International Journal of Computer Trends and Technology (IJCTT) – volume 16 number 2 – Oct 2014

Network Coding an Secure and Efficient Process for Content Distribution

A. Purushotham [1], Bhaludra Raveendranadh Singh [2], K.Laxmi [3] Moligi Sangeetha [4]

[1] pursuing M.Tech (CSE) , [2]Principal , [3]Assistant Professor(CSE), [4]Associate professor & HOD (CSE)

[1,2,3,4]Visvesvaraya College of Engineering and Technology (VCET), M.P Patelguda, Ibrahimpatnam (M), Ranga Reddy (D)-501510, India

***Abstract*— Content distribution in network sometimes may be vulnerable henceforth unauthorized users can inject "bogus" data to corrupt the content during its distribution process in order to deplete the network resource. The direct application of network coding may be insecure. Content verification is the important and practical issue when maintaining integrity of the content. While random linear networking coding is used, it is impracticable for the source of the content to sign all the data, and also the traditional methods such as "hash and sign" are no longer applicable. This is achieved by on the fly verification, which employs a classical homomorphic hash function. In this method content is spitted and hashing is applied. Hashed content is sent to the destination through peer to peer network. Key is sent to the destination directly from the source. Without the key unauthorized users finds difficult to modify the content. Hence the content is secured, however this technique is very complex to be applied to network coding because of high computational and communication overhead. We analyzing this issue further by carefully for different types of overhead, and we propose methods to reducing both communication and computational cost, and also to achieving that providing provable security at the same time.**

Keywords: Content Distribution, bogus data, Security, network coding, verification, computational cost, Peer to Peer network

I. INTRODUCTION

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. From the recent years, there has been an increasing interest on the application of network coding on file distribution. Different techniques have considered the benefit of using network coding on peer to peer networks for file distribution and multimedia streaming, while other techniques have considered using network coding around the Internet for massive distribution of network, operating system updates and software patches. The security issues involving in content distribution schemes using network coding, and how to achieve the security efficiently. An important issue while content delivery in distributed network is how to manage the integrity of the data. We have the problems such as link failures, transmission errors, hardware and software faults and unauthorized attackers.

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 resilience to attacks and eavesdropping, as compared to traditional methods of OSI model or TCP/IP model. Network Security is nothing but providing 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 services like access, confidentiality, authentication, integrity, non-repudiation. Bogus data is to insert fake data to the original data by unauthorized users. If the malicious attackers are able to modify the data in transmission, or inject arbitrary bogus data into the network, they may be able to greatly slow down the content distribution, or even prevent users from getting correct data entirely. Some of 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, firewall, intrusion detection systems, anti malware software and Scanners and secure socket layer.

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. The recent technology in the network security is the biometrics and smartcard which greatly reduces the unauthorized access of secure systems. Content verification means verifying the contents with its strength to check whether the received content is modified by unauthorized users. Even though we provide high security to the data there is still possible of hacking the data. Thus achieving 100% security in the network is not possible.

II. RELATED WORK

In existing, content is sent from the source to destination. The traditional content distribution scenarios, data integrity can be checked using a "hash and sign" paradigm. In this the Source employs a collision resistant hash function H

to compute hash values of the original data X and signs the hash value H(X) using a digital signature scheme S with a signing key K.. By applying hash technique we get hashed content and key. The signature $S_k(H(X))$ is then used to verify received data Y. 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. But these methods are not applicable in practical network coding based content distribution schemes.

If the nodes in the network can perform coding instead of simply forwarding information, multiple sinks in a multicast session can achieve their maximum network flow concurrently. This technique is called as network coding. Some classical experiments provide important insights, would be difficult to apply in practice since they require the knowledge of the network topology during code construction, and require the link failures to follow certain predefined pattern for the code to be reliable. Hence content distribution network can be very dynamic in terms of the topology, membership, and failures. Random linear network coding provides a solution to those problems by allowing each node in the network to make local decisions. But traditional "hash-and-sign" techniques cannot be easily applied with random linear network coding and also in the classical digital signature schemes, only the sender can produce the correct signature of any random combination of data. So, the sender would have to compute and distribute the signatures for all possible linear combinations, which is infeasible.

Hence it is very difficult to detecting malicious modifications at intermediate nodes, especially when it is infeasible for the sender to sign all the data being transmitted, is sometimes referred to as on-the-fly fault detection. So it is considered as the problem in the context of large content distribution using rate less erasure codes and proposed a technique using homomorphic cryptographic hash functions. However, both the computational and communication overhead is high. There is need of simpler and more efficient verification method called secure random checksum (SRM) was proposed, which comes at the price of weaker security that depends on the secrecy of user-specific parameters. Although computationally efficient, this scheme poses certain limitations on the distribution scenarios.

Basic Verification scheme:

The on-the-fly detection technique presented in which we refer to as the KFM scheme is illustrated in Fig. 1. In this scheme, the content X is divided into n blocks x1; . . . ; xn, and each block xi is further divided into m sub blocks xi;1; . . . ; xi;m, where each xi;j can be represented by an element in the multiplicative group Z^*_p for some large prime p. A hash function H is then applied on each block to obtain the hash values h1; . . . ; hn. In particular, the hash function uses m generators g1; . . . ; gm € Z^*_p, and the hash value hi of the ith block is computed as $hi=\pi^m_{j=1} g^x ij_j$ mod p. Clearly, the hash function H is homomorphic in the sense that for any two blocks X_i and X_j, it holds that H (X_i) H (X_j) =H (X_i+X_j).These hash values are distributed to all the nodes reliably in advance. It is suggested that the same technique can be used recursively on the hash values until the final hash value is small enough to be distributed without coding. After receiving a coded block X, which is a linear combination of the original n blocks with coefficients C={c1; . . . ; cn}, a node will be able to verify the integrity of X using X, C, and the hash values h1; . . . ; hn, making use of the homomorphic property of H. In particular, the node checks if the following holds

$$H(X) = \prod_{i=1}^{n} h_i^{c_i} \text{ Mod P}$$

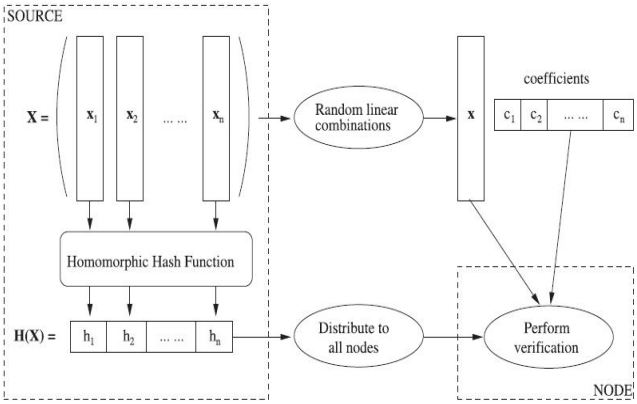

Fig. 1. System Architecture

III. PROPOSED SYSTEM

The proposed system investigates the security and efficiency issues in large content distribution based on network coding. Our proposed scheme consists of two algorithms, namely, the encoding algorithm where the original data are prepared for distribution and the verification algorithm, which is used by individual nodes to verify the integrity of the received data.

Some of the features and advantages of proposed system:

- On-the-fly verification of the integrity of the data in transit.
- Low cost for the computation and communication cost incurred during the content distribution process.
- Able to achieve reasonable speed and the sparse variant performs just as well as the random network coding using typical parameters.

- Completely Secured against traffic analysis in networks.
- In addition, flow tracing or any other such kind of threats cannot be launched networks.
- With homomorphic encryption, the proposed scheme offers two significant privacy-preserving features, packet flow untraceability and message content confidentiality, for efficiently thwarting the traffic analysis attacks.

In this proposed scheme explains the following things

- On-the-fly Byzantine fault detection network creation
- Enhanced Encoding Security Scheme
- Sparse Random Linear Network Coding
- Verification Algorithm
- Attackers.

On-the-fly Byzantine fault detection network creation

In this module first we create the environment of on-the fly Byzantine fault detection network, to propose our technique of An Efficient Content Distribution system via Network Coding using A Faster Homomorphic Hash Function technique. The network creation module will be as shown in Fig: 2.

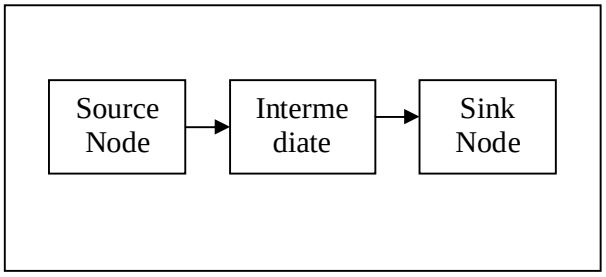

Fig.2. On-the-fly Byzantine fault detection

So, first we create network module with Source node, intermediate nodes and sink node. In this network environment we are going to perform our technique of An Efficient Content Distribution system via Network Coding using A Faster Homomorphic Hash Function.

Enhanced Encoding Security Scheme

Homomorphic Encryption Functions (HEFs) have the property of homomorphism, which means operations on plaintext can be performed by operating on corresponding cipher text. If $(\cdot)$ is a HEF, $(x + y)$ can be computed from (x) and $E(y)$ without knowing the corresponding plaintext x and y.

The main objective of this module (shown in Fig 3) is to protect the message content of outgoing packets from the sender side. The attackers may perform more advanced traffic analysis/flow tracing techniques, including size correlation, time correlation, and message content correlation. Adversaries can further explore these techniques to deduce the forwarding paths and thus to compromise user privacy.

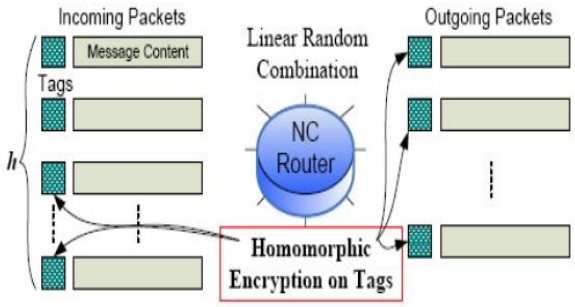

Fig 3: Enhanced Encoding Security Scheme

Sparse Random Linear Network Coding Module

The computation overhead involved in the content distribution consists of two parts. The first part is the cost due to the verification of the packets, and the second part is the cost due to the need to compute random combinations of the data blocks. The preceding sections of this paper focus on the first part of the cost, which can be reduced through the use of more efficient hash functions and batch verification techniques as we have discussed. Nevertheless, the second part of the cost also plays a very important role in practice, especially when the content is large (e.g., in the order of gigabytes), and it has a significant impact on the choice of parameters.

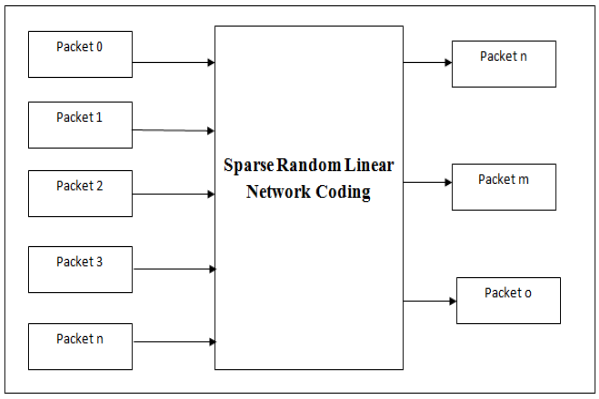

Fig.4 Sparse Random Linear Network Coding

Here, we propose a simple yet powerful alternative to avoid high computation cost when computing the random combinations (as shown in Fig 4). We will refer to this method as Sparse Random Linear Network Coding. The idea is that, instead of computing a random combination of all the n data blocks, we can instead randomly select only one of

them and compute a random combination of only those n blocks.

Verification Algorithm

In this module, we perform batch verification algorithm, where the objective is to verify that the data received at destination is same as that of data sent from source. When the data packets are generated, a coefficient is created. The coefficient value will be checked at the destination. In case, if there are any packet corrupted then the coefficient value is not generated so the verification fails. In case, if there are no packet corrupted then the verification passes.

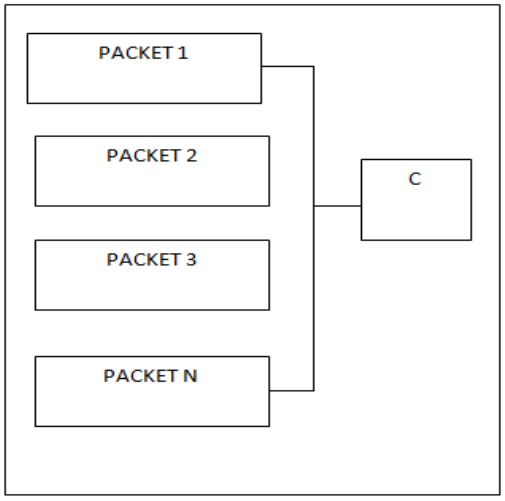

Fig.5: Verification Algorithm

Attackers

The generation number of a packet can be hidden in the secure routing scheme through link-to-link encryption. In this way, attackers cannot find the generation number of a packet for their further analysis. Notice that secure routing paths are only required to be established at the beginning of each session; during the packet transmission, secure routing paths are not required to change or re-established for each new generation. In this module, we show the evaluation of our system by using two scenarios: Normal and attacker. In case, if the scenario is normal then the data content is revealed to the user, in case if the scenario is attacker then the data contents are blocked.

IV. CONCLUSION

In this paper we consider the problem of on-the-fly verification of the integrity of the data in transit. While a previous scheme based on homomorphic hash functions is applicable, it was mainly designed for server side coding only, and will be much less efficient when it is applied on random network coding. We propose a new on-the-fly verification scheme based on a faster homomorphic hash function, and proved its security. We also consider the computation and communication cost incurred during the content distribution process. We identify various sources of the cost, and investigate ways to eliminate or reduce the cost. In particular, we propose a sparse variant of the classical random linear network coding, where only a small constant number of blocks are combined each time. Furthermore, we discuss some possible enhancements under certain conditions of the parameters, and ways to trade off among different cost.

REFERENCES

1) Qiming Li, John C.S. Lui, and Dah-Ming Chiu, "On the Security and Efficiency of Content Distribution via Network Coding", IEEE Transactions On Dependable And Secure Computing, Vol.9, No.2, March/April 2012.

2) Abarna.R, Krupahari.A.L and PraveenKumar.R, Information Technology, Anand Institute of Higher Technology, Chennai " Enhancing the Security of Content Distribution using On-the-Fly Verification via Network Coding" IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 1, March, 2013.

3) 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.

4) 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.

AUTHOR PROFILE

Mr. A.Purushotham is currently pursuing M.Tech in the Department of Computer Science & Engineering, Visvesvaraya College of Engineering and Technology, M.P Patelguda, Ibrahimpatnam (M), Ranga Reddy(D), India. Her research interests include Network Security.

Sri Dr. Bhaludra Raveendranadh Singh working as Associate Professor & Principal in Visvesvaraya College of Engineering and Technology. He obtained M.Tech, Ph.D(CSE)., is a young, decent, dynamic Renowned Educationist and Eminent Academician, has overall 20 years of teaching experience in different capacities. He is a life member of CSI, ISTE and also a member of IEEE (USA).

Ms. K.Laxmi working as Asst. Professor (CSE) in Visvesvaraya College of Engineering and Technology, M.P Patelguda, Ibrahimpatnam (M), Ranga Reddy(D), India.

Ms's. Sangeetha M working as Assoc. Professor & HOD (CSE). She has completed bachelor of technology from Swamy Ramananda Theertha Institute of Science & Technology and Post-graduation from Jawaharlal Nehru Technological University,Kakinada campus and is having 12 years of teaching experience.