A PACKET SCHEDULING MECHANISM FOR WIRELESS PEER-TO-PEER CONTENT DISTRIBUTION

A Dissertation

by

YAO LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Alex Sprintson
Co-Chair of Committee,	I-Hong Hou
Committee Members,	Srinivas Shakkottai
	Evdokia Nikolova
Head of Department,	Chanan Singh

December 2013

Major Subject: Computer Engineering

ABSTRACT

This thesis studies the problem of content distribution in wireless peer-to-peer networks with selfish nodes. In this problem a group of wireless nodes need to exchange a set of files over a lossless broadcast channel. Each node aims to maximize its own download rate and minimize its upload rate. We propose a distributed protocol that provides incentives for selfish nodes to participate in the content exchange. Our protocol does not require any exchange of money and reputation and hence can be easily implemented without additional infrastructure.

Then, we will analyze the performance of our protocol by focusing on the important case in which the system contains two files that need to be distributed. We derive a closed-form expression of Nash Equilibrium and characterize the corresponding system performance in discrete time. Furthermore, we propose a distributed mechanism where the strategy of each node is only based on the observed history of the system and not on the private information of other nodes.

We also study the performance characteristics of the systems that employ network coding to facilitate data exchange. We show that, due to the free rider problem network coding does not necessary improve the performance of the system and, in some cases, may lead to worse system performance. We propose a novel approach to this problem based on random coding. The performance of the network coding algorithms is validated by performing extensive simulation study.

DEDICATION

To my parents, grandparents and aunt.

ACKNOWLEDGEMENTS

First of all, I would like to extend my heartfelt gratitude to my advisor Dr. Sprintson and co-advisor Dr. Hou. I was very lucky that Dr. Sprintson accepted me as his master student and introduced me to his group not long after I began my study at Texas A&M University. So I was able to start my research early. Then, I was attracted by Dr. Hou's research topic and took part in the study. With his constant guidance, I completed the analysis of the protocol in discrete time scenario. Dr. Hou helped me a lot during my study of game theory which is one of the important part of this research. Later on, I took Dr. Sprintson's network coding class from which I got my idea of the randomized algorithm. This algorithm helps to maximize the efficiency of the network coding scheme.

Second, I will thank to Shuangshuang Liang whose research focuses on parameters tuning and modeling. With her help, I am able to simulate my randomized algorithm in different scenarios. I use these result to verify my mathematic analysis of my algorithm.

Third, I also would like to thank Dr. Shakkottai and Dr. Nikolova for being my committee members, and for their suggestions on this research.

Moveover, I will thank to the financial support of this research. My thesis is based upon work partially supported by the NSF under grant CNS-0954153 and by the AFOSR under contract No. FA9550-13-1-0008.

Last but not least, thanks to all my friends and all the professors I met at Texas A&M University.

NOMENCLATURE

GT	Game Theory
NC	Network Coding
NE	Nash Equilibrium
RA	Randomized Algorithm

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

1. NON-MONETARY PROTOCOL FOR WIRELESS PEER-TO-PEER CONTENT DISTRIBUTION*

1.1 Introduction & Background

Recently, there has been a significant interest in using wireless peer-to-peer (P2P) networks to distribute information between mobile devices. The peer-to-peer content distribution has significant performance benefits. For example, in cellular networks mobile phones can retrieve the required information from their peers instead of downloading it from remote base stations. Since exchanging data between local devices requires significantly less power and results in less interference with other devices, such an approach has the potential of reducing power consumption and increasing spatial reuse.

Many existing studies (e.g., [4], [8] and [10]) have demonstrated the benefits of wireless P2P networks. However, these studies have assumed that all nodes are cooperative and do not require additional incentives to cooperate. In practice, nodes may be selfish and have little incentive to help other nodes to obtain the data they need. Therefore, a major challenge for wireless P2P networks is to provide incentives to nodes in the network so that they be willing to contribute to the network by sharing their data with other nodes. While there are many studies, such as [1], [6] and [7], on this topic for wired P2P networks, they cannot be applied to wireless P2P networks due to unique characteristics of wireless medium. In particular, due to the broadcast nature of the wireless medium, when a node transmits a packet, all nodes within the proximity of that node are able to receive the packet. Therefore, in wireless P2P networks, data exchange involves many

*Part of this chapter is reprinted with permission from "A Non-Monetary Protocol for Peer-to-Peer Content Distribution in Wireless Broadcast Networks with Network Coding" by I-Hong Hou, Yao Liu and Alex Sprintson, 2013, WiOpt.

nodes within the system, rather than only two nodes as in wired P2P networks.

In this thesis research, we study wireless P2P networks with selfish nodes. We first provide a model that considers the broadcast nature of wireless transmissions and the incentives of selfish nodes. Each node in the system aims to increase its download rate and decrease its upload rate, so as to reduce its own power consumption. We then propose a protocol for content distribution for this setting. Our protocol does not require the exchange of money, reputation, etc., and hence can be implemented without the need of additional infrastructure. This non-monetary feature further distinguishes our work from other studies that rely on additional infrastructure to set prices or payoffs [7], [2] and [3] or to punish uncooperative nodes [5]. Moreover, our protocol can be easily modified to employ network coding.

We provide a detailed performance analysis for our protocol in a discrete time scenario. For the practically important case with two files in the system, we derive closed-form expressions for each node's strategies under a series of Nash Equilibria. We also derive the prices of anarchy under these Nash Equilibria, both from a node's perspective and the whole system's perspective.

We then integrate a network coding scheme into our protocol to improve the throughput. Our results indicate is that in some settings using network coding may lead to smaller per-node download rate. This problem arises due to the free riders which are introduced by the network coding scheme. We design a randomized coding strategy which discourages the nodes from adopting the free rider strategy and is able to utilize the network coding technique to maximize the throughput.

1.2 System Model and Protocol Overview

We consider the *direct data exchange problem* [9] in which a group of wireless nodes are collaborating to exchange the set of files $X = \{A, B, C, \ldots\}$. Each node has a subset of files in X available to it and needs to obtain all other files in X. The nodes use a lossless broadcast channel to transmit files to other nodes. We assume that the files are very large, hence a large number of packets need to be broadcasted over the channel to deliver every file to other nodes. For clarity of presentation, we assume that each file contains infinitely many packets. For long files, this assumption incurs a very small penalty in terms of the accuracy of the obtained results.

In this thesis we focus on the special case in which each node n requires a single file, denoted by X_n, and has all other files $X \backslash \{X_n\}$ available to it. This is an important case that captures many of the salient features of the problem at hand. This case is of a significant practical importance since in many settings wireless nodes need to recover only a small number of packets, (e.g., packets that are lost due to fading or interference).

The broadcasting nature of wireless transmissions may result in a "*free rider*" problem as illustrated in the following example. Due to the free rider problem, the existing protocols designed for wired networks cannot be applied directly to the wireless setting.

Consider a system with four nodes and two files where $X_1 = X_2 = A$ and $X_3 = X_4 = B$. Suppose that node 1 and node 3 exchange data, that is, node 1 transmits packets of file B, and node 3 transmits packets of file A in return. As all nodes can receive all transmissions, node 2 and node 4 can obtain packets of A and B without transmitting any packet. Therefore, we say that nodes 2 and 4 are *free riders*. Figure 1.1 shows free riders in a networks of 4 nodes,

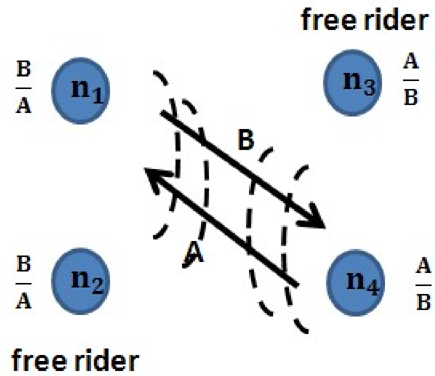

Figure 1.1: Free rider

In addition to being unfair, the possibility of being a free rider may prevent selfish nodes from transmitting data and contributing to the network. In the above example, each node may refrain from transmitting data, in the hope that other nodes participate in exchanging data, making itself a free rider.

In this thesis, we propose a non-monetary protocol for P2P content distribution and study its performance when all nodes are selfish.

1.3 Protocol Description

The transmission process consists of *rounds* such that during each round one packet from A and one packet from B are broadcasted over the channel. At the beginning of a round, each node n secretly picks a back-off time, τ_n. Node n then waits and listens to the channel for τ_n time. If no transmissions take place in τ_n time, node n transmits a control packet that contains the name X_n of the file required by n. The control packet can be interpreted as an obligation of node n to transmit the next packet from a file held by

n if any other node transmits the next packet from file X_n (the *next* packet refers to the packet which has not been broadcasted over the channel). Since the control packets are very small, we assume that their transmission does not incur any cost and that the time required for transmission of such packets is negligible. We refer to the time between the beginning of the round and transmission of the control packet as the *initiation phase*.

After node n transmits the control packet, every node m that has file X_n secretly and randomly picks a back-off timer $\hat{\tau}_m$. Every node m then waits and listens to the channel for $\hat{\tau}_m$ time. If no other nodes transmit in $\hat{\tau}_m$ time, node m transmits a data packet of X_n, and piggybacks its value of X_m. Upon receiving the data packet from node m, node n responds with a packet of X_m, as promised in its control packet. The round ends after node n completes broadcast of packet X_m and a new round begins. In this round, only nodes that need X_n and those that need X_m receive a packet they need, and all other nodes do not receive any useful packets. We refer to the time interval between the end of the initiation phase and the transmission of a packet by node m as the *response phase*. Note that τ_n and $\hat{\tau}_m$ are lengths of the initialization and response phase, respectively. The protocol execution is demonstrated in Figure 1.2,

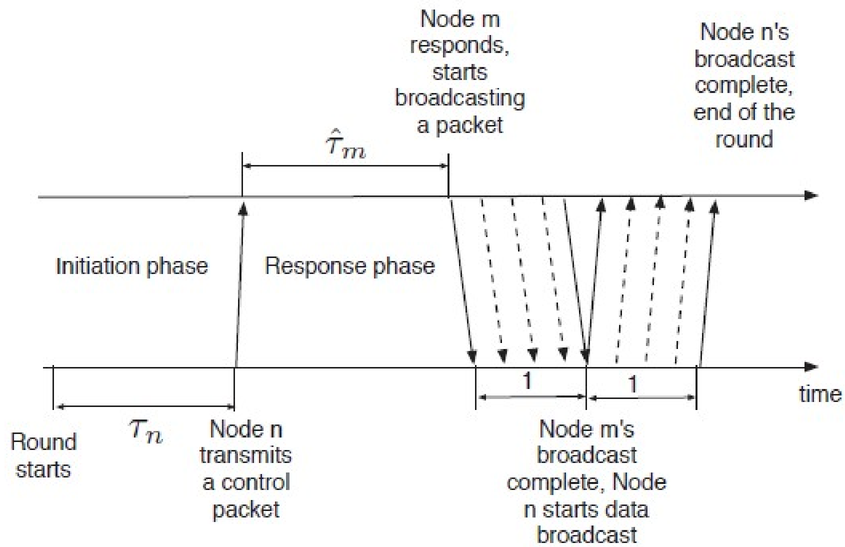

Figure 1.2: An example of a 3 phases round

Intuitively, when a node n chooses large values of τ_n and $\hat{\tau}_n$, it is likely that node n does not transmit, which increases its chance of being a free rider and reduces its transmission cost. However, large values of τ_n and $\hat{\tau}_n$ might result in large waiting times for n, which, in turn, might result in a large waiting cost. By taking waiting costs into account, our protocol provides incentive for the nodes to choose reasonably small values of τ_n and $\hat{\tau}_n$, and hence enables the nodes to exchange data in an efficient manner. We also note that our protocol is non-monetary and can be easily implemented for modern wireless networks without the need of additional infrastructure.

Finally, the protocol can be modified for the settings in which the network coding technique is used. The modified protocol is very similar to the one described above. Still, at each round, exactly two data packets will be transmitted. With the network coding technique, the packet broadcasted at the response phase will be a linear combination of the packets available at the transmitting node. For example, suppose that X contains three files A, B, C and node n requires file A, i.e., $X_n = A$. Then, in the response phase, node n can send a combination of the next packet from B and the next packet from C (the combination can be a linear operation underlying finite field). With network coding, all nodes receive a packet they need in each round.

2. ANALYSIS UNDER DISCRETE TIME SCENARIO

In this section, we analyze the protocol in a discrete time scenario. There are two key points should be considered at the beginning. First, in the discrete time model, a given node has two potential actions in a specific time slot which are transmit a packet or to be idle. This means the action of randomly picking up a back off time reduced to this binary choices. Thus, rather than the random back off time, strategies for a node in the discrete time scenario is the probability for it to transmit a packet in a specific time slot. The second important fact should be noticed is that in the discrete time model, collisions may happens in a time slot that causes the failure of transmission in that slot. Thus, extra time slot in the same phase is triggered in order to let the transmission to be succeeded.

Since the analysis to the first phase and second phase are similar that can be thought of a multiple players game, we only analyze a general scenario in which multiple nodes race to transmit the same packets to a common receiver. We will start from the simplest case in that involves only two nodes. Then, we will consider the more complicated ones such as large file transmissions and multiple nodes compete to transmit packets with the same transmit cost per time slot. Finally, we will come to the most general case in which there are arbitary number of nodes and each of them has its own distant transmission costs.

We analyze the performance of our protocol under Nash Equilibrium. We consider the performance from the system's perspective. We will consider the throughout of the system under Nash Equilibrium and the *price of anarchy of system* which is defined as follows:

Definition1: The *price of anarchy* for system under a Nash Equilibrium is the maximum throughput of the system under a cooperative scenario over the throughput under the Nash Equilibrium.

We denote N as the number of nodes, m as the transmission cost per time slot, K as the number of slots each transmission takes, Γ as the throughput of the networks, Ω as the price of anarchy of the whole system. Also, for simplicity, we assume the waiting cost for all the nodes are same which is 1 per time slot.

2.1 Two Nodes with Distant Transmission Cost

In this section, we consider the case of $N = 2$. We assume that n_1 and n_2 are two nodes involved in the transmission and the transmission cost for each nodes in a single time slot are m_1 and m_2 respectively. Further, we assume $K = 1$. Each node has the probability p_i to choose the transmit policy at Nash Equilibrium. Figure 2.1 illustrates the game for n_1,

costs (n_2 / n_1)	transmit	idle
transmit	$(m_1 + 1) + [p_1 C_{t1} + (1 - p_1) C_{i1}]$	$(m_1 + 1) + 1$
idle	1	$1 + [p_1 C_{t1} + (1 - p_1) C_{i1}]$

Figure 2.1: Costs for n_1 given $N = 2$ and $K = 1$

where, C_{t1} and C_{i1} are the costs for n_1 under the transmit policy. We can get the following equations at Nash Equilibrium,

$$
\begin{aligned}
C_{t1} &= p_2[m_1 + 1 + [p_1 C_{t1} + (1-p_1)C_{i1}]] + (1-p_2)(1+m_1) \\
C_{i1} &= p_2 + (1-p_2)[[p_1 C_{t1} + (1-p_1)C_{i1}] + 1] \\
C_{t1} &= C_{i1}
\end{aligned}
$$

Thus, we get,

$$p_2 = \frac{1}{m_1 + 2}$$

Similarly, we can get

$$p_1 = \frac{1}{m_2 + 2}$$

Let us consider the throughput and price of anarchy,

$$
\begin{aligned}
\Gamma &= p_1(1-p_2) + p_2(1-p_1) \\
&= \frac{m_1 + m_2 + 2}{(m_1+2)(m_2+2)}
\end{aligned} \tag{2.1}
$$

$$\Omega = \frac{(m_1+2)(m_2+2))}{m_1 + m_2 + 2} \tag{2.2}$$

2.2 Large Packet Transmission

In this section, we consider the transmission of large packets which mean a success transmission takes multiple time slots. Hence, unlike the case in last section, $K > 1$. The rest remains the same.

The Figure 2.2 shows the game for n_1,

costs (n_1 \ n_2)	transmit	idle
transmit	$K(\mathrm{m}+1) + [p_1C_{t1} + (1-p_1)C_{i1}]$	$K(\mathrm{m}+1)$
idle	K	$1 + [p_1C_{t1} + (1-p_1)C_{i1}]$

Figure 2.2: Costs for n_1 given $N = 2$ and $K > 1$

Denote C_t and C_i as the transmission cost for any node when under the transmit and idle policy separately and p as the probability for any node to chooses the transmit policy at Nash Equilibrium, then,

$$C_t = p\left[K(m+1) + [p_1C_{t1} + (1-p_1)C_{i1}]\right] + (1-p)(m+1)K$$
$$C_i = pK + (1-p)[1 + [p_1C_{t1} + (1-p_1)C_{i1}]]$$

We also have the following condition at Nash Equilibrium,

$$C_t = C_i$$

Using (8), (9) and (10), we get

$$(K-1)p^2 + (mK+2)p - 1 = 0$$

Hence,

$$p = \frac{-(mK+2) + \sqrt{m^2K^2 + 4(m+1)K}}{2(K-1)} \tag{2.3}$$

Now, we analyze the throughput and price of anarchy,

$$\Gamma = 2p(1-p) \tag{2.4}$$

$$\begin{aligned} \Omega &= \frac{\max_{\alpha \in (0,1)} 2\alpha(1-\alpha)}{2p(1-p)} \\ &= \frac{1}{4p(1-p)} \end{aligned} \tag{2.5}$$

2.3 Arbitary Number of Nodes, Same Transmission Cost

In this case, the number of nodes may be larger than 2. All the nodes share the same transmission cost per time slot. Figure 2.3 illustrates this network.

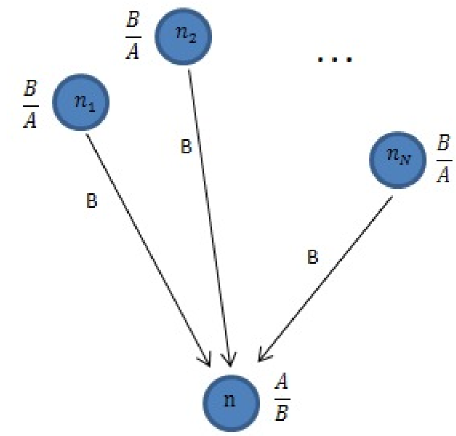

Figure 2.3: Network model

In this network, all the nodes from n_1 to n_N want have packet B and want packet A. Only node n have packet A and want packet B.

Because of symmetry, all the nodes have the same strategy at Nash Equilibrium. That is, each node transmit packet with the same probability, denoted by p, in each time slot. Figure 2.4 and 2.5 show the game for n_1,

costs — $n_2 \sim n_n$ / n_1	only one node transmits	other cases
idle	1	$1 + [p_1 C_{t1} + (1 - p_1) C_{i1}]$

Figure 2.4: Costs for n_1 under action of transmitting given $N > 2$, symmetric case

costs $n_2 \sim n_n$ / n_1	$n_2 \sim n_n$ *idle*	at least one node transmit
transmit	$m_1 + 1$	$m_1 + 1 + [p_1 C_{t1} + (1 - p_1) C_{i1}]$

Figure 2.5: Costs for n_1 under action of being idle given $N > 2$, symmetric case

Let us first consider the case that n_1 chooses the transmit strategy. The probability that the nodes $\{n_i : i = 2, \cdots, N\}$ keep idle is $(1-p)_{n-1}$ while at least one node transmit is $1 - (1-p)^{n-1}$. Further, we denote the expected cost for n_1 under transmit strategy at Nash Equilibrium is C_t (since symmetry, all the nodes have the same cost),

$$C_t = (1-p)^{n-1}(m+1) + \left[1 - (1-p)^{n-1}\right](m + 1 + C_t)$$

Now, we think of the case that n_1 chooses the idle strategy. The probability that only one node among $\{n_i : i = 2, \cdots, N\}$ transmits is $(n-1)p(1-p)^{n-2}$. We denote the expected cost for n_1 under idle strategy at Nash Equilibrium is C_i, then,

$$C_i = (n-1)p(1-p)^{n-2} + (1 + C_i)\left[1 - (n-1)p(1-p)^{n-2}\right]$$

At Nash Equilibrium, we also have

$$C_t = C_i$$

Using (14), (15) and (16), we come to the result,

$$p = \frac{1}{mn - m + n} \tag{2.6}$$

Then, let us discuss the throughput and price of anarchy at Nash Equilibrium

$$\begin{aligned}\Gamma &= np(1-p)^{n-1} \\ &= \frac{n}{m(n-1)+n}\left[\frac{(m+1)(n-1)}{m(n-1)+n}\right]^{n-1}\end{aligned}$$

Then, let us consider the case n goes to infinity,

$$\begin{aligned}\lim_{n\to+\infty} throughput &= \lim_{n\to+\infty}\frac{n}{m(n-1)+n}\left[\frac{(m+1)(n-1)}{m(n-1)+n}\right]^{n-1} \\ &= \lim_{n\to+\infty}\frac{1}{m+1}\left[1-\frac{1}{(m+1)(n-1)+1}\right]^{n-1} \\ &= \frac{1}{m+1}\lim_{n\to+\infty}\left[\left[1-\frac{1}{(m+1)(n-1)}\right]^{(m+1)(n-1)}\right]^{\frac{1}{m+1}}\end{aligned}$$

Since

$$\lim_{n\to+\infty}\left[1-\frac{1}{(m+1)(n-1)}\right]^{(m+1)(n-1)} = e^{-1}$$

We come to the conclusion,

$$\lim_{n\to+\infty} throughput = \frac{1}{m+1}e^{-\frac{1}{m+1}} \tag{2.7}$$

Also, we can get

$$\begin{aligned}\Omega &= \frac{\max_{\alpha\in(o,1)} n\alpha(1-\alpha^{n-1})}{np(1-p)^{n-1}} \\ &= \frac{1}{p(1-p)^{n-1}}\frac{(n-1)^{n-1}}{n^n}\end{aligned}$$

If n goes to infinity,

$$\begin{aligned}\lim_{n\to+\infty}\Omega &= \lim_{n\to+\infty}\frac{1}{p(1-p)^{n-1}}\frac{(n-1)^{n-1}}{n^n}\\ &= \lim_{n\to+\infty}\frac{1}{np(1-p)^{n-1}}\lim_{n\to+\infty}\frac{(n-1)^{n-1}}{n^{n-1}}\\ &= (m+1)e^{\frac{1}{m+1}}e^{-1}\end{aligned}$$

Hence,

$$\lim_{n\to+\infty}\Omega = (m+1)e^{-\frac{m}{m+1}} \tag{2.8}$$

2.4 Arbitary Number of Nodes, Different Transmission Costs

We extend the case in last section to one that nodes hold different transmission costs per time slots. Assume that there are n nodes in the networks. Each node has a unique transmission cost: $m_i : i = 1, 2, ..., N$. In each time slot, node i has the probability $p_i : i = 1, 2, ..., N$ to transmit its packets. Figure 2.6 and 2.7 show the game for n_1.

costs / n_1 \ $n_2 \sim n_n$	only one node transmits	other cases
idle	1	$1 + [p_1 C_{t1} + (1-p_1)C_{i1}]$

Figure 2.6: Costs for n_1 under action of transmitting given $N > 2$, asymmetric case

costs \ $n_2 \sim n_n$ / n_1	$n_2 \sim n_n$ *idle*	at least one node transmit
transmit	$m_1 + 1$	$m_1 + 1 + [p_1 C_{t1} + (1 - p_1) C_{i1}]$

Figure 2.7: Costs for n_1 under action of being idle given $N > 2$, asymmetric case

We can find that, actually the forms above are the same with those in the last section. Let us first consider the case that n_1 chooses the transmit strategy. The probability that the nodes $\{n_i : i = 2, \cdots, N\}$ keep idle is $\prod_{j=2}^{N}(1 - p_j)$ while at least one node transmit is $1 - \prod_{j=2}^{N}(1 - p_j)$. Further, we denote the expected cost for n_1 under transmit strategy at Nash Equilibrium is C_{t1}, then

$$C_{t1} = (m_1 + 1) \cdot \prod_{j=2}^{N}(1 - p_j) + (m_1 + 1 + C_{t1}) \cdot \left[1 - \prod_{j=2}^{N}(1 - p_j)\right]$$

Now, we think of the case that n_1 chooses the idle strategy. The probability that only one node among $\{n_i : i = 2, \cdots, N\}$ transmits is $\sum_{j=2}^{n} \prod_{k \neq 1,j}(1 - p_k)$. We denote the expected cost for n_1 under idle strategy at Nash Equilibrium is C_{i1}, then,

$$C_{i1} = \sum_{j=2}^{N} p_j \prod_{k \neq 1,j}(1 - p_k) + (1 + C_{i1}) \left[1 - \sum_{j=2}^{N} p_j \prod_{1 \neq 1,j}(1 - p_k)\right]$$

At Nash Equilibrium, we also have

$$C_{t1} = C_{i1}$$

Using equations (20), (21) and (22), we come to the following equation,

$$(m_1+1)\sum_{j\neq 1} p_j \prod_{k\neq 1,j}(1-p_k) = \prod_{j\neq 1}(1-p_j)$$

Similarly, we get such equations for nodes $n_i : i = 1, 2, ..., N$ combing the for n_1, we get the equation set,

$$\begin{cases} (m_1+1)\sum_{j\neq 1} p_j \prod_{k\neq 1,j}(1-p_k) = \prod_{j\neq 1}(1-p_j); \\ (m_2+1)\sum_{j\neq 2} p_j \prod_{k\neq 2,j}(1-p_k) = \prod_{j\neq 2}(1-p_j); \\ \quad\vdots \\ (m_N+1)\sum_{j\neq N} p_j \prod_{k\neq N,j}(1-p_k) = \prod_{j\neq N}(1-p_j). \end{cases}$$

Divide the left hand side expression in each equation by the one at the right hand side, we get

$$\begin{cases} (m_1+1)\sum_{j\neq 1} \frac{p_j}{1-p_j} = 1; \\ (m_2+1)\sum_{j\neq 2} \frac{p_j}{1-p_j} = 1; \\ \quad\vdots \\ (m_N+1)\sum_{j\neq N} \frac{p_j}{1-p_j} = 1. \end{cases} \tag{2.9}$$

Denote $A_j = \frac{p_j}{1-p_j}, j = 1, 2, \ldots, N$, we get,

$$\begin{cases} (m_1+1)\sum_{j\neq 1} A_j = 1; \\ (m_2+1)\sum_{j\neq 2} A_j = 1; \\ \quad\vdots \\ (m_N+1)\sum_{j\neq N} A_j = 1. \end{cases}$$

The equation set above is easy to solve, where

$$A_j = \frac{1}{n-1}\sum_{l=1}^{N}\frac{1}{m_l+1} - \frac{1}{m_j+1}, \qquad j = 1, 2, ..., N$$

Thus, at Nash Equilibrium, the probability p_j for each node to choose the transmit policy in any time slot is

$$p_j = \frac{A_j}{Aj+1}, \quad where \;\; A_j = \frac{1}{n-1}\sum_{l=1}^{N}\frac{1}{m_l+1} - \frac{1}{m_j+1}, \tag{2.10}$$

for all $j = 1, 2, \ldots, N$.

Even though the price of anarchy does not make much sense in this case, we still discuss it as follows,

$$\Gamma = \sum_{j=1}^{N} p_j \prod_{k\neq j}(1-p_k) \tag{2.11}$$

$$\Omega = \frac{\max_{\alpha_j\in(0,1)} \sum_{j=1}^{N} \alpha_j \prod_{k\neq j}(1-\alpha_k)}{\sum_{j=1}^{N} p_j \prod_{k\neq j}(1-p_k)} \tag{2.12}$$

where, p_j are the ones in (5).

3. PROTOCOL WITH RANDOMIZED NETWORK CODING SCHEME

3.1 Protocol with Network Coding

In the previous section, all the sender in the second phase expect the same packet from the receive. Hence, all of them will be benefited in the third phase since wireless networks is broadcast. We now extend this case to the scenario that senders want different packet. So in the protocol we just proposed, only one of them will be benefit. The question here is how to improve the efficiency? We will now design a protocol with network coding scheme in this section to maximize the throughput in this more complicated case. The basic idea is the the sender in the third phase will send a packet which coded all the packets needed by the other nodes together.

In this section, we discuss the scenario where there are multiple files in the system. N-odes are separated into groups $\{n_{a1}, n_{a2}, \dots\}$, $\{n_{b1}, n_{b2}, \dots,\}$, $\{n_{c1}, n_{c2}, \dots\}$, $\dots$, where a node in the group $\{n_{a1}, n_{a2}, \dots\}$ needs the file A and possesses all other files, B, C, $\dots$, etc. This means any nodes in the system misses one of the three files in the system. It wants the missing file from the networks and is able the transmit the files it has to other nodes.

The protocol with network coding can be shown in the following Figure 3.1,

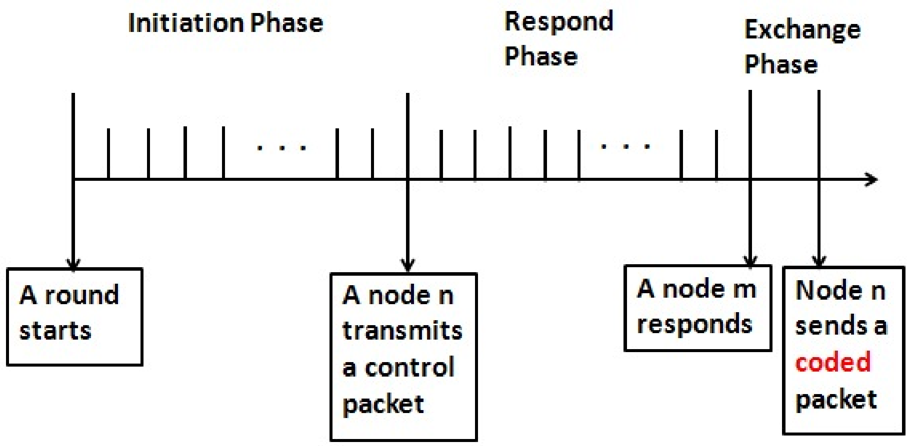

Figure 3.1: Three phases protocol with network coding

Initiation Phase: At the beginning of a round, each node n competes for sending a control packet that contains its value of X_n to others in the same networks. This packet is interpreted as a promise from node n saying that if any node transmits a packet of X_n, node n will respond with a packet that n has. Each transmission takes one single time slot. Collision happens if more than one packets transmitted in the same slot. As a result of collision, all the packets transmitted are discarded. Initiation phase ends when there is only one node, say n, transmits the control while others be idle.

Respond Phase: After node n transmits the control packet, every node m that has the file X_n starts to competes for transmitting a data packet of X_n. Similarly, this phase ends when there is only one node responds n with packet X_n. Also, m piggybacks its value of X_m. Then, the last phase is triggered.

Exchange Phase: Upon receiving the data packet form node m, node n responds with a coded packets which benefits all nodes other than n in the networks. This will end the current round and start the next one.

Drawback of this protocol exists which prevents it reaching the maximum throughput.

The basic idea here is that a node has chance to be a freerider if it chooses to be idle in the respond phase. This comes from the fact that node n benefits all the other nodes with a coded packet in the exchange phase. The intuition to solve this problem is to punish the free rider in the third phase. This leads us to come with the randomized network coding scheme.

3.2 Protocol with Randomized Network Coding Scheme

The difference between the protocol proposed above and the one with randomized network coding scheme lies in the exchange phase. In the old protocol, node n benefits all the other nodes while only node m is guaranteed to be benefited in the new protocol. And nodes other than n and m are benefited with a probability p such that $0 < p < 1$.

The randomized network coding algorithm is illustrated in the following Figure 3.2,

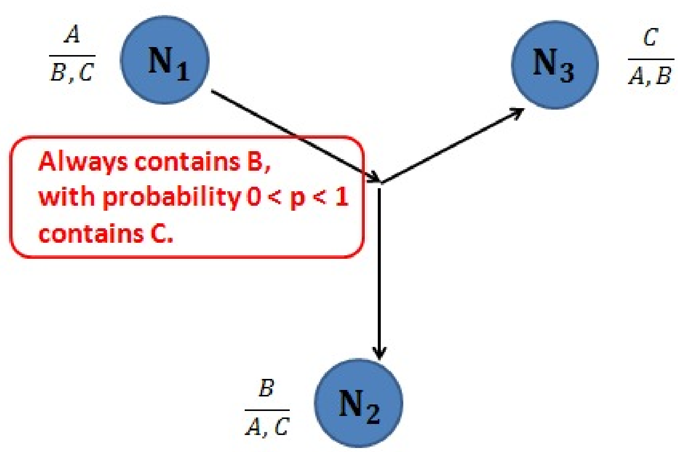

Figure 3.2: Randomized algorithm

For simplicity, we assume that there is only one node in each of the group, that is n_1, n_2 and n_3 miss packets A, B and C separately. But they have the other two packets. Suppose n_1 sends the control packet in the first phase and n_2 responds n_1 with what it wants in the second phase. In the protocol with network coding in the last subsection, n_1 will benefit both the other two in the third phase while in the randomized algorithm, n_1 will benefit n_2 with 100 percentage but n_3 with a probability p.

The intuition here is that any freeride is punished in the exchange phase since there is probability (1 - p) it does not receive the packet it wants. This fact increases nodes' willingness to contribute to the networks in the respond phase and as a result improves the overall performance of the protocol.

3.3 Protocol Performance Analysis

Now, let us analyze this protocol with game theory. Take the simplest case as an example. Suppose there are 3 nodes in the networks, n_1, n_2 and n_3. Each node misses one packets and has two different packets other nodes want. All the nodes transmit packet with the save cost, that is m per time slot whatever it is a data packet or a control one. And waiting cost for nodes is 1 per time slot. Total cost equals to the transmission cost plus the waiting cost. The goal for each node is to download the packet it wants with the lowest cost.

In the game, strategy for each node is the probability it chooses to transmit a packet in each time slot which can be shown in Figure 3.3,

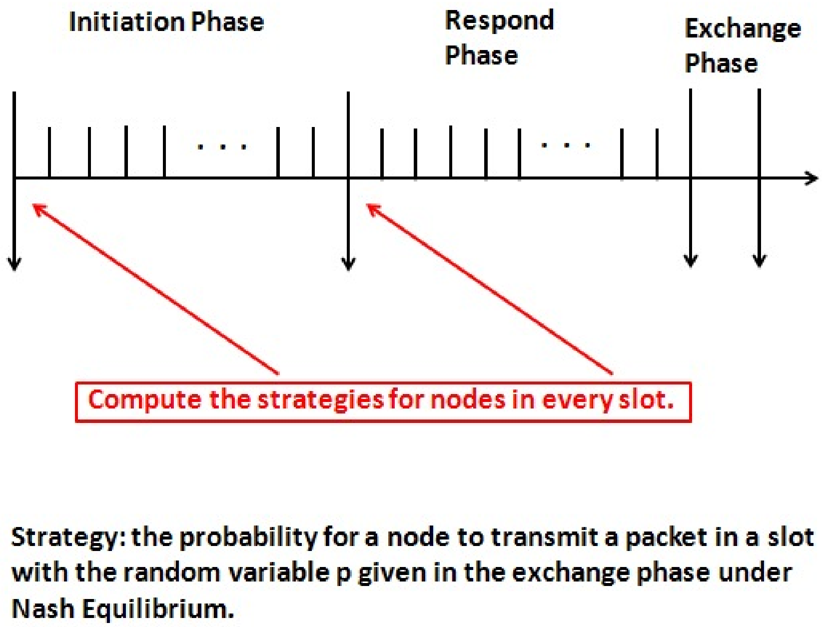

Figure 3.3: Strategy for each node

Let me introduce the notations needed in the analysis. We denote $\hat{P}_1 = Pr\{n_1\, does\, not\, transmit\, in\, the\, initiation\, phase\}$ and $\hat{P}_2 = Pr\{n_1\, does\, not\, transmit\, in\, the\, respond\, phase \mid n_1\, does\, not\, transmit\, in\, the\, initiation\, phase\}$. We should notice that $\hat{P}_1 = \frac{2}{3}$ and $\hat{P}_2 = \frac{1}{2}$ due to symmetry.

p_1 and p_2 denote the probabilities that any node chooses to transmit a packet in a slot in the first or second phase separatively. C_{t1} and C_{i1} are costs for a node from the current slots in the initiation phase to the end of the round if it chooses to transmit or be idle in the current slot under NE. C_{t2} and C_{i2} are the costs for nodes from the slots belongs to respond phase to the end of round under NE. Figure 3.4 and 3.5 show those four kinds of costs,

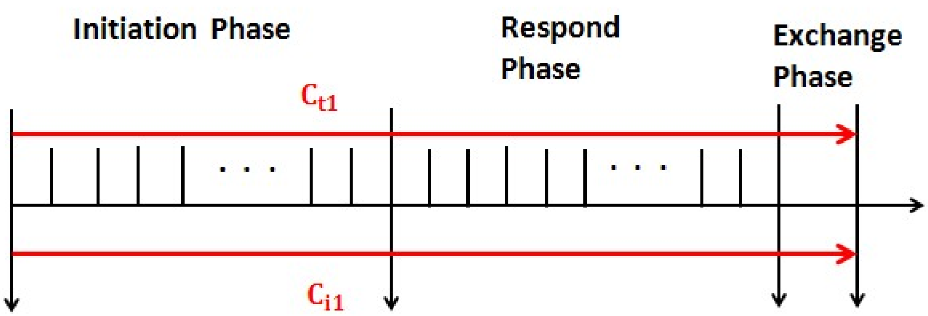

Figure 3.4: C_{t1} and C_{i1}

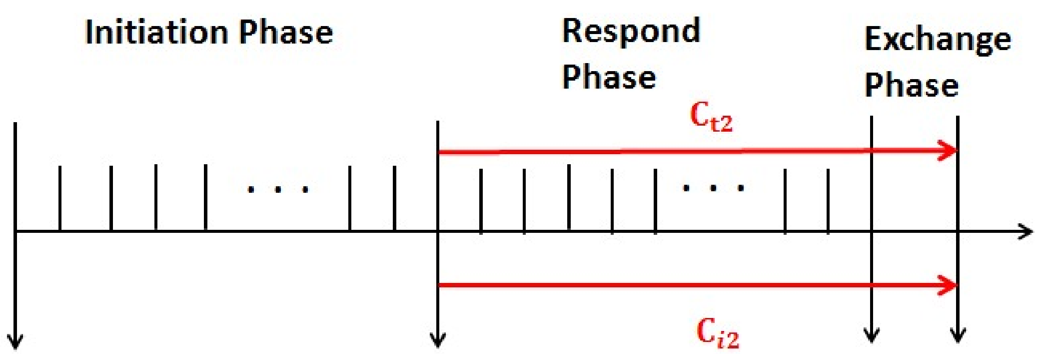

Figure 3.5: C_{t2} and C_{i2}

We also denote C_1 and C_2 to be the expected cost for a single node to successfully get the packet it wants start from the beginning of the first phase and second phase respectively. Moreover, T_2 is the expected time of the respond phase, and obviously,

$$T_2 = \frac{1}{2p_2(1-p_2)} \tag{3.1}$$

Since the analysis of respond phase depends on that of initiation phase, let us first consider the seconde phase. Figure 3.6 shows the expected costs for n_2 in the respond phase under specific policies of n_2 and n_3 given n_1 transmits control packet in the initiation phase.

costs (n_2 \ n_3)	transmit	idle
transmit	$(m+1)+\ [p_2C_{t2}+(1-p_2)C_{i2}]$	$(m+1)+1$
idle	$1+1$	$1+\ [p_2C_{t2}+(1-p_2)C_{i2}]$

Figure 3.6: Expected costs in respond phase

First, let us consider C_{t2}. In this case, n_2 transmit in the first time slot of the second phase. If n_3 also transmits in this slot of which probability is p_2, collision happens. According to the Figure 3.6, cost for n_2 from the current time instant to the end of the transmission round is $(m+1)+[p_2C_{t2}+(1-p_2)C_{i2}]$. Another case is n_3 being idle in this slot of which probability is $(1-p_2)$, and cost for n_2 is $(m+2)$, hence, the expected cost for n_2 if it chooses to transmit in the current time slot is,

$$C_{t2} = p_2[m+1+p_2C_{t2}+(1-p_2)C_{i2}]+(1-p_2)(m+2) \tag{3.2}$$

Similarly, the expected cost for n_2 if it chooses to be idle in the current time slot is

$$C_{i2} = 2p_2+(1-p_2)(1+p_2C_{t2}+(1-p_2)C_{i2}) \tag{3.3}$$

Now, let us consider C_2 under all the possible strategies of n_2. If it transmits in the current slot, its cost from now to the end of the round is C_{t2}. In the meantime, n_3 will also transmit with probability p_2 which causes collision and triggers a new slot in the second phase. Then with probability $\hat{P}_2$, n_3 finally responds n_1 in this phase, and the randomized NC causes n_2 will not be benefited in the 3rd phase with probability $(1-p)$. Recursively, it will cost n_2 another C_1 to get the packet it wants. Hence,

$$C_2 = C_{t2} + p_2 \hat{P}_2 (1-p) C_1 \tag{3.4}$$

Similarly,

$$C_2 = C_{i2} + \left[p_2 + (1-p_2)\hat{P}_2\right](1-p)C_1 \tag{3.5}$$

Then, let us analyze the initiation phase. Figure 3.7 and 3.8 illustrate the expected costs for n_1 in the first phase under specific actions of all the nodes.

costs: n_1 \ (n_2, n_3)	(t,t) (t,i) (i,t)	(i,i)
transmit	$(m+1) + [p_1 C_{t1} + (1-p_1) C_{i1}]$	$(m+1) + T_2 + (m+1)$

Figure 3.7: Expected costs in init phase under action of transmitting

costs: n_1 \ (n_2, n_3)	(t,t) (i,i)	(t,i) (i,t)
idle	$1 + [p_1 C_{t1} + (1-p_1) C_{i1}]$	$1 + T_2(1 + p_2 m) + 1$

Figure 3.8: Expected costs in init phase under action of being idle

Let us first consider C_{t1}. In this case, n_1 transmits in the current time slot. This transmission will succeed only in the case that all the other nodes remain idle. The probability of this event is $(1-p_1)^2$. According to the Figure 3.8, cost of n_1 from now to the end of the round is $2m+2+T_2$. In all the other cases in which there will be at most one of the other nodes transmit, collision happens. More slots are needed in this phase. Thus,

$$C_{t1} = \left[1-(1-p_1)^2\right]\left[m+1+p_1C_{t1}+(1-p_1)C_{i1}\right]+(1-p_1)^2(2m+2+T_2) \quad (3.6)$$

Similarly,

$$C_{i1} = \left[1-2p_1(1-p_1)\right]\left[1+p_1C_{t1}+(1-p_1)C_{i1}\right]+2p_1(1-p_1)\left[T_2(p_2m+1)+2\right] \quad (3.7)$$

Now, let us move to consider C_1 under all the possible strategies of n_1. If it transmits in the current slot, its cost from now to the end of the round is C_{t1}. If any of the other nodes also transmit at the same time of which probability is $1-(1-p_1)^2$, collision happens and everything re-start from the beginning. Hence,

$$C_1 = C_{t1} + \left[1-(1-p_1)^2\right]C_1 \quad (3.8)$$

With the same approach, we get,

$$C_1 = C_{i1} + \left[2p_1(1-p_1)\hat{P}_2(1-p)+\left[1-2p_1(1-p_1)\right]\hat{P}_1\hat{P}_2(1-p)\right]C_1 \quad (3.9)$$

Further, we should notice that throughput is expressed as follows,

$$\Gamma = \frac{2+p}{(\frac{1}{3p_1(1-p_1)^2} + \frac{1}{2p_2(1-p_2)} + 1) * 3} \tag{3.10}$$

3.4 Simulation

Figure 3.9 - 3.14 are the simulation results of throughput verse p under different transmission costs which range from 0 to 5,

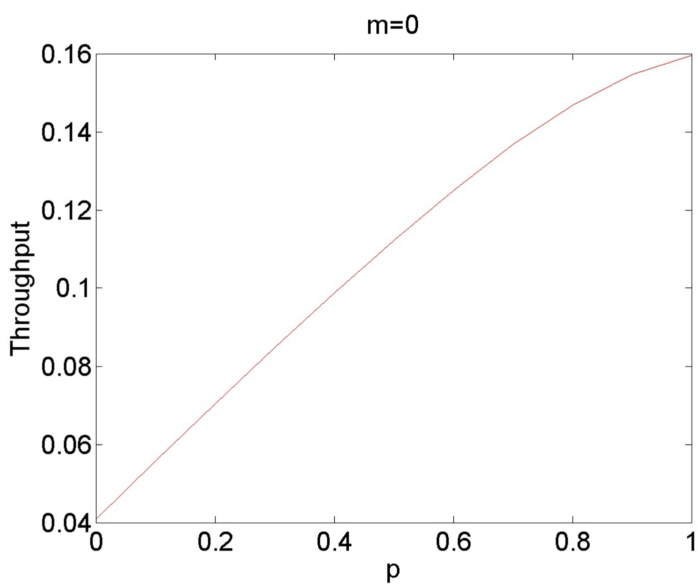

Figure 3.9: Throughput vs. p when m = 0

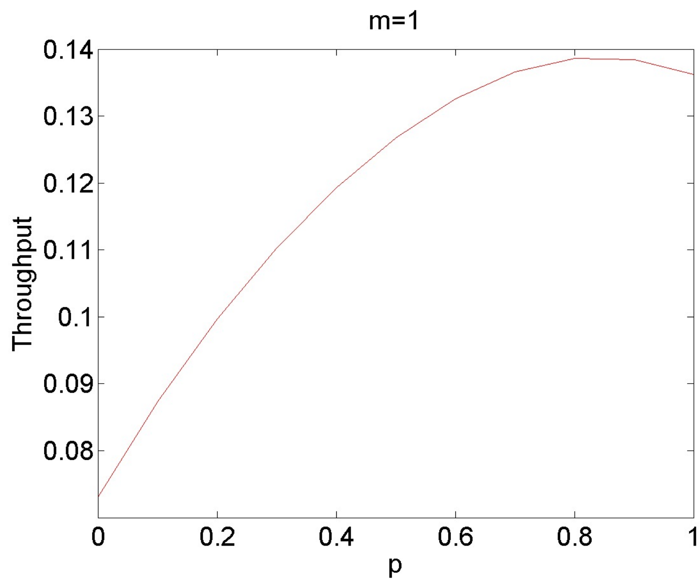

Figure 3.10: Throughput vs. p when m = 1

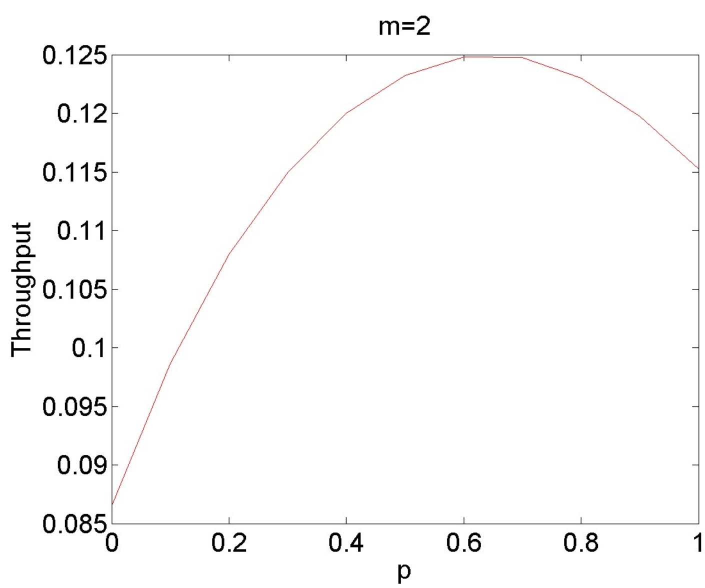

Figure 3.11: Throughput vs. p when m = 2

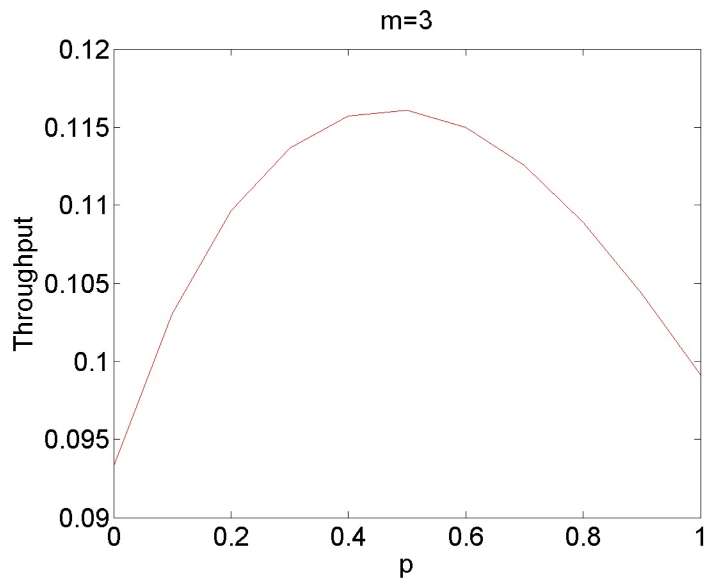

Figure 3.12: Throughput vs. p when m = 3

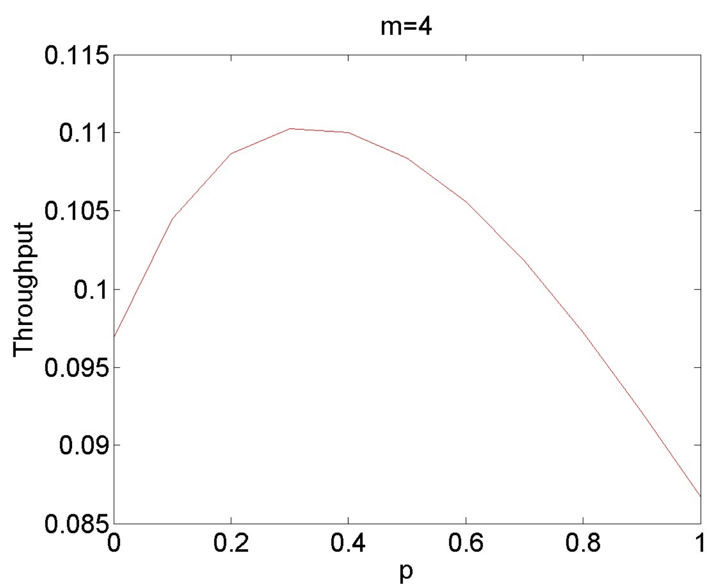

Figure 3.13: Throughput vs. p when m = 4

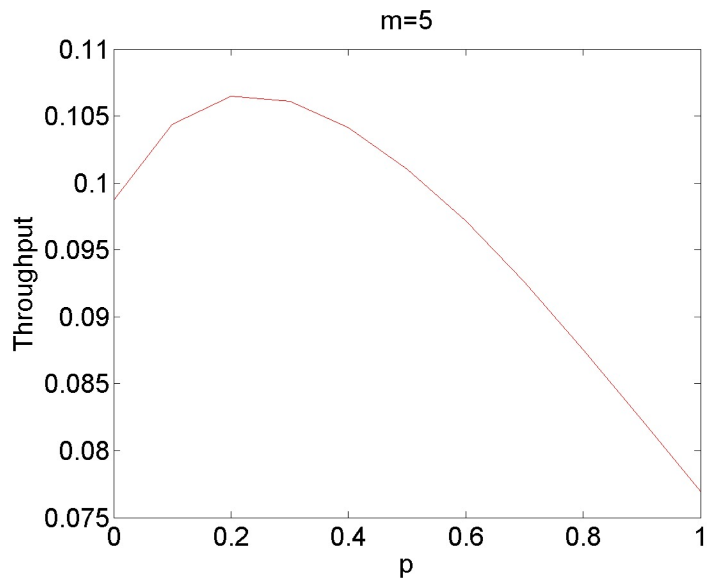

Figure 3.14: Throughput vs. p when m = 5

We can see that the optimal strategy which is the p maximum the throughput lies between points 0 and 1 which. We can get other two observations,

(1) p_{max} decreases with m going up;

(2) We benefit more from the randomized network coding scheme with a higher transmission cost.

The intuition here is that with transmission costs increasing, nodes are less willing to transmit packet. Hence, we need to punish the freerider more to achieve high throughput. Also, we lose more if we benefit all the freeriders in the exchange phase.

Figure 3.15 is the maximum throughput under different transmission cost,

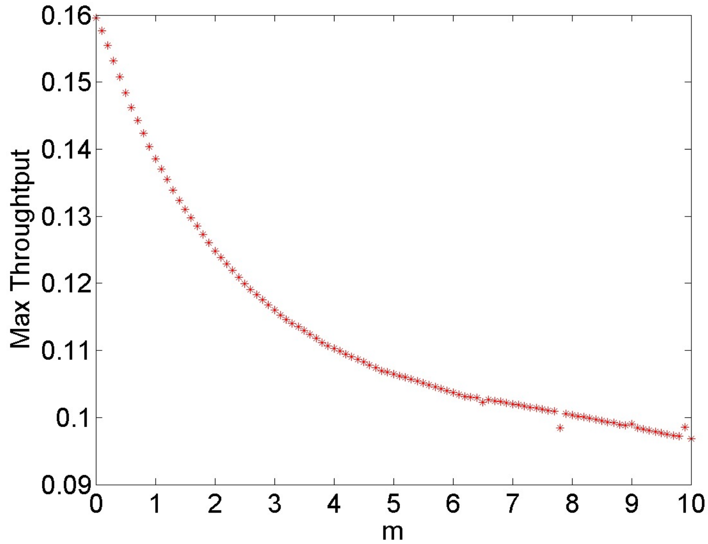

Figure 3.15: Max throughput under different transmission costs

Figure 3.15 shows that the maximum throughput decreases when m increases. This makes sense since nodes has less willingness to transmit a packet under a larger transmission cost.

Further, Figure 3.16 shows the gap between the maximum and minimum throughput when transmission cost changes,

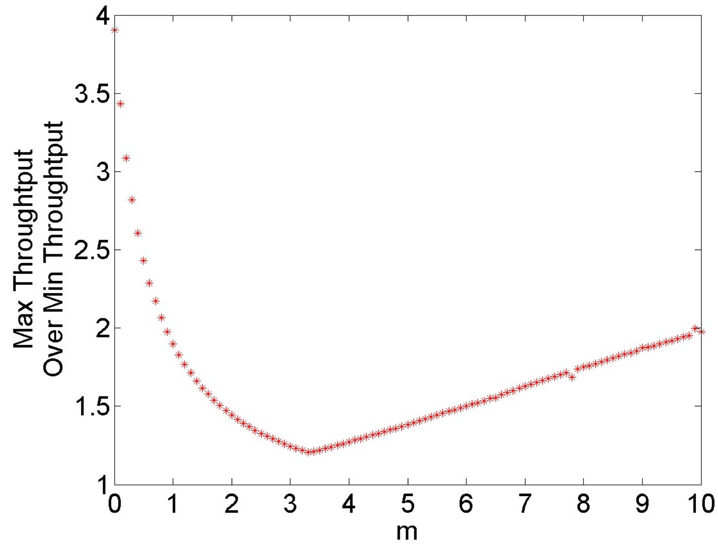

Figure 3.16: Gap between maximum and minimum throughput under different m

4. CONCLUSIONS

This thesis considers the problem of content distribution in wireless P2P networks and proposes a model that captures both the broadcast nature of wireless medium and the incentives of nodes. The thesis presents a non-monetary protocol for content distribution in this model. The protocol provides incentives for selfish nodes to contribute to the network. We have studied the performance of our protocol when all nodes are selfish. For systems with only two files, we have derived closed-form expressions for Nash Equilibria and prices of anarchy.

For systems with more than two files, we study the system performance under both scenarios where network coding is and is not employed. For each scenario, we derive a procedure for finding the Nash Equilibrium. We also propose a randomized algorithm to solve the free rider problem which introduced by network coding scheme. While it is hard to derive a closed-form expression for the throughput for the protocol with randomized network coding scheme under Nash Equilibrium, we provide simulation result instead.

REFERENCES

[1] Christina Aperjis, Ramesh Johari, and Michael J Freedman. Bilateral and multilateral exchanges for peer-assisted content distribution. *IEEE/ACM Transactions on Networking (TON)*, 19(5):1290–1303, 2011.

[2] Alberto Blanc, Yi-Kai Liu, and Amin Vahdat. Designing incentives for peer-to-peer routing. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 1, pages 374–385. IEEE, 2005.

[3] Zhu Han and H Vincent Poor. Coalition games with cooperative transmission: a cure for the curse of boundary nodes in selfish packet-forwarding wireless networks. *Communications, IEEE Transactions on*, 57(1):203–213, 2009.

[4] Hung-Yun Hsieh and Raghupathy Sivakumar. On using peer-to-peer communication in cellular wireless data networks. *Mobile Computing, IEEE Transactions on*, 3(1):57–72, 2004.

[5] Wenjiao Li, Jing Chen, and Bing Zhou. Game theory analysis for graded punishment mechanism restraining free-riding in p2p networks. In *Computer Science and Society (ISCCS), 2011 International Symposium on*, pages 262–266. IEEE, 2011.

[6] W Sabrina Lin, H Vicky Zhao, and KJ Ray Liu. Incentive cooperation strategies for peer-to-peer live multimedia streaming social networks. *Multimedia, IEEE Transactions on*, 11(3):396–412, 2009.

[7] Vishal Misra, Stratis Ioannidis, Augustin Chaintreau, and Laurent Massoulié. Incentivizing peer-assisted services: a fluid shapley value approach. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 215–226. ACM, 2010.

[8] Michael J Neely and Leana Golubchik. Utility optimization for dynamic peer-to-peer networks with tit-for-tat constraints. In *INFOCOM, 2011 Proceedings IEEE*, pages 1458–1466. IEEE, 2011.

[9] Alex Sprintson, Parastoo Sadeghi, Graham Booker, and Salim El Rouayheb. Deterministic algorithm for coded cooperative data exchange. In *Quality, Reliability, Security and Robustness in Heterogeneous Networks*, pages 282–289. Springer, 2012.

[10] Jing Zhao, Ping Zhang, and Guohong Cao. On cooperative caching in wireless p2p networks. In *Distributed Computing Systems, 2008. ICDCS'08. The 28th International Conference on*, pages 731–739. IEEE, 2008.