Delay Optimization in Multi-Hop Wireless Networks with Network Coding and Successive Interference Cancellation

Mohammadhossein Alvandi

A Thesis
In the Department
of
Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Electrical and Computer Engineering) at
Concordia University
Montreal, Quebec, Canada

September 2014

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Mohammadhossein Alvandi

Entitled: Delay Optimization in Multi-Hop Wireless Networks with Network Coding and Successive Interference Cancellation.

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Van Soung Hoa ________ Chair

Dr. Catherine Rosenberg ________ External Examiner

Dr. Chun Yi Su ________ External to Program

Dr. M. Reza Soleymani ________ Examiner

Dr. Dongyu Qiu ________ Examiner

Dr. Mustafa K. Mehmet Ali & Dr. Jeremiah F. Hayes ________ Thesis Supervisor

Approved by

____
Chair of Department or Graduate Program Director

__________ ________________________
Dean of Faculty

ABSTRACT

Delay Optimization in Multi-Hop Wireless Networks with Network Coding and Successive Interference Cancellation

Mohammadhossein Alvandi, Ph.D.

Concordia University, 2014

Wireless networks consist of a number of nodes, which communicate with each other over wireless channels. Unlike wired-networks, wireless networks have limited bandwidth, and are much more susceptible to environmental effects such as wireless interference. As a result, it is difficult to transmit information reliably at high data rates. The problem is further compounded by the quality of service (QoS) requirements such as minimum delay and maximum throughput imposed by current and future applications. That said, recent advances in coding techniques, communication protocols and architectures give the promise of future wireless networks that will proliferate high quality wireless applications.

Network coding (NC) and successive interference cancellation (SIC) have been shown to improve the throughput of multi-hop wireless networks (MWNs). NC enables a node to transmit multiple packets concurrently as a single coded packet, while SIC allows multi-packet reception (MPR) by removing interference. However, emphasis of the work done so far has been determining maximum throughput of such networks without giving consideration to QoS requirements. Maximization of the throughput may lead to paths in the network that experiences very high packet delays. The objective of this thesis is the minimization of average packet delay in a MWN for a given traffic demand matrix with joint application of NC and SIC techniques.

We formulate a cross-layer optimization that performs scheduling, routing, and more importantly capacity allocation in a way that the average packet delay is minimized. Our optimization model considers thoroughly all feasible NC and MPR opportunities in the network and allows nodes to encode up to 4 packets together.

We consider a network that uses conflict-free scheduling and has multi-path routing capability. The method is valid both in the presence and the absence of opportunistic listening on any wireless network topology and any pattern of traffic. We present numerical results to evaluate the performance of the proposed scheme. The results are also compared to that of the previous studies that treat NC and SIC separately. Our findings indicate that significant throughput improvement can be achieved by a winning combination of NC and SIC techniques.

TABLE OF CONTENTS

List of Figures……………………………………………………………...……… **vii**

List of Tables……………………………………………………………..………… **viii**

Glossary…………………………………………………………………….……... **ix**

Abbreviations…………………………………………………………...……… **xii**

1. Introduction **1**

1.1. Problem Statement and Objectives………………………………………………. 5

1.2. Contributions…………………………………………………………………….. 8

1.3. Related Work…………………………………………………………………….. 10

1.4. Thesis Organization……………………………………………………………… 14

2. Unicast Communications **15**

2.1. Network Model………………………………………………………………….... 15

2.2. Wireless Interference…………………………………………………………….. 16

2.3. Wireless Interference Models……………………………………………………. 17

2.3.1. Physical Model……………………………………………………………. 17

2.3.2. Protocol Model……………………………………………………………. 18

2.4. Routing Constraints……………………………………………………………… 19

3. Broadcast Communications **20**

3.1. Network Model and Network Coding Notation…………………………………. 21

3.2. Network Coding-Aware Routing Constraints…………………………….……... 22

3.2.1. Routing Constraints Without Opportunistic Listening………………….……… 22

3.2.2. Routing Constraints With Opportunistic Listening……………………….. 25

4. Delay Optimization with Network Coding **32**

4.1. Network Model……………………………………………………………... 33

4.2. Network Coding-Aware Routing…………………………………………… 34

4.2.1. Without Opportunistic Listening……………………………………….. 35

4.2.2. With Opportunistic Listening…………………………………………... 36

4.3. Capacity Assignment and Scheduling……………………………………… 39

4.4. Optimization Framework…………………………………………………… 42

4.5. Performance Evaluation…………………………………………………….. 43

4.6. Conclusion…………………………………………………………………... 46

5. Delay Optimization with Network Coding and Successive Interference Cancellation **48**

5.1. Network Model……………………………………………………………... 51

5.2. Capacity Assignment……………………………………………………….. 52

5.3. Scheduling Constraints……………………………………………………... 55

5.3.1. Scheduling Constraints without NC and SIC…………………………... 55

5.3.2. Scheduling Constraints with NC……………………………………….. 57

5.3.2.1. Primary Interference Constraints with NC………………………. 58

5.3.2.2. Secondary Interference Constraints with NC……………………. 58

5.3.3. Scheduling Constraints with NC and SIC…………………………………… 59

5.3.3.1. Successive Interference Cancellation……………………………. 59

5.3.3.2. Primary Interference Constraints with NC and SIC…………….. 60

5.3.3.3. Secondary Interference Constraints with NC and SIC………….. 61

5.4. Routing Constraints……………………………………………………… 63

5.4.1. Without Opportunistic Listening……………………………………….. 64

5.4.2. With Opportunistic Listening…………………………………………... 66

5.5. Cross-Layer Optimization Framework…………………………………... 68

5.5.1. Minimization of the Frame Length……………………………………... 70

5.5.1.1. Offline Generation (OG) Method……………………………... 71

5.5.1.2. Column Generation (CG) Method…………………………….. 73

5.5.2. Minimum Delay………………………………………………………… 75

5.6. Power Control Constraints……………………………………………….. 77

5.7. Numerical Results……………………………………………………….... 78

5.8. Conclusion………………………………………………………………….. 92

6. Conclusions and Future Work **93**

6.1. Conclusion………………………………………………………………….. 93

6.2. Future Work………………………………………………………………… 95

6.2.1. Backbone Routing…………………………………………………….. 95

6.2.2. Probabilistic Routing………………………………………………….. 96

References……………………………………………………………………... **98**

LIST OF FIGURES

1.1. Illustrative scenario: NC without opportunistic listening……………………….. 3

1.2. Illustrative scenario: NC with opportunistic listening…………………………... 4

2.1. Primary interference…………………………………………….……………….. 16

2.2. Secondary interference………………………………………………………….... 17

3.1. Network coding models…………………………………………………………. 23

3.2. First NC example………………………………………………………………… 27

3.3. Second NC example……………………………………………………………… 27

3.4. Third NC example……………………………………………………………….. 28

3.5. Fourth NC example……………………………………………………………… 29

4.1. 3-node network………………………………………………………………….... 41

4.2. Random network topology………………………………………………………. 44

4.3. Average packet delay versus the per node demand…………………………… 45

4.4. Average packet delay versus the per node demand…………………………… 46

5.1. The general case of SIC technique at receiving node *i*………………………….. 60

5.2. Scenario A: 2-node network…………………………………………………….. 79

5.3. Scenario B: 3-node network…………………………………………………….. 81

5.4. Scenario C: 9-node network…………………………………………………….. 83

5.5. Scenario C: The routing of NC+SIC scheme…………………………………… 86

5.6. Scenario D: 21-node MWN……………………………………………………... 87

5.7. Scenario D: Average packet delay versus the offered traffic……………………. 89

5.8. Scenario E: 25-node MWN……………………………………………………… 90

5.9. Scenario E: Performance evaluation in terms of scheduling time using CG problem ………………………………………………………………………………. 91

6.1. An example of NC comparing shortest path routing and backbone routing…… 96

LIST OF TABLES

4.1. Source-Destination Pair of Flows…………………………………………….. 44

5.1. Scenario A: Results obtained from P1 and P2 problems……………………...… 80

5.2. Scenario B: Results obtained from P1 problem under symmetric traffic……...... 81

5.3. Scenario B: Scheduling results obtained from P1 problem under symmetric traffic for the scheme of NC+SIC…………………………………………………. 81

5.4. Scenario B: Results obtained from P1 problem under non……………..………. 82

5.5. Scenario B: Scheduling results obtained from P1 problem under non…..……… 82

5.6. Scenario B: Scheduling results obtained from P2 problem under non…..……… 82

5.7. Scenario C: Node's Transmission Power…………………………..…………. 83

5.8. Scenario C: Minimum delay procedure for NC+SIC……………….…………… 84

5.9. Scenario C: Minimum delay procedure for SIC………………………………… 84

5.10. Scenario C: Minimum delay procedure for NC…………………..……………. 84

5.11. Scenario C: Minimum delay procedure for w/o NC+SIC……….…………….. 84

5.12. Scenario C: Minimum delay results…………………………….……………… 85

5.13. Scenario D: OG results for NC+SIC by varying the number of given ISs…..… 88

5.14. Scenario D: CG results for NC+SIC scheme…………………………..…….. 88

GLOSSARY

Term	Definition
T_s	Frame length
T	Set of indices of $\|T\|$ slots in the frame
τ	Slot duration (seconds)
r_i	Transmission range of node i
R_i	Interference range of node i
C	Channel capacity (packets per second)
C_i	Capacity assigned to node i
A_i	Total output flow of node i (packets per second)
μ	Mean packet length
$m_i(t)$	Probability density function of service time of a packet at node i
$\tilde{m}_i(t)$	Probability density function of service time of a packet arriving at an empty queue at node i
$\hat{m}_i(t)$	Probability density function of service time of a packet arriving at a non-empty queue at node i
$\tilde{M}_i(s)$	Laplace transform of $\tilde{m}_i(t)$
$\hat{M}_i(s)$	Laplace transform of $\hat{m}_i(t)$
$E[m_i]$	Mean service time of a packet at node i
$E[\tilde{m}_i]$	Mean service time of a packet arriving at an empty queue at node i
$E[\hat{m}_i]$	Mean service time of a packet arriving at a non-empty queue at node i
$E[\tilde{m}_i^2]$	Second moment of $\tilde{m}_i(t)$
$E[\hat{m}_i^2]$	Second moment of $\hat{m}_i(t)$
I_i	Set of nodes which may have interference with node i
θ_i	Subsets of busy nodes in I_i
φ_i	Subsets of idle nodes in I_i
$F(s)$	Laplace transform of the transmission time of a packet

π_i	Probability that node i's packet is selected for transmission
$Q_i(k)$	Probability that node i's packet is selected for transmission on the k^{th} transmission
u_i	Number of slots allocated to link i
b_i	Average number of packet arrivals per frame for link i
$\bar{d}_i$	Average queuing delay at link i
$\bar{D}$	Average packet delay in the network
ρ_i	Utilization of link i
$\rho_{\max}$	Max.link utilization
L	Set of unicast links
N	Set of network nodes
e_{ij}	Directed link from node i to node j
a_{ij}	Traffic of link e_{ij}
d_{ij}	Distance between nodes i and j
F	Set of flows in the network
$s(f)$	Source node of flow f
$d(f)$	Destination node of flow f
$\lambda(f)$	Average arrival rate of flow f (packets per second)
$\Lambda(f)$	Average arrival rate of flow f packets (packets per frame)
P^f	Set of available paths for flow f
$r^f(p)$	The amount of traffic on path p from flow f
P_i	Transmission power of node i
$P^i_{\max}/_{\min}$	Max./Min. transmission power of node i
P_{ij}	Received signal power of node j from node i
α	Path loss factor
N_0	Ambient noise power level
β	SINR threshold
$\Delta_i[t]$	State of transmission of node i during slot t
$\Lambda_i[t]$	Number of packets received by node i at slot t
χ_i	Max. number of packet reception by node i at a time slot

w_i	Node degree
L_b	Set of links including unicast and broadcast links
$e_{i,B}$	Broadcast link from node i to node set B
$a_{i,B}$	Traffic of link $e_{i,B}$
$a_{i,B}^k$	Traffic associated with coding opportunity $Z_{i,B}^k$ over link $e_{i,B}$
$x_{i,B}^k[t]$	State of coding opportunity $Z_{i,B}^k$ over link $e_{i,B}$ during slot t
$x_{i,B}[t]$	State of transmission of link $e_{i,B}$
$\Upsilon_{i,B,(q)}[t]$	State of concurrent transmissions of node q and node i (over link $e_{i,B}$) in slot t.
$v[t]$	Status of slot t
$Z_{i,B}^k$	NC opportunity
$K_{(i,B)}$	Number of different NC opportunities which arise over link $e_{i,B}$.
s	Coding element of a NC opportunity: (k,hij,v)
S	Coding structure of a NC opportunity
$b(S)$	Set of next-hop nodes of each coding element in S
Γ_i	Set of all possible coding structures at node i
$y_i(s)$	Number of packets associated with s at node i
$u_i^f(p)$	Number of flow f packets on path p transmitted as native packets from node i
γ	Total traffic generated in the network
Y_i	Independent set (IS) i
c_i	Number of slots allocated to IS i (Y_i)
M	The set of all feasible ISs in the network
M'	A subset of M that covers all links in L_b
E_i	Number of times that a same IS is found
$E_{\max}$	Max. limit of E_i
$\sigma_{i,B}$	Dual variable of the capacity constraint of link $e_{i,B}$
$\Theta_{h,i}^q$	Equal to 1 if $P_{qi}>P_{hi}$; otherwise 0
$\Omega_{h,i}^q$	Equal to P_q if $P_{qi}\leq P_{hi}$; otherwise 0

ABBREVIATIONS

NC	Network coding
SIC	Successive interference cancellation
MWN	Multi-hop wireless network
MPR	Multi-packet reception
QoS	Quality of service
SINR	Signal to interference and noise ratio
LP	Linear programming
NLP	Non-linear programming
MILP	Mixed integer linear programming
MINLP	Mixed integer non-linear programming
OG	Offline Generation
CG	Column Generation
TDMA	Time division multiple access
MAC	Media access control
RAM	Random access memory
IEEE	Institute of electrical and electronics engineers
IS	Independent set
WiMAX	Worldwide interoperability for microwave access
LHS	Left hand side
RHS	Right hand side
RMP	Restricted master problem
PP	Pricing problem
OPL	Optimization programming language

CHAPTER 1
INTRODUCTION

Wireless communication is an essential component of modern telecommunications. This technology has shown tremendous growth in a very short time period and has impacted our way of life profoundly. Wireless networks have extended the services given to the wireline users to the mobile users. Wireless networks are more constrained by limited bandwidth than wired-networks, and are much more susceptible to environmental effects such as fading and interferences. As a result, it may be difficult to transmit information reliably at high data rates.

Capacity is a precious resource in wireless networks because of limited spectrum availability. This resource has become even more valuable with the increasing popularity of wireless networks. Wireless capacity has to be used more efficiently to enhance the overall performance of a wireless network. This implies that bandwidth should not be wasted especially on links, which carry little traffic, and effective capacity allocation among wireless links carrying different amounts of traffic should be performed. The underlying philosophy is to allocate capacity in response to traffic demand.

The capacity of a wireless radio is limited by the physical-layer technology that is used. Also, the wireless medium is a shared medium and the effective achievable capacity is limited by interference in the network [1]. The unique nature of wireless networks necessitates a cross-layer approach, involving the physical, MAC and the network layers, to the capacity assignment problem as opposed to the traditional network-layer-based solution.

One significant advance in coding theory in the past decade has been *network coding* (*NC*). NC has been introduced to improve the throughput of communication networks. In NC, a node may combine the packets within a window in its queue into a single outgoing packet. The packets that are combined are referred to as *native* packets. The outgoing packet, which is a combination of native packets, is referred to as a *coded* packet. The destination nodes can decode the received packet if they already have the other native packets either from the previous transmissions or opportunistic listening. When all the destination nodes decode the packet, then coding window moves to a new set of packets.

Multi-hop wireless networks (MWNs) provide good opportunities for NC because of the broadcast nature of the medium [2], [3]. Research on NC can be divided into two main categories: intra-session, where coding is done over the packets belonging to the same session or flow [4]-[6], and inter-session, where coding is applied to packets from different sessions or flows. In this thesis, we take into account inter-session NC.

Based on how the next-hop nodes obtain the other native packets, inter-session NC may be divided into two subcategories. In the first subcategory, next-hop nodes only use their previous transmissions to decode the coded packet; typically, this category is referred to as *NC without opportunistic listening*. In the second one, next-hop nodes in addition to the previous transmissions may use the overheard packets to decode the coded packet; this category is referred to as *NC with opportunistic listening*.

Next to NC, *successive interference cancellation* (*SIC*) has gained much popularity to save bandwidth in wireless networks; this technique has attracted an increasing interest to improve performance of higher layers in MWNs [7], [8]. SIC is a physical-layer technique that improves the performance by exploiting interference in lieu of avoiding it; i.e., SIC allows multi-packet reception (MPR) by removing interference. SIC enables decoding of multiple signals in a sequential manner to either remove interfering signals or receive multiple packets simultaneously. More specifically, SIC decodes the interfering signals stronger than the intended transmission. The decoded signal is then cancelled to mitigate the interference to the packets which are not yet decoded.

As explained above, NC and SIC are two techniques that have been shown to improve the throughput of wireless networks. NC enables a node to transmit multiple packets concurrently as a single coded packet, while SIC enables reception of multiple packets simultaneously by a node from different transmitters. To illustrate the basic idea of NC and SIC, we use a simple scenario shown in Fig. 1.1. In this figure, wireless nodes *a* and *c* want to send their packets p_1 and p_2 to each other through node *b*. For simplicity, we shall assume that time axis is slotted and packet durations are fixed. Using standard techniques of packet forwarding, this process needs 4 time slots as shown in Fig. 1.1a.

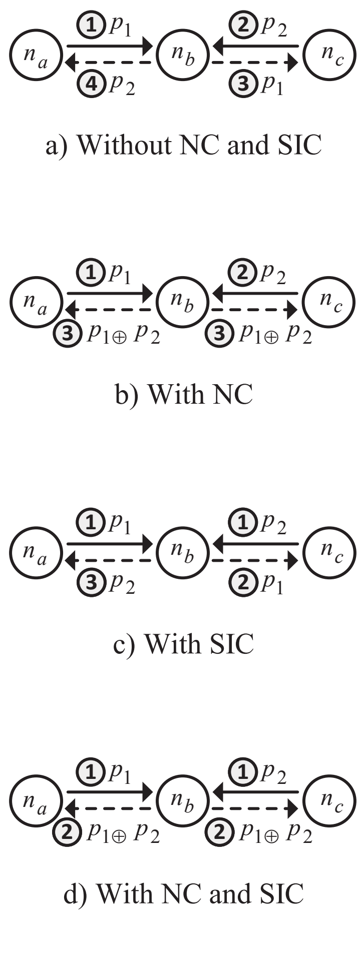


Figure 1.1. Illustrative scenario: NC without opportunistic listening

Now consider Fig. 1.1b, where a simple form of NC is shown; with NC, node *b* XORs two received packets, and then transmits the XORed packet. Due to the broadcast property, the coded packet can be received by the next-hop nodes. Now, nodes *a* and *c* can obtain the desired packet by XOR-ing the coded packet with their own packet (i.e. previous transmission). Thus, this process requires 3 time slots. Indeed, Fig. 1.1b demonstrates the case of NC without opportunistic listening.

Next, consider Fig. 1.1c. Assuming the power level of the signals (from nodes *a* and *c*) received by node *b* are different, node *b* may be able to receive both packets p_1 and p_2 in the first time slot by using SIC (under certain conditions as explained in chapter 5). In this case, node *b* first tries to decode the stronger signal, and then subtracts it from the aggregate signal. After that node *b* can decode the weaker signal as well. Hence, by using SIC, 3 time slots are needed as shown in Fig. 1.1c.

Finally, using both NC and SIC techniques, only 2 time slots may be needed to accomplish the process as shown in Fig. 1.1d. Accordingly, throughput gain via SIC or NC alone is 4/3=1.33 in this scenario; interestingly, the throughput gain via both NC and SIC is 4/2=2, which is very remarkable compared to 1.33.

In addition, NC with opportunistic listening can be used jointly with SIC technique. For example, consider Fig. 1.2. As in the previous scenario, nodes *a* and *c* want to exchange their packets through node *b*; similarly, nodes *d* and *e* want to exchange their packets as well. Clearly, without NC and SIC, this process needs 8 time slots.

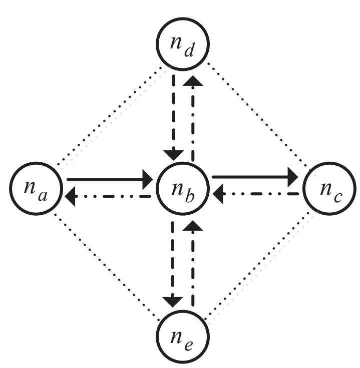


Figure 1.2. Illustrative scenario: NC with opportunistic listening

Next let us consider the case of NC. It will take 4 timeslots for the nodes to send their native packets to node *b*. Then assuming perfect overhearing (i.e., as shown in Fig. 1.2, nodes *a* and *c* can overhear nodes *d* and *e*, and vice versa) node *b* can transmit a single coded packet by XORing all of the 4 packets. Finally, each node would be able to decode their intended packets by XOR-ing the coded packet with the overheard packets. Therefore, with NC the packet transfers are completed using 5 time slots. In the case of ideal SIC, node *b* can receive 4 packets in the first time slot, and needs 4 further time slots to send the packets to the destinations. Thus, this process needs 5 time slots with SIC. Finally, using ideal NC and SIC, only 2 time slots may be needed to complete the process. Accordingly, in this scenario throughput gain via NC or SIC alone is 8/5=1.6, and the throughput gain via both NC and SIC is 8/2=4, which is very considerable compared to 1.6.

As we have seen, NC and SIC are two promising techniques that enable a node to take advantage of concurrent transmissions and receptions, respectively. This approach leads to significant throughput improvement in MWNs, although in practice, the gains may tend to be lower due to sometimes of NC and SIC opportunities.

1.1. Problem Statement and Objectives

In this thesis, we investigate the potential benefits of NC and SIC in MWNs that result in significant performance improvement. As explained earlier, MWNs provide good opportunity for NC because of the broadcast nature of the medium. Besides, SIC is a promising technique to mitigate the wireless interference particularly when multiple users are active within interference range of each other. The past work on optimization of MWNs, particularly with NC, has mainly addressed the problem of determining the maximum throughput, by maximizing a scaling factor of all traffic flows, maximizing the minimum traffic flow, maximizing sum of the traffic flows, or minimizing the length of scheduling under given traffic [9]-[12]. In [9], the authors address the maximizing throughput for coding-aware network routing and define the throughput as the maximum multiplier, or scaling factor λ such that all demands with their values multiplied by λ can be feasibly routed by the network. The authors in [10] study a weighted optimization problem for $max - min$ throughput in the presence of NC in MWNs. In this work the minimum traffic flow in the network is maximized; further, they consider a weighting factor for each flow to ensure that downlink flows receive higher throughput than uplink flows (from gateway). A cross-layer optimization problem maximizing the sum of the weighted traffic flows has been studied in [11] in the presence of SIC. In this work the authors define the throughput as the sum of traffic flows in MWNs and assume that each flow has been assigned a weight. In [12] the authors present an optimization problem that minimizes the length of scheduling in MWNs with NC. The optimal schedule determines the minimum number of time slots allocated to wireless links, needed to satisfy the traffic demands of all flows in the network.

However, the approach of the maximizing throughput has a drawback since it does not deal with QoS requirements of the users. When the throughput is maximized the traffic of a link may approach its capacity, which would lead to unacceptable packet delays. Thus, there may be paths in the network for which the packet delay is prohibitively high. The objective of our work is the optimization of MWNs such that the average packet delay in the network is minimized for a given traffic demand matrix. We propose two solution methods for this problem.

In the first method [13], we focus on the optimization of MWNs when only NC is used. Indeed, we address the following question: given a specific placement of wireless nodes and traffic demand matrix, what is the optimum capacity allocation that minimizes the end-to-end delay of the network? We model each wireless node as an M/G/1 queue with service interruptions. It is assumed that at each service epoch the server chooses the next node to serve randomly according to their traffic loads among the nodes within the transmission range of each other. The service time of a packet at a node increases by the amount of the service given by the server to the nodes within its transmission range. The effect of wireless interference on the performance of network has been taken into account by including scheduling constraints in the optimization framework. Further, the routing paths of flows are chosen by routing constraints in such a manner to further reduce the mean packet delay in the network. We incorporate multi-path routing in order to create more coding opportunities in the network, which provides a method for computing source-destination routes and utilizing the best coding opportunities from available ones.

Finally, we perform simulation to determine the accuracy of our analytical model. The results show that NC leads to reduced end-to-end delay in the network and more importantly, extends the stable operating region of the network. Note that considering each wireless node as an M/M/1 queue [14] may not capture properly the features of a wireless node; such a model has the limitation of modeling the wireless nodes within the transmission range of each other as parallel servers working simultaneously with slower rates. However, in a wireless system, nodes within the transmission range of each other receive service with interruptions.

In the second solution method [15], we extend our previous work by considering the joint application of NC and SIC under spatial TDMA MWNs. In spatial TDMA, links with sufficient spatial separation may use the same time slot for transmission [16]. As shown in [17], the spatial TDMA method performs better in MWNs. Under the assumption of Poisson arrival of packets, the average packet delay of a TDMA queuing system has a closed-form expression [18], [19]. An important feature of the model is that multiple slots can be assigned to a link in the network. We use this model to find the minimum average packet delay in the network. We note that there has been work that optimizes the MWN through the minimization of the TDMA frame length for given traffic demands [12], [20], [21]. However, minimization of the frame length does not necessarily lead to the solution with the minimum packet delay. Further, there may be many solutions with the same minimum frame length. The methods solving this type of problems return typically anyone of these solutions, which may not correspond to the optimal packet delay.

When inter-session NC is used, the coding is done over packets from different sessions or flows at a node in which the flows cross each other. To fully exploit NC, the routing of the flows should be close to each other. However, this may lead to a high delay in bottleneck nodes due to the increasing level of interference. In this thesis, we combine NC with SIC technique, which alleviates the interference. Our goal is to provide a model that fully exploits the benefits of concurrent transmissions and receptions. In other words, we address the following fundamental question: given a specific placement of wireless nodes and a set of traffic demands, what is the optimum routing, scheduling and more importantly, capacity assignment for the network links that minimize the average delay of packets in the system. To answer the above question, we derive a cross-layer optimization model, involving physical, MAC, and network layers, to make capacity assignment, scheduling and routing decisions more effectively. We take into account the effect of wireless interference through incorporating scheduling constraints into the model. In addition, the multipath-routing constraints create more NC and SIC opportunities in the network and we provide a method for utilizing the best opportunities from available ones.

In summary, we formulate the problem of delay minimization of MWNs with NC and SIC as a mixed integer non-linear programming (MINLP) problem. Due to the non-linearity of the objective function, the problem is only solvable for small-sized networks by the state-of-the-art software. Then we propose a method that uses a linearly objective function which determines the TDMA scheduling frame length. In fact, this method finds the minimum delay iteratively by finding the minimum scheduling length under restricted link utilization, which is in the format of mixed integer linear programming (MILP). Then for larger networks we present two optimization models, namely offline generation (OG) and column generation (CG), which are derived by the decomposition of the MILP problem. In addition, we compare the performance of OG and CG models with each other. Finally, to increase the SIC opportunities in the network we present power control constraints which enables nodes to adjust their transmission powers.

We note that in the second solution method capacity is assigned to the links while in the first method it is assigned to the nodes. Further, in the second method in order to incorporate SIC into the optimization framework, we use physical instead of protocol model of the channel to capture more accurately wireless interference in the network.

1.2. Contributions

Next, the main contributions of this research are summarized below. The first two contributions are due to the first solution method, while the remainder is from the second solution method.

- Presenting a cross-layer optimization model that determines the minimum average packet delay in MWNs with NC. Each wireless node is modelled as an M/G/1 queue with service interruptions with variable length packets. The theoretical framework is presented as a joint multi-path routing and conflict-free scheduling problem. Node assignment and protocol model is employed in this model.
- We performed simulations to determine the accuracy of our analytical model. The results show that NC leads to reduced average packet delay in the network and more importantly, extends the stable operating region of the network; thus increasing per node throughput.
- Formulation of a cross-layer optimization to determine the minimum packet delay in TDMA-based MWNs with the combined use of NC and SIC. The theoretical framework is presented as a joint multi-path routing and conflict-free scheduling problem. Further, the power control constraints are presented as an extension. Link assignment and physical model for interference is employed in this solution method. The numerical results show that the average packet delay and traffic handling capacity of a network using w/o NC+SIC, NC, SIC and NC+SIC schemes improves from left to right. Traffic capacity of NC+SIC is double of the w/o NC+SIC. Thus combined utilization of NC and SIC techniques results in significant performance improvement.
- For the optimization model using joint application of NC and SIC, two decomposed model, OG and CG problems, are presented for large-sized networks. Further, we compare the performance of these two methods.
- Our analysis is applicable to any given MWN topology with any pattern of concurrent traffic flows; further, it is valid both with and without opportunistic listening. The optimization models consider thoroughly all possible NC and MPR opportunities in the network. We consider all feasible NC models in which the coding node is allowed to encode up to 4 packets.

We note that SIC indirectly improves the performance of NC by allowing the routing of the flows to be closer to each other, which results in an increase in number of NC opportunities in the network. SIC achieves this by allowing MPR and mitigating the interference in the network.

1.3. Related Work

The capacity assignment problem is one of the most important topics in communication networks. A number of papers have been published on the problem of capacity allocation which minimizes the delay in a network. In the classic capacity assignment problem [22], Kleinrock addressed the capacity allocation for wired-networks which minimizes the average packet delay in the network subject to a cost constraint. In this problem, the network topology and routing; i.e., the loads on different links are given. The problem is to allocate capacities to different links in order to enhance the overall network performance. This problem was found to have a simple closed-form solution. However, the problem in wireless networks is more complex than from wired-networks since the wireless medium is a shared medium and the effective achievable capacity is limited by interference in the network.

Research has been conducted on the capacity of a class of wireless networks; viz. Ad-hoc networks [23]-[25]. In [24], physical-layer capacity enhancement techniques for ad-hoc networks have been proposed to satisfy certain delay constraints. The capacity allocation problem is different because it is concerned with the proper allocation of the available capacity, which is provided by the physical layer. In [26] the authors have studied the capacity allocation in MWNs; they, in fact, proposed a cross-layer approach for capacity allocation in wireless mesh networks that minimizes the average packet delay. More specifically, they extended the work of Kleinrock by considering the impact of wireless interference on the network. In stationary multi-hop networks, the node locations are fixed by the service provider based on market research. Topology control algorithms [27]-[29] can be used to determine the best topology having certain desirable properties based on the node locations and the current network state. The desired topology so obtained is then practically realized by the wireless physical layer. Long-term traffic pattern in the network can be monitored by the network administrator who can then allocate wireless capacity to different links for efficient use of the limited wireless capacity [26]. The drawback of [26] is that the authors use M/M/1 queuing system to model wireless links; however, as mentioned before, this assumption has the limitation of modeling the wireless links interfering with each other as parallel servers working simultaneously with slower rates. Hence, their model falls short in handling wireless interference.

As mentioned before, research on NC can be divided into two main categories: intra-session NC and inter-session NC; i.e., NC can be employed for both multicast and unicast traffic in the network and it can increase the overall throughput of networks from different aspects.

Intra-session NC has been extensively studied, beginning with the pioneering paper [30]. The authors show that having the routers mix information from different messages allows the communication to achieve multicast capacity. In [4] the authors show that for multicast traffic linear codes are sufficient to achieve the maximum capacity bounds. At the same time, an algebraic approach proposed in [31] and showed that coding and decoding can be done in the polynomial time. [32], [33] presented the concept of random linear NC, which makes NC more practical, especially in distributed networks such as wireless networks. In the last few years, many researchers have made efforts to develop viable NC techniques in wireless networks [34], [35]. A great deal of attention has been focused on dealing with practical issues and developing implementable protocols with NC [36], [37], [38]. In particular authors in [39] study intra-session NC and show that the problem of minimizing the communication cost can be formulated as a linear program and solved in a distributed manner.

The benefit of inter-session NC has been demonstrated by COPE [2]; COPE has shown the capability of NC for increasing the throughput, and developed a practical approach that bridges the gap between the theory of NC and its implementation in practice. By combining what one neighbor wants with what other neighbors have, a router with COPE can transmit multiple packets to different neighbors in a single transmission.

Although [31] showed that in general inter-session NC is very difficult, COPE circumvents the complicated issues by decoding at each hop and was demonstrated to provide three to four times the throughput improvement over traditional routing packets through the MWN.

A number of recent works, including [9], [10], [12], [40], [41] addressed the problem of maximizing throughput by formulating joint routing and scheduling problems using NC. In [9], a NC-aware routing scheme in MWNs have been presented for both with/without opportunistic listening mechanisms. Further, in [9] clique constraints are employed to schedule wireless transmissions free of interference in the network similar to [42]; However, the authors only considered coding the maximum of two packets together at a time. The authors in [10] presented a joint routing and scheduling and NC formulation based on the physical interference model in a network where all nodes use the same transmit power and the same modulation/coding scheme. In [12] the authors consider MWNs with WiMax-based backhaul links. They present a cross-layer optimization problem that minimizes the TDMA scheduling length under joint routing, scheduling and power allocation formulation with NC. In [40], an analytical model for computing the maximal throughput of unicast flows that can be achieved by co-operative NC in multi-rate MWNs has been proposed. K-tuple coding studied in [41]; by this method, wireless nodes do not require overhearing under certain conditions. The proposed model is formulated under 2-hop wireless interference.

SIC has been also studied in the literature. SIC is a physical layer technique; a classic reference on interference cancellation is [43]. More details and new advances of some important interference cancellation techniques may be found in SIC [44], parallel interference cancellation [45], and iterative interference cancellation [46], which all intend to enable a wireless receiver to decode multiple signals simultaneously, and reject interference from other unintended transmitters. Recently, a study on the application of interference cancellation technique in cellular systems [47], rated SIC as one of the most promising techniques to reduce interference because of its simplicity and effectiveness.

SIC technique, but not NC, has been also considered for designing of routing and scheduling schemes in MWNs [8], [11], [48]. In [8], the link scheduling problem in ad-hoc networks with SIC has been analyzed. Independent set based greedy scheme has been studied for determining a feasible schedule; further, to reduce the complexity at the network-layer, fixed routing has been used. [11] presented a throughput maximization framework for joint interference exploitation and avoidance with SIC technique. They proposed a cross-layer model to handle scheduling and routing problem in MWNs. In [48], Joint scheduling and routing under the SIC scheme for maximizing network throughput has been studied; SIC scheme has been included in scheduling constraints by allowing concurrent receptions in independent sets. The independent sets are generated offline.

Very recently, [49] provided an analysis of the combined use of NC and MPR under IEEE 802.11 MAC protocol. The authors show that NC+MPR gain decreases when the wireless medium is congested because the current 802.11 MAC is fair to the nodes and not to the flows. Their analysis considers a 5-node NC model in which only the center node is allowed to perform NC. Further, they consider 2 MPR models, namely CSMA/CA and MPR-adapted CSMA. The first model limits the number of receptions to 2 while the second model limits the number of receptions to 4.

Finally, note that the SIC technique used in this work differs from the new forms of interference cancelation such as analog NC [50] and ZigZag decoding [51]. These schemes are not blind and require knowledge of some bits in one of the colliding packets. We should add that in this work we consider digital NC, and not analog NC. Note the analogy between digital NC and its analog counterpart. In digital NC, senders transmit sequentially, and routers (coding nodes) XOR the content of the packets and broadcast the coded version. In analog NC, senders transmit simultaneously. The wireless channel naturally mixes these signals. Instead of forwarding coded packets, routers amplify and forward the mixed signals they receive [50]. Note that analog NC is more suitable for 2-way relaying scenarios, and in this scheme the coding node can only receive 2 packets simultaneously [52].

It appears that the problem of average packet delay optimization with joint application of NC and SIC in MWNs have not been studied in the literature until present time, which has been taken as the main goal of this work.

1.4. Thesis Organization

The rest of this thesis is organized as follows. In chapter 2, we review the traditional unicast communications. We briefly study the wireless interference model under both protocol and physical models. Furthermore, we show the formulation of routing constraints in the form of linear programming. In chapter 3, we generalize the models used in unicast communications to broadcast communications where NC is used. In chapter 4, we address the problem of minimum average packet delay in MWNs with NC. We propose a cross-layer optimization problem with/without opportunistic listening and model each node as an M/G/1 queue. The theoretical formulation is formulated as a conflict free scheduling and multi-path routing problem. We further compare the numerical results with simulation results in this chapter. In chapter 5, we study the problem of average packet delay in TDMA-based MWNs with the joint application of NC and SIC techniques. For large-sized networks we propose two linear optimization frameworks, namely, columns generation and offline generation problems. Finally, chapter 6 contains the future work and conclusions.

CHAPTER 2

UNICAST COMMUNICATIONS

In this chapter, we study wireless interference models and the issues pertaining to routing in unicast communications, where each transmitted packet is destined to a single node. Thus, in unicast communications all transmissions are unicast and NC is not used in the network. We shall show the formulation of routing constraints in linear programming (LP) format.

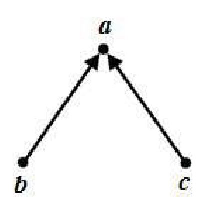
2.1. Network Model

The MWN can be represented by a connectivity graph $G(N,L)$ where N is a set of vertices denoting the nodes, and L is a set of directed edges denoting the unicast links. We let n_i denote node i and e_{ij} a unicast link from node i to node j. We will have $e_{ij} \in L$ if node j can successfully receive a packet from node i. We assume that nodes will communicate in half-duplex mode, so they cannot transmit and receive simultaneously.

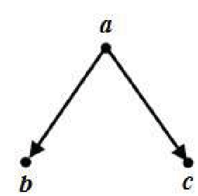
The packets between each source-destination node pair will form a flow in the network. Packets of a flow may have to travel multiple hops between source and destination. Letting F denote the set of flows in the network, for a flow $f \in F$, we will let $s(f)$ denote the source node, $d(f)$ the destination node, and $\lambda(f)$ the arrival rate of packets per second. A flow may be routed through multiple paths in the network. Let P^f denote the set of available paths from source $s(f)$ to destination $d(f)$ for flow f. For instance, one may choose the K-shortest distance paths from $s(f)$ to $d(f)$ as the set P^f. Let routing variable $r^f(p)$ denote the amount of traffic on path p for flow f. For a path p and nodes h and i we will use $e_{hi} \in p$ to denote that link e_{hi} is on path p. We also use a_{ij} to denote the arrival rate of traffic from node i to node j in the network.

2.2. Wireless Interference

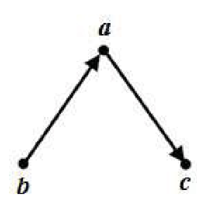
The multi-hop nature of the network makes spatial reuse possible in the sharing of the channel; hence, multiple nodes/links can transmit simultaneously, if their transmissions do not interfere with each other. In the sequel, we will describe how transmissions may collide in wireless networks. In particular, collision may occur in two ways: first, as shown in Fig. 2.1, it can happen when a node has to perform more than one activity at the same time. This is because, the nodes cannot transmit and receive simultaneously and cannot transmit/receive more than one packet at the same time. This interference is typically referred to as *primary* interference.



a. A node receiving multiple packets



b. A node transmitting multiple packets



c. A node transmitting and receiving simultaneously

Figure 2.1. Primary interference

Second, as shown in Fig. 2.2 a collision may occur when a receiver *a* tunes to a particular transmitter *b* but it is within the range of transmitter *c* whose transmissions, though not intended for *a*, interfere with the transmissions of *b*. This interference is referred to as *secondary* interference [53].

We note that in the case of link assignment (or equivalently link scheduling), where the capacity is assigned to links, multiple links may reuse the channel if their transmissions do not interfere with each other. Similarly, in the case of node assignment, where the capacity is assigned to nodes, multiple nodes may reuse the channel if they do not have interference with each other.

In the next section, we shall review two models for the secondary interference.

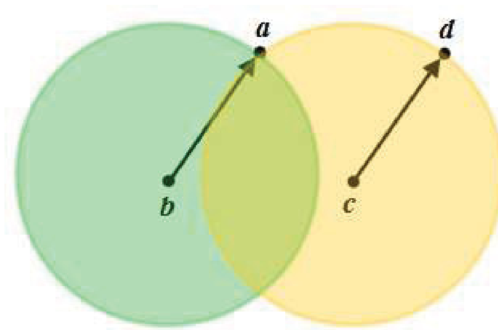


Figure 2.2. Secondary interference.

2.3. Wireless Interference Models

2.3.1. Physical Model

Let P_h denote the transmission power of node h, and d_{hi} denote the distance between nodes h and i. Then using free path loss model, the received signal power of node i from the transmission of node h is given by $P_{hi}=P_h d_{hi}^{-\alpha}$, where α is the path loss factor. In an urban environment a typical value of α is three.

Let N' denote the subset of nodes in N which are active at the time of transmission of node h. Then according to the physical model [1], the transmission from node h is successfully received by node i if

$$SINR_{hi} = \frac{P_{hi}}{N_0 + \sum_{m \in N', m \neq h} P_{mi}} \geq \beta, \tag{2.1}$$

where β is the signal-to-interference-and-noise ratio (SINR) threshold and N_0 is the ambient noise power level. Under the above condition, receiver i treats all the interfering signals from the other ongoing transmissions as noise. Note that under physical model, the link e_{hi} from node h to node i exists only if $P_{hi}/N_0 \geq \beta$.

2.3.2. Protocol Model

Protocol model is a simplification of the physical model. In this model a node is assigned transmission and interference range. Transmission range of node h, denoted by r_h, is the maximum distance under free path loss model that its transmission may be received by a node i in the absence of any interference. Further, in the protocol model, it is assumed that a transmitting node cannot interfere with a receiving node if the distance between them is higher than interference range. Let R_h denote interference range of node h, typically, $r_h \leq R_h \leq 2r_h$.

Let d_{hi} denote the distance between nodes h and i. Then according to the protocol model, transmission of node h to i will be successful if $d_{hi} \leq r_h$ and any node $m \in N$, such that $d_{mi} \leq R_m$ is not transmitting [42].

We point out that in both physical and protocol models what matters is the interference at the receiver and not at the transmitter.

2.4. Routing Constraints

Next, we present multi-path routing constraints for unicast communications for MWNs.

$$\sum_{p \in \mathrm{P}^f} r^f(p) = \lambda(f) \qquad \forall f \in F \tag{2.2}$$

$$a_{ij} = \sum_{f \in F} \sum_{p \in \mathrm{P}^f, e_{ij} \in p} r^f(p) \qquad \forall e_{ij} \in L \tag{2.3}$$

Constraints (2.2) ensures that sum of the traffic routed on available paths of flow f equals to the total traffic generated by flow f; i.e., $\lambda(f)$.

Constraints (2.3) determine the total traffic that is unicast from node i to node j. Note that this traffic is composed of two types of traffic. The first type is the traffic that is generated by node i and the second type is the transit traffic traversing nodes i and j through path p.

CHAPTER 3

BROADCAST COMMUNICATIONS

In this chapter, we generalize the unicast communications to broadcast communications in order to enable implementation of NC. In other words, we take into consideration concurrent transmissions using NC in the network. NC indirectly enables transmission of multiple unicast packets simultaneously to different destinations by a node through coding them into a single packet. Consider m packets p_1, p_2,…, p_m that are received at node n_i from distinct previous-hop nodes (not necessarily during the same time slot). Suppose that the above packets also need to be transmitted to distinct next-hop nodes n_1, n_2,…, n_m respectively. We note that the sets of the previous-hop and next-hop nodes do not need to be the same. By NC, coding node n_i can XOR all the packets together and broadcast a coded packet to all the next-hop nodes, therefore such a transmission is referred to as a broadcast transmission. Each of the next-hop nodes must be able to decode the coded packet in order to recover the intended packet. A node will be able to decode the coded packet, if it already has all the packets except for the intended packet to itself. A node will have all the other packets either from its previous transmissions or through snooping on the transmissions in the medium. The above process ensures that each next-hop node is able to decode the coded packet and extract the intended packet.

In this chapter, similar to the unicast communications we derive the routing constraints in the form of LP formulation when NC is used.

3.1. Network Model and Network Coding Notation

The network model is identical to that of the previous chapter, but nevertheless for the sake of completeness it will be repeated here.

We consider a MWN which is represented by a connectivity graph $G(N,L)$ where N is a set of vertices denoting the nodes, and L is a set of directed edges denoting the unicast links. We let n_i denote node i and e_{ij} a unicast link from node i to node j. We will have $e_{ij} \in L$ if node j can successfully receive a packet from node i. Each node is equipped with a single radio using an omni-directional antenna. We assume that nodes will communicate in half-duplex mode, so they cannot transmit and receive simultaneously.

The packets between each source-destination node pair will form a flow in the network. Letting F denote the set of flows in the network, for a flow $f \in F$, we will let $s(f)$ denote the source node, $d(f)$ the destination node, and $\lambda(f)$ the arrival rate of packets per second.

A flow may be routed through multiple paths in the network. Let P^f denote the set of available paths from source $s(f)$ to destination $d(f)$ for flow f. For instance, one may choose the K-shortest distance paths from $s(f)$ to $d(f)$ as the set P^f. Let routing variable $r^f(p)$ denote the amount of traffic on path p for flow f. For a path p and nodes h, i, and j we will use $e_{hi} \in p$ to denote that link e_{hi} is on path p, and $e_{hi}e_{ij} \in p$ to denote that path p contains links e_{hi} and e_{ij} in consecutive order.

Let B denote a subset of nodes within the transmission range of node i; then we define a broadcast link $e_{i,B}$ as a set of outgoing links from node i to node set B. A node will have a different broadcast link for each subset of nodes within its transmission range. As will be seen later on, the broadcast links are formed when nodes use NC. It will be assumed that each broadcast link has its own queue, which may be real or virtual. A broadcast link includes as a special case unicast links where a transmission is intended only to a single node. By taking into consideration the broadcast transmissions, we are able to construct a generalized connectivity graph denoted by $G(N, L_b)$, where L_b is the set of all broadcast links in a network. Clearly, L_b includes as a special case unicast links when node set B consists of a single node.

Thus each set of incoming and outgoing links at a node may form a NC opportunity. We use $Z_{i,B}^{k}$ to represent a NC opportunity at node i over link $e_{i,B}$, where k is the NC number. NC number k is used to distinguish among the incoming links of the packets to be coded for transmission over link $e_{i,B}$ as will be explained later on.

3.2. Network Coding-Aware Routing Constraints

In this section, we present routing constraints, which are NC aware. In general, NC is classified into two categories as coding with and without opportunistic listening. In NC without opportunistic listening, a node decodes a coded packet only using previous transmissions, while in NC with opportunistic listening it may also use the overheard packets in decoding. We present the constraints for both with/without opportunistic listening with multi-path routing.

Let variable $a_{i,B}^{k}$ denote the arrival rate of traffic to link $e_{i,B}$ for to the opportunity $Z_{i,B}^{k}$. Then, we have

$$a_{i,B} = \sum_{k \in K_{(i,B)}} a_{i,B}^{k} \qquad \forall e_{i,B} \in L_b, |B| \geq 2 \tag{3.1}$$

where $K_{(i,B)}$ is the number of different NC opportunities which arise over link $e_{i,B}$. In (3.1), $a_{i,B}$ determines the total traffic transmitted over link $e_{i,B}$. We note that $a_{i,B}$ corresponds to the arrival rate of coded packets, when $|B| \geq 2$, where $|B|$ denotes the number of nodes in set B.

3.2.1. Routing Constraints without Opportunistic Listening

When there is no opportunistic listening, a network node is not required to overhear packets transmitted by its neighbors to decode a coded packet. The NC model corresponding to this case is shown in Fig. 3.1a.

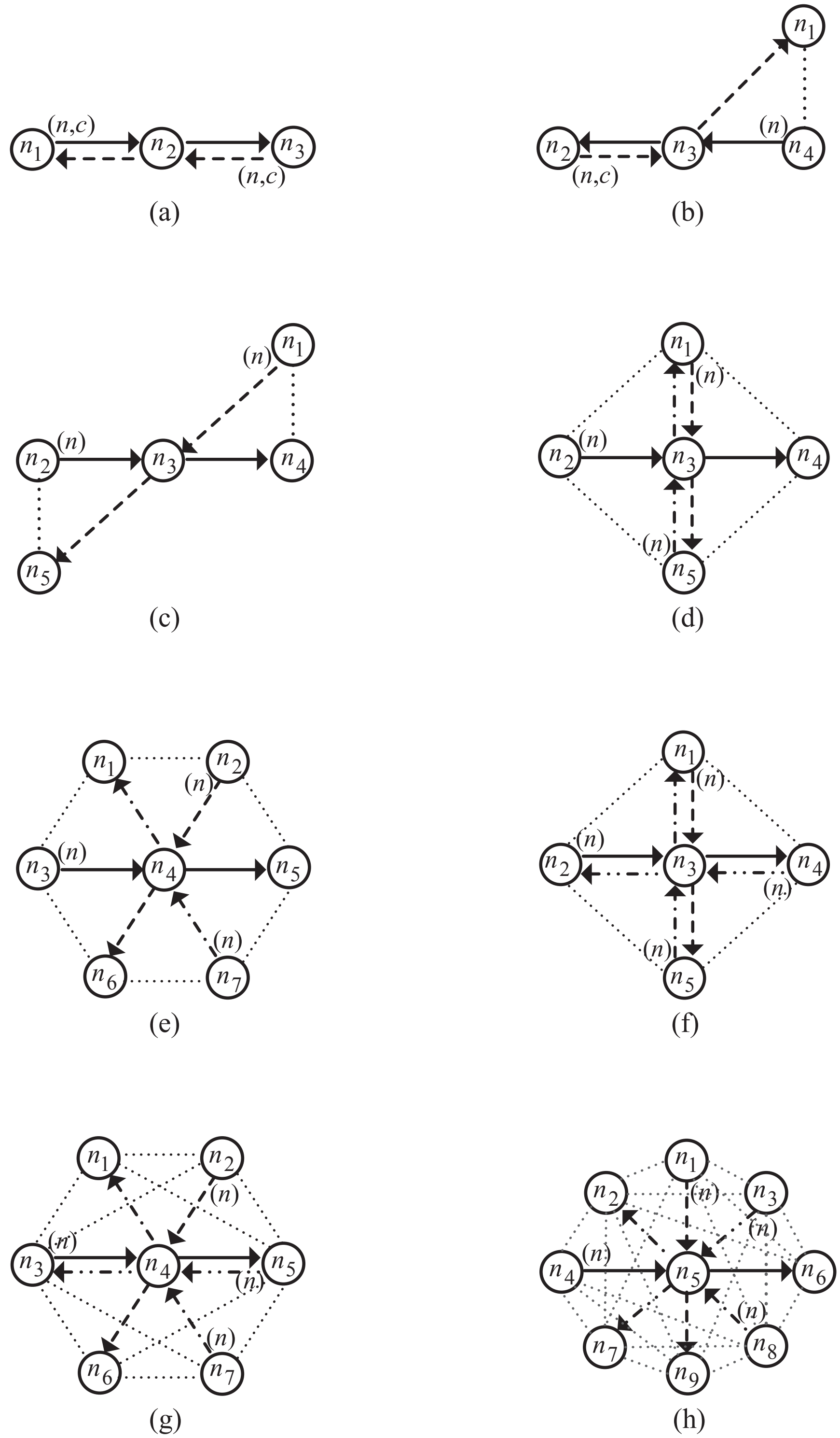


Figure 3.1. Network coding models. (a), (b) and (c) are the models encoding two packets. (d) and (e) are the models encoding three packets. (f), (g) and (h) are the models encoding four packets together.

In this figure, n,c stand for native packet and coded packet respectively, which denote the type of the packet transmitted to the coding node from the previous-hop node. Since opportunistic listening is not used in this NC model, the next-hop nodes need to use packets in their buffer for decoding the coded packet. That is why the packets can be received by the coding node (from the previous-hop nodes) as native packet (n) or coded packet (c). As shown in Fig. 3.1a, the packets enter and leave the coding node using the same links but in opposite directions in this NC model. This means that the coding node XOR exactly two packets from its neighbors and not more. Thus, the number of nodes in node set B is limited to be at most two and for ease of explanation we may use $\{n_1,n_2\}$ instead of B. Note that $|B|=1$ corresponds to the unicast transmissions where a native packet is transmitted to a single node in B.

Let us denote nodes n_1, n_2 and n_3 by h, i and j in Fig. 3.1a respectively. We point out that in the case of without opportunistic listening only one NC opportunity may arise over link $e_{i,\{h,j\}}$ and (3.1) simplifies to $a_{i,\{h,j\}}=a^1_{i,\{h,j\}}$. In other words, $a_{i,\{h,j\}}$ can be obtained by determining the coded traffic of $Z^1_{i,\{h,j\}}$ transmitted over link $e_{i,\{h,j\}}$. Next, we present the routing constraints for NC without opportunistic listening.

$$\sum_{p\in \mathrm{P}^f} r^f(p) = \lambda(f) \qquad \forall f \in F \tag{3.2}$$

$$a_{i,\{h,j\}} \le \sum_{f\in F}\ \sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} r^f(p) \qquad \forall e_{i,\{h,j\}} \in L_b \tag{3.3}$$

$$a_{i,\{h,j\}} \le \sum_{f\in F}\ \sum_{p\in \mathrm{P}^f, e_{ji}e_{ih}\in p} r^f(p) \qquad \forall e_{i,\{h,j\}} \in L_b \tag{3.4}$$

$$a_{ij} = \sum_{f\in F, s(f)=i}\ \sum_{p\in \mathrm{P}^f, e_{ij}\in p} r^f(p) + \sum_{e_{hi}\in L_B} \Big[\sum_{f\in F}\ \sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} r^f(p) - a_{i,\{h,j\}}\Big] \qquad \forall e_{ij} \in L_b \tag{3.5}$$

Constraints (3.2) ensure that sum of the traffic routed on available paths of flow f equals to $\lambda(f)$.

Constraints (3.3) and (3.4) determine the maximum amount of coded traffic that can be broadcast on link $e_{i,\{h,j\}}$. The RHS of (3.3) is the total traffic traversing nodes h, i

and j on path p and the RHS of (3.4) is the total traffic traversing nodes j, i and h in the opposite direction; hence, the traffic that can be encoded by node i is at most the minimum of these opposing traffic.

Constraints (3.5) give the total traffic that is unicast from node i to node j. This traffic is composed of two parts, which appear on the RHS. The first part is the traffic that is generated by node i. The second part is the traffic on the path $e_{hi}e_{ij}$, (where $e_{hi} \in L_b$) which could not be encoded with other traffic.

3.2.2. Routing Constraints with Opportunistic Listening

In the presence of opportunistic listening, each next-hop node may need to overhear packets to decode its packet correctly. Assuming coding nodes can encode at most 4 native packets, all possible NC models are shown in Fig. 3.1. In this figure, NC models (a), (b) and (c) are the models encoding two packets; (d) and (e) are the models encoding three packets; (f), (g) and (h) are the models encoding four packets together.

In Fig. 3.1, we see that the next-hop nodes in the NC models need to overhear native packets transmitted by some of the previous-hop nodes to decode the coded packet, except for model (a). As studied in the previous section, model (a) corresponds to the case of NC without opportunistic listening.

Note that multiple NC opportunities can arise simultaneously at a node. More importantly, from the models that encode more than two packets, other models encoding fewer packets can be derived. For example, from model (h), six models of kind (c) or four models of kind (e) can be derived. In this work, we consider all possible NC opportunities that can arise at a node. Accordingly, each pair of incoming and outgoing links of a coding node may belong to several NC opportunities at a node.

In general, a node can only encode received native packets because these are the packets seen by the neighboring nodes. In all the models, shown in Fig. 3.1, we assume that the center node performs NC. However, in addition to the center node, some previous-hop nodes are also allowed to perform NC in these models. We specify those previous-hop nodes by the type of the packet transmitted to the coding node. As mentioned before n,c stand for native and coded packets, respectively. From the figure, we see that only previous-hop nodes in model (a) and node n_2 in model (b) are allowed to transmit a packet to the coding node as a coded packet. The reason is that in such models the next-hop nodes do not need to overhear packets to decode the coded packet. Thus only in these two models, two subsequent nodes may encode the same packet. Note that although in the mentioned models the packet can be received at the coding node as a coded packet, the coding node first recovers the native packets and then performs the coding.

Clearly, in the other models in Fig. 3.1 all the previous-hop nodes should transmit the packet as a native packet, since the packet should be overheard by the other next-hop nodes; thus in these models coding are not allowed in the previous-hop nodes. As a consequence, along the path of a packet only alternating nodes will be able to encode a packet.

To handle opportunistic listening mechanism, we need a structure which is capable of modeling all the features of a NC opportunity $Z_{i,B}^{k}$. This structure can be comprised of elements that specify the nodes involved in coding, type of the packet received by the coding node, and the NC number; therefore, a NC opportunity can be completely specified by a *coding structure* S which is the combination of *coding elements* of the form $s=(k,hij,v)$, where k is the NC number, h is the previous-hop node, i is the coding node, j is the next-hop node, and $v=c,n$ is the type of the packet received by coding node i. Thus each coding structure S corresponds to a NC opportunity $Z_{i,B}^{k}$. We shall let $b(S)$ denote the set of next-hop nodes of all the elements in a structure S and Γ_i denote the set of all possible coding structures at node i. Let $Z_{i,B}$ denote the set of coding structures at node i with the same set of next-hop nodes, $b(S)$, then $Z_{i,B} = \{Z_{i,B}^{1}, Z_{i,B}^{2}, \dots, Z_{i,B}^{k}\}$.

To clarify our notation, in the sequel we shall give four illustrative examples. In each example, based on the network topology and given routing, we shall specify all possible NC opportunities which might arise at a node. In addition, we show the corresponding coding structure S for each coding opportunity. Note that, for the sake of simplicity of discussion we assume that the demand of all flows in each scenario is exactly one packet per frame which is routed over a single path.

Example 1:

Consider the scenario shown in Fig. 3.2, where there are two concurrent flows f_1: 2→5 and f_2: 5→1 in the network. The coding opportunities are as follows.

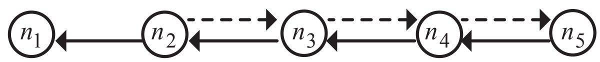


Figure 3.2. First NC example.

$Z^1_{3,\{2,4\}}$: $\{(1,432,n),(1,432,c),(1,234,n)\}$
$Z^1_{4,\{3,5\}}$: $\{(1,543,n),(1,345,n),(1,345,c)\}$

In this scenario, the flow f_2 packet is generated by node 5 to be routed to node 1. Along the path this packet is transmitted by node 4 to node 3. This transmission can be received by node 3 as a coded packet if node 4 performs NC; otherwise, the packet is received as a native packet. We notice that the originating traffic cannot be encoded at the source nodes and cannot be transmitted as a coded packet.

Example 2:

Now, consider the second example shown in Fig. 3.3, where there are three concurrent flows f_1: 1→3, f_2: 3→1 and f_3: 4→1 in the network. Then we have

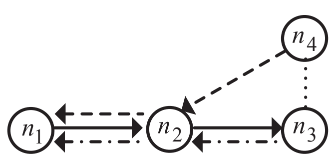


Figure 3.3. Second NC example.

$Z^1_{2,\{1,3\}}$: $(1,321,n),(1,123,n)$
$Z^2_{2,\{1,3\}}$: $(2,421,n),(2,123,n)$

We notice that in this scenario two different NC opportunities arise at node 2 with the same next-hop nodes. These two NC opportunities differ from each other in their previous-hop nodes and are separated by the NC numbers.

Example 3:

Next, consider the third example shown in Fig. 3.4, where there are four concurrent flows f_1: 4→6, f_2: 4→2, f_3: 6→1 and f_4: 2→5 in the network.

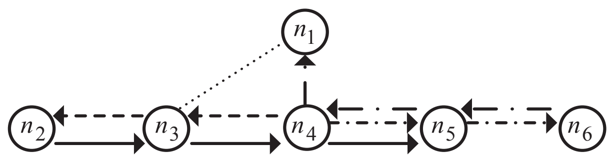


Figure 3.4. Third NC example.

$Z^1_{4,\{1,5\}}$: $(1{,}541, n), (1{,}541, c), (1{,}345, n)$
$Z^1_{3,\{2,4\}}$: $(1{,}432, n), (1{,}234, n)$
$Z^1_{5,\{4,6\}}$: $(1{,}654, n), (1{,}456, n)$

Notice that in this scenario, nodes 4 and 3 are not allowed to perform NC both together, because node 4 is only allowed to perform NC if it receives the packet from node 3 as a native packet; i.e., the packet (of f_4) transmitted by node 3 should be overheard by node 1. However, this is not the case if node 3 performs NC. We note that node 4 may receive the packet from node 5 either as a coded packet or a native packet. The optimal solution can only obtained by solving an optimization problem.

Example 4:

Finally, consider the network topology shown in Fig. 3.5, where there are three concurrent flows f_1: 2→6, f_2: 5→1 and f_3: 6→2 in the network. Then we have

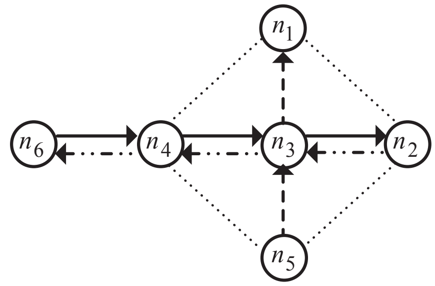


Figure 3.5. Fourth NC example.

$Z^1_{4,\{3,6\}}$: $(1{,}643, n), (1{,}346, n), (1{,}346, c)$
$Z^1_{3,\{1,2,4\}}$: $(1{,}531, n), (1{,}432, n), (1{,}234, n)$
$Z^1_{3,\{2,4\}}$: $(1{,}432, n), (1{,}432, c), (1{,}234, n)$
$Z^1_{3,\{1,4\}}$: $(1{,}531, n), (1{,}234, n)$
$Z^1_{3,\{1,2\}}$: $(1{,}531, n), (1{,}432, n)$

As shown above, in this scenario node 3 can perform NC in four different ways. It might encode three packets together or, alternatively, might encode two packets in three different ways. In the first case, all the packets should be native-received, but in the latter one, this may not be the case depending on which opportunity is chosen. This scenario shows clearly model (d) at node 3 and the three derived models which are one model of kind (a) and two models of kind (c).

Given the set of flows F and the set of available paths for each flow P^f in a network, set of coding structures Γ_i (where $i \in N$) and set of links L_b can be obtained in a straightforward manner. Let ω_i denote the number of neighbors of node i with one hop distance (i.e., the degree of node i); then the number of different coding elements $s=(k,hij,v)$ at node i is at most $2k\omega_i(\omega_i - 1)$. However, in practice the elements that form valid coding structures are much smaller, and coding structures S are generated relatively fast. By having all valid S at node i, Γ_i is obtained. In set Γ_i, the coding structures with the same B forms the broadcast link $e_{i,B}$ at node i.

In addition to the routing variables used earlier, we shall introduce the following variables to express the routing constraints in the presence of opportunistic listening. Let $y_i(s)$ denote the traffic associated with coding element s in S at node i; i.e., this is the traffic amount associated with $e_{hi}e_{ij}$ link-pair which participates in structure *S*. Note that the traffic associated with the same element participating in different coding structures are considered separately. Let $u_i^f(p)$ denote the native traffic of flow *f* which is transmitted by node *i* on path *p*. Then, the routing constraints with opportunistic listening consist of constraint (3.2) and the following additional constraints:

$$\sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} u_h^f(p) \qquad \forall e_{hi}, e_{ij} \in L_b, i \in N \tag{3.6}$$

$$\sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} [r^f(p) - u_h^f(p)] \qquad \forall e_{hi}, e_{ij} \in L_b, i \in N \tag{3.7}$$

$$\sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} r^f(p) = \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} u_i^f(p) + \sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s) + \sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s) \qquad \forall e_{hi}, e_{ij} \in L_b, i \in N \tag{3.8}$$

$$u_{s(f)}^f(p) = r^f(p) \qquad \forall p \in P^f, f \in F \tag{3.9}$$

$$u_i^f(p) \le r^f(p) \qquad \forall i \in p - \{s(f), d(f)\}, p \in P^f, f \in F \tag{3.10}$$

$$a_{ij} = \sum_{f\in F} \sum_{p\in P^f, e_{ij}\in p} u_i^f(p) \qquad \forall e_{ij} \in L_b \tag{3.11}$$

$$a_{i,B} = \sum_{k\in K_{(i,B)}} \sum_{s:(k,hij,v)\in S, S\in\Gamma_i, b(S)=B} y_i(s) \qquad \forall j \in B, e_{i,B} \in L_b \tag{3.12}$$

Constraints (3.6) state that for each coding element *s* at node *i*, the portion of transit traffic that take part in coding as native-received traffic, is at most the amount of traffic which was received as native traffic (by node *i* from *h*). Similarly, (3.7) state that the portion of transit traffic that take part in coding as coded-received traffic is at most the amount of traffic received as coded traffic by node *i* from *h*.

Constraints (3.8) state the traffic conservation at each node for each combination of its incoming and outgoing unicast links. The LHS is the sum of the transit traffic along the path $e_{hi}e_{ij}$. i.e., this traffic enters through node *h* and exits through node *j* at node *i*. The first summation on the RHS is the amount of transit traffic that exits as native traffic so it does not take part in any coding. The second summation is the amount of transit traffic that takes part in coding as native-received traffic, and the third one is the amount of transit traffic that takes part in coding as coded-received traffic.

Constraints (3.9) ensure that for a given path, the source node of each flow transmits the entire traffic on that path as native traffic, since coding opportunities are not available for originating traffic at source nodes. Constraints (3.10) ensure that for a given path, the amount of native traffic at each transit node (except for the source and destination nodes) is at most the total traffic on that path.

Constraint (3.11) determines the total traffic that is transmitted as native traffic over link e_{ij}, and (3.12) determines the traffic that is transmitted as coded traffic over link $e_{i,B}$. In fact, (3.12) is the same as (3.1). The second summation on the RHS gives the coded traffic of $a_{i,B}^{k}$ which is the traffic associated with coding opportunity $Z_{i,B}^{k}$. Note that the $a_{i,B}^{k}$ is the minimum traffic of the coding elements which form $Z_{i,B}^{k}$.

CHAPTER 4
DELAY OPTIMIZATION WITH NETWORK CODING

As explained before, NC has been shown to improve the throughput of MWNs. Prior work on performance modeling of NC mainly addresses the maximization of throughput. However, this work fails to capture the complete picture since there may be paths in the network for which end-to-end packet delay is prohibitively high.

In this chapter, we address the problem of delay minimization in MWNs with NC. The objective has been assignment of wireless node capacities in a way that the average packet delay is minimized for a given network topology and the traffic demand matrix. Indeed, we address the following question: given a specific placement of wireless nodes and traffic demand matrix, what is the optimum capacity allocation that minimizes the end-to-end delay of the network? In a previous work [14], we studied the capacity allocation problem in MWNs with NC that optimizes the average delay in the network. We modelled each wireless node as an M/M/1 queue; the effect of wireless interference on the performance of network has been taken into account by including linear constraints in the optimization framework. That work had the limitation of modeling the wireless nodes within the transmission range of each other as parallel servers working simultaneously with slower rates. However, in a wireless system, nodes within the transmission range of each other receive service with interruptions. Also, the previous work has not dealt with the order of service among the nodes. In this chapter, we extend that work by modelling each wireless node as an M/G/1 queue with service interruptions. It is assumed that at each service epoch the server chooses the next node to serve randomly according to their traffic loads among the nodes within the transmission range of each other. The service time of a packet at a node increases by the amount of the service given to the nodes within its transmission range which provides more accurate modeling of the broadcast nature of wireless channel.

In [14], we had also assumed that routing path of each flow is known in advance. From that information, we formulated the assignment of node capacities in a way that the end-to-end delay in the network with NC is optimized. In this work, the solution is extended to the flows that routing path has not been given [13]. The routing paths of flows are chosen by routing constraints in such a manner to further reduce the mean packet delay in the network. The multi-path routing creates more coding opportunities in the network, which provides a method for computing source-destination routes and utilizing the best coding opportunities from available ones.

In short, we develop a performance analysis of the system, which models network nodes as M/G/1 queues and takes into account wireless interference. The proposed model is valid both with and without opportunistic listening for any wireless network topology. The model also incorporates network coding-aware routing that routes the flows in a manner that increases coding opportunities. We present numerical results, which show that NC reduces the average packet delay in the network and extends the stable operating region of the network. We also perform simulation to determine the accuracy of our analytical model.

4.1. Network Model

A MWN can be represented by a connectivity graph $G(N,L)$ where N is a set of vertices denoting the nodes, and L is a set of directed edges denoting the links. Each node is equipped with a single radio with transmission range r_i and interference range R_i. Let d_{ij} denote the distance between nodes i and j; then based on protocol model there is a directed link e_{ij} from node i to node j if $d_{ij} \leq r_i$. Further, recall that in accordance with protocol model the transmission of node i to node j is successful if any other node $m \in N$, such that $d_{mi} \leq R_m$ is not transmitting. Using this interference model, we find the set of nodes which interfere with each node. We shall use I_i to denote the set of nodes which may have interference with node i.

Let F denote the set of flows in the network. A flow $f \in F$ has source node $s(f)$, destination node $d(f)$, and traffic rate $\lambda(f)$ packet per second. In such a MWN, intermediate nodes forward not only their own traffic but also forward traffic from other nodes. Let P^f denote the set of available paths from source $s(f)$ to destination $d(f)$ for flow f. For instance, one may choose the K-shortest distance paths from $s(f)$ to $d(f)$ as the set P^f. Let routing variable $r^f(p)$ denote the amount of traffic on path p for flow f. For a path p and nodes h, i, and j we will use $e_{hi} \in p$ to denote that link e_{hi} is on path p, and $e_{hi}e_{ij} \in p$ to denote that path p contains links e_{hi} and e_{ij} in consecutive order.

Let the total output flow and the capacity assigned to node i be denoted by A_i packets per second and C_i bps, respectively. We note that A_i is the sum of all the flows routed through node i and the capacity C_i will be managed by node i among the different output links. We will let C denote the wireless channel capacity, and μ denote the mean packet length.

4.2. Network Coding-Aware Routing

The coding aware-routing constraints are similar to the one explained in chapter 3. We repeat the constraints here for the sake of completeness. As before, when NC is used, we need to take into account the broadcast transmissions in the network. We use the notation defined in chapter 3 for broadcast transmissions. Let B denote a subset of nodes within the transmission range of node i; then we define a broadcast link $e_{i,B}$ as a set of outgoing links from node i to node set B. A node will have a different broadcast link for each subset of nodes within transmission range. By taking into consideration the broadcast transmissions, we are able to construct a generalized connectivity graph denoted by $G(N, L_b)$, where L_b is the set of all broadcast links in a network. Clearly, L_b includes as a special case unicast links when node set B consists of a single node. Accordingly, two broadcast links $e_{i,B1}$ and $e_{j,B2}$ interfere with each other if a node $m \in B_1$ is within the interference range of node j, or $m \in B_2$ and it is within the interference range of node i. From now on, unless otherwise stated, link will refer to broadcast links and not to a physical link between two nodes.

We use $Z_{i,B}^k$ to represent a NC opportunity at node i over link $e_{i,B}$, where k is the NC number. Let variable $a_{i,B}^k$ denote the traffic for the opportunity $Z_{i,B}^k$. Then, we have

$$a_{i,B} = \sum_{k \in K_{(i,B)}} a_{i,B}^k \qquad e_{i,B} \in L_b, |B| \geq 2 \tag{4.1}$$

In (4.1), $a_{i,B}$ determines the total traffic of link $e_{i,B}$. We note that $a_{i,B}$ corresponds to the coded traffic, when $|B| \geq 2$. In the sequel, we present routing constraints for both with and without opportunistic listening.

4.2.1. Without Opportunistic Listening

When there is no opportunistic listening, a wireless node is not required to overhear packets transmitted by its neighbors. Thus a coding node encodes exactly two received packets from its neighbors as shown in Fig. 3.1a. In this case, the number of nodes in node set B is limited to be at most two and for ease of explanation we may use $\{h,j\}$ instead of B; we note that $|B|=1$ corresponds to the unicast transmissions where a native packet is transmitted to a single node in B. In the following equations, we present the routing constraints for NC without opportunistic listening.

$$\sum_{p \in \mathrm{P}^f} r^f(p) = \lambda(f) \qquad \forall f \in F \tag{4.2}$$

$$a_{i,\{h,j\}} \leq \sum_{f \in F} \sum_{p \in \mathrm{P}^f, e_{hi}e_{ij} \in p} r^f(p) \qquad \forall e_{i,\{h,j\}} \in L_b \tag{4.3}$$

$$a_{i,\{h,j\}} \leq \sum_{f \in F} \sum_{p \in \mathrm{P}^f, e_{ji}e_{ih} \in p} r^f(p) \qquad \forall e_{i,\{h,j\}} \in L_b \tag{4.4}$$

$$a_{ij} = \sum_{f \in F, s(f)=i} \sum_{p \in \mathrm{P}^f, e_{ij} \in p} r^f(p) + \sum_{e_{hi} \in L_B} \left(\sum_{f \in F} \sum_{p \in \mathrm{P}^f, e_{hi}e_{ij} \in p} r^f(p) - a_{i,\{h,j\}} \right) \qquad \forall e_{ij} \in L_b \tag{4.5}$$

$$A_i = \sum_{e_{i,B} \in L_b} a_{i,B} \qquad \forall i \in N \tag{4.6}$$

$$A_i \mu \le C_i \qquad \forall i \in N \tag{4.7}$$

We note that constraints (4.2)-(4.5) are the same as (3.2)-(3.5). Further, as explained in chapter 3, for the case of without opportunistic listening $a_{i,B}$=$a_{i,B}^1$. Next, we give briefly explanations of the above constraints.

In the above, (4.2) ensures that the total traffic routed on the available paths for flow f must equal the traffic value $\lambda(f)$ of the flow. Note that routing variable $r^f(p)$ is the value of flow f on path p.

Constraints (4.3) and (4.4) determine the maximum amount of coded traffic $a_{i,\{h,j\}}$ that can be broadcast on broadcast link $e_{i,\{h,j\}}$. The RHS of (4.3) is the total traffic traversing nodes h, i, and j on path p and the RHS of (4.4) is the total traffic traversing nodes j, i, and h in the opposite direction; the coded traffic, hence, is at most the minimum of these amounts.

Constraints (4.5) give the total amount of traffic a_{ij} that is unicast to node j from node i. This traffic is composed of two parts which appears on the RHS. The first part is the traffic that is generated at node i. The second part is the amount of traffic on the path $e_{hi}e_{ij}$, (where $e_{hi} \in L_b$), which might not be encoded with other traffic flows.

Constraints (4.6) give the total output flow of node i, which includes both coded and uncoded traffic. Constraints (4.7) ensure that the total output traffic of a node is less than the assigned capacity to the node. As defined before μ is the average packet length.

4.2.2. With Opportunistic Listening

When opportunistic listening is used, each next-hop node needs to overhear packets to decode its packet correctly. As explained in chapter 3, two packets heading to the same next-hop node cannot be encoded. Moreover, a coding node cannot encode its native packets in any coding opportunity. All the models in Fig. 3.1 are opportunistic listening models except for Fig. 3.1a.

To handle opportunistic listening mechanism, we use the coding structure defined in chapter 3. Recall that a NC opportunity, $Z_{i,B}^{k}$, can be completely specified by a *coding structure S* which is the combination of *coding elements* of the form $s=(k,hij,v)$, where k is the NC number, h is the previous-hop node, i is the coding node, j is the next-hop node, and $v=c,n$ is the type of the packet received by coding node i. As before, we shall let $b(S)$ denote the set of next-hop nodes of all coding elements in S and Γ_i denote the set of all possible coding structures at node i.

In addition to the routing variables used earlier, we shall use the following variables to express the routing constraints in the presence of opportunistic listening. Let $y_i(s)$ denote the traffic associated with coding element s in S at node i; i.e., this is the traffic amount associated with $e_{hi}e_{ij}$ link-pair which participates in structure S. Let $u_i^f(p)$ denote the native traffic of flow f which is transmitted by node i on path p. Then, the routing constraints with opportunistic listening consist of the following constraints:

$$\sum_{p\in \mathrm{P}^f} r^f(p) = \lambda(f) \qquad \forall f \in F \tag{4.8}$$

$$\sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F}\sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} u_h^f(p) \qquad \forall e_{hi}, e_{ij} \in L_b, i\in N \tag{4.9}$$

$$\sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F}\sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} [r^f(p) - u_h^f(p)] \qquad \forall e_{hi}, e_{ij} \in L_b, i\in N \tag{4.10}$$

$$\sum_{f\in F}\sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} r^f(p) = \sum_{f\in F}\sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} u_i^f(p) + \sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s) + \sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s) \qquad \forall e_{hi}, e_{ij} \in L_b, i\in N \tag{4.11}$$

$$u_{s(f)}^f(p) = r^f(p) \qquad \forall p \in \mathrm{P}^f, f\in F \tag{4.12}$$

$$u_i^f(p) \le r^f(p) \qquad \forall i \in p - \{s(f), d(f)\}, p\in \mathrm{P}^f, f \in F \tag{4.13}$$

$$a_{ij} = \sum_{f\in F}\sum_{p\in \mathrm{P}^f, e_{ij}\in p} u_i^f(p) \qquad \forall e_{ij}\in L_b \tag{4.14}$$

$$a_{i,B} = \sum_{k\in K_{(i,B)}}\sum_{s:(k,hij,v)\in S, S\in\Gamma_i, b(S)=B} y_i(s) \qquad \forall j\in B, e_{i,B}\in L_b \tag{4.15}$$

$$A_i = \sum_{e_{i,B} \in L_b} a_{i,B} \qquad \forall i \in N \tag{4.16}$$

$$A_i \mu \leq C_i \qquad \forall i \in N \tag{4.17}$$

Constraints (4.8) are the same as (4.2). Constraints (4.9) state that for each combination of incoming traffic from node *h* and outgoing traffic to node *j* at node *i*, the portion of transit traffic that takes part in coding as native-received flows is at most the amount which was received as native traffic by node *i* from node *h*. Similarly, (4.10) states that the portion of transit traffic that takes part in coding as coded-received flows is at most the amount which was received as coded by node *i* from *h*.

Constraints (4.11) state flow conservation at each node for each combination of its incoming and outgoing unicast links. The LHS is the total transit traffic along the path $e_{hi}e_{ij}$. i.e., this traffic enters through node *h* and exits through node *j* at node *i*. The first portion on the RHS is the amount of transit traffic that exits as native so it does not take part in any coding. The second portion is the amount of transit traffic that takes part in coding as native-received flows. The third portion is the amount of transit traffic that takes part in coding as coded-received flows.

Constraints (4.12) ensure that the source node of every path transmits the entire traffic on that path as native traffic, since coding opportunities are not available for originating traffic at the source node. Constraints (4.13) ensure that for a given path, the amount of traffic transmitted as native traffic at each node (except the source and the destination nodes) is at most the total traffic on that path.

Constraints (4.14) determine the total amount of traffic that is transmitted as native traffic from node *i* to node *j*, and (4.15) determines the total amount of traffic that is transmitted as coded traffic from node *i* to node set *B*.

Constraints (4.16) give the total output flow of node *i*, including both unicast and broadcast transmissions, similar to (4.6). Constraints (4.17) are the same as (4.7).

4.3. Capacity Assignment And Scheduling

We assume an N-node MWN where each wireless node can inject traffic flows into the network. Based on a Poisson process each node generates packets, destined for at least one of the network nodes. We assume that the packets are variable length governed by the exponential distribution with average number of bits per packet given by μ. We also assume that at each hop the length of each packet is generated independently. As in the classic study by Kleinrock [22], if the number of traffic flows is high enough there is a smoothing effect that justifies this assumption.

The nodes interfering with each other will share the channel (the server) with each other. From the perspective of a given node, the time that the server takes to transmit the packets of the other nodes will add to the service times of its packets. We will model each node as an M/G/1 queueing system. The service time of a packet will begin with its arrival to the head of its queue and will be completed following its transmission. We note that the service time of a packet arriving at a busy queue will always coincide with the beginning of a new packet transmission. On the other hand, the service time of a packet arriving at an empty queue will probably begin during the transmission of a packet. Thus, in our model the service time of a packet arriving at an empty queue will differ from that of an arriving to a busy queue. Let $\tilde{m}_i(t)$ and $\hat{m}_i(t)$ denote the probability density function of service time of a packet arriving at an empty and non-empty queue at node i, respectively; the corresponding Laplace transforms of message service time densities are denoted by $\tilde{M}_i(s)$ and $\hat{M}_i(s)$. From these transforms, moments can be found in the usual fashion. In [54], the analysis of such an M/G/1 queueing system is presented, from there, the average packet delay at node i is given by,

$$\bar{d}_i = \frac{E(\tilde{m}_i)}{1 - A_i\left[E(\hat{m}_i) - E(\tilde{m}_i)\right]} + \frac{A_i E(\hat{m}_i^2)}{2\left[1 - A_i E(\hat{m}_i)\right]} + \frac{A_i\left[E(\tilde{m}_i^2) - E(\hat{m}_i^2)\right]}{2\left\{1 - A_i\left[E(\hat{m}_i) - E(\tilde{m}_i)\right]\right\}} \qquad \forall i \in N \qquad (4.18)$$

Then using Little's formula [22], the average packet delay in the network is given by,

$$\bar{D} = \frac{1}{\gamma}\sum_{i=1}^{N} A_i \bar{d}_i \tag{4.19}$$

where γ is the total traffic generated by all nodes and is given by

$$\gamma = \sum_{f \in F} \lambda(f) \tag{4.20}$$

Let ρ_i denote the probability that node i is busy; then the mean service time of a packet, $E[m_i]$, at node i is given by

$$E[m_i] = E[\hat{m}_i]\rho_i + E[\tilde{m}_i](1-\rho_i) \qquad \forall i \in N \tag{4.21}$$

From the definition of ρ_i, we also have

$$\rho_i = E[m_i]A_i \qquad \forall i \in N \tag{4.22}$$

From (4.21) and (4.22), ρ_i is given by

$$\rho_i = \frac{E[\tilde{m}_i]A_i}{\left(1 - E[\hat{m}_i]A_i + E[\tilde{m}_i]A_i\right)} \qquad \forall i \in N \tag{4.23}$$

The transmission time of a packet will be exponentially distributed with mean μ/C, with its Laplace transform given by

$$F(s) = \frac{C}{\mu s + C} \tag{4.24}$$

Let θ_i and φ_i denote subsets of busy and idle nodes in I_i, respectively. It will be assumed that following the completion of an interfering transmission, the node i's packet may be selected for transmission with probability π_i. This probability will be a function of the ratio of channel rate of node i to the total channel rate of the nodes within its transmission range, as given below

$$\pi_i = \left(\frac{C_i}{C_i + \sum_{j \in \theta_i} C_j} \right) \left(\prod_{j \in \theta_i} \rho_j \right) \left(\prod_{k \in \varphi_i} (1 - \rho_k) \right) \qquad \forall i \in N \tag{4.25}$$

To explain π_i more clearly, next we give a simple example. Let us consider the 3-node network depicted in Fig. 4.1; then, π_a for node a is given by,

Figure 4.1. 3-node netwok.

$$\pi_a = \frac{C_a}{C_a + C_b + C_c} \rho_b \rho_c + \frac{C_a}{C_a + C_b} \rho_b (1 - \rho_c) + \frac{C_a}{C_a + C_c} \rho_c (1 - \rho_b) + (1 - \rho_b)(1 - \rho_c) \tag{4.26}$$

Expression (4.26) determines the probability that node a's packet is selected on the next service time; the terms on the RHS correspond to both nodes b and c being busy, one of them being busy and none of them being busy. Similarly, π_b and π_c can be found.

Clearly, the nodes that have interference with each other cannot transmit simultaneously but have to be given access individually, so we have

$$\sum_{k \in I_i} C_k \le C \qquad \forall i \in N \tag{4.27}$$

Let $Q_i(k)$ denote the probability that node i's packet will be selected for transmission on the k^{th} transmission, then

$$Q_i(k) = (1 - \pi_i)^{k-1} \pi_i \qquad \forall i \in N, k \ge 1 \tag{4.28}$$

From (4.24) and (4.28), we have Laplace transform of the service time of a message arriving to a busy queue at node i,

$$\hat{M}_i(s) = \frac{F(s)\pi_i}{1 - F(s)(1-\pi_i)} \qquad \forall i \in N \tag{4.29}$$

As mentioned earlier, if the packet has arrived to an empty queue, its service will be different. It may receive immediate service if the system is empty, so the Laplace transform of the service time of a message arriving to an idle queue at node i is given by,

$$\tilde{M}_i(s) = \left[\left(\prod_{\substack{k \in I_i \\ k \neq i}} (1-\rho_k) \right) + \left(1 - \prod_{\substack{k \in I_i \\ k \neq i}} (1-\rho_k) \right) \hat{M}_i(s) \right] F(s) \quad \forall i \in N \tag{4.30}$$

In the above, the first term on RHS corresponds to the case that a packet arriving at an empty queue receives service immediately and the second term otherwise. In the second case, its service will be extended compared to a packet arriving at a busy queue by the residual transmission time of the packet in transmission. This extension will also be given by a regular transmission time due to memoryless property of the exponential distribution. We note that from (4.29) and (4.30) we can determine the moments of service times needed in (4.18).

4.4. Optimization Framework

We are now ready to formulate the optimization problem that will determine the minimum end-to-end delay in MWNs which uses NC. In the absence of opportunistic listening, the problem can be expressed as the following non-linear optimization problem with continuous variables. The solution of the optimization problem results in the minimum average packet delay for the given traffic demand and capacity assigned to each node.

$$\min\ \bar{D} = \frac{1}{\gamma}\sum_{i=1}^{N}\Lambda_i \bar{d}_i$$

subject to

routing constraints: (4.1)-(4.7)

capacity assignment & scheduling constraints: (4.23),(4.25),(4.27),(4.29),(4.30)

In the above optimization problem, some of the constraints are non-linear. We note that in the above constraints, the number of nodes in the node set B (in variable $a_{i,B}$) is limited to be at most two. In the presence of opportunistic listening, the optimization problem remains the same, except that the routing constraints should be replaced by constraints (4.1),(4.8)-(4.17); clearly, in this case the node set B is no longer limited to two nodes (i.e., $|B|)\geq 1$).

The above problems do not have a closed-form solution. We employed a numerical method such as active-set algorithm to solve the problems by using MATLAB [55].

4.5. Performance Evaluation

We now present numerical and simulation results regarding the analytical model proposed in this chapter. For this, as shown in Fig. 4.2, we generated a 9-node network with random topology in a 150x250 unit area. The transmission range and interference range of nodes are set to 100 units. Two nodes which are within the transmission range of each other are considered to have a link between them. We employed the protocol interference model, which had been described in chapter 2.

It is assumed that the mean packet length (i.e., μ) is 125 bytes and the channel capacity (i.e., C) is 54 Mbps. The set P^f of available paths for each flow f was also chosen to consist of the K-shortest distance paths for at most K=3.

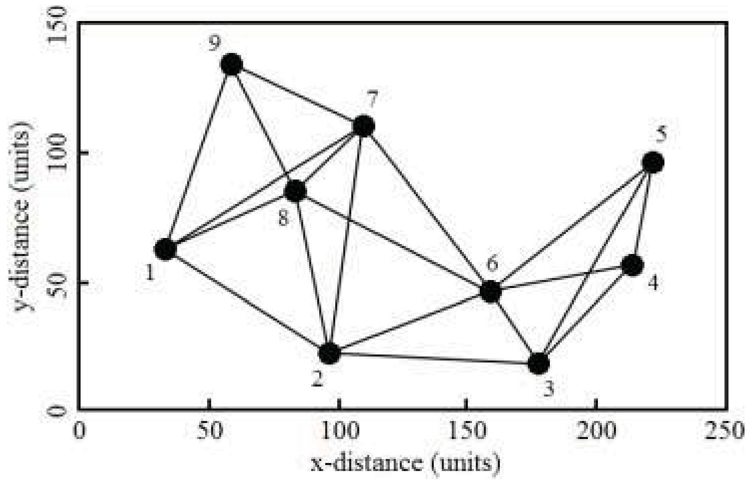


Figure 4.2. Random network topology.

We assumed that each node in the network generates a single traffic flow, and destination node of that flow is chosen randomly. Thus, each node is a source of a single flow but it may be destination of multiple flows as shown in Table 4.1. All flows generate equal amounts of traffic. Without NC the traffic load at each node is determined by summing up all the flows that are routed through that node; with NC some portion of the transit traffic can be encoded, so the traffic load may be reduced at some nodes.

TABLE 4.1 SOURCE-DESTINATION PAIR OF FLOWS.

Source Node *s*(*f*)	1	2	3	4	5	6	7	8	9
Destination Node *d*(*f*)	5	9	8	1	2	9	3	4	3

In the sequel, we study the performance of the generated network under the following schemes: multi-path routing (MPATH), network coding-aware multi-path routing (MPATH-NC) both with/without opportunistic listening.

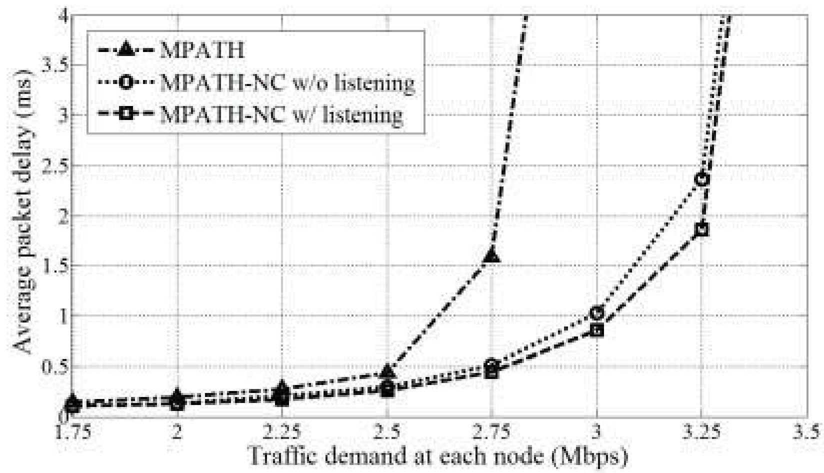


Figure 4.3 Average packet delay versus the per node demand.

Figure 4.3 shows the average packet delay of the network versus traffic demand at each node obtained from the numerical results. It may be seen that utilization of NC reduces the average packet delay in the network. More significantly, NC extends the operating region of the network. This is because as the demands increase, the number of coding opportunities increases which themselves depend on the diversity of the packets in the queue of a node. Without NC congestion is reached at 2.75 Mbps per node traffic demand. However, with NC the network can support traffic demand of around 3.25 Mbps per node before running into congestion where a small increase in traffic gives a very large increase in delay. Thus, using NC results in a throughput improvement of almost 18% for the given flows in the network.

From Fig. 4.3, we also observe that for this network topology allowing opportunistic listening results in a better performance compared to the case of without opportunistic listening, since the number of coding opportunities increases at some nodes. Clearly, the higher the node degree is, the more listening opportunities are at the nodes.

We had also simulated the network under consideration using MATLAB to determine accuracy of our analysis. We used the numerical results to set the parameters of the simulation. In simulation, each node generated packets, according to a Poisson process, destined for one of the network nodes (see Table 4.1). The access probability of each node has been set according to the capacities assigned to the nodes in the numerical results. We note that higher is the capacity allocated to a node, higher will be the probability of access to the channel assigned to that node. We employed static routing based on the routing results of the numerical analysis. We consider spatial reuse in the network, so multiple nodes might transmit at the same time, if their transmissions do not interfere with each other. We consider the principle of NC as described in COPE [2]. Accordingly, when the wireless channel is available, the node takes the packet at the head of its output queue, checks which other packets in the queue may be encoded with this packet, XORs those packets together, and broadcasts the XOR-ed version. If there are no coding opportunities, the node does not wait for the arrival of a matching packet [2].

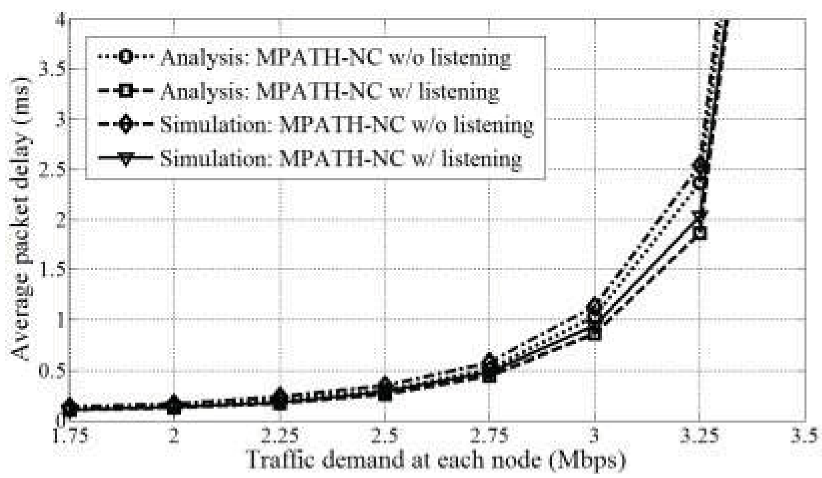


Figure 4.4. Average packet delay versus the per node demand.

In Fig. 4.4, we present both numerical and simulation results for NC scheme with/without opportunistic listening. It may be seen that numerical and simulation results are close to each other with numerical results slightly lower than simulation results.

4.6. Conclusion

The previous work on NC only studies maximization of the throughput without giving consideration to the average packet delay in the network, which is an important performance measure. In this work, we have presented a performance modeling that minimizes the average packet delay of MWNs based on modeling of a node as an M/G/1 queueing system with exponentially distributed packet lengths.

The theoretical framework provides a systematic method to take full advantage of benefits associated with NC, and is applicable to any given network topology with any pattern of concurrent traffic flows. We compared the performance of NC both with and without opportunistic listening, and showed that NC reduces the average packet delay in the network; more importantly, NC extends the stable operating region of the network.

The strength of this model comes from its handling of variable length packets and no need of tight synchronization in the network. On the other hand, it has weakness of giving service to nodes interfering with each other in random order. In the next chapter, we address this weakness.

CHAPTER 5

DELAY OPTIMIZATION WITH NETWORK CODING AND SUCCESSIVE INTERFERENCE CANCELLATION

The continuous rapid growth of the wireless services and introduction of new services are increasing the demand for bandwidth and more efficient wireless communication techniques. NC and SIC are two such techniques that have attracted attention in the recent years. These two techniques have been shown to improve the throughput of MWNs.

Next to NC, SIC has gained much popularity to save bandwidth in wireless networks; this technique has attracted an increasing interest to improve performance of higher layers in MWNs [7], [8]. SIC is a physical-layer technique that improves the performance by exploiting interference in lieu of avoiding it; i.e., SIC allows multi-packet reception by removing interference. SIC enables decoding of multiple signals in a sequential manner to either remove interfering signals or receive multiple packets simultaneously. More specifically, SIC decodes the interfering signals stronger than the intended transmission. The decoded signal is then cancelled to mitigate the interference to the packets which are not yet decoded.

In this chapter, we study the potential benefits of NC and SIC in MWNs that result in significant performance improvement. When inter-session NC is used, the coding is done over packets from different sessions or flows at a node in which the flows cross each other. To fully exploit NC, the routing of the flows should be close to each other. However, this may lead to a high delay in bottleneck nodes due to the increase of interference. In this work, we combine NC with SIC technique, which alleviates the interference. Our goal is to provide a model to fully exploit the benefits of concurrent transmissions and receptions.

In chapter 4, we studied the capacity allocation problem in MWNs with NC that optimizes the average packet delay in the network [13]. That work modeled each node as an M/G/1 queueing system with server interruptions. In this work, we extend that work by considering joint application of NC and SIC under spatial TDMA-based networks. In spatial TDMA, links with sufficient spatial separation may use the same time slot for transmission [16]. Under the assumption of Poisson arrival of packets, the average packet delay of a TDMA queuing system has a closed-form expression [18]. An important feature of the model is that multiple slots can be assigned to a link in the network. We use this model to find the minimum average packet delay in the network.

The objective of this work is minimization of packet delay in a spatial TDMA-based MWN for a given traffic demand matrix with the joint application of NC and SIC techniques. We assume conflict-free scheduling and allow multi-path routing with/without opportunistic listening. We formulate a cross-layer optimization that assigns time slots to different wireless links in a way that the average packet delay is minimized. Our optimization model considers thoroughly all feasible NC and MPR opportunities in the network and allows nodes to encode up to 4 packets together. Further, to increase the MPR opportunities we present a power control extension, which enables nodes to adjust their transmission powers. The problem formulation results in a difficult mixed integer non-linear programming (MINLP). This optimization problem can only be solved for very small-sized networks by the state-of-art software. For large-sized networks we develop a heuristic approach that iteratively determines the optimal solution. We present numerical results to evaluate the performance of the proposed scheme. The results are also compared to that of the previous studies that treat NC and SIC separately. Our findings indicate that significant performance improvement can be achieved by a winning combination of NC and SIC techniques.

In short, the main contributions of this chapter are as follows:

- Formulation of a cross-layer optimization to determine the minimum packet delay in TDMA-based networks. The theoretical framework is presented as a joint multi-path routing and conflict-free scheduling problem. Further, the power control constraints are presented as an extension.
- Two decomposed model, OG and CG problems, are presented for large-sized networks. Further, we compare the performance of these two methods.
- Our optimization model considers thoroughly all possible NC and MPR opportunities in the network. We consider all feasible NC models in which the coding node is allowed to encode up to 4 packets.
- The analysis is applicable to any given MWN topology with any pattern of concurrent traffic flows; further, it is valid both with and without opportunistic listening.

Note that SIC indirectly improves the performance of NC by allowing the routing of the flows to be closer to each other, which results in an increase in number of NC opportunities in the network. This benefit of SIC comes from allowing MPR and mitigating the interference in the network.

We point out that in our mathematical formulation we assume that each node has a single buffer which is connected to a number of virtual queues; i.e., we consider one virtual queue for each active broadcast link. Note that a broadcast link exists only if it belongs to an IS selected by the scheduling process.

There is a wide variation of links for transmission of packets at a node. Indeed, to optimize the network performance our formulation finds the best links from available ones at a node. In other words, we study the queuing delay of those virtual queues in a TDMA-based system. We assign time slots (capacity) to the active links in a frame in such a way that the average packet delay is minimized.

Finally, the main differences between the work of this and previous chapter are that previous model handles variable length packets with exponential distribution, the capacity is assigned to the nodes and nodes within transmission range of each other are served on a random order. The present model handles fixed length packets, capacity is assigned to broadcast links at each node and as a result allows finer control of the network.

5.1. Network Model

We consider a MWN which is represented by a connectivity graph $G(N,L)$ where N is a set of vertices denoting the nodes, and L is a set of directed edges denoting the unicast links. We let n_i denote node i and e_{ij} a unicast link from node i to node j. We will have $e_{ij} \in L$ if node j can successfully receive a packet from node i.

We assume that TDMA system is used for channel access. The time-axis is slotted and slots are organized into frames. The multi-hop nature of the network makes spatial reuse possible in the sharing of time slots, so a schedule can assign time slots to different wireless links in a way that multiple transmissions can occur in the network simultaneously.

The network model is same as before but it will be repeated here for the sake of cohesion. We will assume that the external arrival of packets to each node is according to a Poisson process, which may be destined to different nodes. The packets between each source-destination node pair will form a flow in the network. Packets of a flow may have to travel multiple hops between source and destination. Letting F denote the set of flows in the network, for a flow $f \in F$, we will let $s(f)$ denote the source node, $d(f)$ the destination node, and $\lambda(f)$ the arrival rate of packets per second. We will also let $\Lambda(f)$ to denote the arrival rate of flow f packets per frame. A flow may be routed through multiple paths in the network. Let P^f denote the set of available paths from source $s(f)$ to destination $d(f)$ for flow f. For instance, one may choose the K-shortest distance paths from $s(f)$ to $d(f)$ as the set P^f. Let routing variable $r^f(p)$ denote the amount of traffic on path p for flow f. For a path p and nodes h, i, and j we will use $e_{hi} \in p$ to denote that link e_{hi} is on path p, and $e_{hi}e_{ij} \in p$ to denote that path p contains links e_{hi} and e_{ij} in consecutive order. The total flow on a link will be sum of all the flows routed through that link, which will be approximated as a Poisson process.

Let P_i denote the transmission power of node i, and d_{ij} denote the distance between nodes i and j. Then using free path loss model, the received signal power at node j from the transmission of node i is given by $P_{ij}=P_i d_{ij}^{-\alpha}$, where α is the path loss factor. In an urban environment a typical value of α is three.

5.2. Capacity Assignment

As stated before, the main objective of this work is determining the capacity assignment to the links for a given traffic demand matrix such that the average packet delay in the network is minimized. As said before, TDMA system is used to access the channel, which requires determining assignment of slots to the links. This problem is complicated because the number of assigned slots is an integer variable, flow rates are continuous variables and average packet delay is a nonlinear function. This results in a very difficult mixed integer nonlinear optimization problem that globally optimal solutions are not known.

Let B denote a subset of nodes within the transmission range of node i; then we define a broadcast link $e_{i,B}$ as a set of outgoing links from node i to node set B. A node will have a different broadcast link for each subset of nodes within transmission range. As will be seen later on, the broadcast links are formed when nodes use NC. A broadcast link includes as a special case unicast links where a transmission is intended only to a single node. By considering the broadcast transmissions, we are able to construct a generalized connectivity graph denoted by $G(N, L_b)$, where L_b is the set of all broadcast links in a network.

We assume a TDMA system with frame length of T_s slots. Let $T=\{1, 2,\ldots, |T|\}$ denote the set of indices of $|T|$ slots in the frame, where $|x|$ denotes the size of set x. Note that we will use T_s to represent the frame length as a decision variable in our optimization formulation. The duration of a slot is denoted by τ and it equals to the packet transmission time, which has fixed-lengths. Assuming that C is the channel capacity in packets per second, then $\tau = 1/C$.

A set of broadcast links may transmit in the same time slot either because they do not interfere with each other as a result of physical separation between them or application of SIC technique removes the interference in reception. As before, we refer such a set as independent set (IS). A TDMA schedule provides assignment of slots to ISs in the frame. Let M denote the set of all feasible ISs in the network and c_m denote the number of slots allocated to IS m.

Next, let u_i denote the number of slots allocated to broadcast link i (where $i \in L_b$) in a frame, and b_i denote the average number of packet arrivals to this link during a frame. Then the utilization of link i, denoted by ρ_i, is given by b_i/u_i. Clearly, the maximum link utilization, denoted by $\rho_{\max}$, is 1. Then from [18], under Poisson arrival of packets the average waiting time of a packet at link i is given by the following result for TDMA systems, that allows assignment of multiple contiguous slots to a link. Certainly, in our model the slots assigned to a link may not be contiguous. However, under heavy loading that is of interest to us, this will affect only the frame that the packet is transmitted. In the non-contiguous case, the last frame may be a partial instead of being whole frame. This is not expected to introduce large difference between the results of the two cases.

$$\overline{d_i} = \frac{T_s\tau}{2} + \frac{T_s\tau}{u_i}\left(\frac{b_i}{2} - \frac{u_i - 1}{2} - \frac{u_i(u_i-1) - b_i^2}{2(u_i - b_i)} + \sum_{n=1}^{u_i-1}\frac{1}{1-z_n} + \frac{u_i - b_i}{b_i}\sum_{k=1}^{u_i-1}\frac{w^k}{1-w^k}\prod_{n=1}^{u_i-1}\frac{w^k - z_n}{1-z_n}\right) \quad (5.1)$$

where $T_s = \sum_{m=1}^{M} c_m$, $w = \exp(j2\pi/u_i)$ and z_n, with $n = 1,2,\ldots,u_i - 1$ are the roots of,

$$z^{u_i} - \exp\left(-b_i(1-z)\right) = 0 \quad (5.2)$$

within the unit circle. The roots of (5.2) can be obtained by the Newton-Raphson method; there is also some simplification available in finding the roots, which is given in [54]. We note that since complex roots appear as conjugates, their imaginary parts will cancel each other out, resulting always in a real value for the delay.

We note that the arrival process to each broadcast link consists of external and internal packet arrivals. It will be assumed that both external and internal arrivals will be according to Poisson process. Clearly, this will be an approximation for internal arrivals. This approximation will improve with increasing number of incoming links. Further, we are looking arrival process over a time frame and a broadcast link may receive an arrival in each slot of a frame. Again approximation will improve with the length of the frame. Thus, this approximation will be probably good in large networks.

From (5.1), it may be seen that the mean waiting time in a TDMA system is a function of the frame length. Thus the frame length plays an important role in minimizing the delay. In (5.1), the first term corresponds to the residual waiting time of a packet in the frame that it arrives and the second term is the waiting time due to serving other packets already in the queue. We note that as b_i approaches zero, the second term is negligible and the waiting time becomes $T_S\tau/2$, which is only a function of the frame length. On the other hand, as b_i approaches u_i, the link utilization ρ_i approaches one and second term dominates with the queueing delay increasing without bound.

For the case where each link is assigned exactly 1 slot (i.e. u_i=1), (5.1) reduces to

$$\overline{d_i} = \frac{T_s\tau}{2} + T_s\tau\frac{b_i}{2(1-b_i)} \tag{5.3}$$

In the sequel, we determine the average packet delay in a MWN. Let γ denote the total arrival rate of the packets per frame to the network. Thus, we have

$$\gamma = \sum_{f\in F}\Lambda(f) \tag{5.4}$$

Then from the Little's formula, the average queuing delay in the network can be expressed as,

$$\bar{D} = \frac{1}{\gamma}\sum_{i=1}^{|L_b|} b_i \overline{d_i} \tag{5.5}$$

The main objective of the work in this chapter is to determine the TDMA schedule that minimizes the average packet delay given by (5.5) in a network that jointly applies NC and SIC techniques. The schedule will determine active broadcast links, active set of ISs and their slot assignments in the frame.

5.3. Scheduling Constraints

Scheduling constraints are incorporated in the optimization framework to take care of primary and secondary interference in the network. In this chapter we use physical model to capture the secondary interference. To explain the scheduling of wireless links in TDMA-based MWNs more clearly, we first formulate the scheduling constraints in unicast communications (without NC and SIC) and then generalize it to the case of broadcast communications (with NC) and finally present the case that NC and SIC are jointly used in the network.

5.3.1. Scheduling Constraints without NC and SIC

In the case of without NC and SIC, all packets are transmitted as a native packet and MPR is not allowed at nodes. Let us binary variable $x_{ij}[t]$ denote the state of transmission over link e_{ij} during slot t, where $x_{ij}[t]$=1,0 denote the presence and the absence of a transmission over the link, respectively. In the following we present linear constraints to take care of primary and secondary interference. First we show the formulation of primary interference in unicast communications.

$$\sum_{e_{ij}\in L} x_{ij}[t] + \sum_{e_{hi}\in L} x_{hi}[t] \le 1 \qquad \forall i \in N, t \in T \tag{5.6}$$

The first term on the LHS determines the number of packets node i will be transmitting during slot t, while the second term determines the number of packets that node i will receive during slot t. Since, the RHS of the inequality is one, the inequality ensures that a node can transmit or receive at most a single packet during a slot and it cannot transmit and receive simultaneously during the same slot.

Next, we study the secondary interference based on physical model. Under physical model, the secondary interference is resolved by making sure that the SINR at a receiving node is high enough for correct reception of the intended transmission. Let binary variable $\Delta_i[t]$ denote whether or not node i is transmitting during slot t, where $\Delta_i[t]$=1,0 means node i is transmitting and not transmitting respectively. Accordingly, it is given by

$$\Delta_i[t] = \sum_{e_{ij} \in L} x_{ij}[t] \qquad \forall i \in N, t \in T \tag{5.7}$$

Clearly when (5.6) is satisfied, then, $\Delta_i[t]$ in (5.7) will have either value of zero or one. By rewriting (2.1), the signal transmitted from node h to node i in time slot t could be successfully decoded if

$$SINR_{hi} = \frac{P_{hi}}{N_0 + \sum_{m \in N, m \neq h} P_{mi} \Delta_m[t]} \geq \beta \tag{5.8}$$

The above expression takes care of the secondary interference in unicast communications. However, (5.8) is not still in the format of LP. Next, we reformulate (5.8) as the following linear constraint

$$P_{hi} - \beta \left[N_0 + \sum_{m \in N, m \neq h} P_{mi} \Delta_m[t] \right] \geq (1 - x_{hi}[t]) \delta_{hi} \qquad \forall e_{ij} \in L, t \in T \tag{5.9}$$

where δ_{hi} is a lower bound of the term on the LHS, which can be set as,

$$\delta_{hi} = P_{hi} - \beta(N_0 + \sum_{m \in N, m \neq h} P_{mi}) \tag{5.10}$$

Constraint (5.9) ensures that the $SINR_{hi}$ is above a threshold when link e_{hi} is active during slot t ($x_{hi}[t]$=1). Note that from the definition of δ_{hi}, (5.9) always holds when link e_{hi} is inactive during slot t ($x_{hi}[t]$=0).

5.3.2. Scheduling Constraints with NC

In this section we derive scheduling constraints in broadcast communications with NC. Recall that a broadcast link $e_{i,B}$ is the set of outgoing links from node i to node set B. We use $Z_{i,B}^{k}$ to represent a NC opportunity at node i over link $e_{i,B}$, where k is the NC number. NC number k is used to distinguish among the incoming links of the packets to be coded for transmission over link $e_{i,B}$ as we saw in Fig. 3.3.

We will let the binary variable $x_{i,B}^{k}[t]$ denote the state of a broadcast packet transmission for NC opportunity $Z_{i,B}^{k}$ over link $e_{i,B}$ during slot t. $x_{i,B}^{k}[t]$=0,1 will denote the absence and presence of such a transmission, respectively. Similarly, we will use binary variable $x_{i,B}[t]$ to denote the state of transmission of link $e_{i,B}$ in time slot t. Accordingly, we have

$$x_{i,B}[t] = \sum_{k \in K_{(i,B)}} x_{i,B}^{k}[t] \qquad \forall e_{i,B} \in L_b, t \in T \tag{5.11}$$

where $K_{(i,B)}$ is the number of different NC opportunities which arise over link $e_{i,B}$. We assume that at each node all the coding opportunities with the same B define a single broadcast link.

In the sequel, we will derive the primary and secondary interference constraints in the presence of NC.

5.3.2.1. Primary Interference Constraints with NC

First, we will determine the effect of primary interference on broadcast transmissions. Let us define integer variable $\Lambda_i[t]$ as follows

$$\Lambda_i[t] = \sum_{e_{h,B} \in L_b, i \in B} x_{h,B}[t] \quad \forall i \in N, t \in T \tag{5.12}$$

The above variable denotes the total number of packets intended to node i during slot t, which includes both coded and uncoded packets.

Next, we let integer variable $\Delta_i[t]$ denote the number of coded/uncoded packets that node i will transmit during slot t, then,

$$\Delta_i[t] = \sum_{e_{i,B} \in L_b} x_{i,B}[t] \quad \forall i \in N, t \in T \tag{5.13}$$

We note that as a result of primary interference a node cannot transmit and receive simultaneously during the same slot, and a node may transmit/receive a single coded or uncoded packet during a slot. Hence, we have the following primary interference constraint for node i,

$$\Delta_i[t] + \Lambda_i[t] \leq 1 \quad \forall i \in N, t \in T \tag{5.14}$$

Since the RHS of (5.14) is one, the inequality ensures that a node will either transmit or receive during a slot. Further, if the node is transmitting it can only transmit a single packet, which may be a coded or uncoded packet. Similarly, if the node is receiving it can only receive a single packet, which can be a coded or uncoded packet; as a result, (5.14) constrains $\Delta_i[t]$ and $\Lambda_i[t]$ to be binary variables.

5.3.2.2 Secondary Interference Constraints with NC

Under physical model, when NC is used the SINR for all the next-hop nodes should be high enough to guarantee a collision-free reception at the receivers. By generalizing (5.9) to broadcast transmissions, the signal transmitted from node h to node set B in time slot t could be successfully decoded if

$$P_{hi} - \beta\left[N_0 + \sum_{m \in N, m \neq h} P_{mi}\Delta_m[t]\right] \geq (1 - x_{h,B}[t])\delta_{hi} \qquad \forall e_{h,B} \in L_b, i \in B, t \in T \qquad (5.15)$$

where δ_{hi} is a lower bound of the term on the LHS, which can be set as,

$$\delta_{hi} = P_{hi} - \beta(N_0 + \sum_{m \in N, m \neq h} P_{mi}) \qquad (5.16)$$

We note that in (5.15), $\Delta_m[t]$ will have either value of zero or one when (5.14) is satisfied. Indeed, here $\Delta_m[t]$ represents whether or not the interfering node m is transmitting during slot t. Hence, (5.15) ensures that the $SINR_{hi}$ (where $i \in B$) is above a threshold when link $e_{h,B}$ is active during slot t.

5.3.3. Scheduling Constraints with NC and SIC

In the sequel, we will derive the primary and secondary interference constraints in the presence of NC and SIC. Before presenting the scheduling constraints we present the properties of SIC technique in more details.

5.3.3.1. Successive Interference Cancellation

We consider the analysis of the MPR by using the SIC technique. SIC method enables a node to receive multiple packets from different nodes in the same time slot. In this technique, the receiver tries to decode multiple received signals successively in stages. At each stage, the strongest signal is decoded, by treating all the remaining signals as interference. If the required SINR is satisfied, the strongest signal can be decoded and removed from the aggregate signal. At the subsequent stage, the next strongest signal is decoded, and the process continues till either all the signals are decoded or a point is reached where the SINR condition is not met anymore; at this point the leading signal may not be decoded, nor can any of the signals in the residue be decoded.

Let us assume that node i receives m packets denoted by $p_1, p_2,\ldots, p_k,\ldots, p_m$ during a certain time slot. As shown in Fig. 5.1, the above packets are received from distinct previous-hop nodes $n_1, n_2,\ldots, n_m$ respectively. It is assumed that the received power level of the signals for these packets (from the m previous-hop nodes) at node i, in decreasing order, are $P_{1i} \geq P_{2i} \geq \ldots P_{ki} \ldots \geq P_{mi}$. Let us consider decoding of the k^{th} signal with power level P_{ki} by node i, where $k \leq m$. In this case, node i needs to decode all the signals stronger than k^{th} signal before decoding the k^{th} signal. For decoding of the k^{th} signal the following inequality needs to be satisfied for all values of $h \leq k$,

$$\frac{P_{hi}}{N_0 + \sum_{q=h+1}^{m} P_{qi}} \geq \beta \qquad h = 1, 2, .., k \tag{5.17}$$

By taking advantage of SIC, a wireless receiver will have capability to receive more than one packet from multiple transmitters simultaneously.

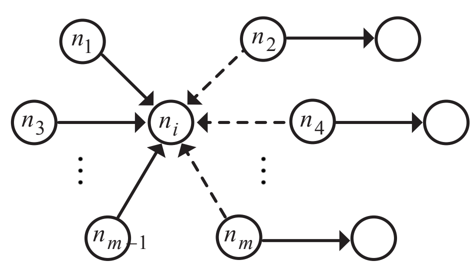


Figure 5.1. The general case of SIC technique at receiving node i.

5.3.3.2. Primary Interference Constraints with NC and SIC

When NC and SIC are jointly used, as a result of primary interference a node cannot transmit and receive simultaneously during the same slot; further, a node may transmit a single coded or uncoded packet during a slot; however, it may receive multiple packets and each packet may be coded or uncoded. Hence, the primary interference constraints for node i consist of constraints (5.12) and (5.13) and the following constraint,

$$\Delta_i[t] + \frac{1}{\chi_i}\Lambda_i[t] \le 1 \qquad \forall i \in N, t \in T \tag{5.18}$$

where χ_i represents the maximum number of packets received by node i in each time slot. To relax the number of receptions at node i we set χ_i to the number of nodes which have node i within their transmission range (i.e. the node degree), clearly $\Lambda_i(t) \le \chi_i$. Since the RHS of (5.18) is one, the inequality ensures that a node will either transmit or receive during a slot. Further, if the node is transmitting it can only transmit a single packet, which constrains $\Delta_i[t]$ to a binary variable. Note that the number of receptions at node i is limited by physical layer constraints as explained in the next section.

5.3.3.3. Secondary Interference Constraints with NC and SIC

We now proceed to deal with the secondary interference. First, as a motivation, let us consider the concept of SIC for unicast transmissions [11]. Accordingly, let us rewrite (5.8) by spliting the summation in the denominator as follows,

$$SINR_{hi} = \frac{P_{hi}}{N_0 + \sum\limits_{\substack{m \in N, m \neq h, \\ P_{mi} \le P_{hi}}} P_{mi}\Delta_m[t] + \sum\limits_{\substack{m \in N, m \neq h, \\ P_{mi} > P_{hi}}} P_{mi}\Delta_m[t]} \ge \beta, \tag{5.19}$$

In (5.19), the summation was split into two parts based on the value of the received power of interfering signals at node i. Clearly, (5.19) in the case of a broadcast transmission, say $e_{h,B}$, should be held for each next-hop node i in B. Derivation of the secondary interference constraint for the combined application of NC and SIC can be carried out in the following two steps.

In the first step, we assume that all the stronger signals (i.e., $P_{mi} > P_{hi}$ where $i \in B$) are already decoded and removed from the aggregate signal successfully. Thus the second summation is zero and the intended transmission is now the strongest signal at node i. Accordingly, by taking into account broadcast transmissions, (5.19) may be reformulated as the following constraint,

$$P_{hi} - \beta(N_0 + \sum_{m \in N, m \neq h, P_{mi} \leq P_{hi}} P_{mi} \Delta_m[t]) \geq (1 - x_{h,B}[t])\delta_{hi} \qquad \forall e_{h,B} \in L_b, i \in B, t \in T \qquad (5.20)$$

where δ_{hi} is a lower bound of the term on the LHS, which can be set as,

$$\delta_{hi} = P_{hi} - \beta(N_0 + \sum_{m \in N, m \neq h, P_{mi} \leq P_{hi}} P_{mi}) \qquad (5.21)$$

Constraints (5.20) ensure that the $SINR_{hi}$ (where $i \in B$) is above a threshold when link $e_{h,B}$ is active during slot t. Note that from the definition of δ_{hi}, (5.20) always holds when link $e_{h,B}$ is inactive during slot t ($x_{h,B}[t]$=0). Further, we point out that (5.20) is valid under the assumption that all stronger signals at node i (where $i \in B$) are already decoded. While it is a necessary condition that all intended signals satisfy (5.20), it is not a sufficient condition if node i receives unintended signals stronger than the intended signals.

We now move on to the second step. In this step, we derive an expression, which ensures that none of the signals, which are stronger than the intended signal at the receiver, are allowed to be active in the same time slot unless they are decodable by the receiver. This ensures that the second summation in the denominator of (5.19) is always zero when using SIC.

Let us consider the case of concurrent receptions, where a node may receive stronger signals which might be unintended in the same time slot. Suppose that a broadcast transmission, say, $x_{h,B}[t]$ where $i \in B$ is active during slot t, and consider another node, say, q which also might be active in the same slot (i.e., $\Delta_q[t]$=1). Then the next-hop node i might receive a stronger signal from node q if P_{qi}>P_{hi}. Note that the stronger signal received from node q may be either an intended transmission or an unintended transmission at some nodes in B. In this situation, node q can be active in the same slot if all nodes in B are able to decode the stronger signal which might be received from node q; i.e.,

$$P_{qi} - \beta(N_0 + \sum_{m\in N, m\neq q, P_{mi}\leq P_{qi}} P_{mi}\Delta_m[t]) \geq (1-\Upsilon_{h,B,(q)}[t])\delta_{qi}$$

$$\forall e_{h,B} \in L_b, i \in B, q \in N, P_{qi} > P_{hi}, t \in T \qquad (5.22)$$

where δ_{qi} is a lower bound of the term on the LHS, which can be set similar to (5.21), and $\Upsilon_{h,B,(q)}[t]$ is a binary variable defined as $\Upsilon_{h,B,(q)}[t]=x_{h,B}[t]\Delta_q[t]$; $\Upsilon_{h,B,(q)}[t]=1$ if and only if $x_{h,B}[t]=1$ and $\Delta_q[t]=1$; otherwise $\Upsilon_{h,B,(q)}[t]=0$. This definition of $\Upsilon_{h,B,(q)}[t]$ may be expressed through the inequality of $\Upsilon_{h,B,(q)}[t]\geq x_{h,B}[t]\Delta_q[t]$. Indeed, $\Upsilon_{h,B,(q)}[t]$ indicates the event of concurrent receptions at node i when $P_{qi}>P_{hi}$, which can be formulated as the following constraints,

$$\{2\Upsilon_{h,B,(q)}[t] \leq x_{h,B}[t] + \Delta_q[t]; \qquad (5.23)$$

$$x_{h,B}[t] + \Delta_q[t] \leq \Upsilon_{h,B,(q)}[t] + 1\} \qquad \forall e_{h,B} \in L_b, i \in B, q \in N, P_{qi} > P_{hi}, t \in T \qquad (5.24)$$

Constraints (5.23) and (5.24) are the linear form of $\Upsilon_{h,B,(q)}[t]\geq x_{h,B}[t]\Delta_q[t]$, and detect the existence of a stronger signal (from node q) for any nodes in B when link $e_{h,B}$ is active (by setting $\Upsilon_{h,B,(q)}[t]$ to 1). As a consequence, the number of concurrent receptions at node i is limited by the above physical layer constraints due to the SIC limitations. Clearly, the maximum number of receptions might be χ_i, as we set in (5.18).

5.4. Routing Constraints

In this section, we present routing constraints, which are NC aware. We use the notation explained in chapter 3. As mentioned before, NC is classified into two categories as coding with and without opportunistic listening. In NC without opportunistic listening, a node decodes a coded packet only using previous transmissions, while in NC with opportunistic listening it may also use the overheard packets in decoding. We present the constraints for both with/without opportunistic listening with multi-path routing. As explained in section 5.1, TDMA schedule assigns slots to the ISs which consist of broadcast links that may transmit in the same time slot.

5.4.1. Without Opportunistic Listening

When there is no opportunistic listening, a wireless node is not required to overhear packets transmitted by its neighbors. A coding node encodes exactly two packets from its neighbors as shown in Fig. 3.1a. In this case, the number of nodes in node set *B* is limited to be at most two and for ease of explanation we may use {*h*,*j*} instead of *B*; we note that |*B*|=1 corresponds to the unicast transmissions where a native packet is transmitted to a single node in *B*. As before, we let binary variable *v*[*t*] to denote the status of the channel during slot *t*. Thus *v*[*t*]=1 if time slot *t* is used by an IS; otherwise 0. Next, we present the routing constraints for NC without opportunistic listening.

$$x_{i,B}[t] \le v[t] \qquad \forall e_{i,B} \in L_b, t \in T \tag{5.25}$$

$$T_s = \sum_{t\in T} v[t] \tag{5.26}$$

$$\{\Lambda(f) = T_s \tau \lambda(f); \tag{5.27}$$

$$\sum_{p\in \mathrm{P}^f} r^f(p) = \Lambda(f)\} \qquad \forall f \in F \tag{5.28}$$

$$\{a_{i,\{h,j\}} \le \sum_{f\in F} \sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} r^f(p); \tag{5.29}$$

$$a_{i,\{h,j\}} \le \sum_{f\in F} \sum_{p\in \mathrm{P}^f, e_{ji}e_{ih}\in p} r^f(p)\} \qquad \forall e_{i,\{h,j\}} \in L_b \tag{5.30}$$

$$a_{ij} = \sum_{f\in F, s(f)=i} \sum_{p\in \mathrm{P}^f, e_{ij}\in p} r^f(p) + \sum_{e_{hi}\in L_B} \Big(\sum_{f\in F} \sum_{p\in \mathrm{P}^f, e_{hi}e_{ij}\in p} r^f(p) - a_{i,\{h,j\}}\Big) \qquad \forall e_{ij} \in L_b \tag{5.31}$$

$$a_{i,B} \le \rho_{\max} \sum_{t\in T} x_{i,B}[t] \qquad \forall e_{i,B} \in L_b, |B| \le 2 \tag{5.32}$$

$$u_l = \sum_{t\in T} x_{i,B}[t] \qquad \forall l : e_{i,B} \in L_b \tag{5.33}$$

$$b_l = a_{i,B} \qquad \forall l:\ e_{i,B} \in L_b \tag{5.34}$$

Constraints (5.25) specify the assigned time slots. Clearly, each time slot can be assigned to a group of links called IS. Constraint (5.26) determines the frame length T_s by summing up all the slots assigned to ISs. Note that the frame length T_s is defined as an integer variable larger than one.

Constraints (5.27) give the traffic of $\Lambda(f)$, which is the average number of packets generated by flow f during a frame, and (5.28) ensures that sum of the traffic routed on available paths of flow f equals to $\Lambda(f)$.

Constraints (5.29) and (5.30) determine the maximum amount of coded traffic that can be broadcast on link $e_{i,\{h,j\}}$. The RHS of (5.29) is the total traffic traversing nodes h, i and j on path p and the RHS of (5.30) is the total traffic traversing nodes j, i and h in the opposite direction; hence, the traffic that can be encoded by node i is at most the minimum of these opposing traffic.

Constraints (5.31) give the total traffic that is unicast from node i to node j. These packets are composed of two parts, which appear on the RHS. The first part is the traffic that is generated by node i. The second part is the traffic on the path $e_{hi}e_{ij}$, (where $e_{hi} \in L_b$) which could not be encoded with other traffic.

Constraints (5.32) state that the total traffic transmitted by each link cannot exceed the number of time slots (i.e. capacity) assigned to that link. These constraints also enable bounding of link utilization in the network.

In constraint (5.33), u_l determines total number of slots allocated to link $e_{i,B}$ (denoted as link l) during a frame, and (5.34) determines packet arrival rate to the same link. We note that b_l corresponds to the arrival rate of coded packets, when $|B|\geq 2$.

5.4.2. With Opportunistic Listening

In the presence of opportunistic listening, each next-hop node may need to overhear packets to decode its packet correctly. Assuming coding nodes encode at most 4 native packets, all possible NC models can be specified as shown in Fig. 3.1.

As explained in chapter 3, to handle opportunistic listening mechanism, we need a structure which is capable of modeling all the features of a NC opportunity $Z_{i,B}^{k}$. A NC opportunity can be completely specified by a *coding structure S* which is the combination of *coding elements* of the form $s=(k,hij,v)$, where k is the NC number, h is the previous-hop node, i is the coding node, j is the next-hop node, and $v=c,n$ is the type of the packet received by coding node i. Thus each coding structure S corresponds to a NC opportunity $Z_{i,B}^{k}$. As before, we let $b(S)$ denote the set of next-hop nodes of each element in S and Γ_i denote the set of all possible coding structures at node i.

In addition to the routing variables used earlier, we shall introduce the following variables to express the routing constraints in the presence of opportunistic listening. Let $y_i(s)$ denote the traffic associated with coding element s in S at node i; i.e., this is the traffic amount associated with $e_{hi}e_{ij}$ link-pair which participates in structure S. Note that the traffic associated with the same element participating in different coding structures are considered separately. Let $u_i^f(p)$ denote the native traffic of flow f which is transmitted by node i on path p. Then, the routing constraints with opportunistic listening consist of constraints (5.25)-(5.28) and the following constraints:

$$\{ \sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} u_h^f(p); \tag{5.35}$$

$$\sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s) \le \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} [r^f(p) - u_h^f(p)]; \tag{5.36}$$

$$\sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} r^f(p) = \sum_{f\in F} \sum_{p\in P^f, e_{hi}e_{ij}\in p} u_i^f(p) + \sum_{s:(k,hij,n)\in S, S\in\Gamma_i} y_i(s)$$

$$+ \sum_{s:(k,hij,c)\in S, S\in\Gamma_i} y_i(s)\} \qquad \forall e_{hi}, e_{ij} \in L_b, i\in N \tag{5.37}$$

$$u^f_{s(f)}(p) = r^f(p) \qquad \forall p \in \mathrm{P}^f, f \in F \qquad (5.38)$$

$$u^f_i(p) \le r^f(p) \qquad \forall i \in p - \{s(f), d(f)\}, p \in \mathrm{P}^f, f \in F \qquad (5.39)$$

$$a_{ij} = \sum_{f \in F} \sum_{p \in \mathrm{P}^f, e_{ij} \in p} u^f_i(p) \qquad \forall e_{ij} \in L_b \qquad (5.40)$$

$$a^k_{i,B} = \sum_{s:(k,hij,v) \in S, S \in \Gamma_i, b(S)=B} y_i(s) \qquad \forall j \in B, e_{i,B} \in L_b \qquad (5.41)$$

$$a_{i,B} = \sum_{k \in K_{(i,B)}} a^k_{i,B} \qquad \forall e_{i,B} \in L_b \qquad (5.42)$$

$$a_{i,B} \le \rho_{\max} \sum_{t \in T} x_{i,B}[t] \qquad \forall e_{i,B} \in L_b \qquad (5.43)$$

$$u_l = \sum_{t \in T} x_{i,B}[t] \qquad \forall l : e_{i,B} \in L_b \qquad (5.44)$$

$$b_l = a_{i,B} \qquad \forall l : e_{i,B} \in L_b \qquad (5.45)$$

Constraints (5.35) state that for each coding element s at node i, the portion of transit traffic that take part in coding as native-received traffic, is at most the amount of traffic which was received as native traffic (by node i from h). Similarly, (5.36) state that the portion of transit traffic that take part in coding as coded-received traffic is at most the amount of traffic received as coded traffic by node i from h.

Constraints (5.37) state the traffic conservation at each node for each combination of its incoming and outgoing unicast links. The LHS is the sum of the transit traffic along the path $e_{hi}e_{ij}$. i.e., this traffic enters through node h and exits through node j at node i. The first summation on the RHS is the amount of transit traffic that exits as native traffic so it does not take part in any coding. The second summation is the amount of transit traffic that takes part in coding as native-received traffic, and the third one is the amount of transit traffic that takes part in coding as coded-received traffic.

Constraints (5.38) ensure that for a given path, the source node of each flow transmits the entire traffic on that path as native traffic, since coding opportunities are not available for originating traffic at source nodes. Constraints (5.39) ensure that for a given path, the amount of native traffic at each transit node (except for the source and destination nodes) is at most the total traffic on that path.

Constraint (5.40) determines the total traffic that is transmitted as native traffic over link e_{ij}, and (5.41) determines the traffic corresponding to coding opportunity $Z_{i,B}^{k}$ that is transmitted over link $e_{i,B}$; Note that $a_{i,B}^{k}$ is the minimum traffic of the coding elements which form $Z_{i,B}^{k}$. Constraint (5.42) determines the total traffic that is transmitted as coded traffic over link $e_{i,B}$.

Constraints (5.43) are similar to (5.32). The only difference is that the next-hop nodes of a broadcast transmission (i.e. B) in (5.43) are not limited to 2 nodes anymore. Constraints (5.44) and (5.45) are the same as constraints (5.33) and (5.34).

5.5. Cross-layer Optimization Framework

We are now ready to formulate the cross-layer optimization problem that will determine the optimal delay in TDMA-based MWNs, which employ both NC and SIC techniques. This optimization will determine the TDMA schedule that minimizes the average packet delay in the network that jointly applies NC and SIC techniques where schedule will determine active set of ISs and their slot assignment in the frame. We note that the solution does not include all the ISs and therefore not all the broadcast links. If a broadcast link is not included in any IS then related coding structures are not being utilized. The coding elements at a node may be encoded into different coding structures. Thus from the available choices optimal solution will choose to activate those broadcast links and consequently those coding structures that result in minimum packet delay.

We note that in practice each broadcast link may have either a virtual or real queue. In the virtual queue approach, all the incoming packets may be stored in a global queue. If during a slot a broadcast link will be served according to the schedule, then if possible a coded packet of that link is encoded from the packets in the global queue and then transmitted. Thus the packets in the global queue to be served by the same broadcast link may be considered to form a virtual queue. In the real queue approach, a node maintains physically separate queue for each broadcast link. As explained in the previous section, a packet may participate in different coding structures each being served by its own broadcast link. In this case, the incoming packets will be directed to the appropriate broadcast queue according to the ratio that a packet participates in coding structures. As packets become available they will be encoded in the broadcast queues.

In the presence of opportunistic listening, the problem can be expressed as program P1 below,

$$[\text{P1}]: \quad \min \ \overline{D} = \frac{1}{\gamma}\sum_{i=1}^{|L_b|} b_i \overline{d_i}$$

subject to

(5.11)-(5.13), (5.18), (5.20)-(5.28), (5.35)-(5.45).

We note that in the case of without opportunistic listening constraints (5.35)-(5.45) should be replaced by (5.29)-(5.34). In this study, nodes with NC capability (with opportunistic listening) are allowed to encode up to 4 packets; accordingly, $1 \leq |B| \leq 4$.

P1 is a mixed integer non-linear program (MINLP) with linear constraints, and does not have a closed-form solution. From (5.1), in the objective function, the number of slots assigned to a link (i.e., u_i) is an integer variable and each of these variables correspond to the number of roots of the equation (5.2); as a result, these decision variables cannot be relaxed to continuous variables. Accordingly, derivative-free algorithms can be used to solve this problem. However, due to the complexity of the problem, we could solve only for networks with very small sizes by using commercial solvers like Midaco [56], and Matlab/GA [57], where the first solver uses an extended ant colony algorithm, and the latter one uses the genetic algorithm.

In the sequel, we propose a heuristic approach to find the minimum delay. Recall that the delay in TDMA-based networks is a function of the length of scheduling time. As discussed in section 5.2 under light traffic loading, frame duration is the dominant factor in the mean packet delay, while under moderate and heavy traffic queueing delay becomes determining factor of the delay. As a result, minimizing the scheduling time results in optimal packet delay only under light loading. However, in this work we are interested in the performance of the network under high loads as TDMA-based systems are more efficient under that regime. We will first derive an optimization framework that finds the minimum scheduling time; then we will propose a heuristic based on minimum scheduling time and maximum link utilization that determines minimum packet delay.

5.5.1. Minimization of the Frame Length

The problem of minimizing the frame length can be easily formulated by changing the objective function in P1 as in P2 below and removing constraints (5.44) and (5.45) from the optimization framework. Hereafter we only present the routing constraints which correspond to the case of NC with opportunistic listening in the optimization framework. For the case of without listening, the corresponding routing should be replaced as in P1.

$$\text{[P2]:} \qquad \min \; T_s = \sum_{t \in T} v[t]$$

subject to

(5.11)-(5.13), (5.18), (5.20)-(5.28), (5.35)-(5.43).

P2 is a MILP problem, since the objective function is also a linear function. This optimization formulation minimizes the number of allocated slots in the frame. Note that each time slot can be used by a set of links, which do not have (the primary and secondary) interference with each other. As stated before, we refer such a set as IS. Indeed, the links involved in an IS can transmit simultaneously free of interference because of two reasons. The first reason is that some links might have a sufficient spatial separation with each other, so there is no interference among their transmissions. The second one is that the transmission of some links might interfere with each other, but the interference will be cancelled by SIC technique. Note that we assume that each IS is a maximal set. A maximal IS is the one which cannot be grown further [42]. The solution of P2 results in a schedule with minimum frame length. The schedule gives the subset of ISs that will be active and their slot assignments in the frame. There may be many minimum length schedules with different IS subsets and slot assignments.

Though P2 is a much simpler problem compared to P1, it still has high computational complexity. The complexity lies in generating the feasible ISs, which grow exponentially with the number of network links. However, very few of these ISs appear in the optimal solution. In the following, we present two methods that handle complexity in the generation of ISs.

The first method removes the constraints that pertain to generation of possible ISs (i.e., the primary and secondary constraints) from the optimization problem P2 and instead gives all possible ISs to the problem as an input. Thus, all possible ISs are generated offline and added to the optimization problem. We refer this problem as offline generation (OG) method.

The second one decomposes the problem into two sub-problems by using column generation (CG) approach [58]. In this method those constraints that generate ISs lie in one sub-problem in such a way that ISs are generated one at a time and added to the other sub-problem. This method adds the most suitable IS at any particular time; hence, the optimal ISs might be obtained without having to enumerate all feasible ISs. We refer this problem as CG method. Next, we describe both of these methods.

5.5.1.1. Offline Generation (OG) Method

In this method, all feasible ISs are generated offline. In the following, as defined before, M denotes the set of all feasible ISs in the network and c_m denotes the number of slots allocated to IS m, represented by Y_m. Then P2 reduces to the following optimization problem,

$$\text{[OG]:} \quad \min \quad T_s = \sum_{m \in M} c_m$$

subject to

(5.27),(5.28), (5.35)-(5.42) and

$$a_{i,B} \le \rho_{\max} \sum_{m \in M, e_{i,B} \in Y_m} c_m \qquad \forall e_{i,B} \in L_b \tag{5.46}$$

Constraint (5.46) is similar to (5.43) and states that the total number of packets transmitted by each link cannot exceed the total number of slots of those ISs to which the link belongs. We note that, each link can be active in multiple ISs. As before, we assume that $\rho_{\max}$ is a constant variable and equals to one.

In the sequel, we give a simple procedure to find all possible ISs in the network. Let G_i denote the set of links in IS i. We define variable E_i to represent the number of times that the IS i has been found during the search, and let $E_{\max}$ denote the maximum limit of an E_i.

Procedure 5.1. Determining ISs through OG

Step 1: Start with an empty set G_i.

Step 2: Choose randomly the first link from L_b, and add it to G_i.

Step 3: Choose randomly a new link if and only if it does not have any interference with any of the links already added to G_i (by taking into account the SIC technique). When all the links have been considered, G_i will contain at least one link.

Step 4: Verify whether this IS is already found or not. If it is new, add it to M; otherwise, discard G_i and E_i=E_i+1.

Step 5: Repeat steps 1-4 while $E_i \leq E_{\max}$.

Note that in step 3, the interference of a new link with the links already added to G_i may be removed by means of SIC technique. Clearly, the procedure terminates when an IS is found $E_{\max}$ times. We shall determine the appropriate value of $E_{\max}$ in the numerical results section.

In this method, different minimum length schedules may be obtained by changing the order of ISs in M.

5.5.1.2. Column Generation (CG) Method

When CG method is used P2 is decomposed into two smaller sub-problems called restricted master problem (RMP) and pricing problem (PP). In this approach, each column represents an IS in the network, and the PP, indeed, generates columns or ISs one at a time in each iteration. The RMP is formulated as follows,

---

$$\text{[CG]: RMP} \quad \min\ T_s = \sum_{m \in M'} c_m$$

subject to

(5.27),(5.28), (5.35)-(5.42) and

$$a_{i,B} \le \rho_{\max} \sum_{m \in M', e_{i,B} \in Y_m} c_m \qquad \forall e_{i,B} \in L_b \tag{5.47}$$

As may be seen the objective function of the RMP is initialized to M' which is a subset of M that covers all links in L_b. In fact, this is the only difference between OG problem and RMP.

The PP can be formulated as follows,

$$\text{[CG]: PP} \quad \max \sum_{e_{i,B} \in L_B} \sigma_{i,B} x_{i,B}$$

subject to

$$\{\Delta_i = \sum_{e_{i,B} \in L_b} x_{i,B}; \tag{5.48}$$

$$\Lambda_i = \sum_{e_{h,B} \in L_b, i \in B} x_{h,B}; \tag{5.49}$$

$$\Delta_i + \frac{1}{\varphi_i} \Lambda_i \le 1\} \qquad \forall i \in N \tag{5.50}$$

$$P_{hi} - \beta(N_0 + \sum_{P_{mi} \le P_{hi}, m \ne h} P_{mi}\Delta_m) \ge (1 - x_{h,B})\delta_{hi} \qquad \forall e_{h,B} \in L_b, i \in B \tag{5.51}$$

$$\{P_{qi} - \beta(N_0 + \sum_{P_{mi} \le P_{qi}, m \ne q} P_{mi}\Delta_m) \ge (1 - \Upsilon_{h,B,(q)})\delta_{qi}; \tag{5.52}$$

$$2\Upsilon_{h,B,(q)} \le x_{h,B} + \Delta_q; \tag{5.53}$$

$$x_{h,B} + \Delta_q \le \Upsilon_{h,B,(q)} + 1\} \qquad \forall e_{h,B} \in L_b, i \in B, q \in N, P_{qi} > P_{hi} \tag{5.54}$$

Note that the constraints used in the pricing problem are similar to (5.12), (5.13), (5,18), (5.20)-(5.24) except for that decision variables are not defined as arrays with the time dimension. Thus, they are not defined over the set of *T* anymore, since only one IS (or column) is generated by the PP at any particular time.

Let $\sigma_{i,B}$ be the dual variable corresponding to capacity constraints (5.47) in the RMP. In order to add a new IS to the RMP we need to verify whether the reduced cost associated with the generated IS is negative or not. Note that, only those ISs which have negative reduced cost can improve the objective function of RMP and the best choice is the one which has the smallest reduced cost. Accordingly, at each iteration PP generates an IS that has the minimum reduced cost. Thus, we have

$$\min\ 1 - \sum_{e_{i,B} \in L_b} \sigma_{i,B} x_{i,B} \tag{5.55}$$

which is equivalent to the objective function of the PP. If the optimal solution of the PP leads to a reduced cost which is non-negative, no improvement to the RMP objective function is possible; as a result, the optimality is reached. However, as long as at each iteration (5.55) is negative, the RMP is re-optimized with a new column added to M'; accordingly the PP is solved to check whether M' should be enlarged with a new IS or not.

We note that when CG is used the RMP should be solved with the LP-relaxation of integer variables, so T_s and c_m where $m \in M'$ become continuous variables; i.e, $c_m \ge 0$

and $T_s \geq 0$. Accordingly, we are not able to use (5.27) in the RMP, (recall that we had previously $T_s \geq 1$); Instead, we use the following constraints

$$\Lambda(f) = T_s' \tau \lambda(f) \qquad \forall f \in F \tag{5.56}$$

where T_s' is a constant that should be equal to T_s, which can be found by the trial and error method.

We point out that at termination the CG method finds the optimal solution of the LP-Relaxation of RMP by finding the optimal ISs. Now that the required ISs are found, we can solve the RMP independent of PP as an integer LP (ILP); note that this solution gives an upper bound of optimal ILP (i.e., the solution of P2) and the solution of the LP-relaxation of the RMP is a lower bound of the optimal ILP. However, in our study most of the time the computed upper bound was equal or very close to the lower bound; as a result, the solution is either optimal or near optimal.

Finally, to find the solution within a reasonable time, one may halt the PP by defining a different condition on the reduced cost [20]. Because when the reduced cost is very close to zero, the improvement in the RM objective function is negligible. Since the pricing model finds the ISs by minimizing the reduced cost at each trial, the primary ISs found by the pricing model has the most improvement on the RMP objective function, and this improvement gradually diminishes at the subsequent iterations. Thus, one may halt the PP when the reduced cost becomes close to zero.

In this method, defining different conditions of termination on the objective function of PP may generate different minimum length schedules.

5.5.2. Minimum Delay

In this section, we will consider determining the minimum average packet delay in the network. In the previous section, we formulated an optimization problem the solution of which determines a subset of ISs that minimizes the frame length. However, minimum packet delay does not occur at minimum frame length except under light traffic loading.

For a moment, let us consider the relationship of frame length to the packet delay. Let us assume that the total traffic demand in the network is fixed, (i.e., certain $\lambda(f)$, $f \in F$) and determine the min frame length for this system under the constraint $\rho_{max} \leq 1$ for all links.

Next, let us consider solution of this problem under a reduced maximum link utilization by the amount of $\Delta\rho$ thus $\rho_{max} \leq 1-\Delta\rho$. Let us consider a link i which had the utilization $1-\Delta\rho \leq \rho_i \leq 1$ for $\rho_{max} \leq 1$. Under $\rho_{max} \leq 1-\Delta\rho$, utilization of this link will be reduced either by offloading some of its traffic load by route changing or increasing the frame length. In the first case utilization and packet delay of link i will be reduced on the other hand, links with offloaded traffic will experience higher utilization and packet delay. In the case of increasing frame length, more slots will be added to the ISs that this link belongs to. From (5.27), when the frame length T_s increases, more traffic is injected to the network by each of the flows during the frame. As a result utilizations and packet delays of all links, except for the links that have received additional slots will increase, while of the latter group will decrease. As a result, overall packet delay in the network will decrease if the reduction in packet delay of link i more than offsets increments of packet delays of other links.

This relationship suggests an iterative procedure to determine the minimum packet delay. Thus we will keep lowering the maximum allowed link utilization in the network as long as the overall delay keeps going down. While this procedure may be applied by choosing as starting point any feasible schedule, it will be most appropriate to choose minimum length scheduling configuration as starting point since it results in optimal packet delay under light traffic load. However, as explained before, there are many minimum frame length schedules with different packet delays, therefore we choose a set of minimum length scheduling configurations as initial point.

Next, we present the heuristic to find the minimum average packet delay in the network. Recall that ρ_{max} represents the maximum allowed link utilization in the network. Thus, we have $b_l/u_l \leq \rho_{max}$ where $l \in L_b$, which is equivalent to capacity constraints (5.46) and (5.47).

Procedure 5.2. Determining the minimum average packet delay.	
Step 1:	Initialize the maximum link utilization to $\rho_{\max}$ =1.
Step 2:	Determine a set of minimum frame length schedules either using OG or CG problem. Compute the delay of each schedule from (5.5) and save the delays in D_{i}.
Step 3:	Update maximum link utilization as $\rho_{\max}$=$\rho_{\max}$−Δρ, where Δρ is the decrement size, and repeat step 2. If it results in a schedule with a larger frame length, then save the delay results in another set, say, D_{i+1}, and go to step 4.
Step 4:	Repeat step 3 as long as min D_{i+1} < min D_i.
Step 5:	The minimum average packet delay in the network is the minimum of the last D_i.

5.6. Power Control Constraints

In this section, we consider an extension to our earlier formulation, where nodes are able to adjust power of their transmission. Accordingly, the SIC opportunities increase by properly adjusting the transmission power of nodes. Here, we study the power adjustment in a way that the routing conditions remain the same. Thus we extend our formulation by only modifying the physical layer constraints. Clearly, this extension can be applied to the CG problem, where ISs are generated in the pricing model. To do this we need to revise (5.51)-(5.54). Note that these constraints become non-linear when the transmission power of nodes is not constant. Hence, we define new variables to make these constraints linear.

Let $P^i{}_{\max}$ and $P^i{}_{\min}$ denote the maximum and minimum transmission power of node i, respectively, and let $\Theta^q_{h,i}$ denote a binary variable; $\Theta^q_{h,i}$ =1 if P_{qi}>P_{hi}; otherwise, $\Theta^q_{h,i}$ =0. Also let $\Omega^q_{h,i}$ represent the power of node q, P_q, when $\Theta^q_{h,i}$=0 (i.e., P_{qi}≤P_{hi}); otherwise, $\Omega^q_{h,i}$ =0. Then we have

$$P_h d_{hi}^{-\alpha} - \beta(N_0 + \sum_{m\in N, m\neq h,i} \Omega^m_{h,i} d_{mi}^{-\alpha}) \geq (1 - x_{h,B})\delta_{hi} \qquad \forall e_{h,B} \in L_b, i \in B \tag{5.57}$$

$$\{P_q d_{qi}^{-\alpha} - \beta(N_0 + \sum_{m\in N, m\neq q,i} \Omega^m_{q,i} d_{mi}^{-\alpha}) \geq (2 - \Upsilon_{h,B,(q)} - \Theta^q_{h,i})\delta_{qi}; \tag{5.58}$$

$$2\Upsilon_{h,B,(q)} \leq x_{h,B} + \Delta_q; \tag{5.59}$$

$$x_{h,B}+\Delta_q \le \Upsilon_{h,B,(q)}+1\} \quad \forall e_{h,B}\in L_b, i\in B, q\in N \tag{5.60}$$

$$\{P_i \le P^i_{\max}\Delta_i; \tag{5.61}$$

$$P_i \ge P^i_{\min}\Delta_i\} \quad \forall i\in N \tag{5.62}$$

$$\{P_q d_{qi}^{-\alpha} - P_h d_{hi}^{-\alpha} \ge -P^h_{\max} d_{hi}^{-\alpha}(1-\Theta^q_{h,i})+\varepsilon; \tag{5.63}$$

$$P_q d_{qi}^{-\alpha} - P_h d_{hi}^{-\alpha} \le P^q_{\max} d_{qi}^{-\alpha}\Theta^q_{h,i}+\varepsilon; \tag{5.64}$$

$$\Omega^q_{h,i} \le P_q + P^q_{\max}\Theta^q_{h,i}; \tag{5.65}$$

$$\Omega^q_{h,i} \ge P_q - P^q_{\max}\Theta^q_{h,i}; \tag{5.66}$$

$$\Omega^q_{h,i} \le P^{\max}_q(1-\Theta^q_{h,i})\} \quad \forall i,h,q\in N, i\ne h\ne q \tag{5.67}$$

Constraints (5.57)-(5.60) are similar to (5.51)-(5.54). The only difference is that the power is no longer constant. δ_{hi} in (5.57) is a lower bound of the term on the LHS similar to (5.21), which can be set as

$$\delta_{hi} = -\beta(N_0 + \sum_{m\in N, m\ne h,i} P^m_{\max} d_{mi}^{-\alpha}) \tag{5.68}$$

The same happens to δ_{qi} in (5.58). Constraints (5.61) and (5.62) limit the transmission power of node i when node i is active. Constraints (5.63) and (5.64) express the definition of $\Theta^q_{h,i}$. Note that ε is a very small constant close to zero, which ensures that $\Theta^q_{h,i}$=0 when the LHS of (5.63) or (5.64) is zero. Finally, (5.65)-(5.67) express the definition of $\Omega^q_{h,i}$ in the form of LP.

The power control extension can be applied to the CG problem by replacing (5.51)-(5.54) by (5.57)-(5.67) in the pricing model.

5.7. Numerical Results

In this section, we present some numerical results regarding the analysis studied in this chapter. First we determine the minimum frame length. Then we determine optimal average packet delay with the corresponding frame length, and slot assignment of the active broadcast links for 5 networks with a given traffic demand matrix. We note that we have two optimization problems, P1 and P2, to determine minimum packet delay and minimum frame length, respectively. We evaluate the performance of the following schemes on five MWNs: (a) w/o NC+SIC; (b) with SIC; (c) with NC; (d) with NC+SIC. Scheme (d) is the theoretical formulation proposed in this work, and the other schemes can be derived easily from (d). In NC scheme the nodes may encode up to 4 native packets together. All the schemes employ interference-aware multi-path routing. It is assume that the power of ambient noise (N_0) is 10^{-6} and the SINR threshold (β) is 1.

In the following, we solve P1 for very small-sized networks, scenario A and B, using MIDACO solver [56] in MATLAB interface [55], while minimum packet delay in larger networks, scenarios C and D, has been determined through application of the proposed heuristic procedure. We set the decrement size $\Delta\rho$ to 0.1 for determining the minimum average packet delay using procedure 5.2.

We solve P2, OG, and CG problems using optimization programming language (OPL) with CPLEX solver [59]. Clearly, P2, OG, and CG problems do not have any feasible solution when the injected traffic into the network is greater than the capacity of the network.

Scenario A:

Consider the 2-node network shown in Fig. 5.2. In this simple network nodes n_1 and n_2 exchange their traffic with each other, where $\lambda(f_1)$=4.5 and $\lambda(f_2)$=1.5 (packet/ms), and the channel capacity is 10 packet/ms. In the following, we study the behavior of the network under two different approaches. In the first approach the links are scheduled in a way that the delay is minimized using P1 problem, and the latter one gives the scheduling under minimum frame length using P2 problem.

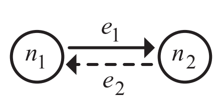


Figure 5.2. Scenario A: 2-node network.

TABLE 5.1.

SCENARIO A: RESULTS OBTAINED FROM P1 AND P2 PROBLEMS FOR THE NETWORK IN FIG. 5.2. $\lambda(f_1)$=4.5 and $\lambda(f_2)$=1.5.

Problem	b_1	b_2	u_1	u_2	ρ_1	ρ_2	T_s	*Delay*
P1	1.35	0.45	2	1	0.68	0.45	3	0.28
P2	0.9	0.3	1	1	0.90	0.30	2	0.79[1]

Table 5.1, compares the results obtained from P1 and P2 problems. From the table, we see that P2 allocates each of the links 1 slot, although the traffic of link e_1 is close to its capacity; as a result, the frame length is 2 slots. Using (5.5), the corresponding delay of this solution is 0.79ms. However, to decrease the delay, P1 allocates 2 slots to link e_1 at the cost of increasing the frame length to 3 slots. The increase of the frame length from 2 to 3, leads to the increase of %50 in the traffic load of each link within a frame, as shown by b_1 and b_2 in the table. As a consequence, the schedule obtained from P1 moderates the channel utilization of the links by decreasing ρ_1 and increasing ρ_2. As we saw, the channel utilization is the most important factor from the delay point of view when the traffic of links is not at the same level.

Scenario B:

Consider the 3-node network shown in Fig. 5.3, where nodes n_1 and n_3 transmit their traffic to each other through node n_2. In this scenario, we study the performance of NC and SIC under both symmetric and non-symmetric traffic. We assume that the channel capacity is 10 packet/ms.

Table 5.2 shows the results of P1 problem under symmetric traffic for $\lambda(f_1)$=2 and $\lambda(f_2)$=2 (packet/ms). From the table we see that the packet delay of NC is less than SIC scheme, since in the case of NC only 3 links are active. However, in SIC scheme there are 4 active links, so 4 transmissions are done over the links.

Note that this scenario is an ideal case since the traffic of links is equal. Under this situation both NC and SIC are very efficient. In addition, the time slots are equally allocated to the links. The scheduling results for the scheme of NC+SIC are shown in Table 5.3, where the third column shows the time slot in which each link is active.

[1] This is the corresponding delay to the solution of P2 problem using (5.5).

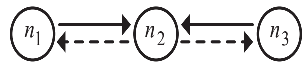


Figure 5.3. Scenario B: 3-node network.

TABLE 5.2.
SCENARIO B: RESULTS OBTAINED FROM P1 PROBLEM UNDER SYMMETRIC TRAFFIC FOR THE NETWORK IN FIG. 5.3. $\lambda(f_1)$=2 and $\lambda(f_2)$=2.

	w/o NC+SIC	**SIC**	**NC**	**NC+ SIC**
Delay	2	0.75	0.56	0.25
T_s	4	3	3	2

TABLE 5.3.
SCENARIO B: SCHEDULING RESULTS OBTAINED FROM P1 PROBLEM UNDER SYMMETRIC TRAFFIC FOR THE SCHEME OF NC+SIC FOR THE NETWORK IN FIG.5.3. $\lambda(f_1)$=2 and $\lambda(f_2)$=2.

Link Number	**Link**	**Time Slot**	b_i	u_i	ρ_i
1	e_{12}	1	0.4	1	0.4
2	e_{32}	1	0.4	1	0.4
3	$e_{2,\{1,3\}}$	2	0.4	1	0.4

Next, we consider the same network under non-symmetric traffic, $\lambda(f_1)$=3 and $\lambda(f_2)$=1. Note that the total traffic injected to the network is the same as the previous case. Table 5.4 shows the results of P1 problem. By comparing Table 5.4 with Table 5.2, we observe that the performances of NC and SIC have deteriorated, and that the scheduling length of each scheme is increased due to non-symmetric traffic. We note that the frame length shown in Table 5.4 is the one corresponding to optimal packet delay and not the minimum T_s. The minimum T_s is obtained by P2 problem under ρ_{max}=1.

The scheduling results obtained from P1 and P2 problems for NC+SIC scheme are shown in Tables 5.5 and 5.6, respectively. By comparing these two tables, we see that the optimal packet delay scheduling takes 5 time slots while the minimum length scheduling takes 3 slots for NC+SIC scheme; in P1 the utilization of links 1 and 4 have been reduced.

Note that as shown in Tables 5.5 and 5.6, the traffic of link 3 is the minimum of the traffic arriving to the coding node (i.e., the traffic of links 1 and 2), and the traffic of link 4 is the residue traffic which could not be encoded. In fact, the rate of encoding at the coding node is the minimum rate of packet arrivals. Indeed, this is one of the drawbacks of NC, specifically in large networks where the arrival rate of packets to a coding node has large variations. However, when SIC is used the rate of packets reception at a node may be the maximum arrival rate of packets, as shown in the tables.

TABLE 5.4.
SCENARIO B: RESULTS OBTAINED FROM P1 PROBLEM UNDER NON-SYMMETRIC TRAFFIC FOR THE NETWORK IN FIG.5.3. $\lambda(f_1)$=3 and $\lambda(f_2)$=1.

	w/o NC+SIC	SIC	NC	NC+SIC
Delay	2.26	1.12	1.3439	0.85
T_s	8	5	7	5

TABLE 5.5.
SCENARIO B: SCHEDULING RESULTS OBTAINED FROM P1 PROBLEM UNDER NON-SYMMETRIC TRAFFIC FOR THE SCHEME OF NC+SIC FOR THE NETWORK IN FIG.5.3. $\lambda(f_1)$=3 and $\lambda(f_2)$=1.

Link Number	Link	Time Slot	b_i	u_i	ρ_i
1	e_{12}	1,2	1.5	2	0.75
2	e_{32}	1	0.5	1	0.5
3	$e_{2,\{1,3\}}$	3	0.5	1	0.5
4	e_{23}	4,5	1	2	0.5

TABLE 5.6.
SCENARIO B: SCHEDULING RESULTS OBTAINED FROM P2 PROBLEM UNDER NON-SYMMETRIC TRAFFIC FOR THE SCHEME OF NC+SIC FOR THE NETWORK IN FIG.5.3. $\lambda(f_1)$=3 and $\lambda(f_2)$=1.

Link number	Link	Time Slot	b_i	u_i	ρ_i
1	e_{12}	1	0.9	1	0.9
2	e_{32}	1	0.3	1	0.3
3	$e_{2,\{1,3\}}$	2	0.3	1	0.3
4	e_{23}	3	0.6	1	0.6

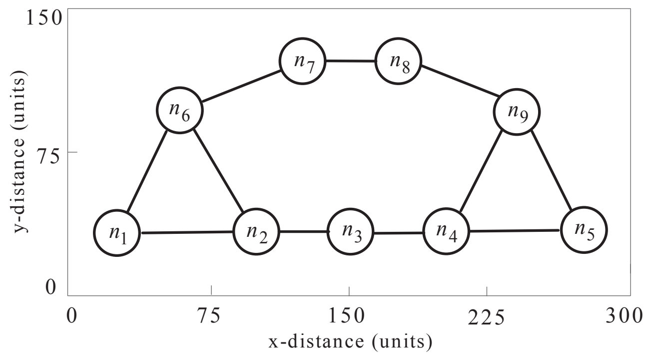


Figure 5.4. Scenario C: 9-node network.

TABLE 5.7.
SCENARIO C: NODE'S TRANSMISSION POWER FOR THE NETWORK IN FIG.5.4.

Wireless Node *i*	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
Transmission Power P_i	1.4	0.5	1	0.9	1.1	0.6	0.9	1.1	1.2

Scenario C:

Consider the 9-node network shown in Fig. 5.4, where there are 4 concurrent flows f_1: 1→8, f_2: 8→2, f_3: 5→2 and f_4: 2→9 in the network. The traffic demand of flows is 3, 2, 1, and 0.5 packet/ms, respectively and the channel capacity is 10 packet/ms. We consider two paths for routing of each traffic flow in the network. The transmission power of the nodes is chosen randomly from the interval of [0.5, 1.5] as shown in Table 5.7.

In this scenario, we obtain the minimum packet delay using procedure 5.2. The results of this procedure for the NC+SIC, SIC, NC, and w/o NC+SIC schemes are presented in Tables 5.8 – 5.11, respectively. In these tables $\rho_{\max}$ shows the maximum allowed link utilization in the network. For example, for the scheme of NC+SIC (see Table 5.8), the first row indicates that in this scenario the variation of $\rho_{\max}$ in the interval of [0.8, 1] does not increase the scheduling length. The last column shows that the minimum packet delay of network is 1.58ms which occurs at T_s=6 slots. It is seen that minimum packet delay does not occur at the frame length, T_s=4.

TABLE 5.8.
SCENARIO C: MINIMUM DELAY RESULTS OBTAINED FROM PROCEDURE 5.2 FOR NC+SIC FOR THE NETWORK IN FIG.5.4. $\lambda(f_1)$=3, $\lambda(f_2)$=2, $\lambda(f_3)$=1, $\lambda(f_4)$=0.5.

D_j	Minimum T_s	$\rho_{\max}$	Minimum Delay of D_j
D_1	4	0.80 – 1.00	2.14
D_2	6	0.70 – 0.79	1.58
D_3	9	0.63 – 0.69	2.03

TABLE 5.9.
SCENARIO C: MINIMUM DELAY RESULTS OBTAINED FROM PROCEDURE 5.2 FOR SIC FOR THE NETWORK IN FIG.5.4. $\lambda(f_1)$=3, $\lambda(f_2)$=2, $\lambda(f_3)$=1, $\lambda(f_4)$=0.5.

D_j	Minimum T_s	$\rho_{\max}$	Minimum Delay of D_j
D_1	5	0.88 – 1	5.54
D_2	7	0.82 – 0.87	3.80
D_3	8	0.80 – 0.81	4.11

TABLE 5.10.
SCENARIO C: MINIMUM DELAY RESULTS OBTAINED FROM PROCEDURE 5.2 FOR NC FOR THE NETWORK IN FIG.5.4. $\lambda(f_1)$=3, $\lambda(f_2)$=2, $\lambda(f_3)$=1, $\lambda(f_4)$=0.5.

D_j	Minimum T_s	$\rho_{\max}$	Minimum Delay of D_j
D_1	5	1	∞
D_2	6	0.90 – 0.99	4.71
D_3	7	0.82 – 0.89	3.65
D_4	8	0.80 – 0.81	4.09

TABLE 5.11.
SCENARIO C: MINIMUM DELAY RESULTS OBTAINED FROM PROCEDURE 5.2 FOR W/O NC+SIC FOR THE NETWORK IN FIG.5.4. $\lambda(f_1)$=3, $\lambda(f_2)$=2, $\lambda(f_3)$=1, $\lambda(f_4)$=0.5.

D_j	Minimum T_s	$\rho_{\max}$	Minimum Delay of D_j
D_1	10	1	∞
D_2	13	0.98 – 0.99	29.39
D_3	16	0.96 – 0.97	19.95
D_4	18	0.90 – 0.95	12.32
D_5	infeasible	0.89	

TABLE 5.12.
SCENARIO C: MINIMUM DELAY RESULTS OBTAINED FROM PROCEDURE 5.2 FOR THE NETWORK IN FIG.5.4.
$\lambda(f_1)$=3, $\lambda(f_2)$=2, $\lambda(f_1)$=1, $\lambda(f_2)$=0.5.

	w/o NC+SIC	**SIC**	**NC**	**NC+SIC**
Optimal Delay	12.32	3.80	3.65	1.58
ρ_{max}	0.9	0.82	0.82	0.7
T_s (optimal delay)	18	7	7	6
T_s (minimum)	10	5	5	4

The performance of different schemes is compared in Table 5.12. This table shows the optimal delay and the corresponding ρ_{max} and T_s obtained by our proposed procedure. The last row shows the minimum scheduling length under ρ_{max}=1. It can be seen that NC+SIC reduces the delay significantly compared to the w/o NC+SIC scheme. Also optimal packet delay occurs at a T_s higher than minimum T_s in all schemes. We should add that the minimum delay of w/o NC+SIC can be also obtained directly by solving P1 problem.

The routing of NC+SIC scheme under minimum delay and minimum frame length are shown in Fig.s 5.5.a and 5.5.b, respectively. From Fig. 5.5, it can be said that routing also differs in minimum packet delay and minimum scheduling length cases. From Fig. 5.5a, we observe that using multi-path routing not only increases NC opportunities in the network but also reduces utilization of the links with high traffic load.

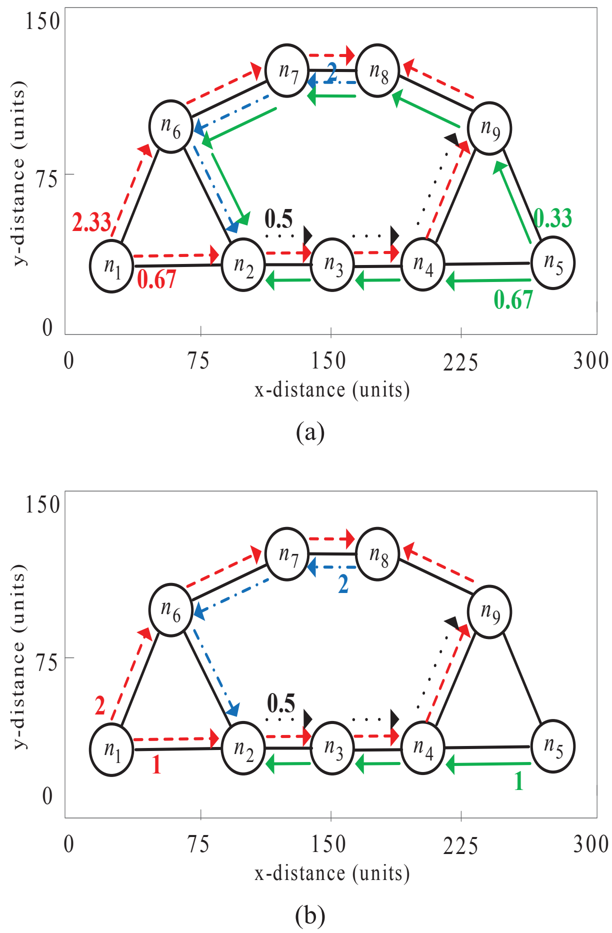


Figure 5.5. Scenario C: The (multi-path) routing of NC+SIC scheme for the network in Fig. 5.4. $\lambda(f_1)$=3 from node 1 to node 8, $\lambda(f_2)$=2 from node 8 to node 2, $\lambda(f_3)$=1 from node 5 to node 2 and $\lambda(f_4)$=0.5 from node 2 to node 9. (a) Solution under minimum delay, $\rho_{\max}$=0.7, T_s=6. (b) Solution under minimum scheduling length, T_s=4.

Scenario D:

In this scenario, we consider the 21-node MWN shown in Fig. 5.6. The transmission power of each node is set to one; since SINR threshold is set to β=1, then from physical model, (5.8), two nodes are within transmission range of each other if and only if their distance is less than 100 units, as shown in Fig. 5.6. Under this assumption, the average node degree is 5.5. There are 20 traffic flows in the network; the source and destination nodes for each flow are chosen at random, and two shortest paths are considered for routing of each flow in the network.

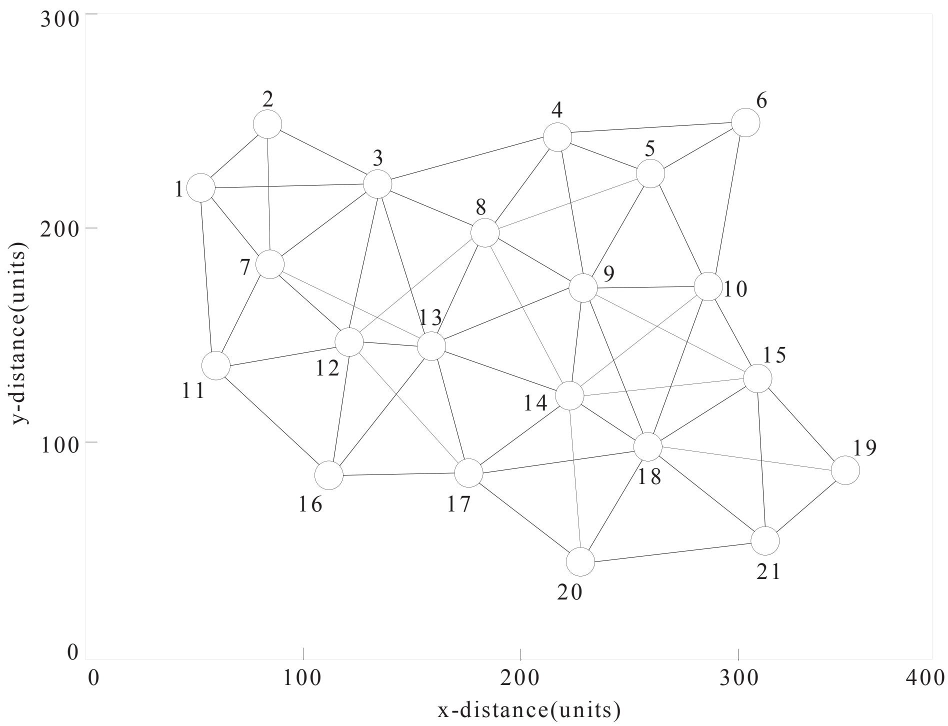


Figure 5.6. Scenario D: 21-node MWN.

In the sequel, first we compare OG and CG methods with each other under NC+SIC scheme. For simplicity, we assume that the traffic of each flow is exactly one packet during the frame length; i.e., $\Lambda(f_i)=1$ or equivalently $\lambda(f_i) =(T_s\tau)^{-1}$. Table 5.13 shows the results of OG method for NC+SIC under variation of $E_{\max}$, where $|M|$ shows the number of ISs in M, the third column shows the time it takes for finding ISs offline using procedure 5.1, and the last column shows the minimum scheduling length at each value of $E_{\max}$ using OG method. We should add that finding ISs for the other schemes takes much less time compared to NC+SIC scheme. In addition, it usually takes a few minutes to solve an optimization problem using OG method for a network like Fig. 5.6.

The results of CG are summarized in Table 5.14. From the table it may be seen that after 398 iterations within 498 minutes the reduced cost (i.e., (5.55)) becomes zero. Further, the optimal value (of *Ts*) of the LP-relaxation is 11.89 and the ILP solution with respect to all the ISs required for solving the LP-relaxation is 13. Comparing the results of OG and CG method, we observe that the optimal solution of ILP is 12. From Table 5.13, we see that depending on the value of $E_{\max}$ a portion of feasible ISs is discovered, and OG method may lead to the optimal ILP solution under a portion of feasible ISs, since only a few ISs are used in the optimal scheduling. As a result, the OG is also an efficient method in terms of time and accuracy. In general, OG is easier method than CG in solving the optimization problem; however, for different schemes, the corresponding ISs should be generated offline separately.

TABLE 5.13.
SCENARIO D: OG RESULTS FOR NC+SIC SCHEME BY VARYING THE NUMBER OF GIVEN ISS FOR THE NETWORK IN FIG.5.6.

$E_{\max}$	$\|M\|$	Time (min)	Minimum T_s
1000	143908	1122	12
100	103933	184	12
50	88044	117	13
10	33740	13	13
5	14355	4	14

TABLE 5.14.
SCENARIO D: CG RESULTS FOR NC+SIC SCHEME FOR THE NETWORK IN FIG.5.6.

Iteration	Time (min)	Minimum T_s (LP)	Minimum T_s (ILP)
398	489	11.89	13

Next, we present optimal packet delay results obtained through the application of heuristic procedure for the network of Fig. 5.6 with 20 flows with randomly chosen source and destination nodes. It is assumed that the channel capacity is 54 (packet/ms) and all flows have the same amount of traffic. The average packet delay of the network versus the offered traffic demand (per flow) is shown in Fig. 5.7.

It can be seen that NC scheme results in lower packet delay than w/o NC+SIC scheme, however, improvement is not significant because of the following reasons. First, the rate of encoding at a coding node is at most the minimum arrival rate of packets which are encoded. In a large network, the arrival rate of packets to a node from different incoming links is variable. Second, when the delay is minimized, the traffic is routed mostly over multiple paths (due to the restriction of links utilization). On one hand, this increases NC opportunities in the network, but on the other hand this increases the number of active links which leads to the increase of number of transmissions. Note that NC does not necessarily reduce the number of transmissions in the network, since the residue traffic, which could not be encoded, should be transmitted over unicast links. Finally, when NC is used not all next-hop nodes are allowed to transmit in the same time slot, so some neighboring nodes are prevented from transmission. As a result, NC is very efficient when the rate of coding is high. As may be seen SIC scheme achieves better performance than NC because the rate of packets reception at a node may be the maximum arrival rate of packets. Finally, NC+SIC gives the best delay performance, since SIC improves the performance of NC by allowing the routing of the flows to be closer to each other, which results in the increase of NC opportunities in the network. With NC+SIC the network can support the per flow traffic of 3.5 (packet/ms) while with w/o NC+SIC only 2 (packets/ms) before delay rises steeply due to congestion.

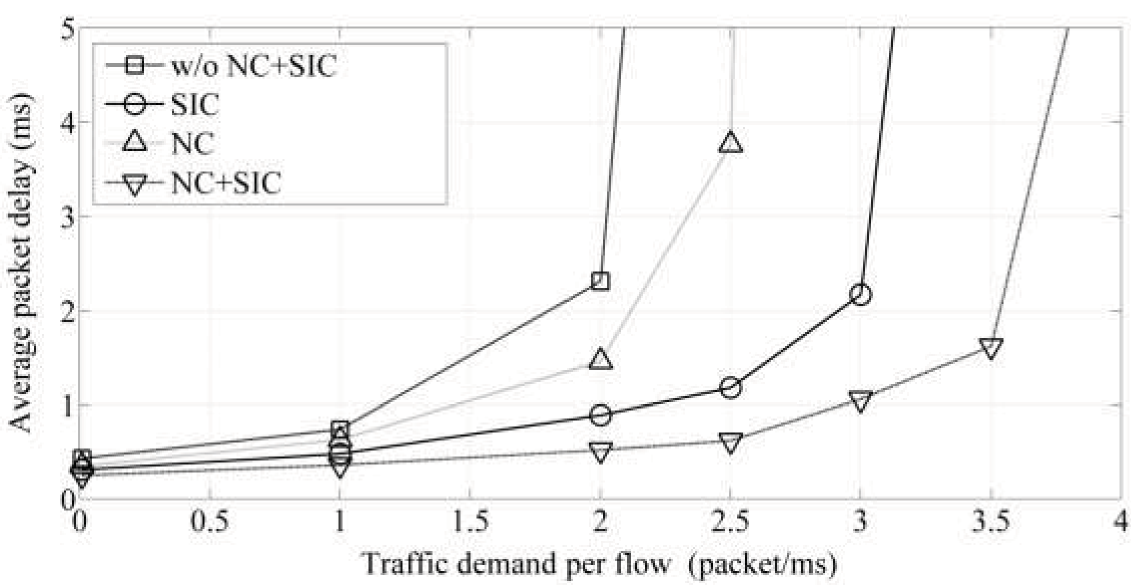


Figure 5.7. Scenario D: Average packet delay (results obtained from procedure 5.2) versus the offered traffic for the network shown in Fig. 5.6.

Scenario E:

Consider a 25-node topology shown in Fig. 5.8. Similar to scenario D, the transmission power of each node is set to one. Under this condition, the average node degree is 5.2. There are 100 traffic flows in the network and the source and destination nodes for each flow are chosen at random. We assume that the traffic of each flow is exactly one packet per frame (i.e., $\Lambda(f_i)=1$). To compare our results with previous work, in this scenario in addition to the previous schemes we consider new schemes. Let MPR(m) represent that m packets can be received by a node without considering the effect of collision (the secondary interference) in a time slot. For example, MPR(1) is similar to w/o NC+SIC scheme but excluding the secondary interference constraints from the optimization framework.

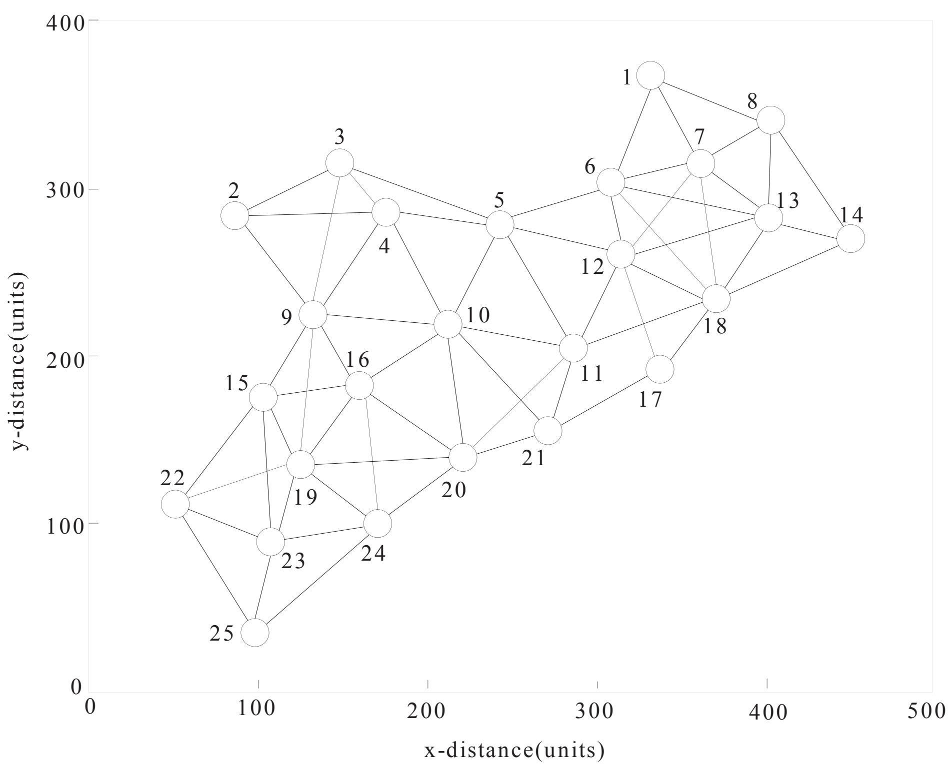


Figure 5.8. Scenario E: 25-node MWN.

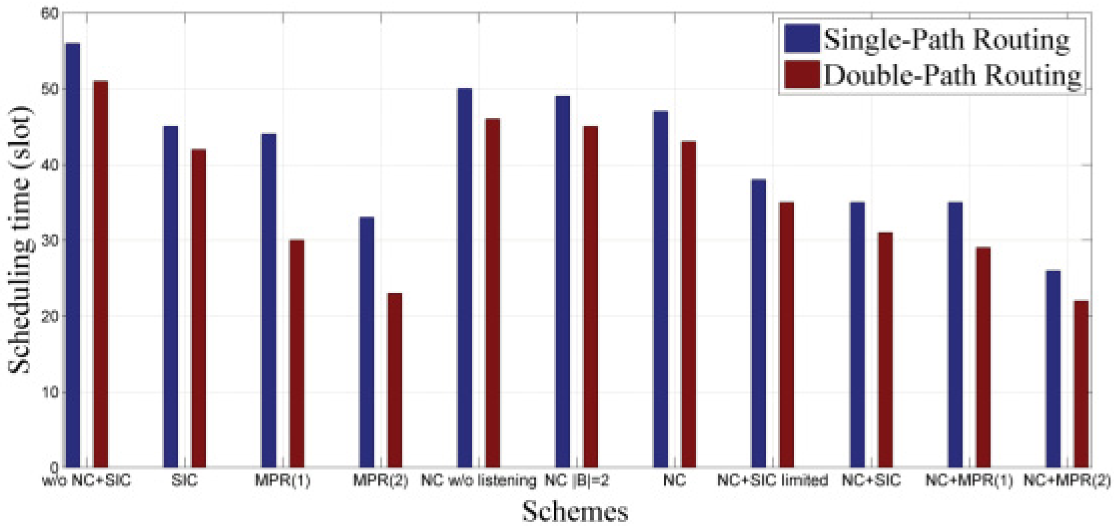


Figure 5.9. Scenario E: Performance evaluation in terms of scheduling time using CG method for the network shown in Fig. 5.8.

The performance of each scheme both under single and double-path routing are summarized in Fig. 5.9. These results are obtained by CG problem under maximum link utilization; i.e., $\rho_{\max}$=1. This figure, in fact, shows the throughput of the schemes in terms of the length of scheduling. The following points may be observed from Fig. 5.9 for this scenario. SIC has the capability of mitigating secondary interference considerably. Interestingly, the performance of SIC is similar to MPR(1) under single path routing for this scenario. Allowing NC up to 4 packets (NC scheme) results in better performance compared to the without listening scheme and the scheme where nodes are restricted to encode 2 packets (NC $|B|$=2). Further, significant performance improvement can be achieved by a combination of NC and SIC techniques (NC+SIC).

In addition, to answer the following question: what will be throughput of the system if only some of the nodes have NC+SIC capability while others do not? Will the performance be degraded significantly? To answer these questions, we study a system that only nodes 4, 12, 16 and 23 have such capabilities. Note that these nodes experience heavy loads in this scenario, and that they are not neighboring nodes; hence, all the traffic received by these nodes is native traffic. The limited NC+SIC results in Fig. 5.9 correspond to this scheme. As may be seen the performance is not degraded considerably compared to NC+SIC scheme.

More importantly, we observe that the performance of NC is close to SIC scheme as opposed to scenario D, where the delay is minimized (see Fig. 5.7). Recall that coding nodes encode packets at the minimum arrival rate of packets. In this scenario, in fact, coding nodes are able to encode packets at the maximum rate. i.e.; at the rate of 1 packet per slot. This is because in this scenario $\rho_{\max}=1$, so the link utilization is not restricted. Accordingly, packets can arrive at coding nodes at the rate of 1 packet/slot; of course the assumption of $\Lambda(f_i)=1$ would help the performance of NC, since under this assumption the traffic rate of each flow is exactly 1 packet per frame.

5.8. Conclusion

This chapter presents an approach that may be useful in the design of MWNs that meets the QoS requirements of the users. We note that the throughput maximization fails to capture a complete picture in the design of MWNs. We have presented a theoretical framework that minimizes the average packet delay in spatial TDMA MWNs that jointly utilizes NC and SIC. We assume interference-free scheduling and allow multi-path routing with/without opportunistic listening. The combined use of NC and SIC increases NC opportunities by alleviating the interference. The nodes with NC capability are able to encode up to 4 packets together. The problem formulation results in a difficult MINLP that might be only solved for very small-sized networks by the state-of-art software. We present a heuristic method that determines optimal packet delay iteratively. At each step of the iteration, procedure reduces the maximum allowed link utilization in the network and determines the corresponding packet delay. The procedure chooses as optimal packet delay, the delay of the last iteration that experienced a reduction.

The numerical results show that optimal packet delay in the networks employing the techniques w/o NC+SIC, NC, SIC and NC+SIC decreases from left to right. Similarly, stable operation region of the network before steep rise of the packet delay increases for the techniques in the previous list from left to right. Thus the combined use of NC and SIC results in a significant performance improvement in the network.

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

In this chapter, we present the conclusions of the research done in this thesis and discuss the future work.

6.1. Conclusions

Prior work on performance modeling of NC and SIC mainly addresses the maximization of throughput. However, these approaches fail to capture the complete picture since there may be paths in the network for which the end-to-end packet delay is prohibitively high. In this research we presented a cross-layer optimization approach to improve the performance of MWNs such that QoS requirements may be met.

We started with unicast communications, where each transmitted packet is destined to a single node. We showed the wireless interference models in MWNs and showed the formulation of the multi-path routing in the format of LP.

We then extended the unicast communications to broadcast communications, where NC is used in the network. To exploit NC with opportunistic listening, NC opportunities have been grouped to NC structures which consist of coding elements. We, in addition, indicated all possible NC coding models that can arise at coding nodes in MWNs, under coding the maximum of 4 packets together at a time. Finally, we determined the formulation of multi-path routing in the presence of NC in MWNs both with/without opportunistic listening.

After having in hand all necessary constraints, we presented the theoretical optimization framework with the application of NC in chapter 4. We addressed the problem of delay minimization in MWNs with NC. The objective had been assignment of wireless node capacities in a way that the average packet delay is minimized for a given network topology and the traffic demand matrix with variable length packets with exponential distribution. We developed a performance analysis of the system, which models network nodes as M/G/1 queues and takes into account wireless interference. The proposed model is valid both with and without opportunistic listening for any wireless network topology. The model also incorporates network coding-aware routing that routes the flows in a manner that increases coding opportunities. We presented numerical results, which show that NC reduces the average packet delay in the network and increases throughput of the network. We, furthermore, present simulation results, which confirm the accuracy of the analysis.

In chapter 5, we presented a comprehensive theoretical framework that minimizes the average packet delay in spatial TDMA MWNs with the joint application of NC and MPR. We allowed MPR at network nodes by employing SIC technique. We assumed interference-free scheduling and allowed multi-path routing with/without opportunistic listening. The nodes with NC capability are again able to encode up to 4 packets together. The problem formulation resulted in a difficult MINLP that might be solved for very small-sized networks. We presented a heuristic method that determines optimal packet delay iteratively. At each step of the iteration, procedure reduces the maximum allowed link utilization in the network and determines the corresponding packet delay. The procedure chooses as optimal packet delay, the delay of the last iteration that experienced a reduction. The numerical results show that optimal packet delay decreases from left to right in networks employing the following techniques w/out NC+SIC, NC, SIC and NC+SIC. Similarly, throughput of the network increases for the list of techniques from left to right.

We then used the heuristic procedure to study the behavior of the network under both optimal packet delay and minimum TDMA schedule. Our theoretical analysis shows that asymmetric loads across a coding node affect the performance of NC, since coding nodes encode packets at the minimum arrival rate of packets, and this happens more often under optimal packet delay where the rate of links is restricted by link utilization. We also showed that SIC technique has the capability of mitigating (secondary) interference considerably, and that the combined use of SIC with NC results in a significant performance improvement in the network. Finally, we studied NC+SIC limited scheme, where only a limited number of nodes have NC+SIC capability. We observed that limiting NC+SIC capability only to the nodes with high traffic loads does not impact the network performance significantly.

Finally, we point out that our proposed optimization framework is applicable to any given network topology with any pattern of concurrent traffic flows. Our solution can be used in the design of scheduling-based networks such as WiMax and LTE. Note that our theoretical formulation provides a centralized TDMA schedule in MWNs, which can be used as a lower bound of a distributed TDMA protocol [60], [61]. The distributed scheduling will be far more difficult in the NC environment because it involves routing of the traffic.

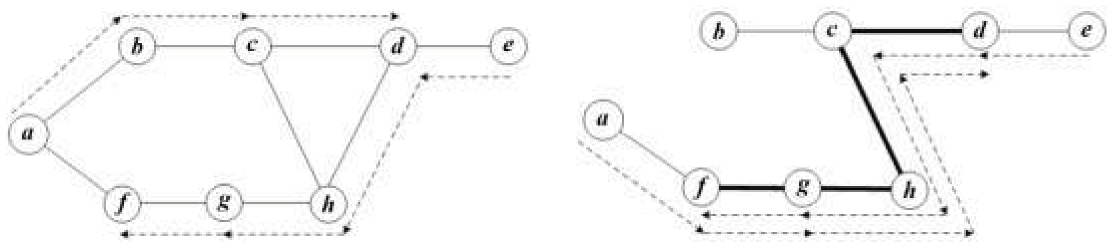
6.2. Future Work

Next, we present few future work proposals.

6.2.1. Backbone Routing

Capacity assignment with fixed routing (i.e., shortest path routing) has been studied in [14]. The drawback of this work is that route selection based on shortest path routing is not efficient in networks that employ NC. Fig. 6.1 clearly illustrates the drawback of shortest path algorithm by a simple example. In this figure, there are two flows in the network, one from node *a* to node *d* and the other from node *e* to node *f*. If the shortest path routing strategy is employed, the paths for the two flows are shown in Fig. 6.1a. We can see that unlike Fig. 6.1b, there is no coding opportunity if using the routes shown in Fig. 6.1a.

On the other hand, using coding-aware routing presented in chapters 4 and 5 complicates the problem; although, this solution results in significant performance improvement in the network.



a) Shortest path routing b) Routing with backbone routing

Figure 6.1. An example of NC comparing shortest path routing and backbone routing

As we saw, using coding-aware routing complicates the problem. Thus as future work one may consider a backbone routing with NC scheme over a MWN, which combines the benefits of both backbone routing and NC techniques. With backbone-based routing, all packets are forced to be transmitted over a constructed backbone. Because of the pre-specified routes in backbone routing, the possibility of coding packets at intermediate nodes can be substantially increased, and thus the benefit of NC is fully exploited. The reason behind is that the performance improvement highly depends on the existence of coding opportunities, which themselves depend on the topology, traffic pattern or offered load.

This approach is motivated by the observation that maximum coding gain could be achieved if the traffic consists of pairs of perfectly overlapping flows going towards opposite directions. Coding gain is the ratio of the number of transmissions required by the current non-coding approach, to the number of encoded transmissions to deliver the same set of packets [2].

6.2.2. Probabilistic Routing

The analysis presented in chapters 3, 4, and 5 implicitly assumes that a node will always have the needed native packets in the buffer to perform the coding. However, because arrival of the packets is random some of the packets needed for coding may not be in the buffer [62], [63]. The solution will be either to wait for the arrival of needed packets or to encode the available packets. These considerations show that the above analysis is an approximation, which would yield optimistic results, particularly for lightly loaded systems. As the loading increases one would expect the approximation to improve.

In [62], [63], this problem has been addressed for an isolated node structure with three nodes without opportunistic listening. Clearly, in NC without opportunistic listening two native packets are combined to give a coded packet. If one of the native packets is missing, the choice is either to transmit the present native packet as is, or wait for the arrival of the other packet to form a coded packet. The former approach increases the network load; on the other hand, the latter approach reduces the load but, increases the delay. Thus, there is a trade-off between the two approaches. In [62], [63], it has been proposed that a solo native packet is transmitted as uncoded according to a Bernoulli trial when a coded packet cannot be formed. They determined the performance of the system as a function of the success probability of the Bernoulli trial.

As future work one may address this problem in the network environment for both with and without opportunistic listening. This process will result in an optimization problem with probabilistic constraints. We think that the solution of the problem may be within the reach of the state-of-art techniques in optimization.

REFRENCES

[1] P. Gupta and P. R. Kumar, “The Capacity of Wireless Networks”, *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, March 2000.

[2] S. Katti, H. Rahul, W. Hu, and M. Medard, “XORs in the Air: Practical Wireless Network Coding”, *IEEE Transactions on Networking*, vol. 16, no. 3, pp. 497-510, May 2008.

[3] J. Le J., J. C. S. Lui, M. D. Chiu, “How Many Packets Can We Encode? -An Analysis of Practical Wireless Network Coding”, in *Proc. IEEE INFOCOM*, 2008, pp. 371-375.

[4] Y. Robert, R. W. Yeung, and N. Cai, “Linear Network Coding”, *IEEE Transactions on Information Theory*, Vol. 49, No. 2, pp. 371-381, February 2003.

[5] A. Eryilmaz, A. Ozdaglar, and M. Medard, “On Delay Performance Gains from Network Coding”, in *Proc. IEEE CISS*, 2006, pp. 864-870.

[6] M. Ghaderi, D. Towsley, and J. Kurose, “Reliability Gain of Network Coding in Lossy Wireless Networks”, in *Proc. IEEE INFOCOM*, 2008, pp. 2171-2179.

[7] P. Pate1 and J. Holtzman, “Analysis of a Simple Successive Interference Cancellation Scheme in a DS/CDMA System”, *IEEE Transactions on Selected* Areas in Communications, vol. 12, no. 5, pp. 796-807, June 1994.

[8] S. Lv, W. Zhuang, X. Wang, and X. Zhou, “Scheduling in Wireless Ad hoc Networks with Successive Interference Cancellation,” in *Proc. IEEE INFOCOM*, 2011, pp. 1282–1290.

[9] S. Sengupta, S. Rayanchu, and S. Banerjee, “Network Coding-Aware Routing in Wireless Networks,” *IEEE/ACM Transactions on Networking*, Vol. 18, No. 4, pp. 1158-1170, August 2010.

[10] S. Shabdanov, C. Rosenberg, and P. Mitran, “Joint Routing, Scheduling, and Network Coding for Wireless Multihop Networks”, in *Proc. IEEE WiOpt*, 2011, pp. 33-40.

[11] C. Jiang, Y. Shi, Y. Hou, W. Lou, and S. Midkiff, "Squeezing the Most Out of Interference: An Optimization Framework for Joint Interference and Avoidance", in *Proc. IEEE INFOCOM*, 2012, pp. 424-432.

[12] J. El-Najjar, H. AlAzemi, and C. Assi, "On the Interplay Between Spatial Reuse and Network Coding in Wireless Networks", *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 560-569, February 2011.

[13] M. Alvandi, M. Mehmet-Ali, and J. Hayes, "Delay Optimization in Multi-Hop Wireless Networks with Network Coding", in *Proc. IEEE WCNC,* 2013, pp. 1381-1386.

[14] M. Alvandi, M. Mehmet-Ali, and J. Hayes, "Delay Optimization of Wireless Networks with Network Coding", in *Proc. IEEE CCECE*, 2011, pp. 1282-1287.

[15] M. Alvandi, M. Mehmet-Ali, and J. Hayes, "Delay Optimization and Cross-Layer Design in Wireless Multi-Hop Networks with Network Coding and successive interference cancellation", *IEEE Journal of Selected Areas in Communications (JSAC); Special Issue on Fundamental Approaches to Network Coding in Wireless Communication Systems*, 2014. (Submitted)

[16] R. Nelson, and L. Kleinrock, "Spatial TDMA: A Collision-Free Multihop Channel Access Protocol", *IEEE Transactions on Communications*, vol. 33, no. 9, pp. 934-944, September 1985.

[17] J. Gronkvist, "Assignment Methods for Spatial Reuse TDMA", in *Proc. IEEE MobiHOC,* 2000, pp. 119-124.

[18] K. Ko, and B. Davis, "Delay Analysis for a TDMA Channel with Contiguous Output and Poisson Message Arrival", *IEEE Transactions on Communications*, vol. 32, no. 6, pp. 707-709, June 1984.

[19] J. Hayes, "Performance Models of an Experimental Computer Communication Network", *The Bell System Technical Journal*, vol.53, no. 2, pp. 225-259, February 1974.

[20] P. Bjorklund, P. Varbrand, and D. Yuan, "Resource Optimization of Spatial TDMA in Ad HOC Radio Networks: A Column Generation Approach", in *Proc. IEEE INFOCOM*, 2003, pp. 818-824.

[21] E. Anderson, C. Phillips, D. Sicker, and D. Grunwald, "Optimization Decomposition for Scheduling and System Configuration in Wireless Networks", *IEEE/ACM Transaction on Networking,* vol. 22, no. 1, pp. 271-284, February 2014.

[22] L. Kleinrock, *Queueing Systems*, Volume 2: Computer Applications, New York, USA, John Wiley and Sons, 1976.

[23] N. Bansal and Z. Liu, "Capacity, Delay and Mobility in Wireless Ad-hoc Networks," in *Proc. IEEE INFOCOM*, 2003, pp. 1553-1563.

[24] C. Comaniciu and H. V. Poor, "On the Capacity of Mobile Ad-hoc Networks with Delay Constraints," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 1536-1276, August 2006.

[25] S. Toumpis and A. J. Goldsmith, "Capacity Regions for Wireless Ad-hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 736-748, July 2003.

[26] V. Ramamurthi, A. Reaz, and B. Mukherjee, "Optimal Capacity Allocation in Wireless Mesh Networks", in *Proc. IEEE GLOBECOM*, 2008, pp. 1-5.

[27] L. Li, J. Y. Halpern, P. Bahl, Y.-M. Wang, and R. Wattenhofer, "A Cone-Based Distributed Topology-Control Algorithm for Wireless Multihop Networks," *IEEE/ACM Transactions on Networking,* vol. 13, no.1, pp. 147-159, February 2005.

[28] L. Ning, J. C. Hou, and L. Sha, "Design and Analysis of an MST-Based Topology Control Algorithm," *IEEE Transactions on Wireless Communications,* vol. 4, no.3, pp. 1195-1206, May 2005.

[29] P. Santi, "Topology Control in Wireless Ad-hoc and Sensor Networks," *ACM Computing Surveys,* vol. 37, no.2, pp. 164-194, June 2005.

[30] R. Ahlswede, N. Cai, R. Li, and R. Yeung, "Network Information Flow", *IEEE Transactions on Information Theory*, Vol. 46, No. 4, pp. 1204-1216, July 2000.

[31] R. Koetter and M. Médard, "An Algebraic Approach to Network Coding", *IEEE/ACM Trans. Networking*, vol.11, no. 5, pp. 782-795, October 2003.

[32] T. Ho, M. Medard, J. Shi, M. Effros, and D. Karger, "On Randomized Network Coding", in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, October 2003.

[33] T. Ho, M. Medard, R. Koetter, D. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast", *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, October 2006.

[34] A. Ramamoorthy, J. Shi, and R. Wesel, "On the Capacity of Network Coding for Random Networks". *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2878-2885, August 2005.

[35] S. Deb, M. Effros, T. Ho, D. Karger, R. Koetter, D. Lun, M. Medard, and N. Ratnakar, "Network Coding for Wireless Applications: A Brief Tutorial", in *Proc. The International Workshop on Wireless Ad-hoc Networks (IWWAN)*, 2005, pp. 196-200.

[36] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading Structure for Randomness in Wireless Opportunistic Routing", in *Proc. ACM SIGCOMM*, 2007, vol. 37, no. 4, pp. 169-180.

[37] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, "Symbol-Level Network Coding for Wireless Mesh Networks", in *Proc. ACM SIGCOMM*, 2008, vol. 38, no. 4, pp.401-412.

[38] X. Zhang, and B. Li, "Dice: A Game Theoretic Framework for Wireless Multipath Network Coding", in *Proc. ACM MobiHoc*, 2008, pp. 293-302.

[39] D. S. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee, "Achieving Minimum Cost Multicast: A Decentralized Approach Based on Network Coding," in *Proc. IEEE INFOCOM*, 2005, pp. 1607-1617.

[40] J Zhang, and Q. Zhang, "Cooperative Network Coding-Aware Routing for Multi-Rate Wireless Networks", in *Proc. IEEE INFOCOM*, 2009, pp. 181-189.

[41] N. Jones, B. Shrader, and E. Modiano, "Optimal Routing and Scheduling for a Simple Network Coding Scheme", in *Proc. IEEE INFOCOM*, 2012, pp. 352-360.

[42] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of Interference on Multi-hop Wireless Network Performance," in *Proc. ACM MobiCom*, 2003, pp. 66-80.

[43] S. Verdu, *Multiuser Detection,* Cambridge Univ. Press, 1998.

[44] P. Frenger, P. Orten, and T. Ottosson, "Code-spread CDMA with Interference Cancellation," *IEEE transaction on selected Areas in Communications*, vol. 17, no. 12, pp. 2090-2095, December 1999.

[45] T.R. Giallorenzi and S.G. Wilson, "Suboptimum Multiuser Receivers for Convolutionally Coded Asynchronous DS-CDMA Systems," *IEEE Transaction on Communication,* vol. 44, no. 9, pp. 1183–1196, September 1996.

[46] X. Wang and H.V. Poor, "Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA," *IEEE Transaction on Communication,* vol. 47, no. 7, pp. 1046–1061, July 1999.

[47] J.G. Andrews, "Interference Cancellation for Cellular Systems: A Contemporary Overview," *IEEE Wireless Communication Magazine,* vol. 12, no. 2, pp. 19–29, April 2005.

[48] P. Mitran, C. Rosenberg, and S. Shabdanov, "Throughput Optimization in Wireless Multihop Networks with Successive Interference Cancellation", in *Proc. IEEE WTS*, 2011, pp. 1-7.

[49] J. Cloud, L. Zeger, and M. Medard, "MAC Centered Cooperation – Synergistic Design of Network Coding, Multi-Packet Reception, and Improved Fairness to Increase Network Throughput", *IEEE Transactions on Selected Areas in Communications*, vol. 30, no. 2, pp. 341-349, February 2012.

[50] S. Katti, S. Gollakota, and D. Katabi, "Embracing Wireless Interference: Analog Network Coding," in *Proc. ACM SIGCOMM*, 2007, pp. 397-408.

[51] S. Gollakota and D. Katabi, "Zigzag Decoding: Combating Hidden Terminals in Wireless Networks," in *Proc. ACM SIGCOMM*, 2008, pp. 159-170.

[52] J. Qureshi, J. Cai, and C. Heng Foh,"Cooperative Retransmissions Through Collisions", in *Proc. IEEE ICC,* 2011, pp. 1-5.

[53] S. Ramanathan, E.L. Lloyd, "Scheduling Algorithms for Multihop Radio Networks", *IEEE/ACM Transactions on Networking*, vol. 1, no. 2, pp. 166-177, April 1993.

[54] J. Hayes, and T. G. Babu, *Modeling and Analysis of Telecommunications Networks*, New Jersey, USA, John Wiley and Sons, Inc., Publication, February 2004.

[55] MATLAB, http:mathworks.com.

[56] Mixed Integer Distributed Ant Colony Optimization (MIDACO) Solver, http://midaco-solver.com.

[57] MATLAB Global Optimization Toolbox/Genetic Algorithm Solver, http:mathworks.com.

[58] M. Lubbecke and J. Desrosiers, "Selected topics in column generation", *Operation Research Journal*, vol 53, no.6, pp.1007–1023, November 2005.

[59] IBM ILOG CPLEX Solver, http://www.ilog.com.

[60] I. Rhee, A. Warrier, J. Min, and L. Xu, "DRAND: Distributed Randomized TDMA Scheduling for Wireless Ad-hoc Networks", in *Proc. ACM International Symposium on Mobile Ad-hoc Networking and Computing*, 2006, pp. 190-201.

[61] T. Herman, and S. Tixeuil. "A distributed TDMA Slot Assignment Algorithm for Wireless Sensor Networks." *Algorithmic Aspects of Wireless Sensor Networks*. Springer Berlin Heidelberg, 2004, pp. 45-58.

[62] W. Chen, K. Letaief, and Z. Cao, "Opportunistic Network Coding for Wireless Networks", in *Proc. IEEE ICC*, 2007, pp. 4634-4639.

[63] R. Zhang, and Q. Qian, "Probabilistic Network Coding with Priority over wireless Networks", in *Proc. IEEE ICMTMA*, 2011, pp. 918-921.