Improving Security of Content Distribution Using Hash Function and Random Linear Network Coding

Akshay Chindarkar, Amey Gujar, Swati Shishupal
Department of Information Technology
Atharva College of Engineering, University of Mumbai, India.
akshay.chindarkar@gmail.com
gujarameya905@gmail.com
shishupal.swati@gmail.com

***Abstract* - The content distribution network (CDN) provides solutions for efficiently and effectively managing content of all types and it also includes core services such as management, and distribution of content. Content distribution networks face many common security and privacy threats like Data modification, Sniffer attack, DDoS attack, man-in-the-middle attack, IP address spoofing, Etc. Therefore, content verification is an important and practical issue in content distribution. When random linear network coding is used, it is impossible for the source of the content to sign all the data, and hence the traditional "hash-and-sign" methods are no longer applicable. We explore this issue further and propose methods to help reducing both the computational and communication cost of content distribution network and provides provable security at the same time.**

Keywords - **Random Linear Network Coding, Hash Function, Content Distribution, Security**

I. INTRODUCTION

For the past few years, there has been an increasing interest on the application of network coding on file distribution. Many researchers have considered the advantages of using network coding on P2P networks for file distribution and multimedia streaming [1] while other researchers have considered using network coding on millions of PCs around the Internet for massive distribution of new OS updates and software patches (For e.g. Avalanche project). What we are interested in is improving security and efficiency aspect of content distribution using network coding.

Network Security is used to provide security to the authorized data which is being distributed from source to destination in the network. It also prevents unauthorized access of data by developing a secure network using security features like access control, confidentiality, authentication, integrity, non-repudiation. Some of the common internet attack methods used to modify the authorized data are eavesdropping, viruses, worms, Trojan, IP spoofing, denial of service. To prevent data from such attacks we use technologies of cryptographic systems, intrusion detection systems, anti-malware software, firewall and Anti-Virus Scanners and secure socket layer. Current development in the network security is the biometrics and smartcard which greatly reduces the unauthorized access of secure systems. Even though we provide high security to the data, there is still possibility of hacking the data. Thus achieving 100% security in the network is not possible.

In existing system, content is sent from the source to destination. By applying hash technique we get hashed content and key. Same key is used for the same content every time. Hashed content is sent to the destination through the centralized server. If the destination does not have the capacity of storing the content which is received from the source then both transmission rate and delay will be too high. Thus by sending same key for the same content unauthorized users easily modify the content.

II. DEFINATIONS

Content distribution is the process of transmitting the messages or data from source to destination. Content Distribution is the act of sharing or circulating content with other websites, directories, or users. Content Distribution is a great means for product companies to circulate their products through various online means. [4]

Network coding is the set of techniques or algorithm for giving security during transmission via networks. Network coding is a technique which can be used to improve a network's throughput, efficiency and scalability, as well as provides better resistance to attacks and eavesdropping, as compared to OSI model or TCP/IP model. [5]

Peer to Peer network is also known as distributed network that interconnects number of systems within the network. It is defined as one computer in the network can act as a client or server for other computers in the network allowing shared access files and other resources such as peripherals and sensors without the need of central servers. [7][8]

Content verification means verifying the contents with its strength to check whether the received content is modified by unauthorized users.

III. LITERATURE REVIEW

April 2005, S.Acedanski, S.Deb, M.Medard, and R.Koetter, Multiple storage locations are available but limited space is consumed. Each storage location chooses a part of the file without the knowledge of what is stored in the other locations. The problem is storing a large file in a distributed manner over a network.

May 2004, M.N. Krohn, M.J. Freedman, and D.Mazieres, The quality of peer-to-peer content distribution can suffer when malicious participants intentionally corrupt content. Using simple block-by-block downloading we can verify blocks with signatures and hashes, but not useful when we use rate less erasure codes.

1998, M. Bellare, J. Garay, and T. Rabin, Many tasks in cryptography call for verification of a basic operation like modular exponentiation is simple and slow.

2005, M.Wang, Y.Zhu, B.Li, Large Volume of data in the overlay network seeks to design and implement the best strategy to disseminate data.

IV. PROPOSED MODEL

In the proposed System, we use three techniques to maintain integrity of the content being distributed from the source to the destination. First we use Random linear network coding [2] which is used to split the content and to store the content in different storage locations randomly. Thus by splitting the content the transmission rate and delay will be less and network traffic will also be avoided. Secure Hash Algorithm (SHA) hash function is the second technique used to hash the splitted content and to generate the hash values randomly. After that Hashed values and Original file are separately but randomly sent through the distributed network. When destination receives the original file, it repeats the procedure done at source to obtain separate hashed values from original file. Finally we use Data verification/ Hash matching method. In which Data verification/ Hash matching process verifies hashed values received from source and hashed values generated at destination to check integrity of received data, If those hashed values are verified/matched successfully then Original file is saved at destination otherwise system generates an error and discards the received contents.

In Fig.1, We are using three algorithms Random Linear Network Coding, SHA Hash Function and Data Verification/ Hash Matching Process.

A. Random Linear Network Coding

Random Linear Network Coding is technique that helps to improve a network's throughput, efficiency and scalability, and also enhances resilience to attacks and eavesdropping. Instead of simply relaying the packets of information they receive, the nodes of a network take several packets and combine them together for transmission. This can be used to achieve the maximum possible information flow in a network. [6]

B. Secure Hash Algorithm (SHA) Hash Function

In this the blocks of previously splitted content are hashed and hashed value is generated. Hashed content values are sent to destination through the distributed network and generated hashed contents values are also stored in the source. [3]

C. Data Verification/ Hash Matching Process

The Data verification/ Hash matching process is done check the integrity of received data. In the data verification/ Hash Matching process the hashed content strength values of original file which are generated at destination and the hashed content strength values sent separately from source are verified to check whether the content is modified during distribution by unauthorized users. If the received data is modified then system gives warning and discards the received file and if it is not modified and both separate hashed values match each other i.e. passes the verification process then proposed system saves original content/ file at destination computer.

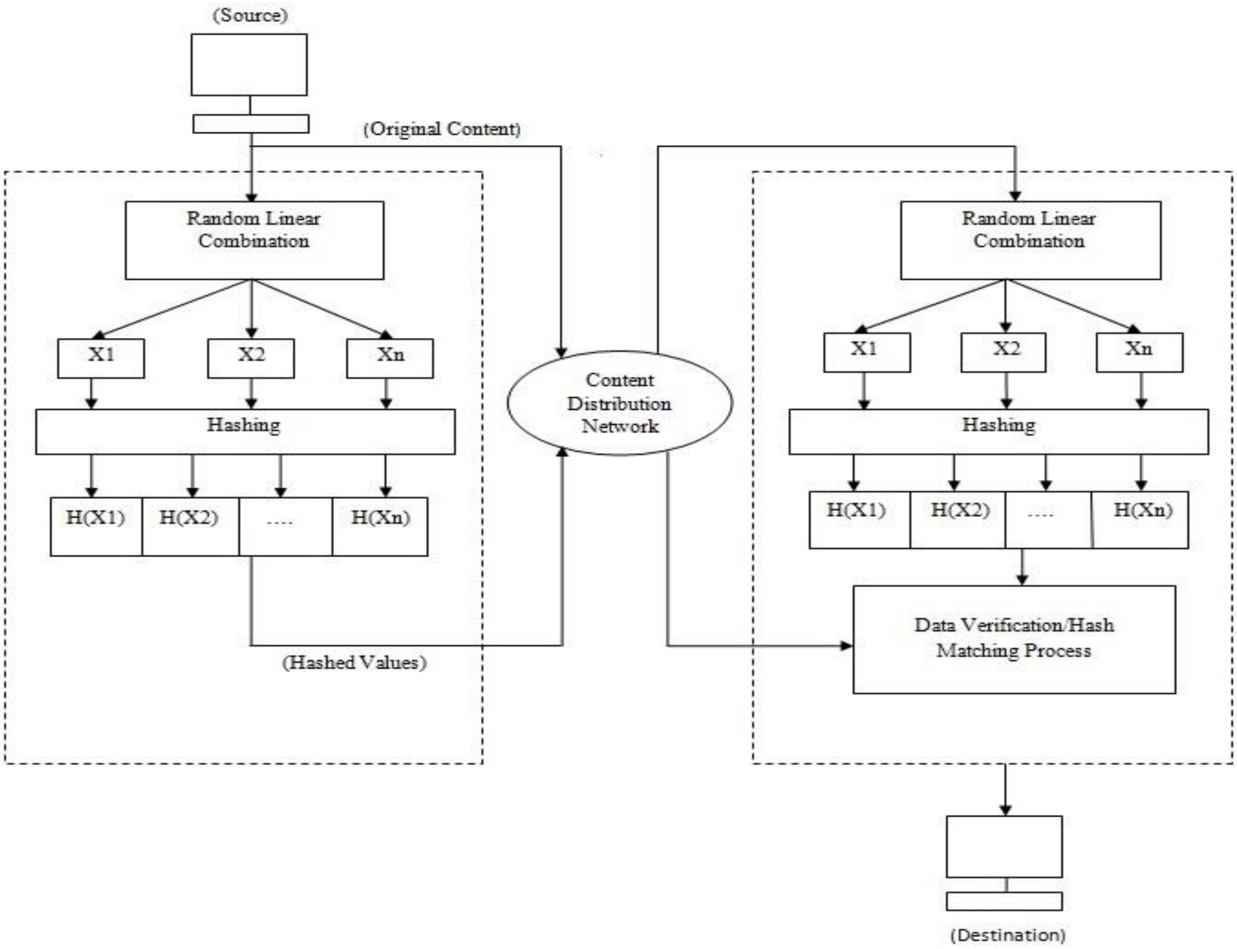

Fig. 1 Proposed System

Step 1: In the source, original file content are split into n parts by using Random Linear Network Coding.

Step 2: Hashing is applied to splitted content. Hashed content of splitted blocks will be generated.

Step 3: Hashed content and original file is sent through the network separately.

Step 4: When original file is received separately at destination, procedure in Step 1 and Step 2 are applied to original file to get separate hashed content of original file at destination.

Step 5: The hashed content received from sender and separate hashed content generated at destination are sent to Data verification/ Hash Matching Process.

Step 6: If both hashed content doesn't match then system discards the received contents.

Step 7: If both hashed contents match each other, then system gives option to save file at destination.

V. CONCLUSION

Network coding is used to improve system throughput or peer to peer networks to improve overall system efficiency. Use of hashing function technique helps in maintaining Security/Integrity of the contents. In proposed system we investigate security and efficiency issues in large content distribution based on network coding. Our proposed system provides security and also helps in reducing network traffic and delay in distributed network environment.

ACKNOWLEDGMENT

This paper describes research done at Atharva College of Engineering in department of Information Technology. We express our gratitude to our project guide Ms. Swati Shishupal for guiding us. We are eager and glad to express gratitude to Head of Dept. Prof. Jyoti Chinchole and all the Project coordinators. We would like to deeply express our sincere gratitude to our respected principal Prof. Dr. Shrikanth Kallurkar and the management of Atharva College of Engineering.

REFERENCES

[1] 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.

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

[3] Chu-Hsing Lin, Chen-Yu Lee, Yi-Shiung Yeh, Hung-Sheng Chien, Shih-Pei Chien, "Generalized secure hash algorithm: SHA-X" EUROCON, 2011.

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

[5] S. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding" IEEE Trans. Inf. Theory, vol. 49, no. 2, pp. 371–381, 2003.

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

[7] C.Gkantsidis, Miller, and P. Rodriguez, "Anatomy of a P2P content distribution system with network coding" Proc. Int'l workshop Peer-to-peer Systems, Feb. 2006

[8] R.T.B.Ma, S.C.M. Lee, J.C.S. Lui, and D.K.Y. Yau, "Incentive and service differentiation in P2P networks: A Game Theoretic Approach", IEEE/ACM Trans. Networking, vol.14, no.5, pp.978-991, Oct. 2006.