ISSN No. 0976-5697

Volume 4, No. 11, Nov-Dec 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Enhancement of Content Distribution and Verification in Network Coding

G.Sukhaveerji[1], S.P.Anandraj[2], S.Poornima[3], K.Jhonson[4]
Department of CSE
SR Engineering College Warangal, India
sukhaveer.ghate222@gmail.com[1], anandsofttech@gmail.com[2], poornima.spa@gmail.com[3], johnson.kolluri@gmail.com[4]

***Abstract*--**Recently network coding has become popular for distributing content over Peer-to-Peer (P2P) and other networks. The reason behind this is that the network coding makes content distribution in large networks easier. However it is vulnerable to attacks when it is applied directly and recursively as adversaries can inject dummy data to spoil the process of content distribution. Unnecessary content might cause the wastage of network resources. For this reason content verification must be employed while using network coding for content distribution. The existing solutions for content verification are computationally high and cause much communication overhead as well. Recently Li et al. employed new methods to reduce the communication overhead and high computational cost. In this paper we implement those methods and build a prototype application that demonstrates the proof of concept. The empirical results revealed that the proposed methods can reduce the communication overhead and computational cost.

***Index Terms*–**Network coding, content distribution, verification, security

I. INTRODUCTION

Network coding has been around for some years for content distribution over large networks. The benefit of the network coding has been realized by researchersas explored in [1], [2], [3], [4], and [5]. It is used in millions of computers over Internet that is involved in massive content distribution. In this paper, we are concerned with the security of the content distribution schemes available. Maintaining integrity of data is an important concern with respect to network coding for content distribution. There are many problems encountered in it including lack of integrity of data, transmission errors, link failures, hardware and software errors besides attacks launched by adversaries. The main problem is that attackers can inject bogus data into the network when network coding is being employed. This will result in many problems including network slowdown, wastage of bandwidth, data integrity and related relayin data distribution. The classical content distribution "hash-and-sign" approach. In this approach the source node applies collision resistant hash function and computes hash values and use digital signatures for security. The signature is used at the receiver end to verify data. However such methods can't be practically applied in network coding kind of content distribution. First of all in [6] it is explored and observed that network can perform coding to make information transfer faster. Later on many researchers studied the problem both with theoretical analysis and empirical results [4], [7], [8], [9]. These schemes are difficult to use in the real world as they need topology information to implement. This is not a feasible solution for this reason. However, content distribution networks are dynamic in nature as there are changes in topology, failures and memberships.

To overcome this problem "Random Linear Network Coding" is proposed in [10] that enable local nodes making decisions on content distribution and network coding. The original data represented by X is split into many blocks such as x1, x2, …,xn. Each code computes the random linear network coding and forwards to the downstream nodes in the network. The random linear combination is computed as follows.

$$\mathbf{p} = \sum_{i=1}^{n} c_i \mathbf{x}_i$$

The pair of p and c (coefficient) is known as a packet now. When there are some packets obtained, then the node can decode to obtain the original content X. However, the verification of the data is the concern. If the data is not verified adversaries may launch attacks to inject bogus content into the packets. For this purpose, the conventional "hash-and-forward" approaches will not do. Secure Random Checksum (SRM) proposed in [11]which is efficient but provides less security as it depends on the parameters specified by users. Based on observations we made we determined to use KFM scheme. There are two problems in using the scheme. First one is computationally expensive while the second one is communication overhead. The first problem is addressed using homomorphic hash functions while the second problem is addressed. The second problem is addressed by analyzing the parameters of the system. The remainder of the system is organized into the following sections. Section II reviews literature. Section III provides information about homomorphic hash function. Section IV provides details about integrity verification scheme. Section V presents experimental results while section VI concludes the paper.

II. PRIOR WORKS

There are many applications of network coding. Maximum capacity is expected in any kindof network. The maximum capacity between the source and sink in a network is the maximum capacity of the network flow carried out between them. Let us take a directed acyclic graph with unit capacity edges, and "f" is known as mincut of the graph. When there are multiple links with a single source node maximizing network capacity performance may not be possible. In [6] it is revealed that network coding is possible to maximize the performance in content distribution.

$$h_i = \mathcal{H}(\mathbf{x}_i) = \prod_{j=1}^{m} p_j^{\alpha x_{i,j}} \mod p$$

The experimental results gave more new insights into the possible networking with respect to coding. Li et al. [12] demonstrated the performance of network coding through intermediate nodes. It is sufficient to use linear network codes to achieve maximal capacity in networks. Each node computes, from its upstream nodes, compute linear combination of information received. However, this scheme should know the underlying topology before applying network coding. This makes it fixed and static while processing content distribution. However, this kind of scheme is not viable for dynamic situations. Linear network coding [13], [14] also considered by extending the scheme provided in [12]. It also considered link failures in the experiments. Checking polynomial identity is the problem in case of the scheme proposed in [15].

III. SPARSE LINEAR NETWORK CODING

The overhead with respect to computation has two parts. They are cost incurred by the verification process and the cost incurred by computation of random combinations of data blocks. To reduce the computation cost, the proposed sparse linear network coding has the following steps.

Step1: choose packets randomly.Let the packets be $(\mathbf{x}_1, \mathbf{c}_1), \ldots, (\mathbf{x}_\theta, \mathbf{c}_\theta)$

Step 2: randomly choose $r_1, \ldots, r_\theta \in \mathbb{Z}_q$.

Step 3: Compute packet (x, c) as

$$\mathbf{x} = \sum_{i=1}^{\theta} r_i \mathbf{x}_i, \quad \mathbf{c} = \sum_{i=1}^{\theta} r_i \mathbf{c}_i.$$

The communication overhead is reduced by suing three factors as part of the content delivery scheme. The factors considered are random coefficients, hash values and the cost of distributing the parameters. More details of the scheme can be found in [16].

IV. PROTOTYPE IMPLEMENTATION

The prototype application is implemented in Microsoft .NET platform. Visual Studio 2010 is the IDE used for the development. The programming language used is C#. The environment used is a PC with 2 GB RAM, core 2 dual processor running Windows XP operating system. The main screen which demonstrates the concept of network coding in content distribution is as shown in figure 1.

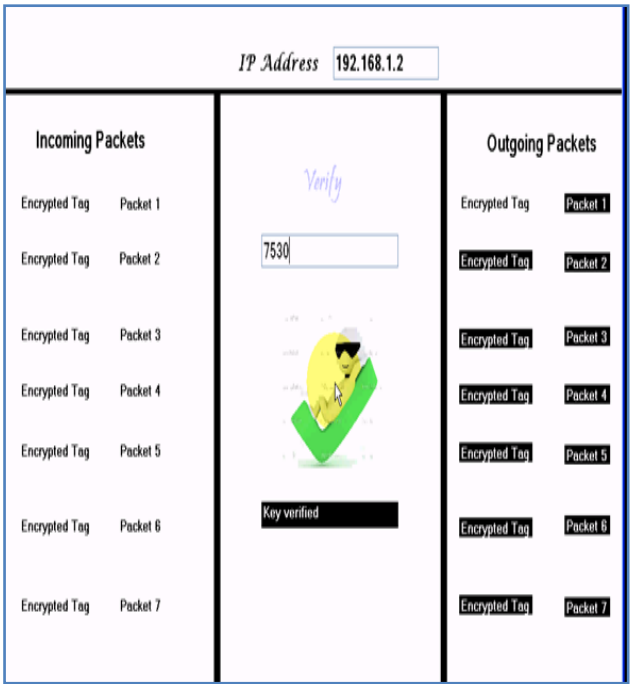

Figure. 1 –One of the interfaces of the prototype implementation

As seen in fig. 1 the application simulates the incoming and outgoing packets while demonstrating the proposed scheme for network coding. The source node and destination nodes along with intermediate nodes are simulated in order to demonstrate the process of network coding. When the verification code is given correctly at the destination, the verification of data integrity takes place.

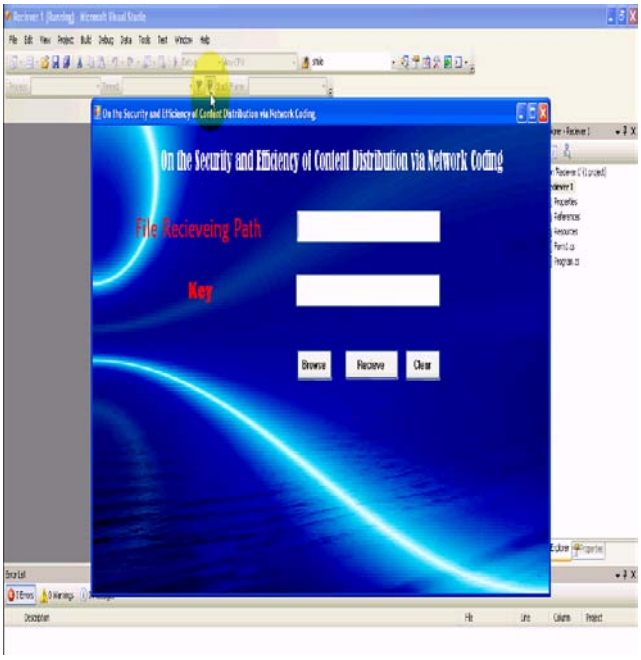

Figure. 2 – File Receiving Path and Key

As seen in fig. 2 the application simulates the incoming and outgoing packets while demonstrating the proposed scheme for network coding. The source node and destination nodes along with intermediate nodes are simulated in order to demonstrate the process file receiving path and key genetared. When the verification code is given correctly at the destination, the verification of data integrity takes place.

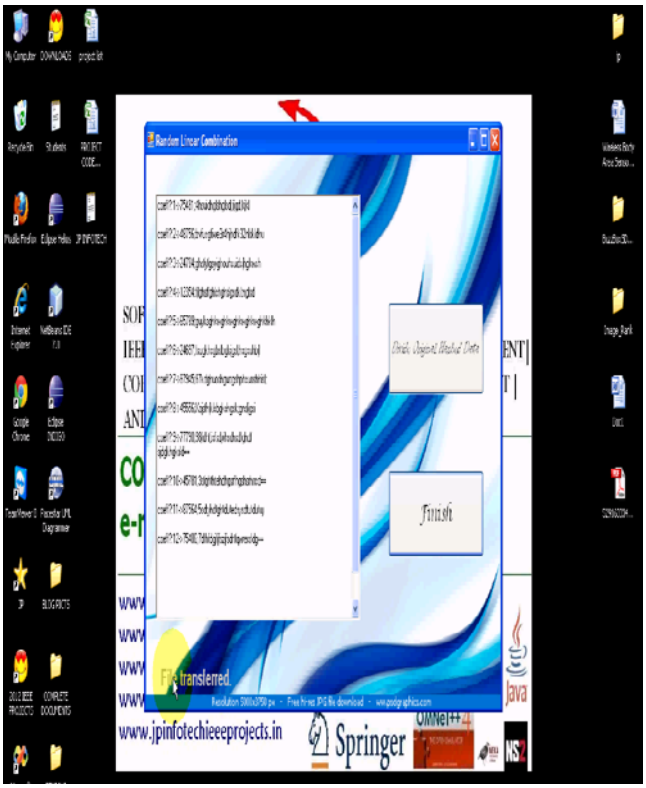

Figure. 3 – Random Linear Combination

As seen in fig. 3 the application simulates the incoming and outgoing packets while demonstrating the proposed scheme for network coding. The source node and destination nodes along with intermediate nodes are simulated in order to demonstrate the process random linear combination.

V. EXPERIMENTAL RESULTS

Experimental results are presented in this section. The computation cost is the combination of hash values and random combinations. For each block when sparse random linear coding is applied for each combined block the proposed computation is more efficient when compared with hash values.

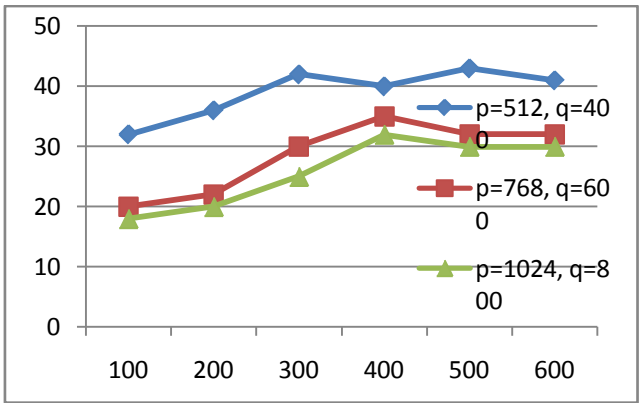

Figure. 4–Computational Efficiency of H1

As can be seen in fig. 4, the horizontal axis represents number of sub blocks while the vertical axis represents throughput. The results differ when p and q values are changed. The throughput is high when p value is 512 and q value is 400. It is least when p value is 768 and q value is 600.

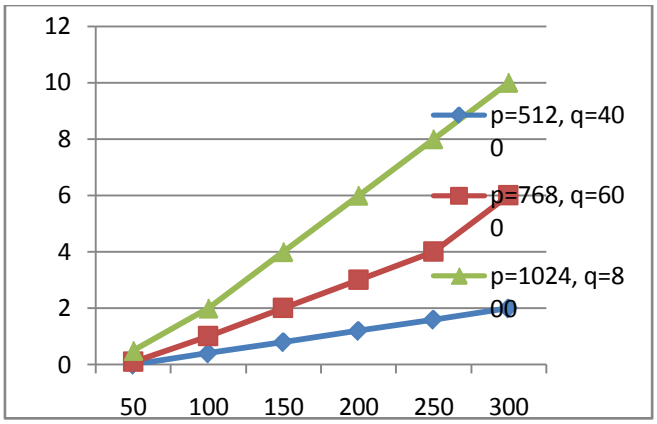

Figure. 5– Computational Efficiency of H2

As can be seen in fig. 5, the horizontal axis represents number of sub blocks while the vertical axis represents throughput. The results differ when p and q values are changed. The throughput is high when p value is 1024 and q value is 800. It is least when p value is 512 and q value is 400.

VI. CONCLUSION

This paper implements the new methods proposed by Li et al. [16] for reducing communication overhead and computational cost for verification process in network coding employed for content distribution. The successful application of network coding has been shown in [17] and [18]. This paper focuses on the security and efficiency issues pertaining to content distribution in large networks. On the fly verification of the network coding process is essential. In this paper we give importance to the data integrity verification when the data is on transit. Our functionality is based on faster homomorphic hash function. We also focus on various factors that cause communication and computation overheads. For this we use sparse variant of network coding. We built a prototype application that demonstrates proof of concept. The empirical results revealed that the proposed methods are effective.

VII. REFERENCES

[1]. Y. Zhu, B. Li, and J. Guo, "Multicast with Network Coding inApplication-Layer Overlay Networks," IEEE J. Selected Areas in Comm., vol. 22, no. 1, pp. 107-120, Jan. 2004.

[2]. M. Wang, Z. Li, and B. Li, "A High-Throughput Overlay MulticastInfrastructure with Network Coding," Proc. Int'l Workshop Quality of Service (IWQoS), 2005.

[3]. C. Gkantsidis and P.R. Rodriguez, "Network Coding for Large Scale Content Distribution," Proc. IEEE INFOCOM, pp. 2235-2245, 2005.

[4]. P.A. Chou, Y. Wu, and K. Jain, "Practical Network Coding,"Proc. Allerton Conf. Comm., Control, and Computing, Oct. 2003.

[5]. S. Acedanski, S. Deb, M. Medard, and R. Koetter, "How Good Is Random Linear Coding Based Distributed Networked Storage," Proc. Workshop Network Coding, Theory and Applications, Apr. 2005.

[6]. R. Ahlswede, N. Cai, S.-Y.R. Li, and R.W. Yeung, "Network Information Flow," IEEE Trans. Information Theory, vol. 46, no. 4, pp. 1204-1216, July 2000.

[7]. C. Gkantsidis, J. Miller, and P. Rodriguez, "Anatomy of a P2PContent Distribution System with Network Coding," Proc. Int'lWorkshop Peer-to-Peer Systems, Feb. 2006.

[8]. T. Ho, B. Leong, R. Koetter, M. Me´dard, M. Effros, and D.R.Karger, "Byzantine Modification Detection in Multicast NetworksUsing Randomized Network Coding," Proc. IEEE Int'l Symp.Information Theory, 2004.

[9]. C. Gkantsidis and P. Rodriguez, "Cooperative Security for Network Coding File Distribution," technical report, Microsoft Research, 2004.

[10]. T. Ho, R. Koetter, M. Me´dard, D.R. Karger, and M. Effros,"The Benefits of Coding over Routing in a RandomizedSetting," Proc. IEEE Int'l Symp. Information Theory, 2003.

[11]. C. Gkantsidis and P. Rodriguez, "Cooperative Security forNetwork Coding File Distribution," Proc. IEEE INFOCOM,pp. 1-13, Apr. 2006.

[12]. S.R. Li, R.W. Yeung, and N. Cai, "Linear Network Coding," IEEETrans. Information Theory, vol. 49, no. 2, pp. 371-381, Feb. 2003.

[13]. R. Koetter and M. Me´dard, "An Algebraic Approach to Network Coding," IEEE/ACM Trans. Networking, vol. 11, no. 5, pp. 782-795, Oct. 2003.

[14]. R. Koetter and M. Me´dard, "Beyond Routing: An AlgebraicApproach to Network Coding," Proc. IEEE INFOCOM, pp. 122-130, 2002.

[15]. S. Jaggi, P. Sanders, P.A. Chou, M. Effros, S. Egner, K. Jain, andL.M. Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction," IEEE Trans. Information Theory,vol. 51, no. 6, pp. 1973-1982, June 2005.

[16]. Qiming Li, John C.S. Lui, and Dah-Ming Chiu, "On the Security and Efficiency of Content Distribution via Network Coding", vol. 9, no. 2, March/April 2012, pp211-221

[17]. J. Le, J.C. Lui, and D.-M. Chiu, "On the Performance Bounds of Practical Wireless Network Coding," IEEE Trans. Mobile Computing, vol. 9, no. 8 pp. 1134-1146, Aug. 2010.

[18]. J. Le, J.C.S. Lui, and D.-M. Chiu, "DCAR: Distributed Coding-Aware Routing in Wireless Networks," IEEE Trans. Mobile Computing, vol. 9, no. 4, pp. 596-608, Apr. 2010.

[19]. Sanjay Anand, and Akshat Verma, "Development of Ontology for Smart Hospital and Implementation using UML and RDF," IJCSI (International Journal of Computer Science Issues,). Vol. 7, Issue 5, pp 7-5-206-212, September 2010. ISSN (Online): 1694-0814.

Short Bio Data for the Authors

Sukhaveerji Ghate, he is pursuing M.Tech (CSE) in SR Engineering College, Warangal, AP, INDIA. He has received B.Tech Degree in Computer Science and Engineering. His main research interest includes Networking.

S. P. Anandaraj received B.E (CSE) degree from Madras University, Chennai in the year 2004, M. Tech (CSE) with Gold Medal from Dr.MGR Educational and Research Institute, University in the year 2007 (Distinction with Honors). Now Pursuing Ph.D in St. Peter's University, Chennai. He has 8 Years of Teaching Experience. His areas of interest are Information security and Sensor Networks He has published papers in International Journal, International Conference and National Conference and attended nearly15 National Workshops/FDP/Seminars etc. He is a member of ISTE, CSI, IEEE, Member of IACSIT and Member of IAENG.

S. Poornima received B.Tech (IT) degree from Anna University in the year 2005, M. Tech(CSE)from JNTUH form 2013. She has 6+ years of experience in teaching field. Her areas of interest include Neural Networks and Wireless Sensor Networks. She has published research papers in various National and International Journals, National and International Conferences. She also attended many National Seminars/FDP/Workshops Etc., She is a life member of ISTE.

K.Johnson received B.Tech (CSE) degree from KITS Warangal, M.Tech (SE) from VNRVJIET Hyderabad. He has 6+ years of experience in teaching field .His areas of interest include software engineering, data mining, computer networks, Neural Networks and Wireless Sensor Networks. She has published research papers in various National and International Journals, National and International Conferences