Improving security in P2P file sharing based on Network Coding for DTN

Submitted By

Marakna Kajol Kantilal

13MCEI10

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Improving security in P2P file sharing based on Network Coding for DTN

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science & Engineering

(Information & Network Security)

Submitted By

Marakna Kajol Kantilal

(13mcei10)

Guided By

Prof. Vishal Parikh

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2015

Certificate

This is to certify that the major project entitled **"Improving security in P2P file sharing based on Network Coding for DTN"** submitted by **Marakna Kajol Kantilal (Roll No: 13MCEI10)**,towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Information & Network Security (CSE) of Institute of Technology, Nirma University, Ahmedabadis the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Vishal Parikh
Guide & Assistant Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sharada Valiveti
Associate Professor,
CSE Department
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Marakna Kajol Kantilal**, Roll. No. **13MCEI10**, give undertaking that the Major Project entitled "**Improving security in P2P file sharing based on Network Coding for DTN**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Prof.Vishal Parikh

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Vishal Parikh**, Assistant Professor Computer Science Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Marakna Kajol Kantilal

13MCEI10

Abstract

Applications related to peer to peer file transfers are used widely in wireless network. However, file transfer in wireless network is difficult due to some characteristics of wireless network like unreliable channel, insecure transmissions, node movement etc. Delay Tolerant Network(DTN) can be used to solve issues related to such applications. It has already been proved that opportunistic networking improves efficiency of peer to peer file sharing systems. Previous research works show that network coding can help when network is built arbitrarily. Hence, peer to peer application are best place to apply it. Peer to peer file transfers are performed using end hosts. Thus they have a very good capacity of encoding and decoding. In past, researchers have proved that peer to peer file sharing system based on network coding or its variant over DTN can be built to improve the quality of file transfer applications.

We have carried out this project in two phases. The first phase includes literature survey for security issues which dwells in delay tolerant network, network coding and peer to peer file sharing system based on network coding. It describes there are many possible attacks in DTN which results into inherent presence of them in peer to peer file sharing system for DTN. The researchers have explored pollution attack, packet drop attacks and its defense schemes as their major concern of their research compared to other possible attacks. The comparison of different defense schemes for pollution attack is elaborated in this thesis. We have shown flaws in each scheme and those schemes which can overcome these flaws. Based on this survey, we have proposed one algorithm which provides defense scheme against pollution attack.

Second phase of our thesis shows implementation of proposed work. Following to it, comparison of our proposed scheme with that of the existing one has been shown. Thus our approach has stood out to be a good defense scheme against pollution attack and packet drop attack. The same can be adopted for future work as that enlisted later.

Abbreviations

DTN	Delay Tolerant Network
NC	Network Coding
P2P	Peer To Peer
RLNC	Random Linear Network Coding
GF	Galois Field
G-by-G	Generation by Generation

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides an introduction to delay tolerant network and its impacts. We provide here the scope of work for this thesis and provide background for the same.

1.1 Background

The traditional file sharing systems were based on client-server architecture. Due to this, they suffered from problem like lake of load balancing and robustness. In addition to it, they were centralized in nature. In order to solve these issues, Peer to Peer (P2P) file sharing systems were developed. P2P file sharing can work in decentralized manner. Underutilization of available resources was observed in traditional network. In order to use these untapped resources and increase the efficiency, Delay Tolerant Network (DTN) was proposed which has the capability of tolerating the delay.

1.2 Scope of Work

The undergoing research has been limited to the development of P2P file sharing system for DTN. This file sharing system is based on Random Linear Network Coding (RLNC). The existing systems, which are based on RLNC, have been suffering from various security threats, whereas we have provided a solution to that. Our approach has been built with an improved security against traditional attacks like pollution attack, entropy attack collusion attack etc.

1.3 Motivation

In the last few years, wireless access to the Internet has ended up with accessibility for an extensive number of individuals, utilizing an assortment of diverse sorts of wireless gadgets to connect with the Network. Fifteen years prior to this, Internet access was the right of wired joined machines and mobile phones were gadgets just ready to make phone calls and send/get instant messages. Today, cell phones are getting to be extremely prominent. These gadgets have advanced from extremely straightforward cellular telephones to gadgets ready to scan the Internet, get to and read messages, and watch feature stream straightforwardly from the web.

All these new conceivable outcomes made new issues with respect to the related appeals for connectivity. While the center of the system is very connected and appropriate for routing through conventional routing algorithms, mobile devices regularly work in situations with bases that experience the ill effects of irregular network and trouble in foreseeing changes in topology. In these connections, the traditional routing algorithm intended for wired, highly connected systems lack of efficiency and different methodologies must be embraced.

P2P(overlay) networks are an impeccable networks to apply Network Coding (NC) because of two reasons: P2P network topology is built self-assertively. It is not difficult to create the topology to encourage NC. The nodes in a P2P network can perform more intricate operations, it can encode-decode, save, and retransmit the data.

In spite of their tremendous potential and notoriety, existing schemes, experience the ill effects of various security dangers which makes them vulnerable.

1.4 Thesis Organization

The whole dissertation work has been divided into appropriate chapters as briefed here.

Chapter 2 provides detailed study pursued for understanding the basic concepts of DTN, P2P, NC and Galois Field.

Chapter 3 is a detailed explanation of security issues related to DTN, P2P, and NC. This chapter provides description for the first phase of our thesis.

Chapter 4 contains defense schemes in the existing techniques against various attack. These schemes have already been developed and used, which is useful in understanding the loopholes to overcome them in our dissertation.

Chapter 5 proposes a defense scheme against packet drop and packet pollution attacks. We have used the basic knowledge to understand their limitations and have described how our proposed work can upgrade them.

Chapter 6 has the implementation results and analysis of the same. We have implemented our work in NS2. Its outcomes and comparative analysis have been shown in this chapter.

Chapter 7 is the conclusion of our proposed work and is enlisted with possible future work.

Chapter 2

Literature Survey

This chapter provides an overview for delay tolerant network,network coding techniques and P2P file sharing systems.It also describes Galois field and generation by generation network coding.

2.1 Introduction to Delay Tolerant Network

In conventional web internet architecture a principle disadvantage confronted amid downloading files is the caused delay. Delay presented amid downloading thwarted the process and at last stopped it. To overcome this, DTN networks were designed. It endures undesirable delay, eventually increasing the capability of the application. This is a developing research area in the field of p2p file sharing- DTN. It is new store, carry and forward network network architecture and protocol suite, which takes an alternate methodology to inter-networking allows administration and permits working in stressed and additionally in exceptionally heterogeneous situations. DTNs have the preference of handing-off information with impermanent associations likewise has the limit and capability of associating those territories, in a comparative manner to the postal network which can't be attained by current networks.[1]

Dtns suit long defers between and inside local networks, to accomplish interoperability between the local networks. It obliges mobility and limited power of evolving wireless networks [7] .[5]

Ordinary Network data exchange utilizes routers that transmit data through wired

connections and The UDP protocol is used to send packets from source to destination needs an impeccable way to recover full information, however communication links among routers are not always reliable. Web routers just discard any packets that can't be forwarded on the grounds that the connection is down. This results in data loss between the sender and receiver if the end-toend connection is not flawlessly balanced.

In DTN, routers are supplanted with DTN nodes that can store and forward data bundles. The DTN nodes have a property to store. On the off chance that a connection is down DTN nodes will hold the bundle until the connections are Up once more. DTN uses store and forward mechanism to handle delay at last to end way, so all bundles can get past. Now and then regardless of the possibility that a router or a DTN node considers a connection to be up, a packet can at present be not available. Hence, both the standard web and DTN has dependability conventions to retransmit missing information.[6]

Internet has TCP for reliability though DTN uses hop by hop custody transfer. In the this technique of DTN, progressive nodes take custody of bundles, and if a bundle is lost and custody is not acknowledge by the following node , the last custody is retransmitted.[1]

- The Internets underlying assumptions
  - Continuous, bidirectional end-to-end path
  - Short round-trips
  - Symmetric data rates
  - Low error rates
- The characteristics of evolving potential networks
  - Intermittent connectivity
  - Long or variable delay
  - Asymmetric data rates
  - High error rates

2.2 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 includes peers, seeds, torrent, and many more.

In [7], three principles underlying P2P networks are identified:

Resource Sharing[7] P2p systems include a provision of resource sharing, for example, disk space network transfer speed or services. By imparting resources, applications can be acknowledged which couldn't be set up by a solitary node.

Decentralization[7]**:** As the information is accessible at distinctive nodes however not at any central element that inevitably causes the decentralization in the network. There is no such capable server to answer the queries for the request asked, and that is the primary issue that such a large number of unlawful aspects are associated with peer to peer file sharing.

Self-Organization[7]**:** Because of decentralization there is no more database defenseless to store the data or to recover the data .therefore its a key necessity that relationship toward oneself must be there in the nodes, in view of the data they contain by regional standards and with the collaboration with the neighbor nodes they could figure out the availibility of the content.

2.3 Introduction to Network coding

In conventional network, an approach to transfer packets to the end was exceedingly singleton i.e. a solitary packet transfer approach was used. In DTN, information is aggregated into chunks, and afterward transferred to next hop. This aggregation obliges a specific scheme which is well-known as network coding scheme.

Network coding is an intriguing scheme which improves throughput and provides a very good vigor in packet networks. It breaks with the "store-and-forward" rule of routine correspondence arranges by permitting any network node to recombine a few data packets into one coded packet, rather than basic forwarding.
A typical Network coding scheme consists of two methods:

LNC- Linear Network Coding

RLNC- Random Linear Network Coding

2.3.1 Linear Network Coding (LNC)

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 GFs. Here in, 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.

2.3.2 Random Linear Network Coding (RLNC)

[1] Because of wireless network's design of node mobility and heterogeneity of network, utilizing dispersed structure rather than incorporated is fairly more suitable. To overcome it, RLNC an enhanced strategy, utilizes era of arbitrary numbers to encode coefficient. The wireless network has a property of evolving topology, prompting channels having high error rate and higher obstruction between channels.Subsequently, conventions intended for wireless network must suffice these circumstance

Overview of Random Linear Network Coding (RLNC)

The figure underneath gives a fundamental diagram of the operations performed in a NC system. On the off chance that aim is plan to encode a large file then it ought to be

part into a few blocks, additionally called generation each one comprising of g symbols. On the off chance that the entire file was viewed as one major block, then the computational complexity nature of the encoding and decoding operations would be very high[4].

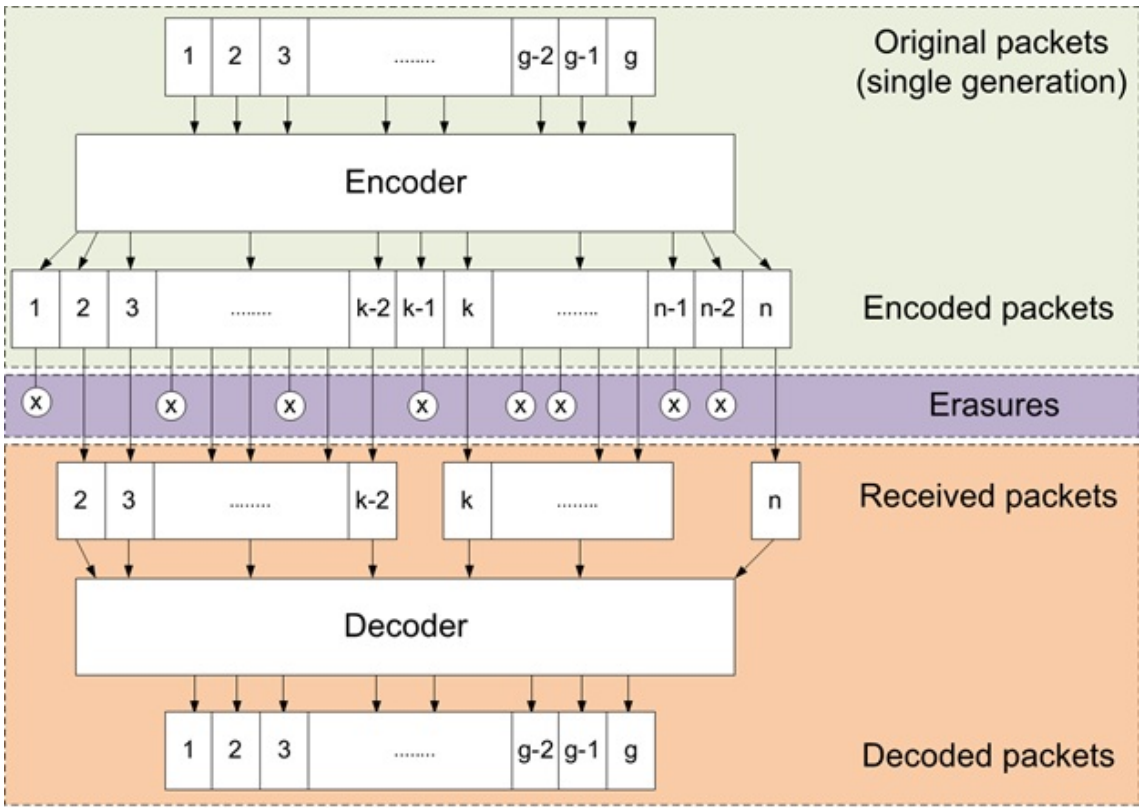

Figure 2.1: RLNC [4]

In the start of the diagram, encoder first produces and transmits linear combination of certain symbols, that are unique. Linear combinations of some symbols may have similar size, because operations like addition and multiplication are performed over a Galois field.For solitary generation, any number of encoded packets may be produced.The middle layer, is a wireless channel, where erasing of packets may happen, depending upon channel conditions. The nodes in the network gets an arrangement of encoded packets that are gone to the decoder (the lower part in the figure) which will have the capacity to reproduce the actual symbol in the wake of getting at any rate g linearly independent packets.

RLNC, works as follows:

- Every node generates its own coding coefficient for every encoded packet.
- These coefficients are then sent to the destination in the packet header.
- This has the advantage to the destination, as it can still decode packet, without the knowledge of network topology, encoding rules even if topology is not fixed.

Results:

- The successful transmissions were measured into two cases:
  - With RLNC.
  - Without RLNC.
- The transmission results showed, that for a distance greater than 500 units, transmissions are very much successful.

In both LNC and RLNC, the overhead increases because of extra overhead of arriving packets. This is due to the mechanism that, transmission of already arrived packets is blocked due to additional packets yet to be arrived.

In RLNC,

- As the nodes increases, congestion decreases.
- Random generation of coefficients, gives high probability a linear independence of the output packets from a node for a sufficiently large size.
- Probability of RLNC in a multicast scenario, is valid is at least $(1 - d/q^n)$[3] , where d is group size i.e. number of destination nodes, q=field size and n number of links.

2.4 Galois field

In customary math i.e. divide, multiplication, addition, subtraction and so forth, operators manage with infinite or unbounded numbers. In any bit by bit error detection, if results are used from computer calculations, then extensive fittings issues may emerge, if register is used to store integer number, . The issue which may emerge is of memory

overflow. To overcome it, Reed Solomon codes were made by manipulation of infinite group of numbers. These number system is called Galois Field, GF(256).This field comprises of all whole numbers in scope of 0-255, in a specific request. The overflow issues are tackled by GF. On the off chance that one could devise arithmetic where the aftereffect of every operation produces an alternate number. The generation(requesting) of the field is key. e.g. a straightforward monotonic series from 0 to 255 is a limited field yet modulo 255 math comes up short commutative tests i.e. certain operations won't invert.A Galois field gf(p) is the component 0 followed by the (p-1)succeeding power of $\rho, 1, \alpha, \alpha^2, \alpha^{p-1}$.

To perform this, an irreducible primitive polynomial is required."Irreducible means it cannot be factored into smaller polynomials over the field." The arithmetic operations like ADDITION and SUBTRACTION in GF(256) can be implemented as following.[8]

```
function Sum(x,y)
begin
result:=x xor y;
end;
```

```
function subtract(x,y)
begin
result:=x xor y; end;
```

```
function multiplication(a,b)
begin
result:=a(x) * b(x) mod p(x);
end;
```

The Multiplication of two numbers are implemented by modulo operation of the two numbers as follows.

The Division of two numbers is a multiplication of dividend and multiplicative inverse of divisor. Ex. 5 2=5 * multiplicative inverse(2)

2.5 Generation by generation network coding[1]

The functionality, provided by the implementation of Galois Field is provided by following example. A node n has to send some packets to a specified destination. Packets are 1,2,3,4. Assuming the generation size to be, all packets are grouped into one generation. An encoding packet, will mix some randomly generated vectors and make four independent copies of mixed packets as shown below. Vector of information is stored in M and generated vectors are stored in G.

$$M = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, G = \begin{bmatrix} 83 & 91 & 151 & 109 \\ 14 & 203 & 121 & 177 \\ 246 & 246 & 63 & 243 \\ 0 & 234 & 173 & 02 \end{bmatrix}$$

Mixed packets are derived from G and M are shown as below :

$$E = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 83 & 91 & 151 & 109 \\ 14 & 203 & 121 & 177 \\ 246 & 246 & 63 & 243 \\ 0 & 234 & 173 & 02 \end{bmatrix} = \begin{bmatrix} 232 \\ 240 \\ 173 \\ 44 \end{bmatrix}$$

E and G would now reach to the intermediate node. Intermediate node will encode these received encoded packets with its own coefficient and generate new coefficient and new encoded packets as shown below. Effective coefficients are determined from received coefficients and newly derived Coefficients as shown below.

$$\begin{bmatrix} 76 & 201 & 166 & 110 \\ 35 & 109 & 224 & 15 \\ 87 & 75 & 12 & 168 \\ 239 & 125 & 226 & 165 \end{bmatrix} \begin{bmatrix} 83 & 91 & 151 & 109 \\ 14 & 203 & 121 & 177 \\ 246 & 246 & 63 & 243 \\ 0 & 234 & 173 & 02 \end{bmatrix} = \begin{bmatrix} 51 & 162 & 108 & 62 \\ 13 & 201 & 19 & 60 \\ 93 & 195 & 42 & 13 \\ 100 & 216 & 0 & 181 \end{bmatrix}$$

At destination node effective coefficients and encoded packets are received from Intermediate node as shown below.

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \begin{bmatrix} 51 & 162 & 108 & 62 \\ 13 & 201 & 19 & 60 \\ 93 & 195 & 42 & 13 \\ 100 & 216 & 0 & 181 \end{bmatrix} = \begin{bmatrix} 38 \\ 71 \\ 140 \\ 39 \end{bmatrix}$$

The sequence generated above is a system of four linear equation. Hence its native packets i.e. x,y,z,w can be decoded if sufficient rank is achieved for received packets. In the case of G-By-G Network Coding if sufficient packets of particular generations are not received then whole generation has to be retransmitted. For G-By-G, to decode the encoded packets and to determine native packets gauss elimination method is to be preferred for large system.

Chapter 3

Security Issues

This chapter provides an is a detailed explanation of security issues related to DTN, P2P, and NC. This chapter provides description for the first phase of our thesis.

3.1 Security issues in DTN

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. As per the analysis, the following conclusions have been noted down.

Possible attacks in DTN :[9]

- **Packet drop:** As per the three major components of security-CIA, due to packet drop , the availability parameter gets affected , thus damaging security. here malicious node keeps on dropping the packets to disrupt data availability.
- **Bogus packet injection:** In this attack, the adversary injects some unwanted packets, known as bogus packets into the network, thus affecting the network resources. Functioning of a network goes smooth if the bandwidth is sufficient enough to carry the network load. Bogus packet injection increases unproductive load ,thus decreasing networks efficiency and throughput.
- **Noise injection:** Noise injection refers to injection of data ,which hinders the integrity of information. Packets are modified and forwarded toward destination by

attacker and its integrity get affected

- **Routing attacks:** In routing attack ,adversary tries to manipulate route of transmitted packets. DTN deals with intermittent and long delays.so attacker can take advantage of this characteristic.
- **Flooding attacks:** To reduce the efficiency of network ,adversary can launch flooding attack, in which unnecessary packets are transmitted to keep communication channel busy and legitimate traffic delayed.
- **Impersonation attacks:** In this attack, attacker tries to impersonate the legitimate node in the system and try to mislead the network.

Study on wormhole attack[10] In this attack, the main target are those distant nodes, in the network, that requires low latency. Here, the victim node, will record and shift the information to alternate victim node. This creates a virtual picture to the nodes which are actually far-away, the feeling of being near.

Wormhole attack compromises the following:

1. Disturb or change the network topology
2. Affect the routing
3. Mechanism to deliver packet may be disturbed.

These schemes cannot be applied to DTN's because of pre-requisites on resources or on network data, connecting these networks.

3.2 Security issues in Network coding

3.2.1 Study of security issues in network coding-based wireless systems[2]

There are two primary ways to apply network coding 1)intra flow network coding 2) inter flow network coding. This approaches are dealing security threats which disrupts data

delivery process.

- Security issues in intra flow network coding:
  - Data forwarding:
    - **Packet pollution:** This attack is very well lnown in wireless networks often written as jamming attack
    - **Packet dropping:** Attacker node starts dropping of packets frequently.
  - Forwarding node selection and rate assignment:
    - **Wormholes:** Wormhole attacks are capable of creating imaginary links between honest nodes, disturbng their knowledge of network topology.
    - **Link quality falsification or modification:** The attacker node may demand false metrics for adjacent links
  - Acknowledgment delivery:
    - **ACK injection or modification :** The attacker inserts a bogus ACK or modifies an ACK packet forcing the source to move in next batch prematurely.
    - **ACK dropping:** If the attacker node lies on the ACK delivery path, it can drop all the ACK packets
    - **ACK delay:** The attacker node delays the delivery of ACK packets, instead of dropping ACK packets completely.
- Security issues in inter flow network coding:
  - Discovery of coding opportunities:
    - **Packet reception information mis-reporting:** An attacker can impersonate honest nodes and report incorrect packet reception information to their neighbors.
    - **Link state pollution:** Localized coding protocols also rely on the link quality between nodes to infer packet reception status at other nodes.
    - **Neighbor set pollution:** A node determines coding opportunities based on the neighboring node set information collected during the route discovery process.

– Transmission of coded packets:

  * **ACK injection or modification:** By injecting bogus ACKs or modifying ACKs, the attacker node can cause premature ending of necessary packet re transmissions in the pseudo-broadcast technique, resulting in the failure of packet reception at next hop nodes.
  * **Packet pollution:** The attacker node injects corrupted packets into the network. Packet over-coding. packets are unnecesarily coded by attacker encoder node.
  * **Packet under-decoding:** This attack is similar to a packet over-coding attack, but is performed by a decoding node.
  * **Packet dropping:** Cattacker node start dropiing the packet unnecessarily

3.3 Security issues in peer to peer file sharing using network coding[3]

As per research, the P2P file share has threat to a) Entropy attacks and b) Jamming attacks.This threats cannot be ignored while building a secure system.[3].

3.3.1 Entropy Attacks

A major drawback of Network Coding scheme is, it is unaware of the block to be transmitted, to the node, alternate to it. Hence, it transmits the entire block. These encoded blocks are of utmost value the node given, because they convey new data. This makes them innovative. Innovation is decided upon the coefficient vector, which produces the block. There can be a change in the encoded block which is downloaded and locally accessible block. The attacker may play smart, by using the blocks which are encoded for sending non-innovative blocks, which for the non-attacker node is an inconsequential linear combination of current blocks.

This is an entropy attack because attackers try to abort the entropy or important qualities in the system, which lowers down the download options for nodes in the system,

leading to a low system rate. Alternate consequences of such attacks are when a user is fooled by sending useful information to adversary, but receieving a non-innovative packet in return. This happens mostly during downloading.

3.3.2 Jamming attacks

In this particular attack, the network is jammed by injecting corrupted or flase chunks of packets, to block the download. This is called jamming attack.

In network architecture scenario, jamming attack prevails when many corrupted packets, continously block the network. The attack is initiated when a malicious node sends a couple of an corrupted encoded block and/or a coefficient vector . The recipient will get tainted data, and, when the reciever uses this data to make encoded blocks it will inject more corrupted blocks in the network. Since recipients have constrained bandwidth they gain an advantage from such strategy that distinguishes corruption as it happens, so they can reach end associations with malicious neighbors and search for fair nodes in a better place in the network.

The other option is to wait for the reciever to complete the entire download and then try to decode the entire file. But, if the file is full of corrupted blocks, it is tough to recognize corrupted blocks during the time of decoding.Downloading additional blocks and performing various deciphering operations with varying combination of blocks to reproduce a legitimate file is too costly.

This is plainly unsatisfactory, particularly for vast downloads where the expense of various interpreting gets to be restrictive. One corrupted block ought not destroy hundreds or a great many legitimate ones.

P2P mechanism provides a source verification mechanism to protect against jamming attacks. Nodes downloading the data, verify individual blocks. This check works fine when the distributer is on the first web. A standard way is to sign all the packets periodically, to protect against servers or set of servers.

The major flaw in jamming attack is every useful block is produced from the corrupted block, thus contaminating the entire network. Hence, every generated block is unique, and the server fails to sign it.

Table 3.1: Possible attacks

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

Dong et al [2] says 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 ,over coding ,under-decoding and drop data attacks.

Chapter 4

Defense Schemes

After identification of major threats to such systems ,study for available defense scheme has been done.there exist various types of defense schemes to defend against various attacks.These schemes have already been developed, which is useful in understanding the loopholes to overcome them in our dissertation.

4.1 Defense against pollution attack

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

- Error correction
- Attack detection
- Attacker location identification

4.1.1 Error correction

There exist multiple approaches in error correction categories. But as per [arXiv:1102.3504] 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[arXiv:1102.3504]. Such schemes are not capable of detecting corrupted packets.

4.1.2 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.[11] ,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 [12].

Agrawal and Boneh [13] 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 [6], 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[18] also suffers from same drawback. Addition to it this scheme require spacemac as its MAC algorithm.

Kehdi and Li [17] 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.

[12] 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.

4.1.3 Attacker location identification

Jafarisiavoshani et al. [19] took advantage the subspace properties of random network coding to approximate attackers location attackers. This scheme works only with fixed directed acyclic graph.

Wang et al. [20] 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. [4] has proposed a scheme which give higher security in less per byte overhead then [20] with the use of TESLA while avoiding the need of checksum dissemination.

Chapter 5

Proposed Work

This chapter proposes a defense scheme against packet drop and packet pollution attacks. We have used the basic knowledge to understand their limitations and have described how our proposed work can upgrade them.

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 approach 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 epensive 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 parents 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 childs end.

The MAC based scheme proposed here, is based on Spaacemac. It is branched into following:

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.

- MAC (k,id,y)
  - Input to the algorithm: Input key k, identifier of source id and vector y.
  - Output: Tag t
- Combine ((y1,t1,a1)up to (Yp,tp,ap))
  - Input: p vectors y1, . . . , yp, their tags t1, . ..,tp, key k, coefficient a1, . . . ap.
  - Output: tag t for vector y.

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

The batch verification procedure is described below.

This way efficient scheme can be built to support p2p file sharing in DTN. The psuedo code is written below.

The batch verification process is a fully imparted process i.e. either it is executed completely or it doesnt initiate. It is executed for the verification stage of the peer-to-peer network system.

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.

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 nave partitioning.

Algorithm 1 MAC based scheme improved with batch verification

```
if running time = 0 AND node = source then
  Generate_sessionid()
  Transmit_sessionid()
else if running time != 0 AND node = source then
  Transmit_EncodedMessage()
else if running time != 0 AND node = intermediate OR receiver then
  Flag = MAC_Verify()
  if Flag = 1 then
    Batch_Verification()
  else
    MAC_Combine()
    Transmit_EncodedMessage()
  end if
end if
```

Chapter 6

Implementation and Results

This chapter has the implementation results and analysis of the same. We have implemented our work in NS2. Its outcomes and comparative analysis have been shown in this chapter.

6.1 Simulation environment

As per the need of the project,the suitable simulation environment is NS-2.NS-2 can be downloaded from http://www.isi.edu/nsnam/ns.It is a discrete event network simulator which support many protocols and craetion of new protocols.Protocols are implemented in c++.

Table 6.1: Simulation environment

Parameter	**value**
Simulator	NS-2
Channel type	Channel/wireless channel
Radio propoggation model	Propogation/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	800*800
Simulation time	700 sec
No.of nodes	10-75
Routing protocol	NCR

6.2 Required implementation

To develop a suitable a scheme for the security of defined network,the following mechanisms are required to be implemented in NS-2.

- A new protocol.
- DTN capabilities.
- Declaration and definition of malicious node in the protocol itself.
- network coding capabilities.
- Algorithm to secure against malicious nodes.

In second phase of this project,the above listed things were implemented.After successful implementation of the algorithm,results were measured under different situations.

6.3 Results

After simulationg 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 packets

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.

6.3.1 Average Throughput

The following graph 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

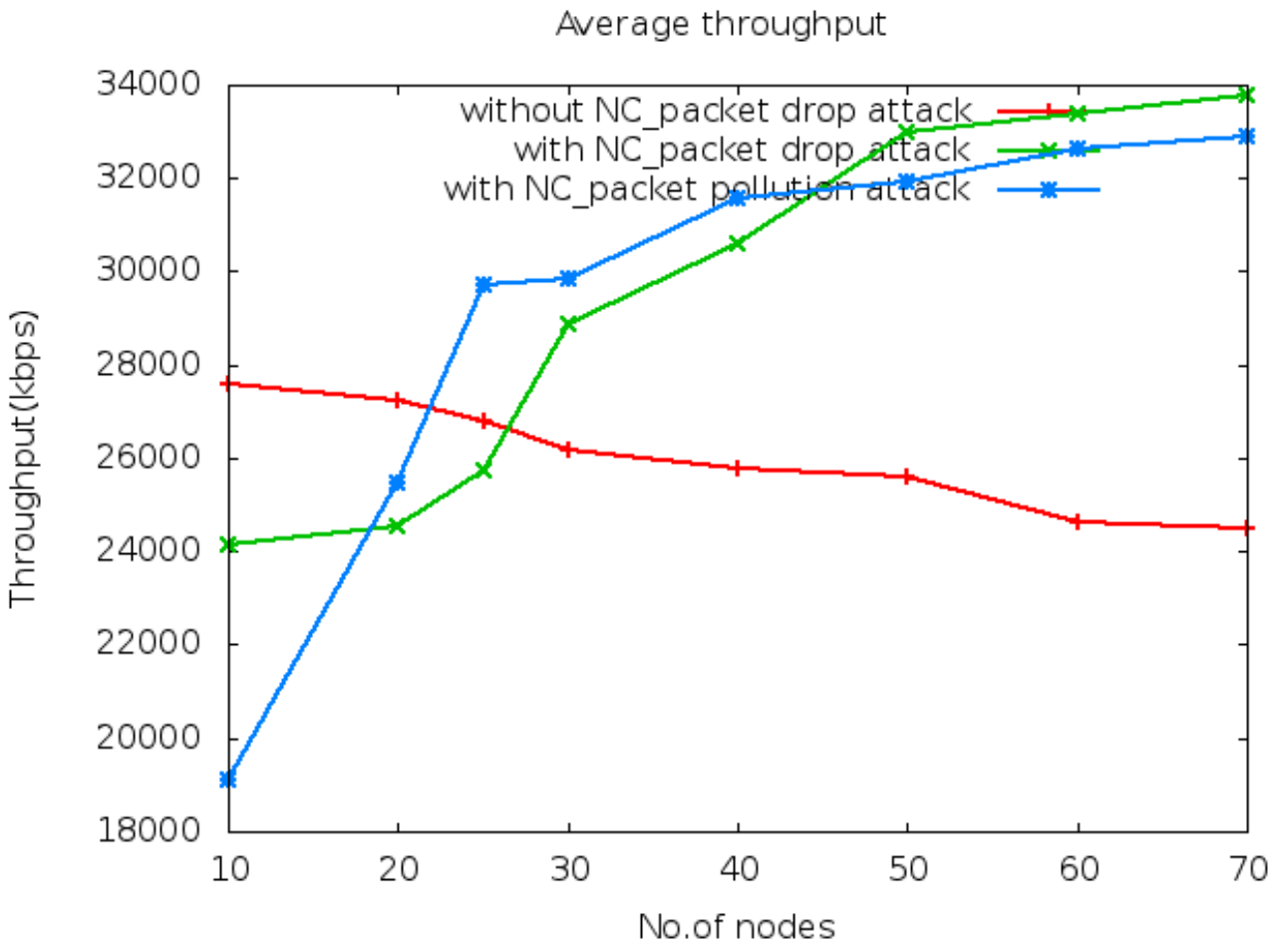

Figure 6.1: Average throughput

6.3.2 End to end delay

The following graph 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.

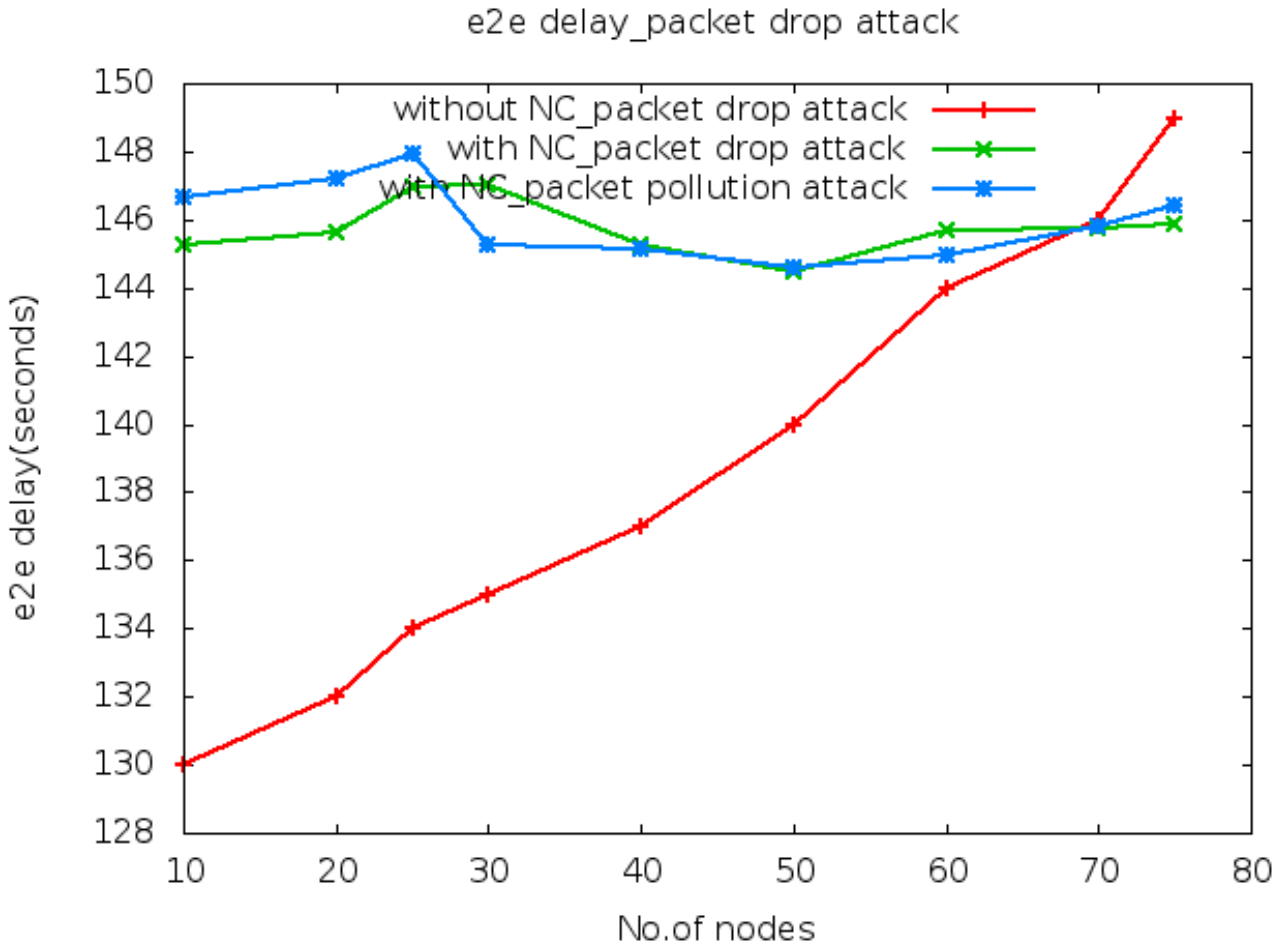

Figure 6.2: e2e delay

6.3.3 Instantaneous average Throughput

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

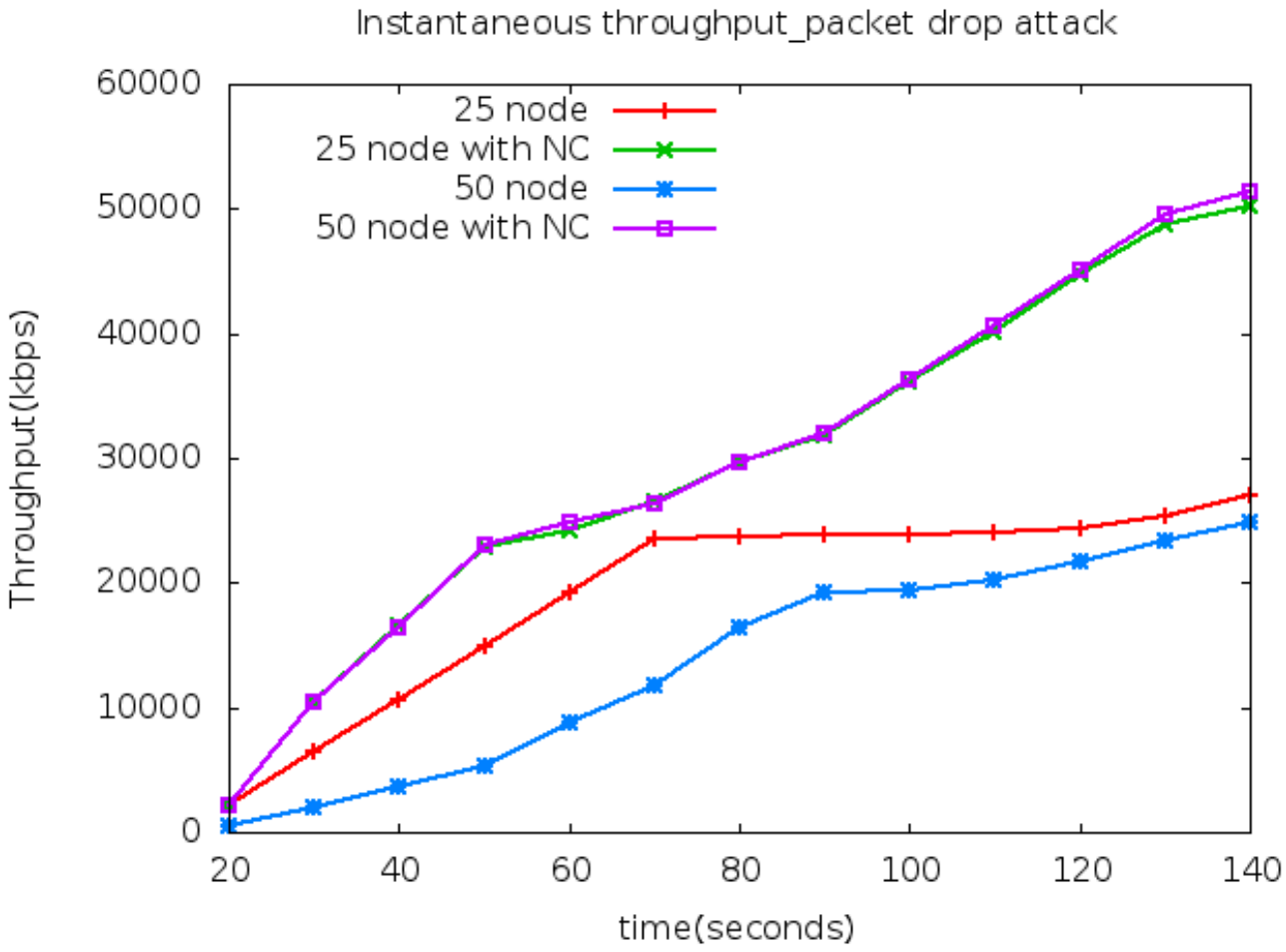

Figure 6.3: Instantaneous average throughput under packet drop attacks

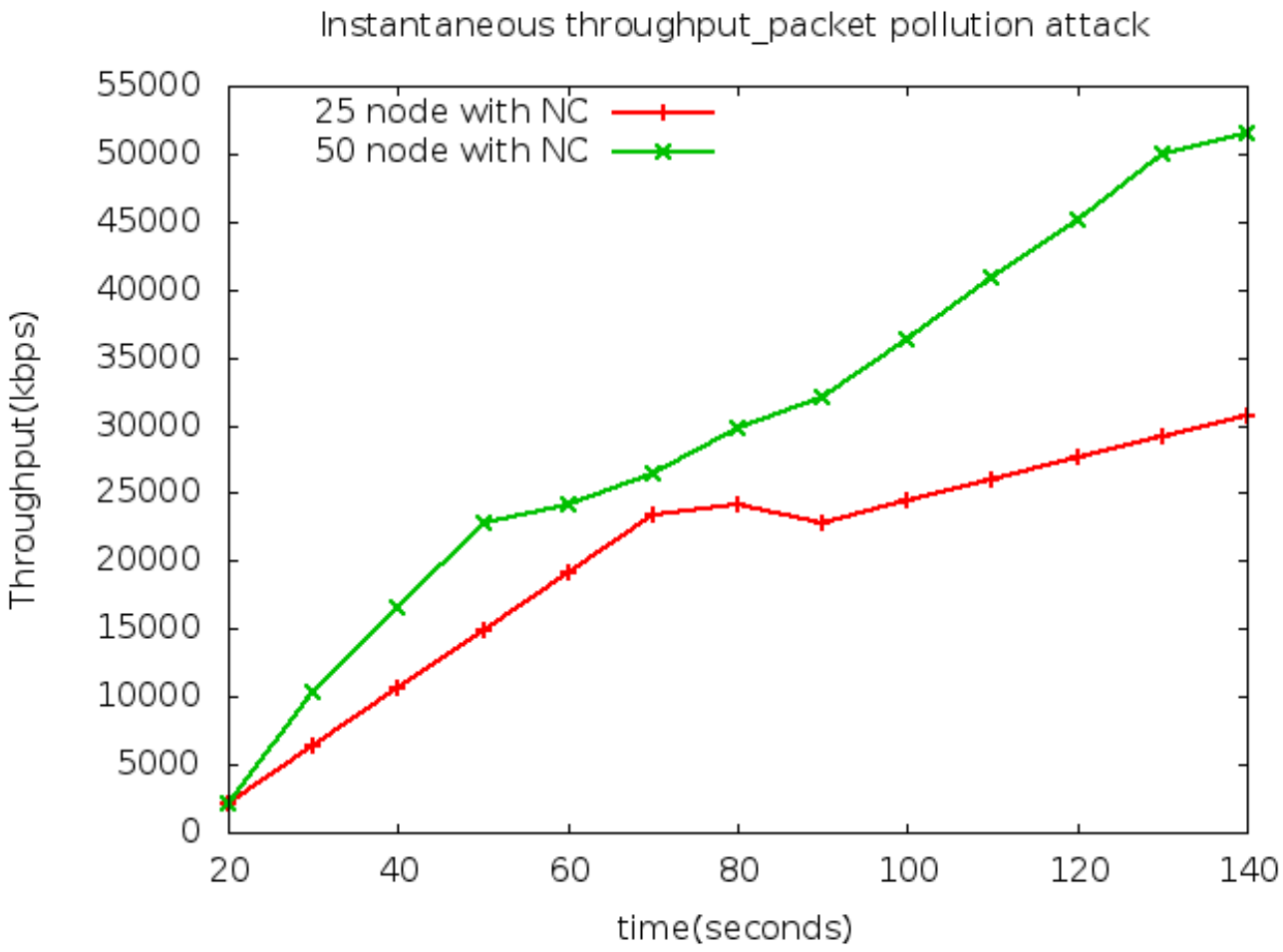

Figure 6.4: Instantaneous average throughput under packet pollution attacks

6.3.4 Ratio of decoded packets to received packets

The following graph 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 upto great extent

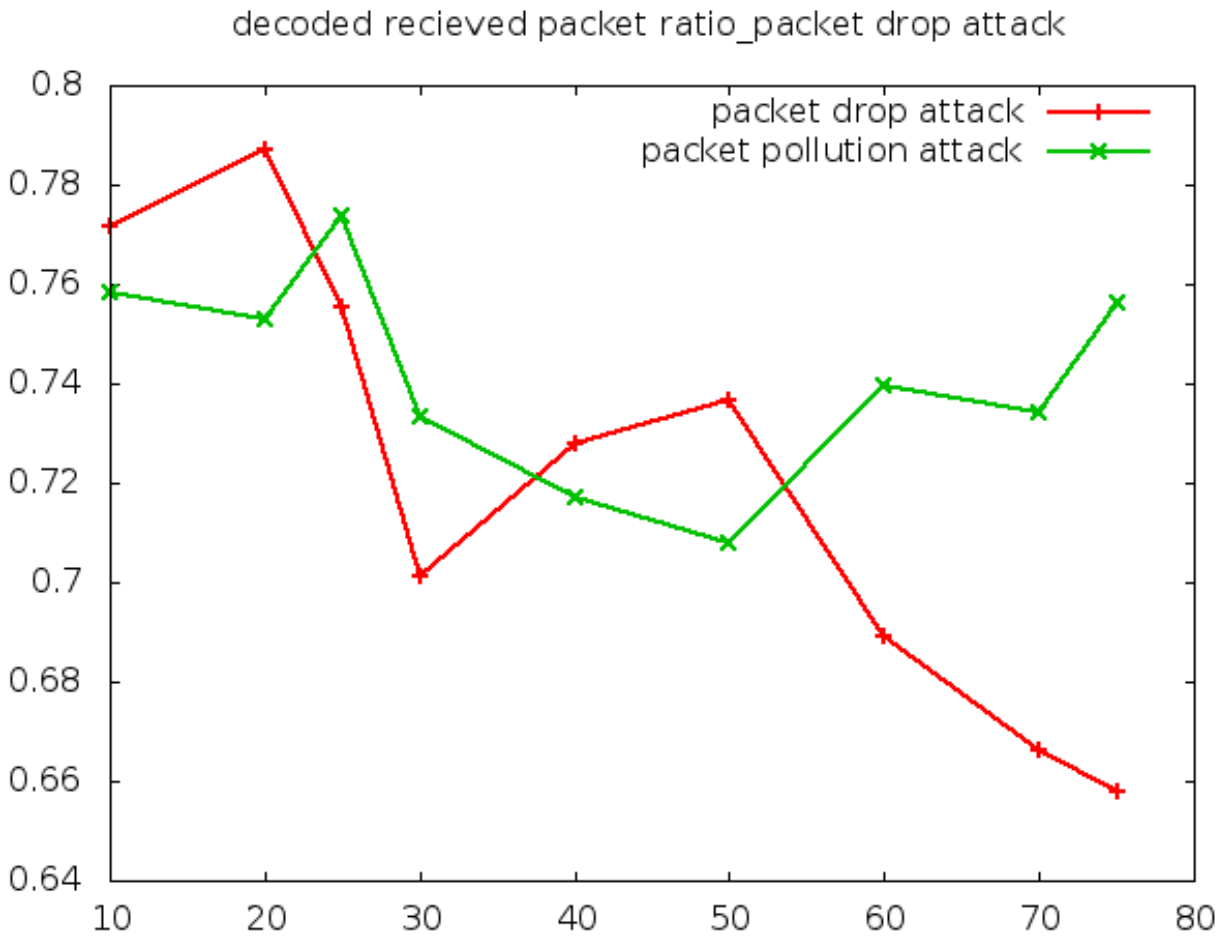

Figure 6.5: decoded packet to received packet ratio

Chapter 7

Conclusion and Future Work

In this thesis, we evaluted the performance of a proposed defence algorithm against packet drop attack and packet pollution attack. These two attack are serious problems to existing schemes. Therefore,a defense scheme was required to defend against these attacks.

7.1 Conclusion

Summarizing, this thesis has contributed the following :

- Implemented delay tolerant network agent in NS-2
- Implemented random linear netwokr coding in NS-2
- Created simulation environment for packet drop attacks and packet pollution attacks in network coding.
- Performed simulation for different number of nodes in absence and existence of packet drop attacks and in existance of packet pollution attack.
- Analyzed the trace files produced after simulation and generated the graphs for various parameters for comparison

Analysis shows that proposed scheme can defend against pollution attack and packet drop attack.however,due to network coding some latency will be created.

7.2 Future work

The following work may be pursued in future.

- Here, we have assumed a fixed number of attackers associated with a specific number of nodes. However, this can be created more dynamic in nature so as to accept possible ratio number of attackers and nodes in a changing scenario.
- There are some more parameters which can be analyzed using trace files generated during simulation.
- A more specific range for number of nodes can be selected for simulation.
- In proposed scheme,MAC scheme can be replaced with other scheme and performance can be measured.

References

[1] V. U. Parikh and Z. Narmawala, "A survey on peer-to-peer file sharing using network coding in delay tolerant networks,"

[2] J. Dong, R. Curtmola, and C. Nita-Rotaru, "Secure network coding for wireless mesh networks: Threats, challenges, and directions," *Computer Communications*, vol. 32, no. 17, pp. 1790–1801, 2009.

[3] 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 on*, pp. 1–5, IEEE, 2012.

[4] kodo, "ncintro."

[5] D. security, "Contemporary survey of dtn security."

[6] youtube, "Dtn."

[7] J. Kangasharju, K. W. Ross, and D. A. Turner, "Optimizing file availability in peer-to-peer content distribution," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pp. 1973–1981, IEEE, 2007.

[8] D. security, "Galois field arithmetic."

[9] E. Ayday and F. Fekri, "An iterative algorithm for trust management and adversary detection for delay-tolerant networks," *Mobile Computing, IEEE Transactions on*, vol. 11, no. 9, pp. 1514–1531, 2012.

[10] Y. Ren, M. C. Chuah, J. Yang, and Y. Chen, "Detecting wormhole attacks in delay-tolerant networks [security and privacy in emerging wireless networks]," *Wireless Communications, IEEE*, vol. 17, no. 5, pp. 36–42, 2010.

[11] J. Dong, R. Curtmola, and C. Nita-Rotaru, "Practical defenses against pollution attacks in intra-flow network coding for wireless mesh networks," in *Proceedings of the second ACM conference on Wireless network security*, pp. 111–122, ACM, 2009.

[12] A. Le and A. Markopoulou, "Tesla-based defense against pollution attacks in p2p systems with network coding," in *Network Coding (NetCod), 2011 International Symposium on*, pp. 1–7, IEEE, 2011.

[13] S. Agrawal and D. Boneh, "Homomorphic macs: Mac-based integrity for network coding," in *Applied Cryptography and Network Security*, pp. 292–305, Springer, 2009.

[14] M. Yang and Y. Yang, "Peer-to-peer file sharing based on network coding," in *Distributed Computing Systems, 2008. ICDCS'08. The 28th International Conference on*, pp. 168–175, IEEE, 2008.

[15] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network information flow," *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, 2000.

[16] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 4, pp. 2235–2245, IEEE, 2005.

[17] R. Di Pietro, S. Guarino, N. Verde, and J. Domingo-Ferrer, "Security in wireless ad-hoc networks–a survey," *Computer Communications*, vol. 51, pp. 1–20, 2014.

[18] Y. Li, H. Yao, M. Chen, S. Jaggi, and A. Rosen, "Ripple authentication for network coding," in *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9, IEEE, 2010.

[19] E. Kehdi and B. Li, "Null keys: Limiting malicious attacks via null space properties of network coding," in *INFOCOM 2009, IEEE*, pp. 1224–1232, IEEE, 2009.

[20] M. J. Siavoshani, C. Fragouli, and S. Diggavi, "On locating byzantine attackers," in *Network Coding, Theory and Applications, 2008. NetCod 2008. Fourth Workshop on*, pp. 1–6, IEEE, 2008.

[21] Q. Wang, L. Vu, K. Nahrstedt, and H. Khurana, "Identifying malicious nodes in network-coding-based peer-to-peer streaming networks," 2009.

[22] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Médard, "Resilient network coding in the presence of byzantine adversaries," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pp. 616–624, IEEE, 2007.

[23] NS-2, "www.isi.edu/nsnam/ns."