Improving Security in P2P File Sharing Based on Network Coding for DTN

Vishal Parikh
Institute of Technology
Nirma University

Kajol Markana
Institute of Technology
Nirma University

Vijay Ukani
Institute of Technology
Nirma University

Malaram Kumhar
Institute of Technology
Nirma University

Abstract: Peer-to-peer file sharing is important due to increasing usage of resources sharing in Internet. This paper targets the background of possible attacks in peer to peer file sharing based on network coding for delay tolerant network. Currently existing security schemes are compared. Comparison shows that a new defense scheme is required to protect the system against pollution attack. Proposed scheme is based on homographic hash scheme and batch verification scheme. It is improvement to the existing spacemac scheme with some added functionalities. In the end, results for the implementation are added.

Keywords: DTN, NC, P2P, security, pollution attack.

.

I. Introduction

Peer-to-peer file sharing using network coding for Delay Tolerant Network (DTN) has brought attention of many researchers because file transmission is very challenging due to the time-varying and unreliable wireless channels. In mobile networking, mobile devices move among various sectors of connectivity, hence providing opportunity to transmit data through wireless interfaces or simply carrying the data toward available connections. To leverage these untapped resources, DTN can be used. In P2P networking it is easier to apply network coding because each node can perform the complex operations like encoding and decoding rather than simply storing and forwarding and it is easier to create the topology for network coding. File transmission in such system is very insecure due to security issues' presence in existing p2p systems, network coding and DTN.so it becomes necessary to develop a secure system for P2P file sharing in DTN.

The rest of the content is organized as follows. Section II continues with background about delay tolerant network, peer to peer file sharing system, network coding, and security issues Section III gives idea of available defense schemes to prevent the attacks having major impact on P2P file sharing system based on network coding in delay tolerant network.. Section IV shows proposed work. In the last, section V shows implementation and results of implemented protocol with existing schemes. Section VI gives conclusion of paper.

II. Background

A. *Introduction to Delay Tolerant Network*

In traditional internet architecture, a main drawback faced during downloading files is the incurred delay. Delay introduced during downloading can hinder the download.it can lead to failure in download. To overcome this, DTN networks are designed. It tolerates unwanted delays, ultimately increasing the potential of the application. DTN works based on new **store, carry and forward** network architecture and protocol suite, which takes a different approach to inter-networking and allows working in highly heterogeneous environments.

In DTN, as discussed by P. Patil and N. Penurkar in [5], routers are replaced with DTN nodes that can store and forward bundles of data. The DTN nodes have a property to store. DTN node have hold the bundle if link is down, until the link is up again. To make sure that all bundles can get through DTN uses store and forward mechanism to handle delay in the end to end path. Sometimes, a packet or a bundle can still be lost even if a router or a DTN node considers a link to be up. For this reason, both the standard internet and DTN has reliability protocols to retransmit missing data. Internet has TCP for reliability (connection oriented-retransmission technique) whereas DTN uses hop by hop custody transfer. For the reliability in DTN, successive nodes take custody of bundles and the last custody has to be retransmitted, if a bundle is lost and custody is not acknowledged by the next node.

B. *Introduction to Peer-to-Peer File Sharing*

Peer to Peer file sharing is a mechanism where nodes can act as a server and client at a same time. There is no centralized server for the content storage and content distribution. Peer to peer file sharing has different components that include peers, seeds, torrent, and many more.

In [2], three principles underlying P2P networks are identified as resource sharing, decentralization and self-organization.

C. *Introduction to Network Coding*

In traditional network, a way to transfer packets to the destination was highly singleton i.e. a single packet transfer approach was used. In DTN, data is aggregated into chunks, and then transferred. This aggregation requires a particular scheme which is well-known as network coding scheme.

DOI: 10.090592/IJCSC.2016.024

When the requirements of throughput improvements and a high degree of robustness in packet networks, network coding is an interesting technique for the same. It uses “store-and-forward” principle of conventional communication networks. Instead of simple forwarding of packets it allows any network node to recombine several input packets into one coded packet.

A typical Network coding scheme consists of two methods:

1) LNC- Linear Network Coding:

Traditional network or components of network like relay node, router etc. follows a mechanism to simply forward the packets it has received. This approach is dumb in nature. A newer technique LNC, follows an aggregative approach. It combines all the packets, that it receives or it generates and aggregates into 1 single outgoing packet, performing multiplication and addition over GF’s. Herein, the linear combination method used is not simple concatenation; because combining packets of length L, would produce an encoded packet of length L. LNC requires a mechanism of coefficients for performing encoding and decoding. Implementing it demands an arrangement of central authority for controlling generation of coefficients. This employs the algorithm to be centralized in nature.

II) RLNC- Random Linear Network Coding:

As discussed by V. Parikh and Z. Narmawala in [3], because of wireless network’s architecture of node mobility and heterogeneity of network, employing distributed structure instead of centralized is rather more suitable. To overcome it, RLNC – an improved technique, uses generation of random numbers to encode coefficient. The wireless network has a property of changing topology, leading to channels having high error rate and higher interference between channels.

Hence, protocols designed for wireless networks must suffice these circumstances.

Table 1: POSSIBLE ATTACKS

ATTACK	*DTN*	*Intra flow NC*	*Inter flow NC*	*P2P system based on NC*
Packet drop	✓	✓	✓	
Bogus packet injection	✓			
Noise injection	✓			
Routing attacks	✓			
Flooding attacks	✓			
Impersonation attacks	✓			
Wormhole attack	✓	✓		
Link quality falsification or modification		✓	✓	
Packet pollution		✓	✓	✓
ACK injection or modification/dropping/delay		✓	✓	
Packet reception information mis-reporting			✓	
Neighbour set pollution			✓	
Packet under-decoding			✓	
Entropy Attack				✓

D. Security issues

DTN gives solution to the problems like intermittent connectivity, asymmetric data rate, long or variable delay and high error rates which generates security issues. Intermittent delays ad long delays create a challenge in routing. To make network resistant to attacks, store and forward mechanism, which is generally applied in DTN, requires secure storage and communication. Study on the possible threats in DTN shows that there exist unavoidable threats in it. Table below shows the possible attacks in DTN.

Dong et al in [1], gives security issues in network coding-based wireless systems. There are two general ways to apply network coding in wireless networks known as intra flow network coding and inter flow network coding. These approaches are dealing security threats which disrupt data delivery process. The table given below shows possible attacks in network coding.

As discussed by P. Patel and J. Bhatia in [4], traditional P2P cooperative architectures suffers from a number of attacks including disrupting the topology. So blocking that legitimate users will disconnect from peers that frequently supply them with corrupted content. Thus, we assume that malicious users will try a mixed strategy in which they serve a mixture of valid and corrupted blocks. Under these assumptions, the P2P cooperative model is vulnerable to Entropy attacks and Jamming attacks. In Entropy attacks malicious clients try to disrupt the diversity of the system and the tit-for-tat exchange balance. While in Jamming attacks where malicious clients try to inject bogus blocks in the content distribution process. Table 1 shows the summary of possible attacks.

Dong et al in [1] states that intra flow network coding has major threats for packet pollution, drop data and drop acknowledgement attacks while inter flow network coding suffers from packet pollution under-decoding and over coding and drop data attacks.

III. DEFENSE SCHEMES

A. Defense against pollution attack

Network coding has one inherent weakness i.e. it is vulnerable to pollution attacks. Defense schemes against pollution attack can be classified in to three categories:

- Error correction:

There exist multiple approaches in error correction categories. But as per A. Le and A. Markopoulou in [11] these approaches are information theoretic and they correct the errors at receiver side. Moreover , these approaches are do not provide security to all type of adversaries .these approaches are developed under assumptions that adversaries will be able to corrupt small number of packets and edges.one scheme suggests to add redundancy at source in order to correct it at receiver, but in increases communication overhead [11]. Such schemes are not capable of detecting corrupted packets.

- *Attack detection*

There exist hash based and signature based approaches for attack detection but these approaches are computationally expensive at intermediate nodes .and that results in to high latency.

Dong et al in [6], has designed a linear transformation checksums which can be used with a time-based authentication scheme to provide in-network detection. This scheme uses public key verification and frequent time synchronization is required among the nodes in network. This scheme suffers from high delay in delivery of data and is vulnerable to denial of service attacks [7].

Agrawal and Boneh in [8], has constructed a homomorphic MAC scheme based on cover free set systems which pre distributes keys to provide in network detection. But this scheme is resistant to c-collusion attack only. Moreover it is vulnerable to tag pollution attack.

Li et al in [9], proposed a scheme known as RIPPLE which is based on homomorphic MAC and it is capable to resist against collusion attack. This scheme is inspired by TESLA and uses time asymmetry. It works only for fixed directed acyclic graph. Scheme proposed in [11] also suffers from same drawback. Addition to it this scheme require spacemac as its MAC algorithm.

Kehdi and Li in [10] proposed a scheme based on NULL KEYS. Intermediate node checks if it belongs to the subspace spanned by source vector. Null keys are used for this verification.

Le Proposed a scheme based on TESLA and inspired by RIPPLE.so this scheme used combination of both homomorphic Mac and time asymmetry. In contrast to RIPPLE, it pre-distributes the tags. This scheme is resistant to collusion attack and is capable of in network detection of attack.

- Attacker location identification:

Jafarisiavoshani et al. in [12] took advantage the subspace properties of random network coding to approximate attacker's location attackers. This scheme works only with fixed directed acyclic graph.

Wang et al. in [13] suggested a non-repudiation protocol, this technique distributes multiple checksums (of all the blocks sent by the source) to all the peers when an attack is detected, which incurs significant communication overhead.

Le et al. [7] has proposed a scheme which give higher security in less per byte overhead then [13] with the use of TESLA while avoiding the need of checksum dissemination.

IV. PROPOSED WORK

Based on the study of these existing schemes, there are some conclusions are derived. These general schemes are not applicable for the application of delay tolerant network. So development of secure scheme specific for delay tolerant network is required. Because delay tolerant network do not need full functionality of digital signatures, some of the functionality can be removed from conventional digital signature schemes.

One of the suitable approaches for making such scheme is to include batch verification along with some functionality of MAC based authentication algorithm.

In network coding scheme, two most prominent attacks are i) Pollution Attack ii) Packet Drop Attack.

Various mechanisms have been devised to design a good peer-to-peer network coding technique. One of the two most used methods is MAC based authentication and Batch Verification technique. In the proposed method, a mixture of both these techniques has been combined.

Batch verification model is expensive in terms of resource utilization but MAC based scheme require public key infrastructure to authenticate the source. However source authentication is not required in DTN so that functionality can be excluded.

A pollution attack is a typical malware attack, in which the attacker injects corrupted packets to the outgoing link of the network and combines with legitimate packets in the downstream nodes. This prevents decoding of the legitimate data, and ultimately runs for a degradable performance.

Hence, it is important to keep a track of the packets being sent at the receiver node. Thus MAC based algorithm is used for signing in of the intermediate packets for keeping account of the expanding of the packets over their time subspaces. The working of the algorithm is as follows.

In the parent's space, if the child has a provision to choose any randomly generated vector yn, then the parent node reverts back to a legitimate child node with a tag y. The entire reception of the tag and its verification is based on homomorphic MAC calculations, to ensure correct tag calculation at child's end.

The MAC based scheme proposed here, is based on Spacemac. It is customized into 3 polynomial time algorithm: 1) Mac 2) Combine 3) Verify. Let K and I denote the key domain and spaces of the sources respectively.

1) MAC (k, id, y)
   - Input to the algorithm: Input key k, source identifier id and vector y.
   - Output: Tag t

2) Combine ((y1, t1, a1) up to (Yp, tp, ap))
   - Input: y1 up to yp are p vectors, t1 upto tp are respective tags to p vectors, k is key, coefficient a1 up to ap are coefficients.
   - Output: tag t for y.

3) Verify (k, id, y, t)
   - Input: k is shared key, id is identifier, y is vector, t defines tag.
   - Output: 0 for reject, 1 for accept.

     The basic difference between proposed scheme and available scheme, is the proposed algorithm uses a fixed space and generates tags for the entire space.

The batch verification process is a fully imparted process i.e. Either it is executed completely or it doesn't initiate. It is executed for the verification stage of the peer-to-peer network system. The necessary condition for it is, none of the block must be corrupted. But, this approach has some disadvantages:

- Due to certain circumstances, if the verification of a block fails, there is no mechanism to inquire on the failed node or peer.
- In a real-time scenario, there are many corrupted packets. Hence, discarding the entire batch due to a single corrupted packet, damages the bandwidth.
- Hence, to maximize bandwidth efficiency, the peer must keep as many non-corrupted blocks as possible.
- To avoid the above problem, another approach is verification of each block individually.
- This may make up for a heavy cost. Because, as the number of blocks may increase, the cost increases too. Hence, this approach may be efficient but not a real time feasible solution.

In the proposed scheme, an efficient way to find the corrupted blocks is constructed.

- At the first step, a bisection search is performed. Hence, the entire sub-batch is divided into two parts. Each part is tested and verified individually.
- The entire process is performed on both the blocks individually and the process continues to find the corrupted blocks.
- Here, in both these blocks, process is being repeated and all the corrupted blocks are collected.
- This approach works best, when the number of corrupted blocks is small compared to size of batch.
- The space complexity is a logarithmic complexity of blocks, composing a batch.

However, this approach cannot be best justified if the number of blocks are more or the communication is a bit fast. Hence certain amendments were needed.

- Before transmission of each block, each block is verified independently.
- Like, if a particular node forwards a packet to the other node, and if it suspects to be malicious, then it is concentrated into smaller subsets.
- Those, sub-sections which verify correctly are passed on or moved further for processing. While those suspected to be corrupted are again processed for detection using a bi-section approach.
- This technique maintains a good balance between bandwidth efficiency and computational overhead.
- This form of partitioning is neighbor dependent. Hence, its complexity grows linearly with the increase in neighbors.
- This form of partitioning is called naïve partitioning.

Algorithm for MAC based scheme with improved batch verification scheme is given below:

```
if running time = 0 AND node = source then
    Generate session id ()
    Transmit session id()
else if running time !=  0 AND node = source then
    Transmit EncodedMessage()
else if running time != 0 AND node = intermediate OR     receiver
then
    Flag = MAC V erify ()

    if Flag = 1 then
    Batch V erification()
    else
    MAC Combine()
    Transmit EncodedMessage()
    end if
end if
```

V. IMPLEMENTATION AND RESULTS

A. *Implementation*

Implementation is done using NS2. Protocols are implemented in C++.

B. *Simulation environment*

Parameter	***Value***
Simulator	*NS-2*
Channel type	*Channel/wireless channel*
Radio propagation model	*Propagation/two way ground wave*
Network interface type	*Phy/wirelessphy*
Mac type	*Mac/802.11*
Interface queue type	*Drop tail*
Link layer type	*ll*
Antena	*Antena/omni antena*
Maximum packets	*150*
Area	*700*700*
Simulation time	*150 sec*
No.of nodes	*10-75*
Routing protocol	*NCR*

Table 2: Simulation Environment

C. *Results*

After simulation the algorithm on NS-2, results were compared using the following parameters

- *Average throughput*
- *end to end delay*
- *instantaneous throughput*
- *ratio of decoded packets to received packet*

Results were measured in three different situations.

- Protocol without network coding scheme and with packet drop attacks.
- Protocol with network coding scheme and with packet drop attacks.
- Protocol with network coding scheme and with packet pollution attacks.

The figure 1 and figure 2 shows measured results for average throughput which proves that proposed scheme performs better than without network coding scheme when number of nodes are 30 or more under packet drop attacks and number of nodes are 25 or more under packet pollution attack.

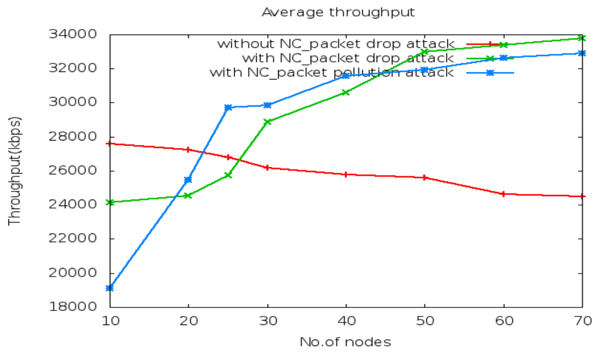

Figure 1

The figure 3 shows measured results for end to end delay which proves that proposed scheme performs almost average in proposed scheme, while end to end delay increases as the number of node increases in the network.

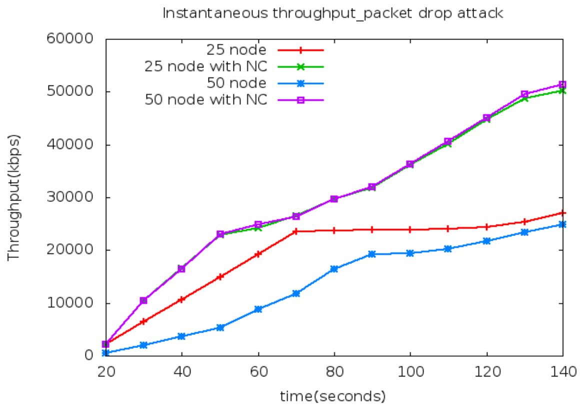

Figure 2

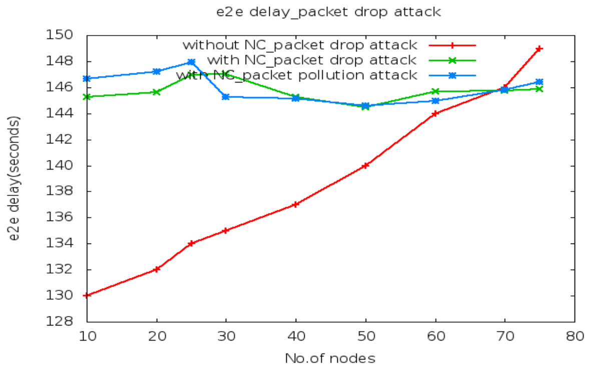

Figure 3

The figure 4 shows measured results for Instantaneous average throughput where it is proven that proposed scheme is good.

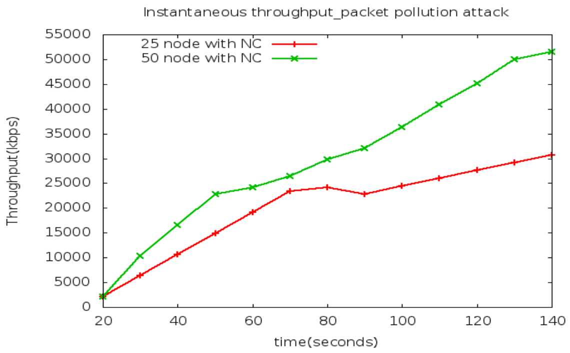

Figure 4

The figure 5 gives idea about decoded packets and received packets in proposed scheme. It shows that packet pollution attack doesn't affect the number of coded packets up to great extent.

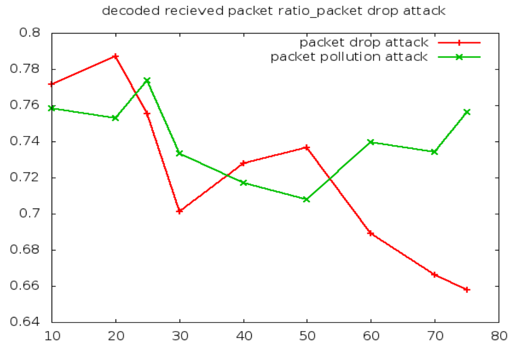

Figure 5

VI. CONCLUSION

Results of the implementation for proposed work shows that compared to existing schemes, proposed scheme is capable of defending against pollution attack and packet drop attacks. However, in depth comparison can be done using other parameters to visualize the result of proposed work.

REFERENCES

[1] Jing Dong, Reza Curtmola, and Cristina Nita-Rotaru. “Secure Network Coding for Wireless Mesh Networks: Threats, Challenges, and Directions.” Computer Communications, pp 1790–1801, Volume 32, Issue 17, 15 November 2009.

[2] Jussi Kangasharju, Keith W Ross, and David A Turner. “Optimizing file Availability in Peer-to-Peer Content Distribution.” In INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, 2007.

[3] Vishal U Parikh and Zunnun Narmawala. “A Survey on Peer-to-Peer File Sharing Using Network Coding in Delay Tolerant Networks.” In IJCSC, pp 74-79 Volume 5 Number 1 March-Sep 2014.

[4] P Patel and J Bhatia. “Review on Variants of Reliable and Security Aware Peer-to-Peer Content Distribution using Network Coding.” In Engineering (NUiCONE), 2012 Nirma University International Conference pp 1-5, IEEE, 2012.

[5] Peeyush Patil and Milind Penurkar. “Congestion Avoidance and Control in Delay Tolerant Networks”. In International Conference on Pervasive Computing (ICPC). IEEE, 2015.

DOI: 10.090592/IJCSC.2016.024

[6] J. Dong, R. Curtmola, and C. Nita-Rotaru, “Practical Defenses Against Pollution Attacks in Intra-Flow Network Coding for Wireless Mesh Networks.” in ACM WiSec’09, 2009

[7] Le, Anh, and Athina Markopoulou. "TESLA-based defense against pollution attacks in p2p systems with network coding." Network Coding (NetCod), 2011 International Symposium on. IEEE, 2011.

[8] S. Agrawal and D. Boneh, “Homomorphic MACs: MAC-Based Integrity for Network Coding.” in ACNS, 2009.

[9] Y. Li, H. Yao, M. Chen, S. Jaggi, and A. Rosen, “RIPPLE Authentication for Network Coding.” in IEEE INFOCOM, 2010

[10] E. Kehdi and B. Li, “Null Keys: Limiting Malicious Attacks Via Null Space Properties of Network Coding.” in IEEE INFOCOM, 2009

[11] A. Le and A. Markopoulou, “Cooperative Defense Against Pollution Attacks in Network Coding Using Spacemac." Selected Areas in Communications, IEEE Journal pp. 442-449 Vol. 30, No. 2, Feb 2012.

[12] M. Jafarisiavoshani, C. Fragouli, and S. Diggavi, “On Locating Byzantine Attackers.” in NetCod, 2008.

[13] Q. Wang, L. Vu, K. Nahrstedt, and H. Khurana, “Identifying Malicious Nodes in Network-Coding- Based Peer-to-Peer Streaming Networks.” in IEEE INFOCOM 2010.

DOI: 10.090592/IJCSC.2016.024