogonal subspaces are given from which one subspace is available for the valid symbols and other is available for invalid symbols (or Null keys symbols). The distribution of null keys symbols are done randomly among the available nodes. Linear algebra provided that the result obtained is zero when we take the product of two vectors from two given orthogonal subspaces. At each node, the incoming symbol is checked by multiplying it with the null key value it has stored. The result will not be zero if the incoming symbols are illegal. Whenever the distribution of randomly taken null key is done, it can prove that the possibility of malicious node to create a fake node is extremely small. The scheme can only detect the errors and if any error occurs, it will fail in forwarding the packets. In proposed technique, two orthogonal subspaces are considered to implement error-correction code and for the correction of errors, the result of the product of obtained input variables and the null keys are used. One can recover the corrupted message by successfully collecting all the syndromes from the intermediate nodes in a network. By doing this, we can compensate all the defects of the null keys algorithm. Vahid et al. [14] provided a survey paper which describes about network coding and its applications along with various key security assumptions of network coding systems. The taxonomy of various security mechanisms and schemes are also given.

3. DESCRIPTION OF THE ERROR DETECTION AND ERROR CORRECTION

This segment of our paper proposes a scheme that is based on error-correction in which (n, k, d) vector is utilized to signify a binary linear code and also a matrix M_{k*n} is generated, here 'k' refers to rank/dimension of the codeword of matrix and D refers to minimum distance (Hamming distance). Minimum distance between any two different code words a and b of C is $d_{min} = \min \{d(a, b) | a, b \in C\}$

Where,

$$d(a, b) = \min| \{a_i \neq b_i \mid a, b \in C\}| \tag{1}$$

The subspace with dimension 'k' over 'n' dimensional space is prepared by 2^k number of total code words. The distance between two code words should be at least d_{min}. A k-tuple message m helps to find all the code words c, c= m.M.

A parity check matrix is given by:

$$H_{(n-k)\times n} \tag{2}$$

also exists with the generating matrix M, in which rows of H are null keys and are linearly independent & satisfying c are terms as $\{h_1, h_2, \ldots\ldots\ldots\ldots, h_{n-k}\}$. For each code word c, we have

$$c.H^T = 0 \tag{3}$$

Messages at every node are checked by using eq. 1. There are maximum chances that after a few hops of transmission, so the scheme with null keys will be able to detect or enquire those messages which are polluted. Whenever some kind of pollution exists in a network, the "check & dump" setup achieves very low efficiency. Hence, communication efficiency also becomes low.

Assume, c is the code word that is to be transmitted & r = c+e received code word, where a malicious node generates n-tuple error message, given by e. Then according to equation 1, for every received code word r, c is orthogonal to H,

Hence,

$$r.H^T = (c + e).\ H^T = 0 + e.H^T = e.H^T \tag{4}$$

Equation 4 is called as syndrome or pattern of 'e' error pattern which is represented as s and here r is code word iff, s = 0. Whenever s =0, e ‡ 0, the main purpose of utmost likelihood decoding is to check the weight error pattern e which is minimum, so that:

$$r.H^T = e..H^T \tag{5}$$

Here in this situation, the new corrected value of receiver 'r' is r + e = c. Within a network, single or multiple null keys are carried by every node from H. For each null key h_i, we can easily derive syndrome bits,

$$s_i^- = r.h \tag{6}$$

Here the received code word is r_i. Then the nodes again forward the syndromes which are non-zero to the succeeding relay nodes. On the basis of the initial code selection, when any intermediate node collects all the syndromes, it is capable of correcting (d-1)/2 errors by using existing error-correcting schemes [8] and can also detect all polluted nodes within a network.

4. PROPOSED LPA-NC SCHEME FOR NETWORK CODING

In this paper, the given notations of [4], [12] are used and network is represented as directed graph i.e., G = (V, E), in which V represents the set of nodes in the network and E represents the collection of connectives between various nodes. In the proposed technique, consider that given v_i is a node from V. To provide security at a high level, and for providing key distribution, homomorphic hashes are used. Because of the use of homomorphic hashes, the additional calculations overhead for the system is reduced. This is totally random process for assigning the keys. Every v_i node will obtain m average keys.

Assume a network in which the source node chooses the messages from (n, k, d) code subspace 'C'. Consider the matrix to be generated is M. Here ' $C^{\perp}$ ' indicates the null space from the code subspace C and $C^{\perp}$ is the combination of all h vectors for which M.h=0, thus

$$\dim(C) + \dim\left(C^{\perp}\right) = n \tag{7}$$

There are 2^k number of (n, k, d) linear block code C code words. These code words generate a subspace of k-dimensional for the vector space of n-tuples over GF (2). The objective of this decoding technique is to divide 2^n received vectors to 2^k subsets that are disjoint. To achieve this, each subset must contain $2^{(n-k)}$ vectors; the division of received vectors is done on the basis of linear structure of the code.

Theorem1 ([8]), every linear code (n, k) can detect $2^n - 2^k$ error patterns and correct 2^{n-k} error patterns. The pollution location is also contained within the errored pattern.

Theorem 2([1], the singleton bound). If (n, k, d) linear code = C, then there must be,

$$k + d \le n + 1 \tag{8}$$

Theorem 3 ([8]), If the minimum distance of linear code (d_{min}) satisfy,

$$d_{\min} \ge +\tau + 1 \tag{9}$$

Then it is proficient of correcting lambda or smaller amount of errors and also detecting τ (τ + λ). As we consider that every node v_i carries m keys on average, then m syndromes can also derived at each node on average. The received code word is error-free if all the syndromes are zeros. For this case, our approach will behave same as the scheme given by Kehdi and Li [12]. In our proposed scheme, a few of syndromes are non-zeros & the intermediate nodes in the network will forward the syndromes to the supplementary nodes for correcting the errors.

Error decoding must be initiated at the node, where the syndrome bit which is last, is obtained so that the message pollution from the network can be minimized. For practical usage, suppose that the polluter will only be able to corrupt a few number of nodes in a network. If the errors obtained is smaller than or equals to (d-1)/2, then by using maximum, likelihood decoding algorithm ([8]), we can easily calculate the error pattern which is of minimum weight and recover the transmitted code word.

The algorithm of the proposed scheme LPA-NC is described below:

Step 1: Algorithm for Key distribution
Step 2: At the source node S
Step 3: For each $v_i \mathcal{E} V$ do
Step 4: h^T_i – rowvector (H)
Step 5: for j=1 to n-k do
Step 6: if rand(1) < m/(n-k) then
Step 7: $v_i \leftarrow h_i^T$
Step 8: End if loop
Step 9: End for loop
Step 10: End for loop
Step 11: Algorithm for Error Correction
Step 12: At each node v_i
Step 13: for each incoming message r do
Step 14: $r.\{h^T\} = \{s\}$
Step 15: If the size of {s} is in n-k then
Step 16: then decode r and assign the result to u
Step 17: End if loop
Step 18: End for loop

5. SIMULATION ANALYSIS AND RESULTS OF LPA-NC SCHEME

Error control techniques are used to analyze the proposed schemes efficiency. Network coding can be detected by null keys scheme if redundancy is added to the long messages of k-bits. K/n represents the error-control code rate for (n, k, d) linear code. The elevated k/n outcomes in potentially increased code rate and higher communication efficiency. Theorem 1 and 2 states that, if k is big, then d_{min} and number of errors which can be identified and corrected are small. The number of errors that can be detected or corrected is d-1 or lower bound of (d-1)/2.

Following example illustrates the comparison of proposed method with the technique given by Kehdi and Li [12].

5.1. Illustrative examples

Following example describes the comparison of above defined schemes and it also shows the effect of k parameter on the efficiency of transmission.

Example 1: For a (8, 6) code, given $d_{min} = 2$, then here every single-bit error can be detected. Assume that 100 messages are transmitting with 80% probability that only one-bit is polluted. Then, from the perspective of probability, the null keys scheme will drop 80% of the messages. Thus, on the whole the efficiency for null keys technique is,

$$6/8 * 20\% = 15\ \%$$

If one more bit, that is parity checking bit, is added to code word, then new code become (9, 6, 5) linear code. After that implementing our proposed LPA-NC technique, two-bit errors can be detected and any of the single-bit error can be corrected. Again, 100 messages are to be transmitted with 80% that only one-bit is polluted. Then, at the sink node, no messages will be dropped. Thus, the overall efficiency become 6/9 = 66.7%. By doing this, the overall efficiency get increased by 4.47 times.

Example 2: Consider that the null keys technique and our proposed technique LPA-NC; both methods are utilized to transmit the messages using a (4, 2, 2) linear block code. Suppose while the transmitting the messages, 10% of messages gets polluted with one-bit error and 10% gets polluted with two-bit errors. The existing scheme based on null keys will drop all the 20 % messages which are polluted by which on the whole transmission efficiency becomes 2/4 * 80% = 40%.

Our proposed scheme LPA-NC is capable of correcting all single-bit error patterns. Hence, 10%, messages that 2-bit errors are dropped. Thus, the overall efficiency becomes 2/4 * 90% = 45%.

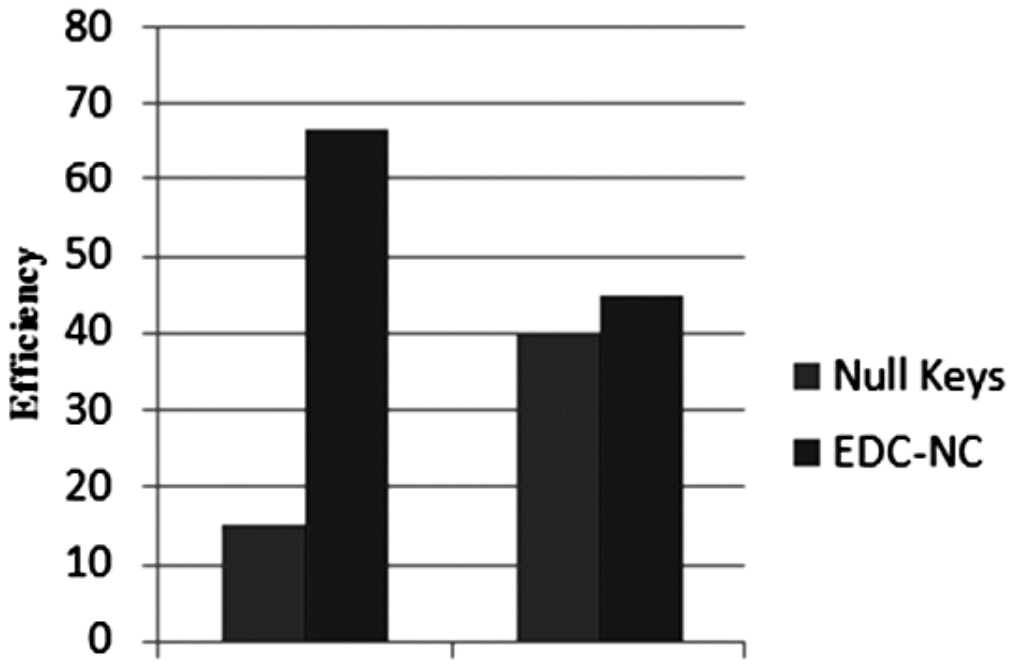

Figure 1: Comparison of Transmission Efficiency

5.2. Results of Simulation

For the simulation point of view both the schemes are adopting the same system setting. The analysis parameter for both the schemes (i.e. PLA-NC and the null key based scheme) is the maximum number of polluted messages forwarded by each hop. Suppose that from the subspace of null key, about m number of random vectors has been carried out by each node. It has been analyzed that when the value of m is increased, the time delay is reduces for the computation of required syndromes. However, larger m will also leads to high energy consumption and computation overhead.

In Figure. 2 and figure.3 , a comparison of proposed LPA-NC and the null key based scheme has been depicted that clearly represents the throughput efficiency of both the schemes. In this scheme each node carries m number of null keys.

Figure. 2 uses a (32, 16, 12) Golay linear code. In this code each node is used to carry six(6) null key values and major part (70%) of the total messages are corrupted with one to three (1-3) errors and minor part of the message i.e. (30%) has been corrupted with four(4) errors. It can be observed the proposed scheme can increase the throughput from 0 % to 70% as compared to the null keys scheme. The previously mentioned null key based scheme dumps all the polluted messages at ten(10) hops, but the LPA-NC scheme corrects all the errors by using only five(5) more steps.

Figure. 3 uses (10, 4, 7) linear code. In this code each node is used to carry two(2) null key values and two-third(2/3) of the total messages are corrupted with single error and minor part of the message i.e. one-third(1/3) has been corrupted with 2 or 3 errors. By using above similar criteria, all the pollution can be eliminated and all the messages that are correct are recovered at about ten(10) another nodes of message forwarding. Thus, throughput can be increases from 0% to 66.7%.

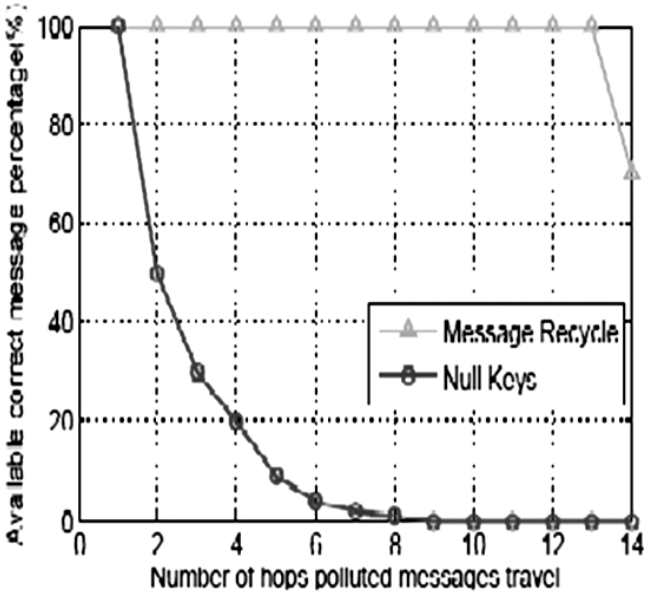

Figure 2: A linear code(32,16,12) depicting the comparison between the proposed scheme LPA-NC and the null key base scheme on the account of Throughput efficiency

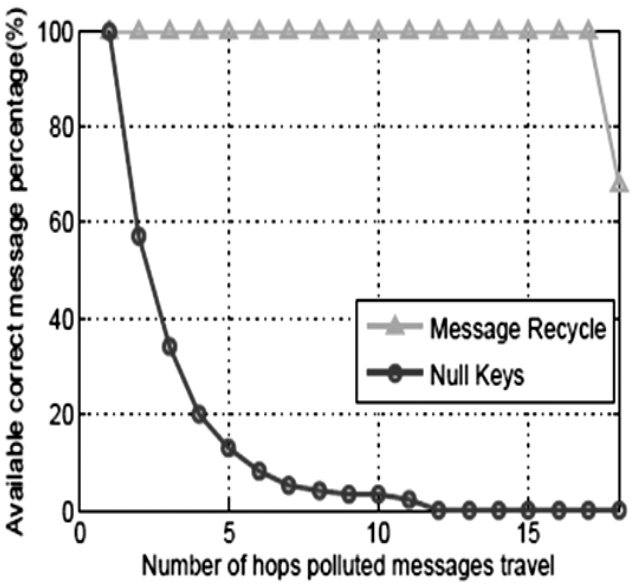

Figure 3: A linear code (10,4,7) representing the comparison between the proposed LPA-NC and the null key based scheme for the throughput enhancement

6. CONCLUSION

The newly proposed network coded technique LPA-NC is quite helpful to limit the network pollution for the robustness of the network. In this scheme the selection of an appropriate linear network coding design reduces the message dropping capability of the technique which is very important in the selection of an error detection correction code. The result shows that, this scheme can efficiently enhance the network performance with moderate network overhead. The proposed method effectively improves the robustness and throughput of the network which interns improve the computational complexity as compared to null keys scheme.

REFERENCES

[1] J Mac Williams and N.J.A.Sloane,"The theory of Error- Correcting Codes", North- Holland, 1977.

[2] R.Ahlswede, N.Cai, S.Y.Li and R.Yeung, "Network information flow", IEEE Transactions on information theory, vol. 46, pp.1204-1216, July 2000.

[3] N.Cai and R.W.Yeung, "Network coding and error correction", in Proc. of IEEE Information Theory Workshop (ITW 2002), pp. 119-122, 2002.

[4] R. Koetter and M.Medard, "An algebraic approach to network coding", IEEE/ACM Transactions on Networking, vol. 11, no. 5, , pp. 782-795 , 2003.

[5] S.Y, Li.R.W.Yeung and N.Cai, "Linear network coding", IEEE Transactions on Information Theory, vol.49, no. 2, pp. 371-381 ,2003.

[6] T.Ho.B.Leong, R.Koetter, M.Medard, M.Effros, and D.Karger, "Byzantine modification detection in multicast networks using randomized network coding", in International Symposium on Information Theory (ISIT), July 2004.

[7] M.Krohn, M.Freedman and D.Mzzerieres,"On-the-fly verification of rate less erasure codes for efficient content distribution", in IEEE Symposium on Security and Privacy 2004, pp. 226-240, May 2004.

[8] S.Lin and D.J.Costello, "Error Control Coding", Prentice Hall, 2nd ed., June 2004.

[9] S.Jaggi, M.Langberg, T.Ho and M.Effros,"Correction of adversarial errors in networks", in Proc. of international Symposium on Information Theory (ISIT 2005), pp. 1455-1459, 2005.

[10] D.Charles, K.Jain and K.Lauter,"Signatues for network coding", in Proc. of CISS'06, pp. 857-863, 2006.

[11] C.Gkantsidis and P.Rodriguez,"Cooperative security for network coding file distribution", in IEEE INFOCOM 2006, pp. 1-3, Apr.2006.

[12] E.Kehdi and B.Li,"Null Keys: Limiting malicious attacks via null space properties of network coding", inIEEE INFOCOM, pp. 1224-1232, Apr.2009.

[13] V.Sangeetha and S.Radhapriya, "Schemes involved in pollution attacks in Network Coding- A Survey", in IJARCET, vol. 3, no.7, pp. 2439-2442, July 2014.

[14] Vahid Nazari and Talooki Riccardo Bassoli," Security concerns & counter measures in Network Coding based communication systems", Elsevier Journal of computer networks, vol. 83, pp. 422-445, June 2015.

[15] Neha Mamidwar and D.B Gothawal," Pollution attack precluding in P2P system with Network Coding using KEPTE scheme", IJIR, vol. 2, pp. 635-641, 2016.