ENHANCING THE PERFORMANCE OF RELAY NETWORKS WITH NETWORK CODING

by

Scott Harold Melvin

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2012

DALHOUSIE UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Enhancing the Performance of Relay Networks with Network Coding" by Scott Harold Melvin in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: August 2, 2012

External Examiner: ______________________________

Research Supervisor: ______________________________

Examining Committee: ______________________________

____

Departmental Representative: ______________________________

DALHOUSIE UNIVERSITY

DATE: August 2, 2012

AUTHOR: Scott Harold Melvin

TITLE: Enhancing the Performance of Relay Networks with Network Coding

DEPARTMENT OR SCHOOL: Department of Electrical and Computer Engineering

DEGREE: PhD CONVOCATION: October YEAR: 2012

Signature of Author

To My Parents and Family

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation examines the design and application of network coding (NC) strategies to enhance the performance of communication networks. With its ability to combine information packets from different, previously independent data flows, NC has the potential to improve the throughput, reduce delay and increase the power efficiency of communication systems in ways that have not yet been fully utilized given the current lack of processing power at relay nodes. With these motivations in mind, this dissertation presents three main contributions that employ NC to improve the efficiency of practical communication systems.

First, the integration of NC and erasure coding (EC) is presented in the context of wired networks. While the throughput gains from utilizing NC have been demonstrated, and EC has been shown to be an efficient means of reducing packet loss, these have generally been done independently. This dissertation presents innovative methods to combine these two techniques through cross-layer design methodologies.

Second, three methods to reduce or limit the delay introduced by NC when deployed in networks with asynchronous traffic are developed. Also, a novel opportunistic approach of applying EC for improved data reliability is designed to take advantage of unused opportunities introduced by the delay reduction methods proposed.

Finally, computationally efficient methods for the selection of relay nodes and the assignment of transmit power values to minimize the total transmit power consumed in cooperative relay networks with NC are developed. Adaptive power allocation is utilized to control the formation of the network topology to maximize the efficiency of the NC algorithm.

This dissertation advances the efficient deployment of NC through its integration with other algorithms and techniques in cooperative communication systems within the framework of cross-layer protocol design. The motivation is that to improve the performance of communication systems, relay nodes will need to perform more intelligent processing of data units than traditional routing. The results presented in this work are applicable to both wireless and wired networks with real-time traffic which exist in such systems ranging from cellular and ad-hoc networks to fixed optical networks.

List of Abbreviations and Symbols Used

The list of abbreviations and acronyms used in this dissertation are as follows:

ANC	analog network coding
BER	bit error rate
CDMA	code division multiple access
DF	decode-and-forward
EC	erasure coding
FEC	forward error correction
IP	internet protocol
LT codes	Luby transform codes
MAI	multiple access interference
MUI	multiple user interference
NC	network coding
OFDM	orthogonal frequency division multiplexing
P2P	peer-to-peer
PDF	probability density function
PLR	packet loss rate
QoS	quality of service

RF	radio frequency
RSSI	received signal strength indicator
RV	random variable
SNR	signal-to-noise ratio
UDP	user datagram protocol
WSN	wireless sensor network
XOR	exclusive-or

The list of symbols used in this dissertation are as follows where bold-face letters denote matrices and vectors and scalar variables are in plain lower-case letters.

P_{packet}	power required to transmit a packet to a destination
$\oplus$	exclusive-or operator
k	number of information packets in a systematic erasure coded group
r	number of redundancy packets in a systematic erasure coded group
n	total number of packets in a systematic erasure coded group
e	erasure recovery capability of a code
(n, k, e)	erasure code parameters
m_i	unit of information from the ith information packet
$\mathbf{m}$	vector of m_i drawn from k information packets
$\mathbf{G}$	generator matrix
$\mathbf{I}$	identity matrix
$\mathbb{P}$	parity matrix
$\mathbf{p}$	vector of erasure coded information units
$\mathcal{L}$	length of a packet

$\mathbf{M}$	matrix of information packets
$\mathbf{P}$	matrix of erasure coded packets
b	number of bits in a symbol
$\lfloor \cdot \rfloor$	floor operator
$\mathbf{P}^k$	matrix of any k received erasure coded packets
$\mathbf{G}^k$	$k \times k$ submatrix of $\mathbf{G}$
$(\cdot)^{-1}$	inverse of a matrix
$\binom{n}{k}$	n-choose-k operator
b_i	bit from the ith information packet
PLR_{eff}	effective packet loss rate
PLR_{raw}	raw link packet loss rate
r	distance between nodes
$P_{\text{Rx}}(r)$	received power as a function of r
$P_{\text{Tx}}(r)$	transmit power as a function of r
α	path loss exponent
r_{AB}	distance between nodes A and B
ϕ	angle between lines connecting relay node R to nodes A and B
P_{AB}^{D}	total power for the exchange of a pair of packets between nodes A and B via direct method
$P_{\text{AB}}^{\text{DF}}$	total power for the exchange of a pair of packets between nodes A and B via relay(s) using DF
$P_{\text{AB}}^{\text{NC}}$	total power for the exchange of a pair of packets between nodes A and B via relay(s) using NC
$\max\{\cdot\}$	maximum value selector
p	raw packet loss rate of a link
$P(\cdot)$	Probability of an event

b_i	data packet i
$b_{j,i}$	data packet i from stream j
$b_{j,F}$	EC redundancy packet for stream j
PLR_{corr}	end-to-end packet loss rate after erasure recovery
PLR_{ee}	end-to-end packet loss rate without erasure recovery
M	number of hops between erasure encoding and erasure recovery nodes
$p_{corr}^{M=i}$	end-to-end packet loss rate after erasure recovery when $M = i$
λ	rate parameter
Δt	time interval
a_i	ith packet from node A
b_i	ith packet from node B
c_i	ith network coded packet from node R
Δ	time interval
t_0	arrival time of first packet
N_{NC}	number of network coded packets to wait before flushing
N_P	number of packets to wait before flushing
L	number of relay nodes
ρ	distance from node R to a equispaced point between nodes A and B
$\min_i(\cdot)$	find the minimum value over the range of i
$\mathcal{O}(\cdot)$	operational complexity
$\lvert \cdot \rvert$	magnitude operator

Acknowledgement

I would first like to thank my supervisor, Dr. Jacek Ilow, for his support and many useful suggestions throughout the course of my research at Dalhousie.

I also wish to express my appreciation to Dalhousie University for the financial support it has provided, to the Natural Sciences and Engineering Research Council of Canada for awarding me a three-year postgraduate research scholarship and to the Communication Networks and Services Research (CNSR) project of the Atlantic Innovation Fund for their continued financial support. This work would not have been possible without their support.

I extend my thanks to all of the faculty members who have taught and helped me, particularly Dr. Zhizhang (David) Chen and Dr. William (Bill) Robertson for their insightful comments on my research as the members of my supervisory committee. I also wish to thank Dr. Cheng Li for serving as the external examiner on my examination committee.

I would also like to express my appreciation to members of my research group for their kind help, namely Ali Al-Shaikhi, Mohan Baro, Aasem Alyahya, Craig Jamieson, Harpal Pabla, Hanhua Zhao, and Shreyas Rangappa.

Finally, I would like to thank my parents, Harold and Sheila, and brothers, David and Paul, for their love and support. Without it, this thesis would not have been completed.

Chapter 1

Introduction

When exchanging data between a source and a destination in the traditional model of a communication network, the intermediate nodes operate as decode-and-forward (DF) packet routers that transmit packets over point-to-point links. To overcome the complexity of exchanging data between the nodes at the edge of the network, communication protocols are developed to allow independent data flows to share network resources, enabling their arrival at the final destinations. However, the data flows are treated separately and the intermediate nodes are not permitted to modify the data as it traverses the network. Generally, all functions in existing networks, including error correction, routing and buffering, are based on this unmodified independent data flows assumption [1], [2], [3].

Networks have a finite amount of resources available to execute data transfers between nodes, so finding methods to maximize the use of those resources is key to maximizing the efficiency of the network. Network resources are generally measured using metrics such as available capacity or maximum data throughput, delay, energy consumption, etc [2], [3]. Different metrics may be used depending on the type of network being employed (such as with wired versus wireless networks), but sometimes there is a common trend among the network types (e.g. delay for real-time applications) that requires special consideration when calculating the metric.

From a network resource utilization perspective, treating the various data flows independently is not an optimal solution. Network coding (NC) moves away from the independent data flow model, opting to allow data from various flows to intermix at the various nodes that exist along the transmission path, forming new data flows [4] – [8]. The nodes in the network not only perform DF on the data packets, but they can also combine several of these packets to form new coded packets to be forwarded on. This is accomplished through a process such that the receiver(s) can recover the original data from the delivered packets. This process of combining the packets has the objective of reducing the total number of packets transmitted throughout the network and can be compared to lossless data compression. The goal is to deliver sufficient data to the destination(s) to allow recovery of the original information. However, in the context of NC, this is done at the packet level rather than at the bit level, as is the case in conventional information compression algorithms [9], [10].

The concept behind NC was first introduced in 2000 by R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung in their paper entitled “Network Information Flow” [4]. Since then, NC has seen significant research due to its ability to greatly enhance data throughput, especially in multicast communications. It has been applied in such areas as ad-hoc sensor networks, peer-to-peer (P2P) networks and even in the area of network security [11] – [15].

Network coding was originally applied in wired networks and later extended to wireless applications allowing it to take advantage of the broadcast characteristics of the wireless medium, where a single packet transmission may be received by multiple network nodes when omnidirectional antennas are used [16], [17]. For both of these communication systems, most of the research work in NC assumes synchronous traffic and uniform length data packets. While the second assumption may be easily dealt with, the asynchronous nature of data traffic is a limiting factor in practical applications of NC. This dissertation provides several options to deal with this problem, while adding additional insight to help improve the network's performance.

A similar paradigm shift has been occurring in the area of data protection and reliability. Packet erasure coding (EC), which is a type of forward error correction (FEC) coding, protects against packet loss by adding redundancy packets to a group of data packets. This allows the destination to recover lost packets under certain conditions and reduces the need for retransmission of data. This is ideal for real-time communication applications as it makes the delay experienced by the data packets more predictable. One category of these types of codes are the fountain codes [18], with Luby transform codes (LT codes) [19] and Raptor codes [20] being two practical examples. Traditionally packet EC has been applied at the information source to provide end-to-end data protection and has been shown to improve the reliability when used with multipath routing in ad-hoc networks [21]. More recently it has been applied on a single outgoing data flow at intermediate nodes allowing hop-by-hop data protection [22].

While EC reduces the throughput of the network by adding additional redundancy packets into the data flow, its ability to reduce the raw packet loss rate (PLR) on the communication links, outweigh the additional traffic. When real-time data traffic is involved, retransmission of lost data is not possible, so the application of FEC techniques such as EC is a very desirable option to improve the reliability of the data transmission.

Considering these two paradigm shifts, a motivation of this research is the prospect of applying NC and EC in combination to enhance the performance of relay networks where the performance is primarily measured through the PLR. The primary focus of this work is to integrate NC and EC with the objectives of improving the reliability of data transmissions while controlling the delay experienced by packets as they traverse the network. Concerns with regard to asynchronous traffic are addressed through the use of special scheduling. The results presented in this area of the dissertation are applicable to both wired and wireless networks.

Additionally, in the context of wireless networks utilizing relay forwarding combined with NC, this work incorporates adaptive power control to provide additional energy savings. By reducing the number of packets transmitted by nodes in a wireless network, NC reduces the energy consumption of the network. By allowing nodes in the network to adjust the power they use to broadcast packets, adaptive power control also reduces the energy consumed by the network. Combining these two techniques can lead to even further improvements in energy savings, which motivated addition research presented in this dissertation.

Adaptive power control introduces new issues that must be addressed to efficiently integrate it with NC. In particular, the calculation of routes and determination of which nodes act as relay nodes requires special treatment. Solutions to these issues are presented as part of this dissertation.

Major contributions in this work involve the integration of NC and EC for improved data reliability and reduced packet delay, along with the deployment of NC combined with adaptive power control for greater energy efficiency. It is expected that the contributions outlined in this work will aid in the design of more advanced, intelligent and efficient communication networks. Specifically, in regard to the integration of NC and EC proposed in this dissertation, this work can be applied to areas such as content distribution and multimedia streaming. It will improve the data reliability and increase the data throughput for source to multiple destination type communications.

The work presented in this dissertation on NC with asynchronous traffic and NC with adaptive power control within wireless networks will have implications for wireless sensor networks (WSN) as well as cellular cooperative networks. Such networks will see improvement in their energy efficiency, which translates to increased life expectancy of sensor nodes or handheld devices.

1.1 Dissertation Objectives, Contributions, and Organization

1.1.1 Objectives

The overall objective of this dissertation is to find efficient methods to manipulate packets in a network so as to improve the overall efficiency of the communication network. Novel applications of packet coding techniques are pursued to achieve this goal; specifically, the use of NC in combination with EC is applied to both wired and wireless networks. Further, in the context of wireless networks, adaptive power control is considered in conjunction with NC to achieve improved efficiency.

Numerous performance metrics exist to evaluate the performance of a network. In this dissertation, the specific performance metrics considered are: (i) effective throughput, (ii) PLR, (iii) power saving and (iv) delay. All proposed schemes were compared based on these four metrics. As well, where applicable, the computational complexity of proposed algorithms are discussed so that better insight in to their operation can be obtained.

Previous attempts to integrate NC and EC have been made, but they have generally been limited to very specific topologies or applications. For example, in Chapter 5 of [23], the authors consider a wireless system where a base station is communicating with a series of wireless nodes. Network coding is used to combine the data being sent to the nodes and erasure protection is provided thought the fact that nodes overhear packets destined for neighbouring nodes due to the broadcast nature of the channel and can rebroadcast them if a neighbouring node has discovered an erasure has occurred.

The single stage nature of this process limits the application of this method to a select set of topologies and specific applications. The work presented in this dissertation is different in that despite the fact it is presented for some specific topologies or applications, the algorithms and methods can be generalized to work in nearly any topology or application.

The current state of research in NC does not adequately address the critical issue of asynchronous traffic. Previous work has dealt with this issue by simply buffering the packets so that they could be matched up and then network coded. As demonstrated in this dissertation, this is not an acceptable solution for real-time applications as the buffering delay may be prohibitively high. To reduce this delay, we propose a premature transmission of some of the buffered packets before they have been paired up and network coded. This will cause extra energy to be expended by the network, but at the same time creates opportunities which we will take advantage of.

Adaptive power control and NC in the context of routing is a new area that has been initiated by the author as previous work has only investigated fixed radius transmissions. This work looks at allowing nodes to adaptively change their transmissions radii to adjust the topology of the network to minimize the energy required to exchange data via relay nodes using NC. There has been some past work in the area of relay power allocation for simple data exchanges where orthogonal frequency division multiplexing (OFDM) is used and frequency selective fading channels are a concern [24]. In that case, the authors worked on power allocation at the subcarrier level of the OFDM signals to minimize the total transmit power while maintaining a specific end-to-end bit error rate (BER). The work presented in this dissertation looks at the problem from the packet level and is concerned with the reduction of the total number of packets needed to complete a data exchange, which will reduce the total transmit power of the network regardless of the signaling scheme being employed.

As a result of the increasing complexity of the algorithms used in this dissertation, it becomes increasingly difficult, or even impossible, to use only analytical methods to evaluate the performance of the algorithms. Therefore, a combination of analytical and simulation results are used. All proposed schemes and algorithms were validated and compared though a large number of computer simulations run in the MATLAB® computing environment. This is an acceptable research methodology to solve networking problems and create new knowledge/designs in this type of research.

1.1.2 Contributions

Results of the research described in this dissertation have been published in the form of conference papers [25] – [28]. In addition, several journal publications are currently in preparation. The details of these publications are outlined below.

Refereed Conference Proceeding Publications

[C-1] **S. H. Melvin** and J. Ilow, "Opportunistic network and erasure coding for asynchronous two-way relay networks," in *IEEE Global Communications Conference 2012 (GLOBECOM)*, Anaheim, CA, December 2012, [Accepted].

[C-2] **S. H. Melvin** and J. Ilow, "Relay selection in network coded communications with power control," in *2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, June 2012, pp. 697 - 702.

[C-3] **S. H. Melvin** and J. Ilow, "Equi-spaced relay selection in cooperative communication networks with power control," in *IFIP IEEE Wireless Days 2011*, Niagara Falls, ON, Canada, October 2011, pp. 1 - 3.

[C-4] **S. H. Melvin** and J. Ilow, "Network and erasure coding for improved packet delivery," in *The IEEE 4th International Conference on Signal Processing and*

Communication Systems (ICSPCS), Gold Coast, Australia, December 2010, pp. 1 - 5.

Papers In Preparation to Refereed Journals

[**IPJ-1**] **S. H. Melvin**, A. Alyahya and J. Ilow, "Power allocation in wireless relay networks with packet-level coding," *IEEE J. Select. Areas Commun. (Special Issue on Virtual MIMO)*, In Preparation.

[**IPJ-2**] **S. H. Melvin** and J. Ilow, "Opportunistic network and erasure coding with asynchronous traffic," *IEEE Trans. Wireless Commun.*, In Preparation.

The research described in each of the papers cited above was initiated and carried out by the author of this dissertation, who was also the principal author of the papers.

The research contributions of this dissertation can be classified into three areas, which correspond to the three main chapters of the dissertation. The specific papers and the chapters that correspond to them are listed below.

Chapter 2: **Combining Network Coding and Erasure Coding for Improved Packet Delivery**

The design of a combined NC and EC framework for improved network throughput and data reliability in wired networks, described in Chapter 2, are presented in: [**C-4**];

Chapter 3: **Wireless Relay Networks with Two-Way Asynchronous Traffic**

The methods developed to limit the delay experienced by packets when NC is applied in wireless relay networks where asynchronous traffic is present, along with the addition of opportunistic EC methods to provide improved data reliability, described in Chapter 3, are presented in: [**C-1**] and [**IPJ-2**];

Chapter 4: **Optimum Selection of Relays via Adaptive Power Control in Network Coded Relay Networks**

The development and analysis of an efficient, near optimal algorithm for the selection of relay nodes and adaptive broadcast power allocation to minimize the total transmit power in network coded wireless relay networks, described in Chapter 4, are presented in: [**C-2**], [**C-3**] and [**IPJ-1**];

1.1.3 Organization

The organization of the chapters of this dissertation are briefly reviewed below.

Chapter 1

This chapter begins with an overview of the research work presented in this dissertation. It then outlines the objectives, contributions, and organization of the dissertation in Section 1.1. The remainder of this chapter is dedicated to a review of the general concepts and elements used throughout the remainder of this dissertation. Section 1.2 reviews the concepts behind NC for both wired and wireless networks. Section 1.3 discusses the concept of FEC and specifically the use of EC for FEC. Section 1.4 provides an overview of power control issues as they relate to wireless networks and what impact they have on NC. The chapter concludes with Section 1.5 providing a brief summary of the material presented in the chapter.

Chapter 2

In the second chapter a framework for the integration of EC with NC is developed. This chapter begins with an analysis of the effect adding a PLR to all network links has on the performance of NC in Section 2.1. Two schemes which integrate EC into the NC framework to improve the reliability of NC are developed in Section 2.2. Analytical tools to evaluate the performance of the proposed combined NC and EC framework and elaboration on which relay nodes should be involved in the EC functions are discussed in Section 2.3. The performance of these methods in terms of their reliability are presented in Section 2.4. The chapter concludes with a summary of the results in Section 2.5.

Chapter 3

In the third chapter the use of NC in wireless relay networks with bi-directional asynchronous traffic is explored. A discussion of the delay introduced as a result of using NC is presented in Section 3.1. This is followed by an analysis of that delay in Section 3.2. Section 3.3 outlines several methods proposed by this dissertation to limit the amount of delay experienced by packets as a result of using NC. As a by-product of the proposed delay limiting methods, there exists some underutilized capacity in the network. How these oppertunities arises and how this work exploits them in a novel way to improve the data reliability of the network is discussed in Section 3.4. Simulation results for the proposed methods are outlined in Section 3.5 and a summary of the work presented in this chapter is given in Section 3.6.

Chapter 4

In the fourth chapter the author develops a computational efficient relay node selection algorithm employing adaptive power control which provides near optimal minimization of total transmit power in a wireless relay network. Section 4.1 provides some additional background material on distance determination and power allocation that is necessary for the further discussion in this chapter. A geometric analysis of the optimum relay node positioning is performed in Section 4.2, which leads to the development of a computationally efficient selection algorithm which is presented in Section 4.3. In Section 4.4, validation of the selection of equi-spaced relay nodes for minimization of total transmit power is performed and results showing the performance of the selection algorithm and its improvement over an exhaustive search method are presented. The chapter concludes with a summary of the results in Section 4.5.

Chapter 5

The fifth and concluding chapter summarizes the major contributions in this dissertation and outlines the impact of the approaches, algorithms and analysis developed as part of this work. It also offers suggestions for future work that could be carried out to expand upon the results presented in this dissertation.

1.2 Network Coding

One can identify two fundamental approaches to network coding: packet level NC and analog network coding (ANC). In packet level NC, routers mix (or code) packets that belong to different data flows that intersect at intermediate points in the network. The coding is done using bitwise operations to combine packets from the various flows into a single coded packet.

Analog network coding, as the name implies, operates on the analog signals rather than packets [29], [30], [31]. If two or more packets are transmitted simultaneously, the resulting interference of the signals representing those packets is a linear combination of those signals. The router merely has to rebroadcast this interference signal and the originating nodes can recover the information using one of the available interference cancelation schemes. In ANC, it is the operations on continuous signals that brings the benefits of "mixed" information transmission and removes the intermediate router's need to decode the data packets and perform the "mixing".

The focus in this work is exclusively on packet level NC, which will be explained first in the context of wireless networks and then in the context of wired network, in Sections 1.2.1 and 1.2.2, respectively.

1.2.1 Wireless Network Coding

Wireless network coding is primarily used in the wireless networking environment where nodes communicate over wireless links by broadcasting data to their neighbours. Most wireless networking devices have a limited communication range dictated by the resources they have available (i.e., battery capacity and transmitter/receiver design) or by device longevity requirements (i.e., conservation of battery life). This is especially true in the case of WSN where nodes are generally small in size, have limited battery capacity and users desire as great an interval as possible between replacement.

As a result of these communication limitations and no fixed or predefined topology, there is a high probability that a random pair of nodes that wish to communicate with each other would not be within each other's range. To allow these nodes to communicate, an intermediate node would have to act as relay device, passing data back and forth between the pair of nodes. It may also be the case that one relay node may be insufficient to create a transmission path between the pair of nodes, requiring several intermediate relay nodes to assist.

In this type of cooperative relay network, energy conservation and a fair selection of which nodes act as relays are essential. As such, wireless NC is applied at the relay nodes to reduce the number of transmissions made by the nodes, thus conserving energy resources and improving data efficiency. We will now look at how this can be achieved.

An example of a three node wireless relay network is shown schematically in Fig. 1.1 a). A and B are two nodes that wish to exchange data packets a and b via the relay node R. In the DF relay strategy, the exchange process proceeds with nodes A and B first broadcasting their packets to node R. Node R then rebroadcasts packet a and subsequently, packet b, which completes the transaction. The order in which A and B transmit to node R and the order in which node R rebroadcasts the packets does not have an impact on the power consumption of the network, but does impact the delay experienced by the packets.

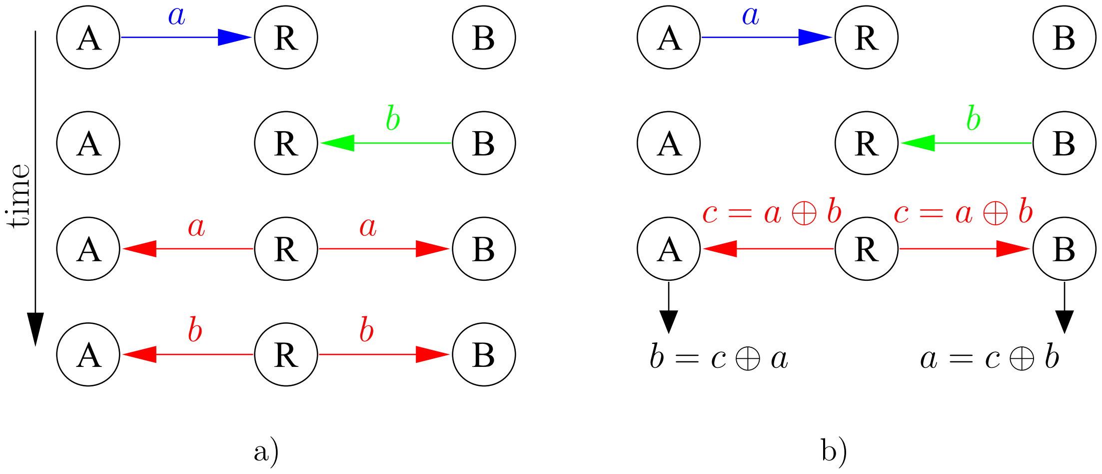

Figure 1.1: DF vs. NC deployment of a relay node

Turning our attention to the application of NC, Fig. 1.1 b) shows the same three node network now deploying wireless NC. Instead of using two separate transmissions to send packets a and b, NC allows the relay node R to encode the two received packets and broadcasts a single, exclusive ORed (XORed) packet $c = a \oplus b$ [4]. Both nodes A and B can recover packets b and a from the c packet by XORing c with their own packets a and b respectively.

Comparing the number of transmissions required to complete the data exchange between nodes A and B, one can observe that for the DF relay strategy case in Fig. 1.1 a) a total of four transmissions is required. This number is reduced to three when NC is applied as shown in Fig. 1.1 b). Assuming that all nodes transmit with the same power P_{packet} and this is sufficient to guarantee reliable reception, then for the DF case, the total power consumption is $4 \cdot P_{\text{packet}}$. For the NC case, this drops to three transmissions and a power consumption of $3 \cdot P_{\text{packet}}$, which is a 25% savings when looking at the total network power consumption, and a 50% savings when looking solely at node R's power consumption. This is the theoretical maximum savings that could be obtained, as in reality the savings will vary depending on the node interdistances.

In the example shown in Fig. 1.1, it was assumed that the data traffic was synchronous. All nodes broadcast at predetermined times or in a predetermined sequence. Because there is generally no centralized control in this type of network, the assumption of synchronous traffic is unrealistic. Nodes A and B are more likely to broadcast at random intervals when they have data available to transmit. The resulting asynchronous traffic can lead to several issues that must be resolved.

The first is the potential for data collisions resulting in transmission errors, as broadcasts from two or more nodes have the potential to overlap. This has an impact on both the DF and NC relay networks. We will not examine this issue as there are a number of methods to mitigate this. Allowing each node to use a different frequency for transmitting would prevent collisions caused by asynchronous traffic is but one of them [32], [33], [34]. This is shown schematically in Fig. 1.1 by the different colours of the arrows that represent the various transmissions.

The second issue that arises as a result of asynchronous data traffic is delay. This is not a concern for the DF network shown in Fig. 1.1 a) as the relay nodes simply forward on packets once they arrive at the relay. For the NC network, since the relay node in Fig. 1.1 b) must wait until it has data from both sides before it can perform the coding, it may need to buffer some packets before they can be network coded, and this will introduce a buffering delay to some packets. This additional delay can be significant, making this a prime motivator for the work in this dissertation and a topic of extensive investigation.

To the best of the author's knowledge, all research on network coding at the packet level has assumed synchronicity of the data, as it simplifies the system analysis and avoids the introduction of random delay as described above. There are some published results dealing with asynchronous arrival for ANC, but they are not related to the work in this dissertation as they work with time variations determined by propagation delays in the wireless medium [35], [36]. Balancing the benefits of NC (increased throughput and energy conservation) and the drawbacks (increased packet delay) is evaluated in Chapter 3 of this work.

1.2.2 Wired Network Coding

Wired NC is the counterpart to wireless NC and operates under the same principles. Because it specifically works in the wired network environment, a number of the concerns that exist in the wireless environment are not present. Since nodes are connected by fixed wired links, the topology of the network is most often fixed. Nodes will likely be capable of performing more complex computations and routing, and are not concerned with power conservation, as the electric grid, as opposed to batteries, power them.

While these were some of the prime motivations for the application of NC in wireless networks, the wired network can still benefit from NC. The data compression that it generates by combining packets together increases network throughput and reduces the bottlenecks that exist, reducing packet delay [37], [38]. This is clearly demonstrated in the following example.

A standard model of a wired network that can benefit from NC is the so-named butterfly network. It typically consists of seven nodes with the configuration shown in Fig. 1.2. In this example there is one source node H, four intermediate nodes A, B, C, D and two destination nodes X and Y.

Alternatively, the source H can be removed to make nodes A and B sources, as

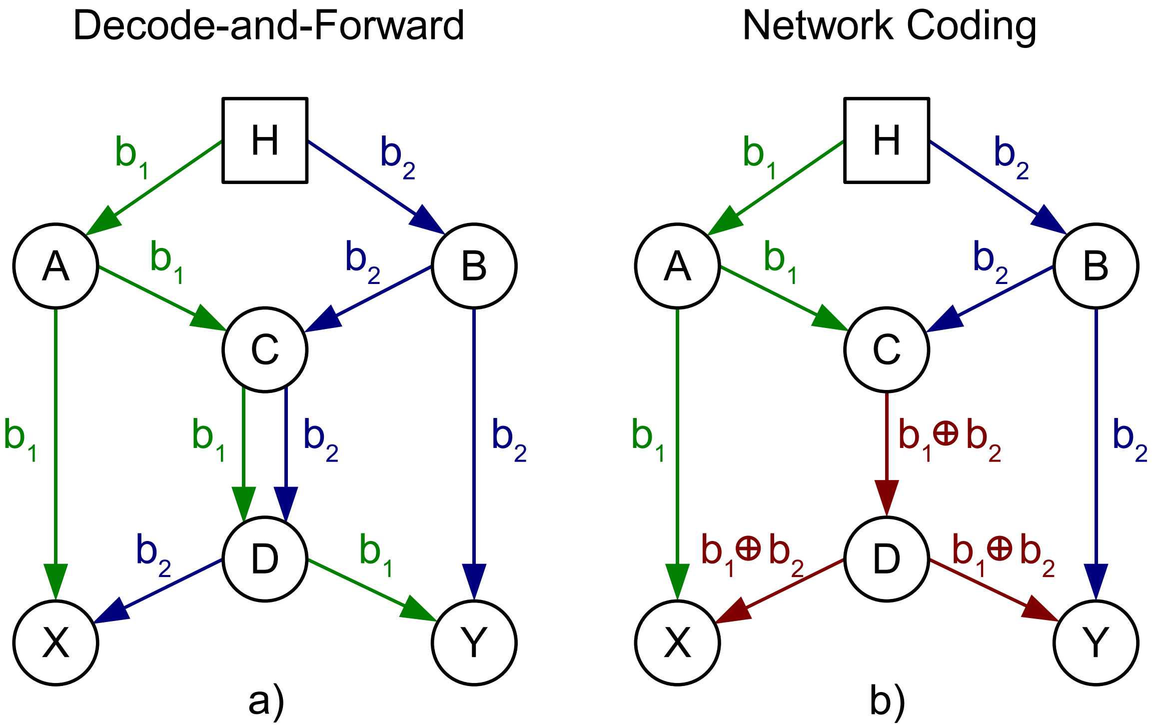

Figure 1.2: Example of DF and NC on a butterfly network

it does not change the routing dilemma that exists. As a result, the discussion here can apply to a single or multi source network in which multicast is desired.

Figure 1.2 depicts two butterfly networks with the network on the left using DF and the network on the right using NC. In both cases, the source H needs to send two data packets, b_1 and b_2 to the destination nodes X and Y. The difference with these two networks is that the channel capacity of the link CD is 2 for the DF case and 1 for the NC case.

For the DF network, in the first stage, H sends b_1 to node A and b_2 to node B. During the second stage, nodes A and B forward their information to sink nodes X and Y respectively, as well as to node C. During stage three, since the capacity of the channel CD is 2, illustrated here by the two lines between nodes C and D, node C can send b_1 and b_2 simultaneously to node D. Finally, during the last stage of transmission, node D sends b_1 to node Y and b_2 to node X and the transmission is complete. Both sink nodes have received both information packets with the same delay.

The issue with DF arises if we reduce the capacity of the link CD to 1. Now, when node C goes to forward the data on to node D, it can only send one packet at a time. It must forward b_1 and store b_2 to be forwarded during the next stage. As a result, node Y receives b_1 at the same time as in the previous case, but node X must wait one more stage before it receives the packet b_2. As a result, the throughput of the network is decreased due to the limit of the channel CD.

For the NC case, however, the fact that the capacity of the channel CD is 1 does not reduce the networks throughput as it does in the DF case. Node C, which has received both packets b_1 and b_2, can form a new packet out of these two packets, $b_1 \oplus b_2$, and transmit this new packet to node D that, in turn, transmits this packet on to the two destination nodes. When this new packet arrives at the destination nodes X and Y, given they both have one of the original packets already, they can recover the second packet from the $b_1 \oplus b_2$ packet. As a result, the same throughput is achieved as in the traditional routing scenario, but now with link CD having a capacity of 1.

It is this ability of NC to improve the throughput of data when limited link capacity exists that has motivated research into wired NC. Regrettably the bulk of research done in this area has assumed no loss on the wired links in the network, resulting in the interest in determining how NC performs in the presence of error prone links [39]. This dissertation therefore examines the affect of adding packet losses to the links in the network and the impact this has on the reliability of the NC.

It should be noted that regardless of whether we are discussing wireless or wired NC, this dissertation assumes we are working with data packets at the forwarding plane level. This is equivalent to working with routed protocols, such as the User

Datagram Protocol (UDP) in the Internet Protocol (IP) stack [2], [3]. As a result, decisions on which nodes are involved in forwarding packets and which perform NC are based on the network topology learned from routing protocols and are therefore not considered in this dissertation [2], [40].

1.3 Erasure Coding

Erasure coding is a technique that has conventionally been applied at the information source, adding redundancy to the data being sent so as to improve the reliability of the transmission. This is accomplished at the packet level using linear block codes by grouping k information packets and $r = n - k$ redundancy packets into a coded packet group of n packets. Every EC technique has an associated erasure recovery capability represented by e, and it is the number of packets in the coded packet block that the code can guarantee to be recovered. The values of n, k and e, which are normally given as (n, k, e), are specified by the erasure code being employed [41].

One method to create systematic erasure codes is via the use of a generator matrix like that of the following. A unit of information, either a bit or a symbol, $m_i, i = 1, \cdots, k$, is taken from each of the k information packets to form a column vector $\mathbf{m}$. The r parity symbols are constructed with the help of the generator matrix $\mathbf{G}$. These are then grouped back in to r redundancy packets and transmitted along with the original k information packets. This operation is expressed mathematically by the linear system of equations:

$$\mathbf{G} \cdot \mathbf{m} \quad = \quad \mathbf{p} \; . \tag{1.1}$$

Here $\mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_k \end{bmatrix}$ is the column vector of k information symbols from k information packets, $\mathbf{G} = \begin{bmatrix} \mathbf{I} \\ \cdots \\ \mathbb{P} \end{bmatrix}$ is the $n \times k$ generator matrix of the systematic code considered, $\mathbf{I}$ is the $k \times k$ identity matrix, $\mathbb{P}$ is the $r \times k$ parity matrix and $\mathbf{p}$ is the column vector of n coded symbols.

The assumption used in this dissertation is that all packets in any coded group are of the same length $\mathcal{L}$. This limitation can be easily overcome by padding the packets to the same length. To simplify the notation, Eqn. (1.1) can be rewritten using the packet version of this relation as:

$$\mathbf{G} \cdot \mathbf{M} \quad = \quad \mathbf{P} \ , \tag{1.2}$$

where $\mathbf{M} = \begin{bmatrix} M_0 \\ M_1 \\ \vdots \\ M_{k-1} \end{bmatrix}$ is a matrix of information packets arranged in rows, each with $\left\lfloor \frac{\mathcal{L}}{b} \right\rfloor + 1$ symbol elements, b is the number of bits per symbol, $\mathbf{P}$ is the corresponding matrix of coded packets, and $\lfloor \cdot \rfloor$ denotes the floor operation.

If there are any lost information packets at the receiver, it must solve for $\mathbf{M}$ the following system of equations:

$$\mathbf{G}^k \cdot \mathbf{M} \quad = \quad \mathbf{P}^k \tag{1.3}$$

or equivalently determine:

$$\mathbf{M} \quad = \quad \left(\mathbf{G}^k\right)^{-1} \cdot \mathbf{P}^k \ , \tag{1.4}$$

where $\mathbf{P}^k$ is the vector of any k received packets, $\mathbf{G}^k$ is a $k \times k$ sub-matrix of $\mathbf{G}$ with rows corresponding to the k received packets as determined by the received packets sequence numbers, and $(\cdot)^{-1}$ represents the matrix inversion operation. This matrix inversion can be achieved using techniques based on Cramer's rule, Gaussian elimination or Gaussian Jordan elimination methods [42], [43], [44].

A challenge in the design of an erasure code is to design $\mathbf{G}$, or in the case of systematic codes, the corresponding coefficient matrix $\mathbb{P}$. The matrix $\mathbf{G}$ should be designed such that any $k \times k$ sub-matrix, $\mathbf{G}^k$, will be invertible (full rank). The total number of such sub-matrices is $\binom{n}{k}$. For the work in this dissertation this is not a concern as there are a number of well documented codes in the literature that could be deployed here, and this work uses one type of code meeting this requirement.

In this dissertation, a parity check code is used for the erasure coding and we define the unit of information we operate on, m_i, as a bit. This means that only one redundancy packet is created and the parity matrix $\mathbb{P}$ in Eqn. (1.1) is a row vector of all 1's. Assuming that a group of $k = 3$ information packets are to be erasure coded using this method, Eqn. (1.1) can be written as:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ p_1 \end{bmatrix}, \tag{1.5}$$

where b_i is a bit from the ith information packet and p_1 is the redundancy bit created. The left side of Fig. 1.3 expands this operation to the packet level and demonstrates how the redundancy packet can be generated though a bit-wise XORing operation.

The right side of Fig. 1.3 demonstrates that to recover an erased packet it is a simple matter of bit-wise XORing the remaining k received packets together to recreate the missing packet. This is a simplistic form of EC and it only offers a limited

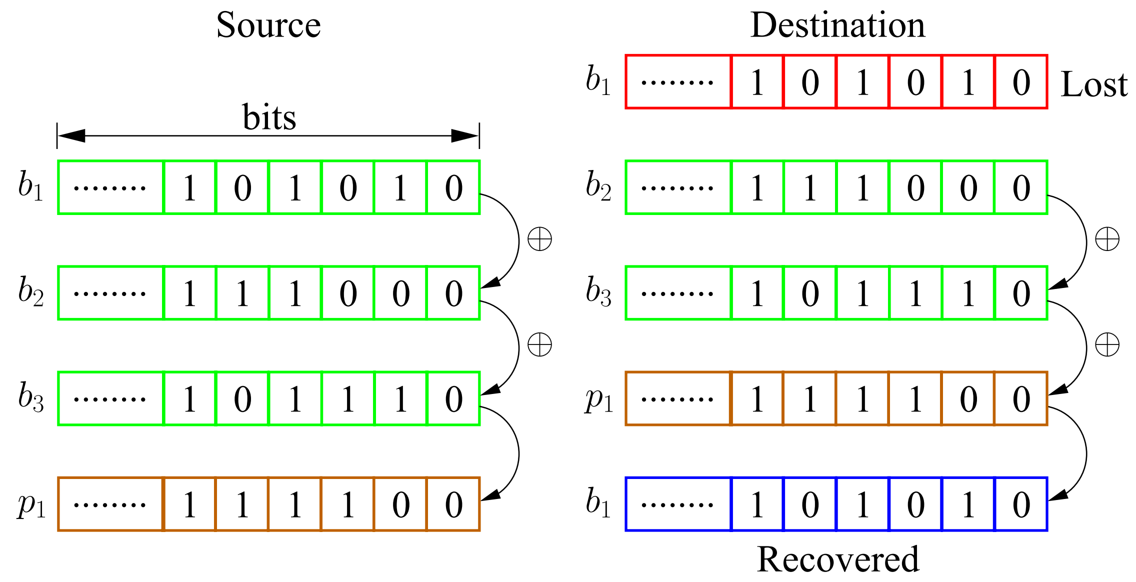

Figure 1.3: Simple erasure packet coding by XORing

amount of protection to the data packets being transmitted. But its simplicity lends itself well to ease of implementation, and makes it compatible with the XOR nature of NC.

This single packet recovery mimics the single bit error correction capability of the Hamming code [41]. The same formula used to calculate the effective bit error rate after correction can therefore be applied to calculate the effective packet loss rate (PLR_{eff}) after erasure recovery. For this specific case, the formula is:

$$PLR_{\text{eff}} = \left[PLR_{\text{raw}} - PLR_{\text{raw}} \cdot (1 - PLR_{\text{raw}})^{(n-1)}\right] , \tag{1.6}$$

where $n = k + 1$, k is the number of packets being protected by the single redundancy packet and PLR_{raw} is the raw packet loss rate.

1.4 Adaptive Power Control in Wireless Networks

Allowing nodes in a wireless network to adjust their transmission power is important so as to guarantee reliable connectivity between the nodes with the minimum power invested. Since average received signal power is inversely proportional to the distance between the nodes, the power they must transmit with is proportional to that distance. Additionally, given a fixed noise power level, nodes must adjust their transmit power to achieve a high enough signal-to-noise ratio (SNR) to maintain an acceptable BER. Adaptive power control can also be used to manage spatial reuse of the spectrum. An example would be in code division multiple access (CDMA) systems where power control is used to overcome near-far problems where, by controlling the power of simultaneously transmitting users, one can not only assure that the received signal of interest is strong enough but also the resulting multiple user interference (MUI) is sufficiently suppressed.

The work in Chapter 4 of this dissertation is primarily concerned with the use of adaptive power control to minimize the transmission power used by all nodes to reduce the overall energy consumption of the network. This is a concern as generally most wireless networking devices are powered by finite energy sources, so a reduction in transmission power would translate into extended device life. Time based adaptive power control can also be used to control the topology of the network so relay paths that consume the least amount of power are formed.

Given that data transmissions are generally the most power consuming functions of a wireless node, the wireless channel the system is operating in determines the overall energy performance of a device. We will therefore examine the transmit power requirements as dictated by the deterministic radio frequency (RF) signal propagation and summarize how this dissertation will address them.

1.4.1 RF Signal Propagation

From the statistical RF propagation models, propagation over a radio channel between a transmitter and a receiver can generally be decomposed into three independent, multiplicative propagation phenomena. The first is a path loss component, which is the result of the attenuation of the signal due to the distance separating the transmitter and receiver. The second is a shadowing component caused by objects blocking the propagation path of the signal, and the third is a multipath component resulting from the signal arriving at the receiver over several independent paths [45]. As we are dealing with relatively short distances and relatively slow mobility of nodes, the last two effects will be minimal compared to the path loss component and are not considered in this work.

The path loss model states that the receive power, $P_{\mathrm{Rx}}(r)$ is a decreasing function of the distance r between the transmitter and the receiver, and can be represented by:

$$P_{\mathrm{Rx}}(r) \sim \frac{P_{\mathrm{Tx}}}{r^{\alpha}} \ , \tag{1.7}$$

where P_{Tx} is the power of the signal at the transmitter and α is the path loss exponent used in the deterministic path loss model [46]. The parameter α depends on the environment and can vary between 2 (in free space) and 6 (in heavily built-up urban areas).

In this dissertation, we assume that to have an acceptable link performance, the received SNR should be controlled through the transmit power. Based on Eqn. (1.7), the transmit power should therefore be given by:

$$P_{\mathrm{Tx}}(r) \sim r^{\alpha} \ . \tag{1.8}$$

We also assume: (i) reciprocal propagation conditions, (ii) use of omnidirectional antennas and (iii) homogenous propagation conditions. These assumptions simplify the notation later in this dissertation, however, careful consideration once all ideas are clarified will demonstrate that they are not necessary conditions, as the proposed routing algorithms could be based on the actual power estimates. Finally we assume that time (or frequency) division based orthogonal channels are utilized in relaying, so we do not consider multiple access interference (MAI).

The routing protocols investigated in this dissertation are distance-based, where the wireless nodes have a database of inter-distances between each other. These distances could be obtained from power estimates in the neighbour discovery phase and/or from the network control plane via periodic transmission of beacon packets [47]. Precise power measurements are most likely not needed as any practical system implementation will use discrete transmit power increments. More details on this are provided in Section 4.1.1

1.4.2 Relay Node Strategies and Power Allocation

In a cooperative wireless network, there are two fundamental questions to be asked when attempting to determine the power efficiency of the network. They are: (i) which node(s) should act as a relay? and, (ii) What relay strategy with what corresponding power allocation should be used? The remainder of this section will explore these two questions for the cases of DF and NC, setting the ground work for the contributions of this dissertation outlined in Chapter 4.

Figure 1.4 illustrates the three node cooperative relay network that was discussed in Section 1.2.1. The radio propagation ranges of nodes A and B, indicated by the concentric circles centered at the nodes, are not large enough for direct communication to occur. To establish communication between A and B, the network will use a relay node R that is within the coverage area of both A and B, and in this case, equidistant from nodes A and B.

Having node R equidistant from nodes A and B is the best case scenario and provides the best performance enhancement when the various relay node strategies

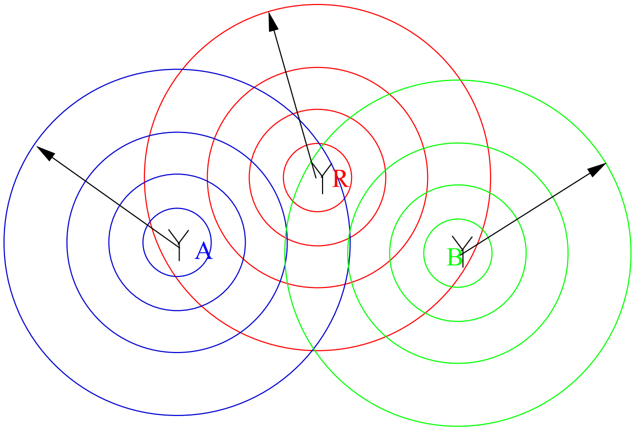

Figure 1.4: Transmission ranges in wireless networks with a single relay node

are applied. To better demonstrate the performance gains achieved using the various strategies, a more realistic network topology will be used. Figure 1.5 shows the same three nodes in a different configuration. Nodes A and B are separated from one another by a distance r_{AB}, while they are each separated from the relay node R by distances r_{AR} and r_{BR}, respectively.

We assume here a free space propagation environment ($\alpha = 2$), so that if A and B were communicating directly, to exchange data between them they would use the total power of:

$$P_{\mathrm{AB}}^{\mathrm{D}} \sim 2 \cdot r_{AB}^{2} \tag{1.9}$$

considering the link reciprocity and with the power control guaranteeing an acceptable BER performance as outlined in Section 1.4.1. If the nodes are transmitting via node R using the DF strategy, the total power used for the same data exchange in both directions would be:

$$P_{\mathrm{AB}}^{\mathrm{DF}} \sim 2 \cdot \left(r_{AR}^{2} + r_{BR}^{2}\right) . \tag{1.10}$$

Node R would adjust its power so that it only used the appropriate amount to reliably reach the destination of the packet it was transmitting (i.e. r_{AR}^{2} when broadcasting to node A and r_{BR}^{2} when broadcasting to node B).

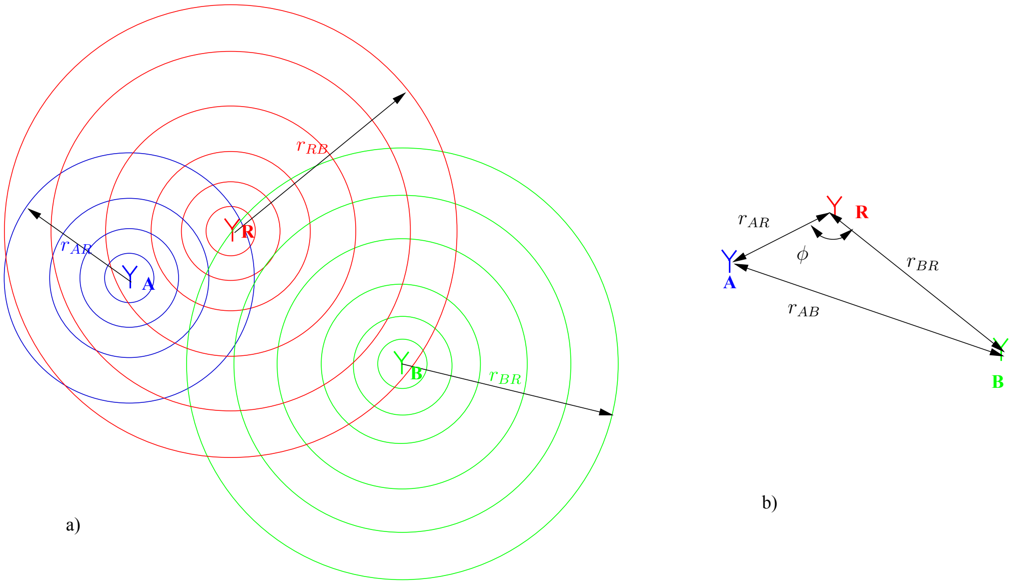

Figure 1.5: Transmission ranges in a system with a single relay node: a) power adjustment and reachability; b) interdistance relations

Considering the topology in Fig. 1.5 b) corresponding to the single relay node scenario, based on the law-of-cosine:

$$r_{AB}^2 = r_{AR}^2 + r_{BR}^2 - 2 \cdot r_{AR} \cdot r_{BR} \cdot \cos(\phi) \tag{1.11}$$

one can conclude that as long as $\frac{\pi}{2} < \phi < \frac{3\pi}{2}$, the power of the DF methods will be smaller than the power of the direct transmission, i.e., $P_{\mathrm{AB}}^{DF} < P_{\mathrm{AB}}^{D}$.

To analyze the power required to exchange data between nodes A and B using NC, we refer back to Figs. 1.1 b) and 1.5 b). From these figures, it is evident that for node A to send packet a to the relay node R, the invested power is $\sim r_{AR}^2$. Similarly, to send packet b from node B to the relay node R, the invested power by node B is $\sim r_{BR}^2$. However, the invested power for the relay node R to broadcast packet c must be $\max\{r_{AR}^2, r_{BR}^2\}$. This is to ensure both nodes A and B reliably receive packet c. Therefore, the total power invested for the bidirectional exchange using NC is:

$$P_{AB}^{NC} \sim r_{AR}^2 + r_{BR}^2 + \max\{r_{AR}^2, r_{BR}^2\} \ . \tag{1.12}$$

1.5 Summary

This chapter has presented an overview of the research work to be presented in this dissertation. It additionally outlined the objectives, contributions, and organization of the dissertation. A review of the general concepts and elements used throughout the dissertation was then provided. This included a review of both wireless and wired NC, the concept of FEC (specifically the use of EC for FEC) and an overview of power control issues in wireless networks and what impact they have on NC.

Chapter 2

Combining Network Coding and Erasure Coding for Improved Packet Delivery

Network coding and erasure coding share a common principle of encoding incoming information packets at network intermediate nodes so that destination nodes can reconstruct the original information packets using a sufficient number of encoded packets collected at the destinations. While NC generally reduces the number of packets sent through the network, EC introduces redundancy packets into the network to aid in the recovery from lost packets. This chapter develops two schemes for integrating NC and EC with the purpose of balancing throughput and PLR requirements [28].

One area of this research which has seen great interest is determining how NC performs in the presence of error prone links. To that end, this chapter first examines in Section 2.1 the effects that the raw PLR in the network has on the performance of conventional NC algorithms. It then considers the addition of EC into the NC framework to allow for improved packet recovery by contributing two schemes outlined in Section 2.2. This strengthens the network codes against the raw packet losses as evaluated first analytically in Section 2.3 and then via simulations in Section 2.4.

The novelty of the research in this chapter is in applying the EC concepts in the context of NC by not allowing the accumulation of lost packets in the network beyond the erasure recovery capability of the EC codes deployed at selected nodes.

2.1 Effect of Link Errors

In the introduction to the basic concepts of NC presented in Chapter 1, this dissertation assumed that all point-to-point links in networks were lossless and that all packets arrive at their destinations after every transmission. Unfortunately this assumption does not hold true for most real world applications as most relays drop some packets. This can be a result of congestion in the network which can lead to buffer overflows, the rejection of packets at lower layers due to bit errors caused by a high BER or by rejection of packets due to late arrival when real-time applications are involved.

This section will examine what effect adding a PLR to every point-to-point link in the standard butterfly network with NC, presented in Section 1.2.2, has on the delivery of the two streams of data packets being sent by the source node. This is with the purpose of motivating the designs presented in the Section 2.2.

2.1.1 Adding a Packet Loss Rate to all Links

For this discussion it is assumed that all point-to-point links have the same raw PLR and we denote this by p. The values p can take on are in the range 0 to 1 and represents the probability of a packet being lost. We also define $q = (1 - p)$ to be the probability a packet is successfully transmitted over a link.

Referring back to the network coded butterfly network in Fig. 1.2, the probabilities of data packets b_1 and b_2 failing to arrive at node X will be derived. By virtue of the symmetry of the network, it can be seen that the same analysis presented here is also valid for node Y.

First, for the data packet b_1 to arrive it must be transmitted without error over links HA and AX. Since these links have an equal PLR, and losses on the links are independent, the probability of data packet b_1 arriving at node X is given by:

$$P(b_1 \textit{ arrives at node } X) = (1-p)\cdot(1-p) = q \cdot q = q^2 \tag{2.1}$$

and conversely, the probability that data packet b_1 fails to arrive at node X is given by:

$$P(b_1 \textit{ does not arrive at node } X) = 1 - q^2 \ . \tag{2.2}$$

Now turning to data packet b_2, it is obvious that the data packet must first arrive at node C and must therefore be successfully transmitted over links HB and BC. Now there are two possibilities that can occur at this node: 1) Data packet b_1 arrives at node C at the same time as data packet b_2 or 2) Data packet b_1 does not arrive at node C.

If the first scenario occurs then node C will XOR packets b_1 and b_2 and transmit $b_1 \oplus b_2$. As a result, for node X to receive data packet b_2, $b_1 \oplus b_2$ must be successfully transmitted over links CD and DX and b_1 must be transmitted error free over link AX. For the second scenario, node C will simply forward data packet b_2 on to node D and thus for node X to receive data packet b_2 it must be successfully transferred over links CD and DX.

Using similar logic as before, the mathematical expression for the probability of data packet b_2 arriving at node X is given by:

$$\begin{aligned} & P(b_2 \textit{ arrives at node } X) \\ = \; & \{(1-p)\cdot(1-p)\}\cdot\{(1-p)\cdot(1-p)\cdot(1-p)\cdot(1-p)\cdot(1-p)\} \\ & +\{1-(1-p)\cdot(1-p)\}\cdot\{(1-p)\cdot(1-p)\cdot(1-p)\cdot(1-p)\} \qquad (2.3) \\ = \; & (q\cdot q)\cdot(q\cdot q\cdot q\cdot q\cdot q)+(1-q\cdot q)\cdot(q\cdot q\cdot q\cdot q) \\ = \; & q^7+q^4-q^6 \end{aligned}$$

and conversely, the probability that data packet b_2 fails to arrive at node X is given by:

$$P(b_2 \textit{ does not arrive at node } X) = 1-(q^7+q^4-q^6)\ . \qquad (2.4)$$

As an example, for a PLR of 5% or $p = 0.05$, Eqns. (2.2) and (2.4) predict that the probabilities of data packets b_1 and b_2 failing to arrive at node X are 9.75% and 22.22% respectively.

While NC has improved the throughput of the network, there is a cost when it comes to reliable transmission of the data to the destination nodes. For this example, if one were to simply transmit data packet b_2 from node H to node X along the same path but without any NC then the probability of data packet b_2 being lost would be given by $1-q^4$ which, for the previous example's parameters, would result in a value of 18.55% or a 3.67% improvement over using NC. This does not appear significant here, but this is a simple network example with only two data packets being network coded together. If more data packets were allowed to be combined, this degradation would increase significantly.

The next section examines how EC can be integrated with NC to improve its performance in these error prone networks.

2.2 Combining Network and Erasure Coding

In Sections 1.2 and 1.3, a brief overview of a linear network coding scheme and an erasure code that protects a group of data packets from single packet erasures were presented. Upon examining these two operations it can be seen that they have essentially the same implementation via XORing and it is only the interpretation of what is being accomplished that distinguishes them. It is this observation that lead to looking at the combining of these two operations to improve the overall reliability of NC.

Two implementations that were developed for combining NC and EC are presented next.

2.2.1 Scheme I

To simplify this discussion, it will initially be assumed that the network is a lossless one. Referring back to the right side of the diagram in Fig. 1.2 for the structure of the Butterfly network, source node H generates two data packet streams $b_{1,i}$ and $b_{2,i}$ which are transmitted on its two output links and where i represents the index of the packet. Node C receives packets $b_{1,i}$ and $b_{2,i}$ on its two input links at every time instance (a result of the lossless network assumption). It then network codes these two packets into a single coded packet and sends it out.

This is shown schematically in Fig. 2.1 and represents the basic building block of NC where a single relay merges two incoming flows into a single network coded outgoing flow. Even though specific topologies are to develop the work in this chapter, the algorithms developed are general enough to be applied to any topology as one can break more complex topologies down into smaller elements similar to the building block shown in Fig. 2.1.

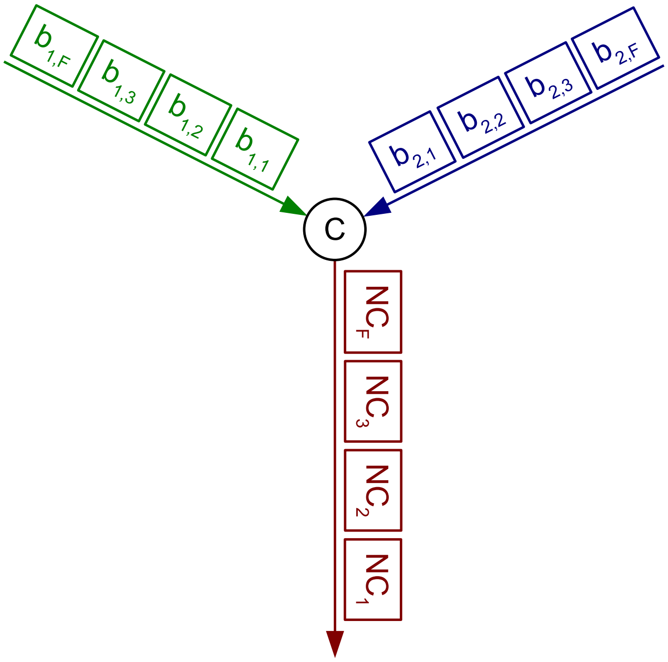

Figure 2.1: Node C of the butterfly network

If it is further assumed that the source node H applies the EC described in Section 1.3 to the two data streams it is generating, then node C will be receiving packets from two groups of erasure coded packets in sequence and NC them together. The advantage here is that the group of network coded packets which were generated from the combination of two groups of erasure coded packets are also erasure coded and protected from single packet erasures. An example of this is shown in Fig. 2.2.

It can be easily shown that if any one of the four network coded packets shown on the right side of Fig. 2.2 is lost, it can be recovered by simply bit-wise XORing the remaining three packets and hence, without any additional processing, node C has erasure coded its network coded packets.

As this result is based on the assumption of a lossless network, it is not very realistic. This can be addressed by requiring node C to buffer all the packets from the two data streams, perform any erasure recovery operations on the erasure coded

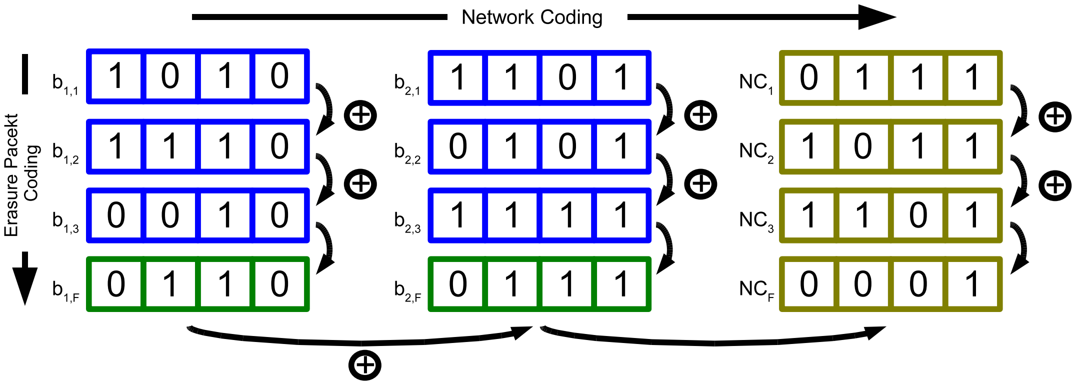

Figure 2.2: Combining erasure coding and network coding - example 1

groups, and then network code the two streams into one. The disadvantage to this method is that a delay in delivery of the packets to the destination is introduced and the nodes need to buffer the data packets. The latter issue is already a necessity if the nodes are expected to recover from any packet erasures.

2.2.2 Scheme II

A more realistic scenario is to allow the NC nodes to combine packets 'on the fly' and deal with the recovery of packets as needed. This results in packets being transmitted out of order, so the NC of packets may not necessarily combine all the packets from an erasure coded group in one stream with all the packets from an erasure coded group in the second stream, as was the case in Fig. 2.2.

This reordering of the data packets does not necessarily break the EC of the network coded packets. As long as all the packets from two erasure coded groups are recovered and network coded together, the EC of the network coded packets holds. An example of this is shown in Fig. 2.3.

In Fig. 2.3, packets $b_{1,3}$ and $b_{2,1}$ are lost. Instead of waiting for all the data packets to arrive and recovering these lost packets, the node simply forwards the received packet (equivalent to XORing with an all zero packet). The resulting network coded

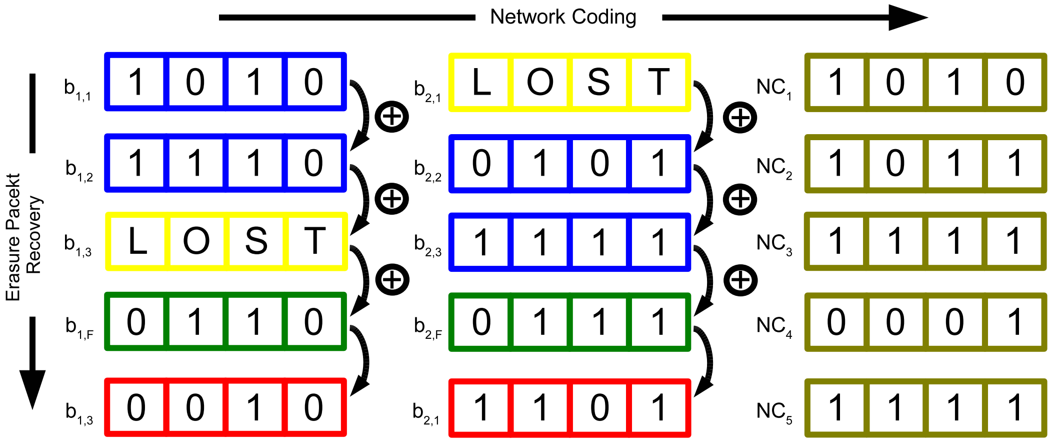

Figure 2.3: Combining erasure coding and network coding - example 2

packets are also erasure coded as in the previous case, but the need for delaying the packets has been removed.

This procedure has had the side effect of increasing the size of the network coded, erasure protected block by 1, which has reduced the overall throughput of the network slightly. There is also a requirement for side information to be stored in the header of the network coded packet to identify which packets were network coded together (this applies to both schemes presented here), but this is easily accommodated using the method described in [48].

The next section will look at the ways to apply this combined coding system and the effects that it has on the reliability of the network coded network using analytical PLR evaluations.

2.3 Analytical Evaluations in Combined Erasure and Network Coding

The purpose of this section is to introduce analytical tools to evaluate the performance of the proposed combined NC and EC framework and elaborate on the position of relays which should be involved in the EC functions. It begins by examining the impact that EC has on the PLR experienced by packets in a multi-hop relay network.

Using the early discussion in Section 1.3, we assume that the erasure code has the parameters (n, k, e). Referring to [22], for a multi-hop network utilizing EC, the averaged end-to-end PLR (here defined as PLR_{corr}) after erasure recovery has been performed and is given by:

$$PLR_{corr} = \frac{1}{n} \sum_{i=e+1}^{n} i \binom{n}{i} PLR_{ee}^{i} (1 - PLR_{ee})^{n-i} \ , \tag{2.5}$$

where:

(i) $PLR_{ee} = 1 - (1-p)^M$ = end-to-end PLR without erasure recovery;

(ii) p = raw link PLR as previously used;

(iii) M = number of links between the "source" and the "destination" where erasure recovery occurs;

(iv) n and e are parameters of the erasure code used.

If erasure recovery is performed at every node in the network then $M = 1$ and PLR_{ee} simply becomes p. Using the parameters from our previous examples where three data packets were erasure coded, $n = 4$, $k = 3$, $e = 1$ and $p = 0.05$, the averaged end-to-end PLR is calculated as $PLR_{corr} = 7.13 \times 10^{-3}$. This value is significantly lower than the original uncorrected PLR of 0.05 used in the previous example. Applying this new value to Eqns. (2.2) and (2.4) one obtains $P(b_1 \textit{ does not arrive at node } X) = 1.42\%$ and $P(b_2 \textit{ does not arrive at node } X) = 3.50\%$.

This results in an increased reliability in the delivery of packets, at node X, from stream b_1 of more than 8% and an almost 19% increase for packets from stream b_2. This, of course, requires that all nodes in the network perform erasure recovery which will use up resources on all the nodes and many not be the most economical solution to the problem. An alternative is to only allow certain nodes to perform the erasure recovery and have the remaining nodes perform simple DF or NC operations.

An example of this would be to only allow nodes which generate or forward network coded packets and those nodes which are destinations to perform the erasure packet recovery. In the butterfly network shown in the right side of Fig. 1.2, this would be nodes C, D, X and Y. For this new configuration, the probability of b_1 failing to arrive at node X would be calculated using Eqn. (2.5) with $M = 2$ and results in a value of $P(b_1 \textit{ does not arrive at node } X) = 2.58\%$. This is a slight reduction compared to performing erasure recovery at all nodes, but is still a more than 7% increase over no erasure recovery.

Turning to the probability of b_2 failing to arrive at node X, Eqns. (2.3) and (2.4) must be modified to incorporate the effects that Eqn. (2.5) will have on the network. For this analysis we will let p represent the raw PLR of a link as used earlier, $p_{corr}^{M=1}$ will be the value of Eqn. (2.5) with $M = 1$ and $p_{corr}^{M=2}$ the value of Eqn. (2.5) with $M = 2$.

The probabilities of packets b_1 and b_2 arriving at node C changes from $(1-p)\cdot(1-p)$ to $(1 - p_{corr}^{M=2})$ as node C performs erasure recovery. The probability of the network coded packet leaving node C and arriving at node X changes from $(1-p)\cdot(1-p)$ to $(1 - p_{corr}^{M=1})^2$ since erasure recovery occurs at nodes D and X. For this analysis, the probability of packet b_1 being successfully transmitted between nodes A and X is left as $(1 - p)$ but this is not the correct value to use. This is because EC is being done between nodes H and X so the reliability would actually be higher. Using this value will therefore represent a worst case scenario. Applying these changes to Eqn. (2.3) results in Eqn. (2.6) below.

$$\begin{aligned}
& P(b_2 \textit{ arrives at node } X) \\
= \;& (1-p_{corr}^{M=2})^2 \cdot (1-p_{corr}^{M=1})^2 \cdot (1-p) + \{1-(1-p_{corr}^{M=2})\} \\
& \cdot\{(1-p_{corr}^{M=2}) \cdot (1-p_{corr}^{M=1})^2\} \qquad (2.6) \\
= \;& (q_{corr}^{M=2})^2 \cdot (q_{corr}^{M=1})^2 \cdot q + (1-q_{corr}^{M=2}) \cdot (q_{corr}^{M=2} \cdot (q_{corr}^{M=1})^2) \\
= \;& q_{corr}^{M=2} \cdot (q_{corr}^{M=1})^2 \cdot (q_{corr}^{M=2} \cdot (q-1)+1)
\end{aligned}$$

Applying these changes to Eqn. (2.4) yields Eqn. (2.7) below.

$$\begin{aligned}
& P(b_2 \textit{ does not arrive at node } X) \\
= \;& 1-[q_{corr}^{M=2} \cdot (q_{corr}^{M=1})^2 \cdot \{q_{corr}^{M=2} \cdot (q-1)+1\}] \qquad (2.7)
\end{aligned}$$

Using the same parameters as the previous examples ($n = 4$, $k = 3$, $e = 1$ and $p = 0.05$), Eqn. (2.7) gives $P(b_2 \textit{ does not arrive at node } X) = 8.64\%$. This is an increase in the loss rate of approximately 5% compared to the case where erasure recovery is performed at all nodes, but it is still a more than 13.5% reduction in the rate compared to the case without EC. By selecting a subset of nodes to perform erasure recovery, this scheme has traded some reliability in the network for a reduction in the amount of processing power required in the network.

There are additional variations on our method that can be considered, two of which will be examined here. The first will be to allow only the nodes which duplicate non network coded packets to perform erasure recovery (which for this example would be nodes A and B) and will be referred to as the duplicating nodes. The second would be the traditional end-to-end erasure recovery which means only nodes X and Y would perform erasure recovery operations. Formulas for these two additional cases were derived in a similar fashion to those already given.

2.4 Reliability Results for Proposed Methods

Figures 2.4 and 2.5 show plots for the probability that b_1 and b_2 fail to arrive at node X respectively for the five different erasure recovery scenarios described earlier for a range of raw link PLRs. The x-axis represents the raw link PLR (which was also denoted by p in earlier discussions) and the y-axis is the end-to-end PLR after erasure recovery has been performed. A wide rand of raw link PLR values was chosen to demonstrate the performance of the methods under a wide range of network scenarios. For example, with a fixed BER, as the number of bits per packet increases, the PLR will increase. So for a network using jumbo packets to reduce the header overhead, even a very low BER could result in a fairly high PLR.

No EC represents the case where no erasure recovery is performed, *EC All Nodes* denotes every node in the network performs packet recovery, *EC End-to-End* is erasure recovery only at the destinations nodes X and Y, *EC NC Nodes* is erasure recovery on all nodes that handle network coded packets (that is nodes C, D, X, and Y), and finally, *EC Duplicating Nodes* denotes only the nodes that replicate non-network-coded data and the destination nodes perform erasure recovery (that is nodes A, B, X, and Y).

Looking first at Fig. 2.4, which is for the path $HA - AX$ that packet b_1 takes, the *EC All Nodes* and *EC Duplicating Nodes* curves are identical and provide the best performance. This was expected since erasure recovery is performed at every node along this path. The *EC End-to-End* and *EC NC Nodes* curves are also identical, as in both cases erasure recovery is only performed at the destination node along this path. All four of these options outperform the unprotected scenario, but the extent depends on the raw link PLR with all four schemes giving very similar performance below a p of 0.05.

Looking at Fig. 2.5, which is for the path $HB - BC - CD - DX$ that packet b_2

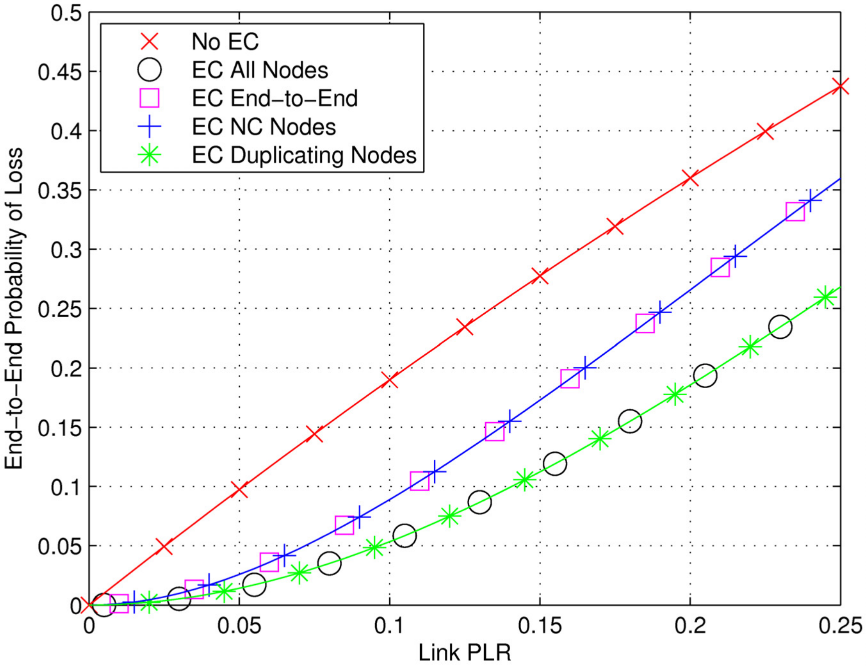

Figure 2.4: Probabilities that packet b_1 fails to arrive at node X for various error recovery scenarios $[(n, k, e) = (4, 3, 1)]$

takes, we see that *EC All Nodes* provides the best improvement over no protection (which was expected) and that *EC End-to-End* provided the lowest improvement. The *EC End-to-End* for this scenario actually converges with the unprotected case at high raw link PLR values. The curves for *EC Duplicating Nodes* and *EC NC Nodes* are very similar but have a cross-over point where the performances of the two schemes switch. This will be examined in more detail shortly. It should be noted that these curves are plotted for one specific set of (n, k, e) and the performance varies as one varies these parameters (which alters the coding rate of the EC).

In Fig. 2.6, the probability curves for the *EC Duplicating Nodes* and *EC NC Nodes* schemes are plotted for two different sets of (n, k, e). The point where the two methods' performances cross-over varies depending on the parameters (n, k, e). The *EC Duplicating Nodes* method performs better before this cross-over point and

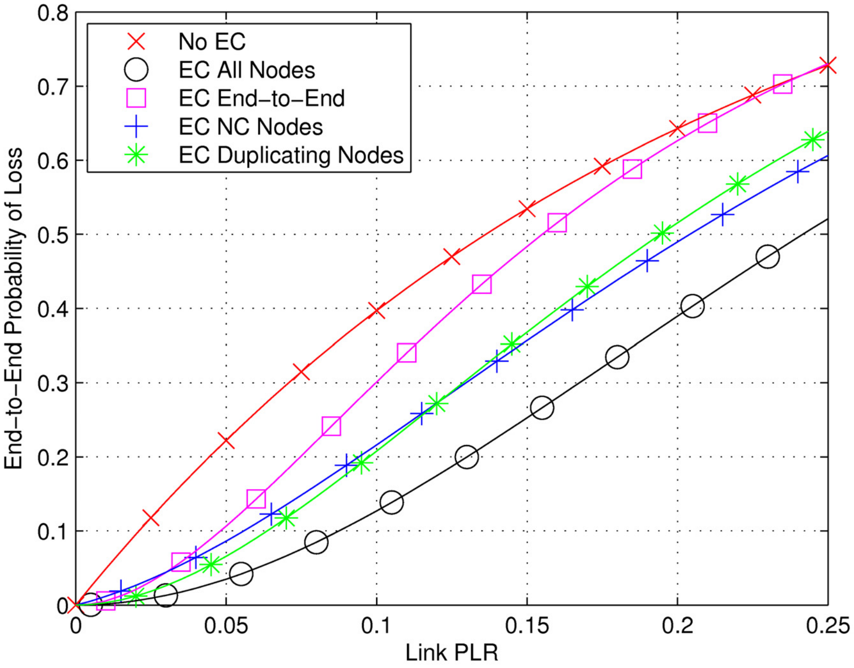

Figure 2.5: Probabilities that packet b_2 fails to arrive at node X for various error recovery scenarios $[(n, k, e) = (4, 3, 1)]$

the *EC NC Node* method is better after the cross-over point. For this example, e is kept constant and n and k are varied. This is to provide performance analysis for different levels of erasure protection against raw PLR. The EC method used is the one described earlier, so $k = n - 1$. It can therefore be stated that as the coding rate k/n increases, the raw link PLR at which these two methods intersect decreases.

Both of these schemes provide a significant performance increase compared to that of an unprotected or end-to-end protected network while consuming fewer resources compared to erasure recovery implemented at every node. By having some knowledge of the PLR of the links in a network, one can adjust the locations where erasure recovery is performed and the coding rate of the EC to achieve a desired system performance while balancing the resource utilization.

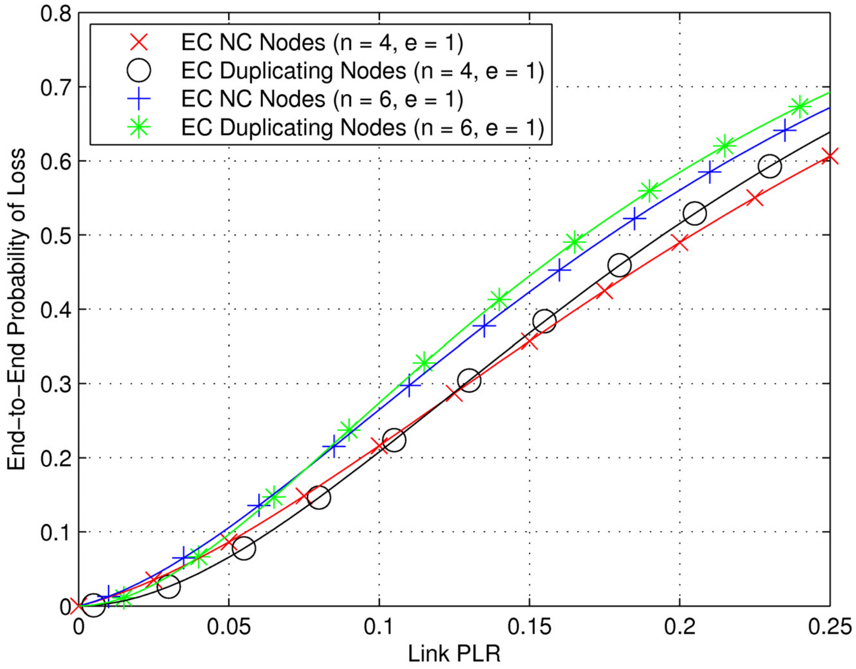

Figure 2.6: Probabilities that packet b_2 fails to arrive at node X for two error recovery scenarios with different (n, k, t) parameters

2.5 Summary

This chapter first discussed how packet loss on all links in the network affects the reliability of the NC. Implementations of EC were presented which could easily be intergraded into the NC framework. It was then shown that applying this EC scheme in conjunction with NC greatly increased the reliability of the network while introducing a minimal amount of overhead and complexity to the network. Several sub-optimal schemes were shown which provided improvements over a network without EC while balancing the amount of resources needed to implement the schemes.

Chapter 3

Wireless Relay Networks with Two-Way Asynchronous Traffic

When deploying wireless network coding with bi-directional exchange of data, the issue of time synchronization requires special attention. If the two terminal nodes exchanging data generate traffic flows with the same average rate and random arrival times, to make network coding operational based on XORing of packets at the relay, buffering of the traffic flows is needed, which may lead to prohibitive delays.

To restrict the amount of delay experienced by any data packet when NC is used, this work proposes to periodically flush the packet buffer at the relay of any unpaired packets [25]. When this flushing occurs, if there are any unpaired packets waiting to be network coded at the relay, ‶single packet″ broadcasts will be used to transmit these unpaired packets. While the extra broadcasts introduced by this flushing will reduce the throughput efficiency of the network, these opportunistic ‶single packet″ broadcasts are used in our work to send erasure coded packets. This is with the purpose of improving the overall reliability of the data exchange.

In particular, two methods to restrict the delay experienced by packets at the relay node are employed. The approaches are to either impose a time limit for buffering the packets, or limit the number of network coded transmissions before flushing the buffer and broadcasting any unpaired packets combined with erasure coded packets. Performance trade-offs between erasure based improvements in PLRs, delays, energy conservation and throughput are documented for traffic with Poisson arrival times.

In this chapter we will examine the delay introduced by NC when deploying it in a three node wireless relay network with bi-directional asynchronous traffic. First, in Sections 3.1 and 3.2 we present the problem and analyze the delay in systems with NC. In Section 3.3 we will outline three methods to restrict the amount of delay introduced by NC and examine what the benefits and costs are for these methods. One of these costs is the broadcast of some packets with underutilized capacity from the NC perspective, providing an opportunity to introduce some reliability protection. Details on this will be discussed in Section 3.4. Results from simulations of the three proposed methods will be outline in Section 3.5 and this will be followed by concluding remarks in Section 3.6.

3.1 Delay Introduced by Network Coding

To analyze the delay introduced by deploying wireless network coding, first a re-examination of the fundamental operation of a relay network with three nodes as shown in Fig. 3.1 will be performed. In this network, nodes A and B exchange data packets a_i and b_i via a relay node R, where i represents the packet index generated by the corresponding source node. These packet indexes can also be interpreted as time values as each packet is created at a specific point in time. For the synchronous case these times would be evenly spaced by Δt and then $i \cdot \Delta t$ would represent the specific time the packet was created. For the asynchronous case the packets are not generated at uniform intervals, therefore the time between successive packet indexes i and $i+1$ is a random time interval instead of Δt.

As was shown earlier in Section 1.2.1 and in Fig. 3.1 b), network coding will allow node R to combine packets a_i and b_i to form packet $c_i = a_i \oplus b_i$. The process introduces a minimal amount of delay assuming the packets a_i and b_i arrive within the time before the next group of two packets (a_{i+1} and b_{i+1}) arrive. This, of course, implies that the two data flows are synchronized.

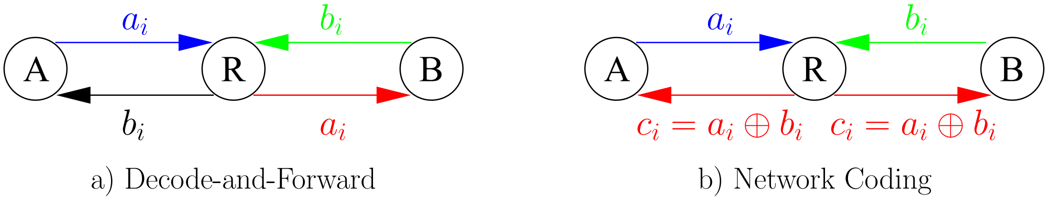

Figure 3.1: Three node relay network

When the two data flows are not synchronized, node R must buffer the packets arriving and only when a pair of packets can be formed does it network code them together and broadcast the new packet. For the remainder of this work, this will be referred to as the Infinite Wait method and the reason for the coining of this name will be provided in the next section based on queuing theory results. This is shown schematically by the timing diagram shown in Fig. 3.2.

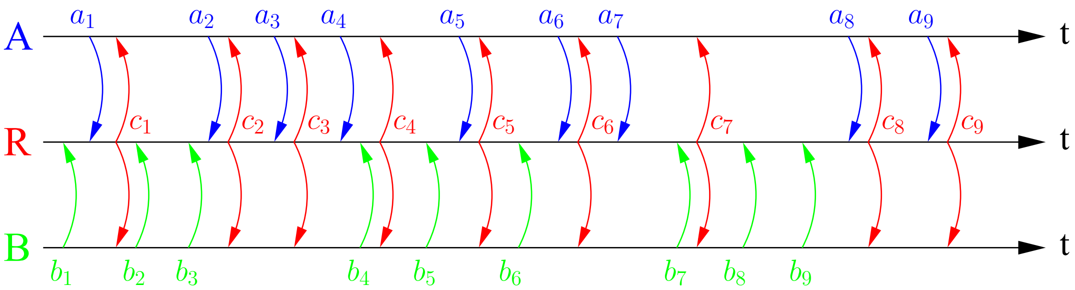

Figure 3.2: Transmission timing diagram, general wireless network coding

The diagram depicts nine packets being broadcast by both node A and B at random times and the resulting broadcasts from Node R. As can be seen, node R must buffer packets from either node A or node B at any given time but never from both at the same time. This means that for some instances, packets from node A are experiencing a delay and at others, packets from node B are delayed. This timing diagram format (with 9 packets being sent by each node) will be used later to show how the various proposed delay reducing methods behave.

3.2 Delay Analysis

Having established the mechanism by which NC introduces delay into the bi-directional data exchange, the properties of this introduced delay within a realistic scenario will be examined to determine how significant the delay is. It will be assumed that nodes A and B will independently broadcast data packets at random intervals. The interoccurrence times between broadcasts from each node will be derived from an exponential distribution with a rate parameter of λ. The rate parameter for both nodes will be the same so that on average, both nodes will generate the same number of packets in a given time interval. It is also assumed that all transmissions from nodes A and B are received at node R without error. This results in node R receiving two data streams consisting of packets a_i and b_i with Poisson arrival times. The Poisson traffic model is a traditional stochastic model for the traffic flows from data sources used during the initial development stages of communication protocols in order to analyze the performance and capacity of networks under study [2]. The methods developed here are general enough that they can handle other traffic models. Additionally the Poisson model used can be modified to simulate other traffic conditions, for example, burstiness can be model by varying the parameter λ over time.

Figures 3.3 and 3.4 show histograms of the delay experienced by packets sent from node A for two different runs of a simulation with $\lambda = 1$. The first was for $100,000$ packets and the second was for $1,000,000$ packets. As can be seen, the

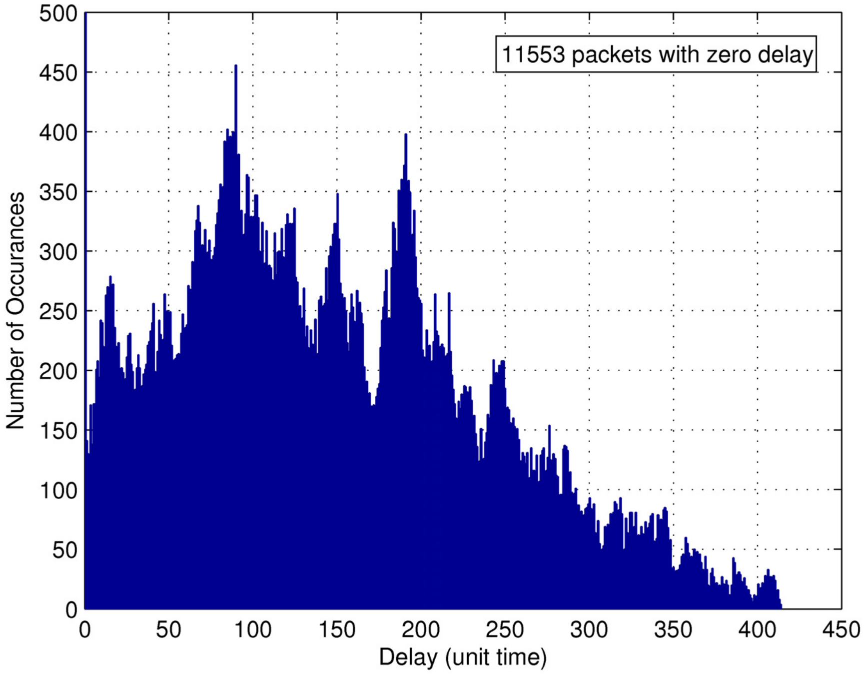

Figure 3.3: Histogram of delay packets from node A experience, $100,000$ packets

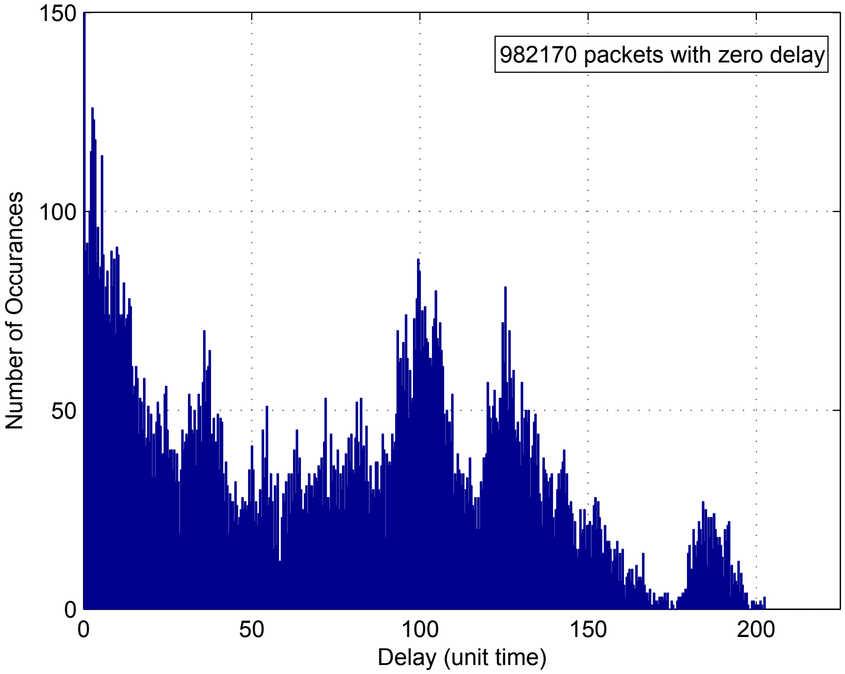

Figure 3.4: Histogram of delay packets from node A experience, $1,000,000$ packets

distribution of the delays are not consistant and do not appear to fit one of the standard distributions. To allow us to view the variation in the delays, the number of packets which experienced zero delay have been removed and are noted in the box at the upper right corner of the plot. The number of packets which experience zero delay also showed a large variation (multiple runs generating the same plots show here were done, but are not included in this dissertation) which would tend to indicate that in some runs the arrival of packets from node A are lagging those of node B, and in others, the reverse is true.

The histogram plots were presented here as the histogram is a tool used to characterize realizations of a random variable (RV) and a way to estimate the probability density function (PDF) of a RV. To provide a good approximation of the PDF using a histogram, many independent realization of the same RV must be generated. The two plots here attempted to use a histogram to characterize the distribution of the wait time packets experienced at the relay node (time spent waiting to be paired). As can be seen from Figs. 3.3 and 3.4, we conclude that either histogram analysis is not an appropriate tool to use to characterize the wait time or the simulations run did not have truly independent realizations of the wait time. Therefore, another approach was taken to examine the wait time and attempt to characterize it. This approach was to look at the problem from a queuing perspective and examine the number of packets waiting to be paired at node R.

The plot in Fig. 3.5 shows the number of packets waiting at node R to be paired and network coded for six simulation runs with $\lambda = 1$. It was generated by assigning a "$+ 1$" to packets arriving from node A and a "$- 1$" to packets arriving from node B and running a cumulative sum. Thus as packets are paired up, they sum to zero and once all packets are paired the total will be zero.

As can be seen in the plot, there is a lot of variability in the queue and at times there can be a substantial number of packets waiting to be paired up. Since the

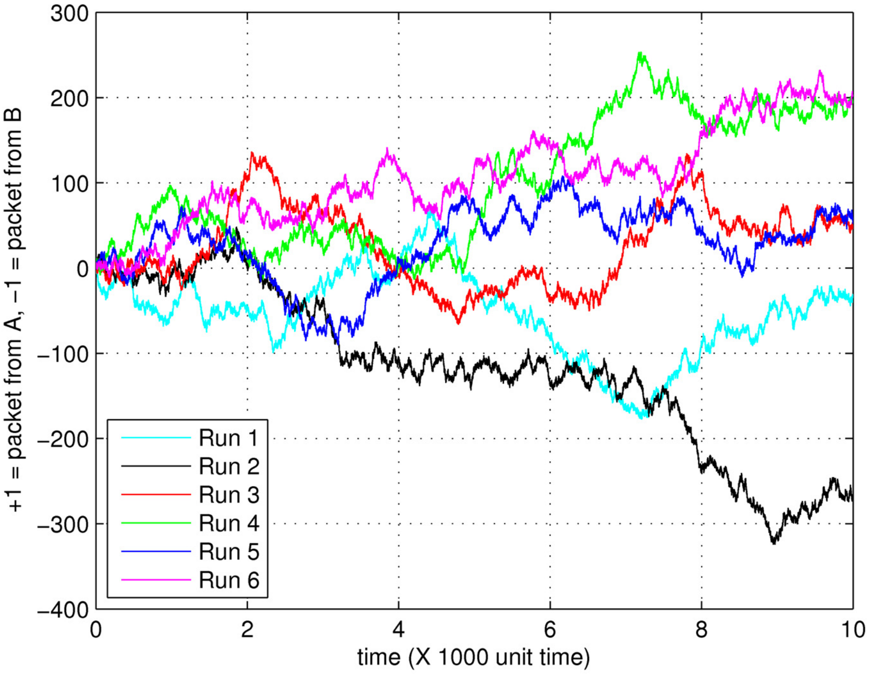

Figure 3.5: Packets buffered at relay node in general wireless network coding

arrival times between packets for each of the two source nodes are independent and are generated using the same rate parameter λ, then any arrival at node R has an equal probability of being from node A or node B. As the arrival of any packet either increments the number of queued packets by "$+1$" or "-1", this size of the queue can be looked at as a form of a simple symmetric random walk.

A random walk is the mathematical representation of the trajectory that results from taking successive random steps. In the case of a simple symmetric random walk discussed here, these random step events occur at uniform intervals in time. Additionally, at each of these events, the random step can take on only one of two values "$+1$" or "-1", and have a equal probability of occurring.

Comparing the scenario presented here for the queue at node R, one can observe that the only difference between it and the simple symmetric random walk is that the interval between events (packet arrivals) is not uniform. This does not significantly affect the mathematical theory and therefore the results for the simple symmetric random walk can still be applied to this problem.

For a simple symmetric random walk, the mean approaches zero and the variance approaches ∞ as $t \to \infty$ [49] – [52]. And this clearly shows the drawbacks to using NC when asynchronous traffic is involved. The variability in the number of packets that must be queued increases as time progresses, and consequently, the variability in the amount of delay a packet experience increases. The observation that the NC behavior in time can be interpreted as a random walk when the incoming traffic follows a Poisson model is a unique contribution of this dissertation. The author expects that this will be exploited to a much greater extent in the future analysis of network coded communication systems.

This level of variability in delay and queue size will be prohibitive factors in a number of applications for wireless relay networks. It is also because of this tendency for the delay to continually increase as time passes that the author has called this case the Infinite Wait method. The next section of this chapter will turn its attention to reducing and limiting the delay experienced by packets and consequently the size of the queue at the relay node also.

3.3 Limiting the Delay

The following outlines three proposed methods which provide a more predictable delay distribution compared to the standard Infinite Wait method.

3.3.1 Maximum Wait Method

The first proposed method utilizes a fixed time interval Δ to determine when node R flushes its buffer of unpaired packets. These flushings occur at the fixed intervals $t_o + n\Delta$, where t_o is the time of arrival of the first packet and $n \in \{1, 2, ..., \infty\}$. An example of this is shown in Fig. 3.6.

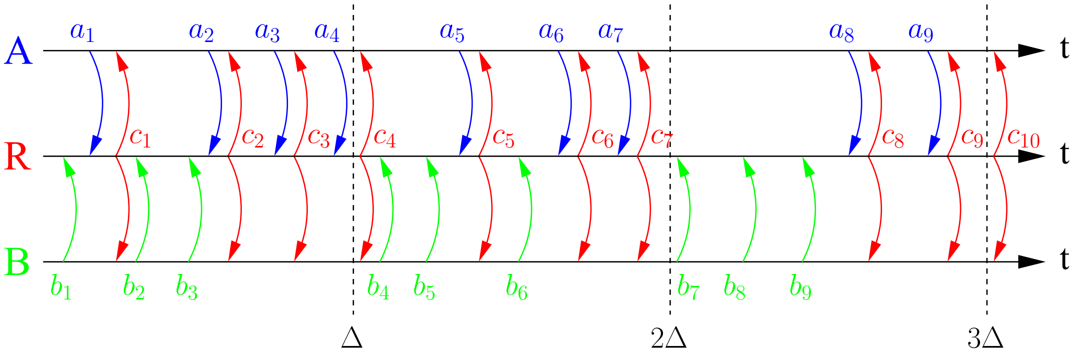

Figure 3.6: Transmission timing diagram, maximum wait time

This means that the maximum delay any packet could experience is limited to Δ. As a consequence of these flushings, node R now broadcasts some packets that have not been network coded, resulting in more broadcasts compared to waiting for all packets to be paired. In the example shown in Fig. 3.6, ten broadcasts are made by node R to complete the exchange (compared to the nine that would be required for the Infinite Wait method). Eight of these broadcasts are network coded packets ($c_1 = a_1 \oplus b_1$, $c_2 = a_2 \oplus b_2$, $c_3 = a_3 \oplus b_3$, $c_5 = a_5 \oplus b_4$, $c_6 = a_6 \oplus b_5$, $c_7 = a_7 \oplus b_6$, $c_8 = a_8 \oplus b_7$, and $c_9 = a_9 \oplus b_8$) and two are single packet broadcasts ($c_4 = a_4$ and $c_{10} = b_9$).

3.3.2 Maximum Pairs Method

In the second proposed method, node R does not flush the queue of unpaired packets at specific, evenly spaced intervals as was the case in the first method. Instead, it opts to perform the flushing after a specific number of network coded packets are broadcast. The number of network coded packets it waits for is denoted by N_{NC} and an example of this is shown in Fig. 3.7 where $N_{NC} = 3$.

As with the example of the previous proposed method, nodes A and B have nine packets to exchange and node R makes ten broadcasts to complete the exchange.

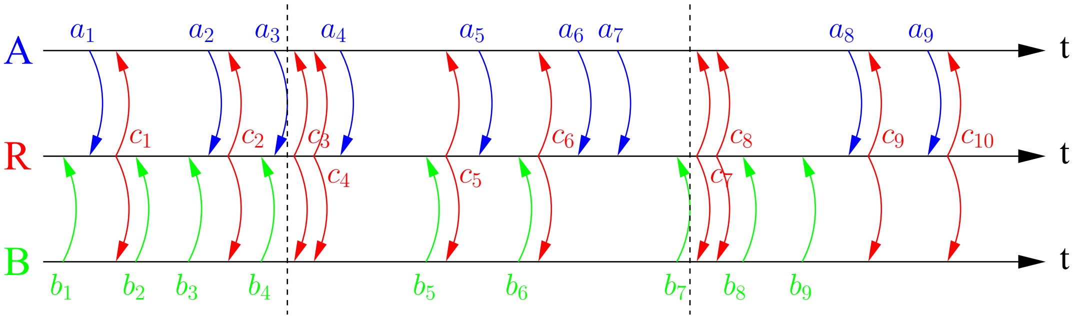

Figure 3.7: Transmission timing diagram, maximum number of pairs

Also as before, eight of these broadcasts are network coded packets ($c_1 = a_1 \oplus b_1$, $c_2 = a_2 \oplus b_2$, $c_3 = a_3 \oplus b_3$, $c_5 = a_4 \oplus b_5$, $c_6 = a_5 \oplus b_6$, $c_7 = a_6 \oplus b_7$, $c_9 = a_8 \oplus b_8$, and $c_{10} = a_9 \oplus b_9$) and two are single packet broadcasts ($c_4 = b_4$ and $c_8 = a_7$).

For the example shown in Fig. 3.7 with only a fixed number of packets to be sent from each source node, all packets from nodes A and B were either paired and broadcast or broadcast as single packet transmissions. Because the flushing mechanism used in this method relies on packet broadcast to trigger a flushing, when there is only a fixed number of packets to exchange there is the possibility that some packets my be stuck waiting at node R and never get broadcast. This issue would have to be dealt with in a similar manner to the previous method by setting a time limit to wait after the last packet arrives and flush the queue of packets after that limit had been reached.

3.3.3 Maximum Packets Method

The third proposed method is a variation on the Maximum Pairs method discussed above. Instead of flushing after a specific number of network coded packets are sent, node R flushes the remaining queued packets once a specific number of packets, N_P, have been received from either source node. This will restrict the number of packets waiting to be paired for network coding to a maximum of N_P and thus cap the maximum size of the queue to that value. A timing diagram example of this is shown in Fig. 3.8.

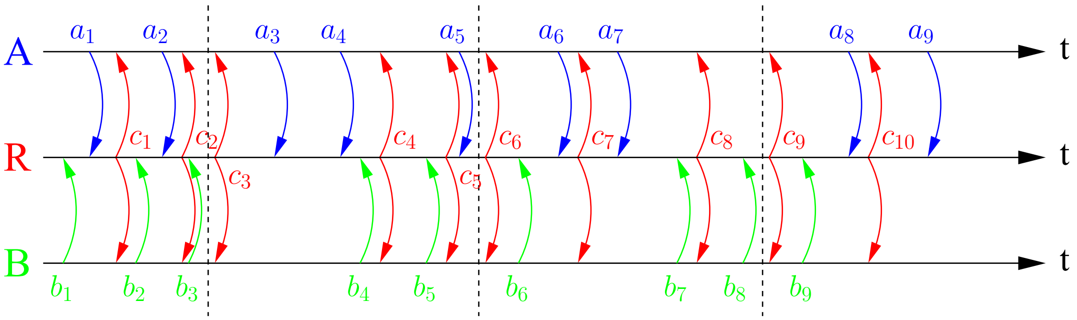

Figure 3.8: Transmission timing diagram, maximum number of packets

In this example, N_P is set to 3 and results in node R once again making a total of ten broadcasts. As was mentioned for the previous proposed method, when there is only a fixed number of packets to be exchanged, some packets may get stuck at node R. For this case, packet a_9 fails to get paired with a packet from node B and is never broadcast. Of the ten broadcasts made by node R, only seven of these broadcasts are network coded packets ($c_1 = a_1 \oplus b_1$, $c_2 = a_2 \oplus b_2$, $c_4 = a_3 \oplus b_4$, $c_5 = a_4 \oplus b_5$, $c_7 = a_6 \oplus b_6$, $c_8 = a_7 \oplus b_7$, and $c_{10} = a_8 \oplus b_9$) and three are single packet broadcasts ($c_3 = b_3$, $c_6 = a_5$, and $c_9 = b_8$).

As can be seen from the three proposed solutions to reduce the delay experienced by packets when NC is employed, this reduction comes at a cost. One of those costs is the broadcast of single packets. The next section will look at using these single packet broadcasts to carry additional information to improve data reliability in the network.

3.4 Underutilized Broadcast Capacity

All of the methods outlined in Section 3.3 rely on some mechanism to trigger node R to flush its queue of any unpaired packets. This results in additional transmissions by node R compared to the conventional Infinite Wait method which waits for all packets to be paired. While this increases the energy consumption of node R as there are more packets being sent, it also provides an opportunity to broadcast additional data within the underutilized capacity to improve the reliability of the communication from node R to the two source nodes.

By NC the unpaired packets which come from one source, with a mixture of previously transmitted packets from the opposite source, error protection can be provided in the form of EC, reducing the effective packet loss rate between node R and the two source nodes. Erasure coding was outlined previously in Section 1.3 and we refer the reader there for details on the specifics of the implementation.

To show a practical application of EC in the scenarios analyzed in this chapter, we consider a segment of the timing diagram for the Maximum Wait method previously shown in Fig. 3.6 (and reproduced in Fig. 3.9). At time Δ, node R broadcasts packet c_4, which is just the single packet a_4, and provides an opportunity to provide erasure protection for packets sent from node B. Referring back to Fig 1.3 in Section 1.3, we see that by XORing packets b_1, b_2, and b_3 from node B, we are able to create a redundant packet p_1. And by replacing the broadcast packet $c_4 = a_4$ from node R with $c_4 = a_4 \oplus p_1$, node B can still recover packet a_4 and node A can now recover any one of the packets b_1, b_2 or b_3 if they were lost, provided it received any two out of b_1, b_2 and b_3 and it received the redundancy packet p_1.

Since at the time of flushing there can only be unpaired packets from one of the source nodes, it is not possible to provide protection for packets from both sources at the time of flushing. It is proposed here that node R tracks all the packets from each source since the last time an opportunity to provide erasure protection for them occurred, and when the next opportunity arises, apply protection to all of those previously unprotected packets. This approach of using the spare capacity introduced by the delay limiting methods to carry erasure recovery data is a unique contribution

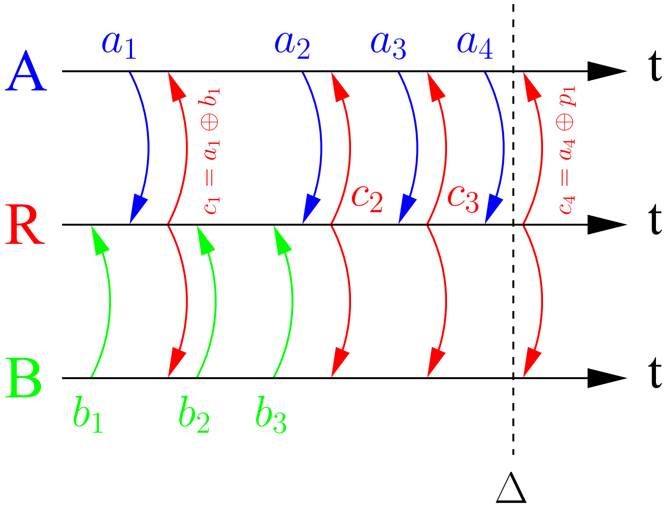

Figure 3.9: Exploiting underutilized capacity of unpaired packets in network coding

of this dissertation.

In the event that more than one unpaired packet exists at the time of flushing, the group of perviously unprotected packets is divided into as close to equally sized groups so that there is one unpaired packet per group to provide protection for that group. If there are more unpaired packets than unprotected ones, a one to one matching is used and the remaining unpaired packets are simply broadcast by themselves.

In the next section, simulation results of the three proposed methods are outlined and a comparison of the benefits and limitations of each are provided.

3.5 Results

To examine the advantages and disadvantages of the three proposed methods outlined in Section 3.3, and integrated with opportunistic EC as discussed in Section 3.4, simulations were performed using the classic Monte Carlo method in MATLAB®. Each source node sent 100,000 packets to the relay node. The interoccurrence times between each packet transmission were generated from an exponential distribution with a rate parameter λ (expressed in events per unit time). Values of the rate parameter λ, raw packet loss rate PLR_{raw}, maximum wait time Δ, maximum number of network coded packets N_{NC}, and maximum number of packets N_P were varied to examine their impact on the performance of the proposed methods. A summary of these is show in Table 3.1.

Table 3.1: Parameters for three node relay network with asynchronous traffic simulations

Network setup	Three node relay network (Fig. 3.1)
Number of packets sent by each source	$100,000$
Values of the rate parameter λ	$1,2,3,4,5$ (events/unit time)
Values of the raw PLR parameter PLR_{raw}	0.01 to 0.1 in steps of 0.01
Values of the max wait time parameter Δ	$1,2,3,4,5$ (unit time)
Values of the max pairs parameter N_{NC}	$1,2,3,4,5$
Values of the max packet parameter N_P	$1,2,3,4,5$

Performance metrics examined in this chapter were the average packet delay introduced, the effective PLR created by the use of EC, and the average increase in the number of transmissions made by the relay node compared to the Infinite Wait method. The next three sections examine each of these metrics for the proposed methods.

3.5.1 Average Packet Delay

Delay is calculated as the time between a packet's arrival at the relay and its broadcast as part of a network coded packet, c in this case. The time to process and combine packets was not considered, so some packets will experience a zero delay. Other packets may experience an infinite delay due to the nature of the methods used, specifically in the cases where the method waits for a specific number of packets or for a pairing of packets. These occur at the end of the simulation as there are only a finite number of transmissions made by the source nodes. These are a termination effect caused by the simulation and are therefore excluded from the calculations.

Figure 3.10 shows the average packet delay, expressed in terms of unit time, experienced by packets arriving at the relay node R from node A for the Maximum Wait method. As λ is increased, the average delay decreases for all values of the maximum wait parameter Δ. As Δ increases, the initial value of the delay experienced by the packets increases, as does the amount of variation over the range of λ values.

As a number of packets will experience zero delay, Fig. 3.11 shows the same data shown in Fig. 3.10 but with the zero delay packets removed. While the values of the average delay have now all increased, the general trend has remained the same and therefore the number of zero delay packets has remained relatively constant over the variation of the parameters.

Results for the average delay when the Maximum Pairs method is used are shown in Fig. 3.12 and Fig. 3.13. Figure 3.12 shows the average packet delay when all packets are taken into consideration. It can be seen that the average delay is initially higher for this method compared to the Maximum Wait method but decreases quite rapidly resulting in comparable average delay values at higher λ values. There is also less variation in the average delay between the various values of the number of network coded pairs parameter N_{NC}.

Figure 3.13 has the zero delay packets removed from the average calculation. The initial average delay value has increased more than in the Maximum Wait method which would indicate that more packets are experiencing a zero delay for the Maximum Pairs method. The rapid decrease in average delay is still present, so at higher values of λ, the average delay is comparable to that in the Maximum Wait method.

Figures 3.14 and 3.15 show the average delay experienced by packets when the Maximum Packet method is used. Figure 3.14 includes the zero delay packets and has the lowest average delay of the three methods. When the maximum number of

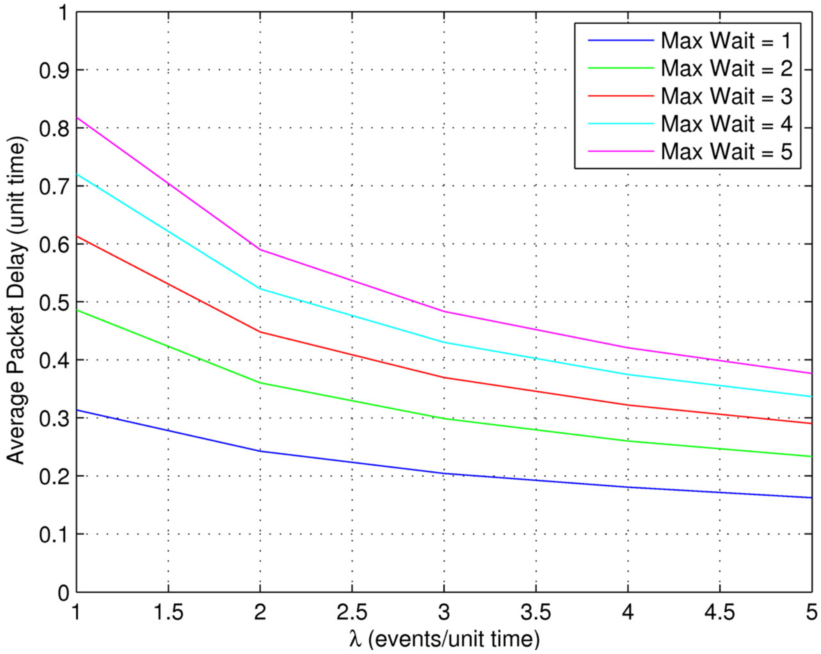

Figure 3.10: Average packet delay for Maximum Wait method

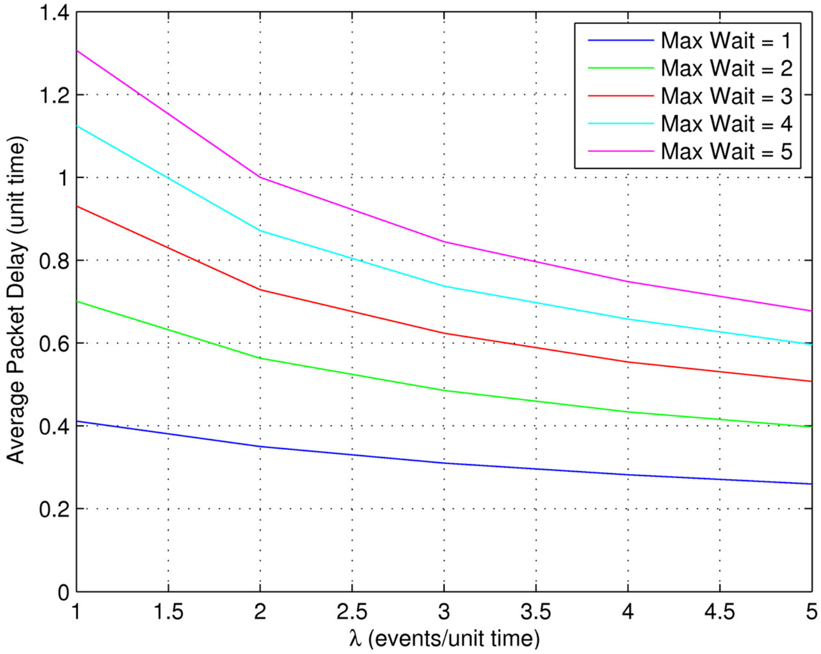

Figure 3.11: Average packet delay for Maximum Wait method when packets with zero delay removed

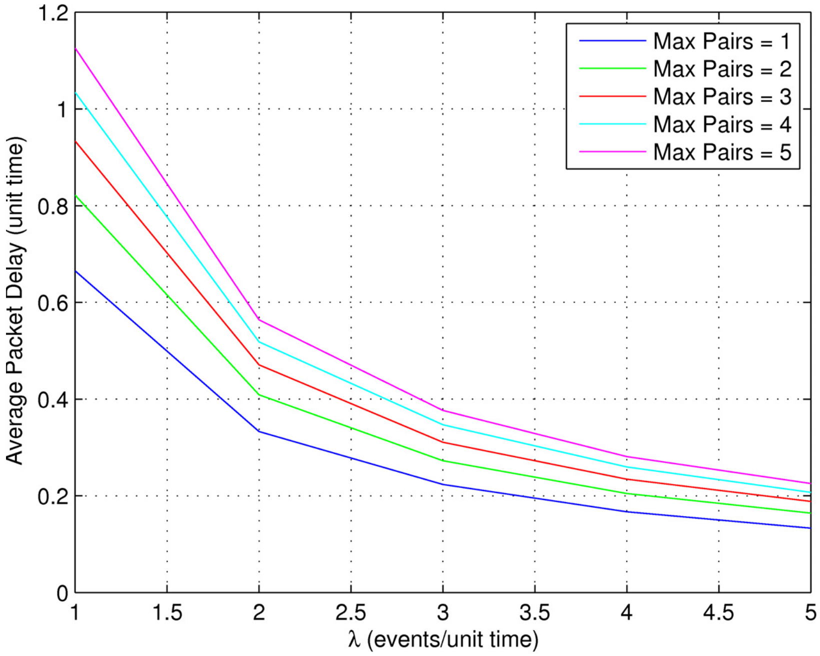

Figure 3.12: Average packet delay for Maximum Pairs method

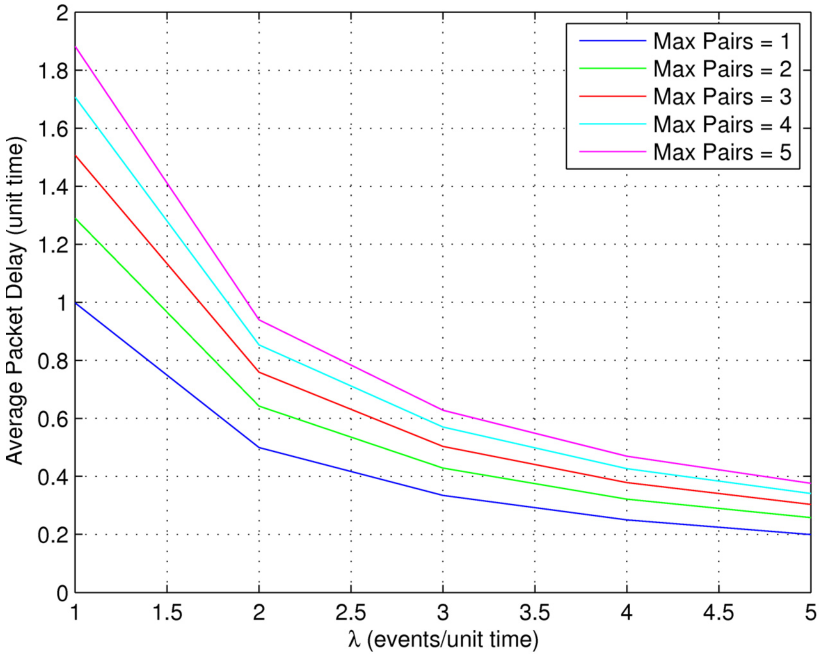

Figure 3.13: Average packet delay for Maximum Pairs method when packets with zero delay removed

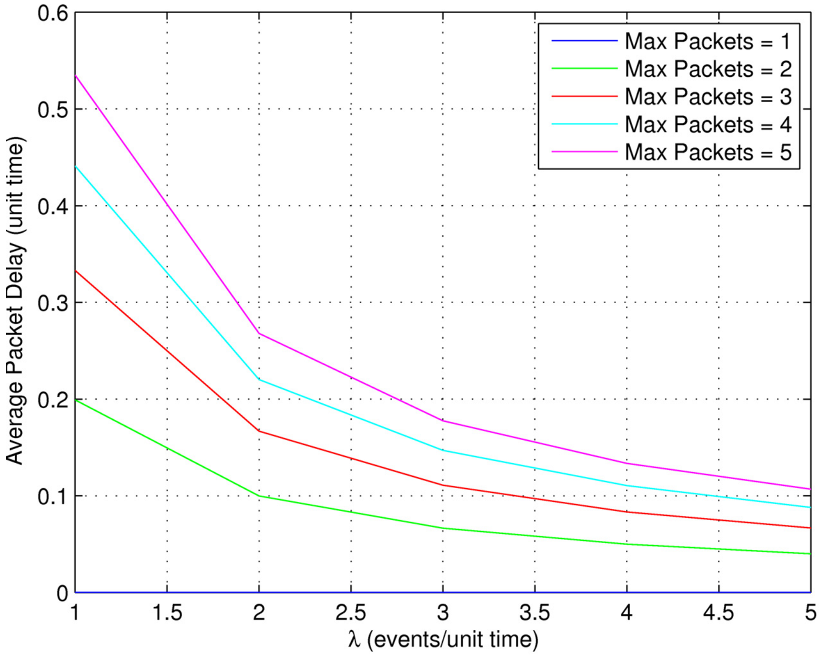

Figure 3.14: Average packet delay for Maximum Packets method

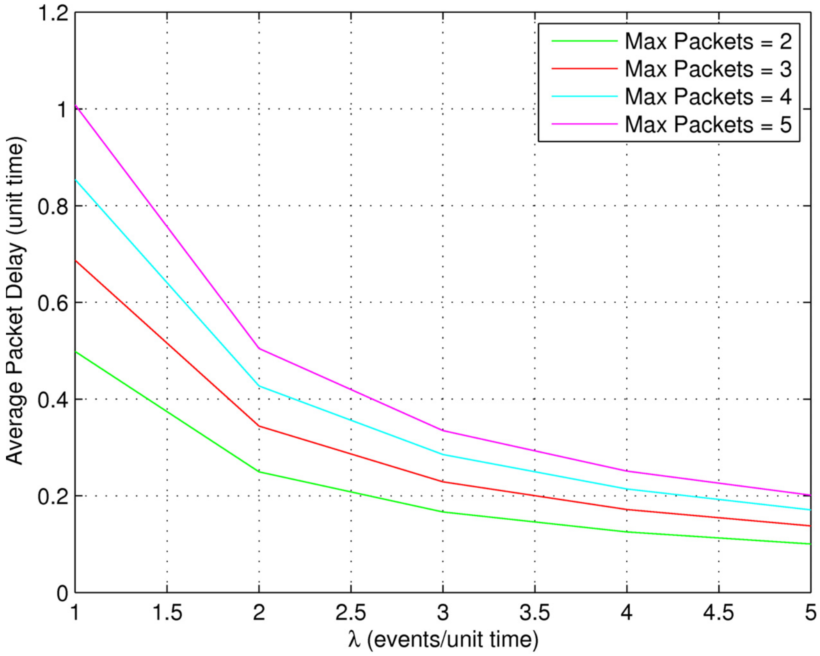

Figure 3.15: Average packet delay for Maximum Packets method when packets with zero delay removed

packets to wait for is set to $N_P = 1$, there is no delay introduced. In such a case, the method reverts to essentially a DF method with no NC.

Removing the zero delay packets results in the plot shown in Fig 3.15. Again the delay increases by a significant factor, indicating that a large number of packets are experiencing zero delay and the plot for $N_P = 1$ disappears as all packets experience zero delay for that case. The average delay still drops off rapidly as λ increases.

Next the impact that adding erasure coded data opportunistically has on the PLR experienced by the packets leaving node R will be examined.

3.5.2 Packet Loss Rate

Following in the same order as the previous section, we will now examine the effective PLR on the links from node R to nodes A and B. This effective PLR is the result of using the unpaired packet transmissions as an opportunity to introduce packet erasure protection.

Figures 3.16, 3.17, and 3.18 show plots of the effective PLR for the Maximum Wait method for $\lambda = 1$, 3, and 5 respectively. As λ increases, the effective PLR for all values of Δ decreases (approaches the diagonal dashed line). In addition, the variation between the curves for the various values of Δ also decreases as λ increases. From these three plots it is evident that the largest improvement in PLR for the Maximum Wait method occur for low values of Δ and λ.

Turning to the Maximum Pairs method, Fig. 3.19 shows the effective PLR when $\lambda = 3$. Only one plot is necessary here as variations in λ have no impact on the effective PLR. The variation in the curves for various values of N_{NC} is low and they provide similar performance to those for the Maximum Wait method when $\lambda = 2$.

Looking at the Maximum Packets Method, Fig. 3.20 shows the effective PLR for various values of N_P when $\lambda = 3$. Again, only one plot is necessary as the effective PLR is not impacted by λ. This method provided the largest reduction in PLR due

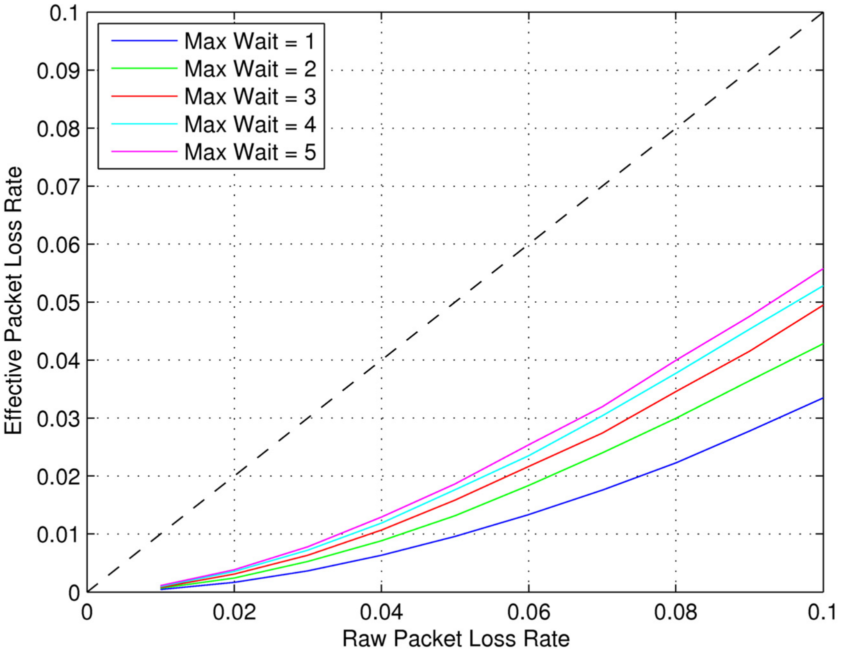

Figure 3.16: Effective packet loss rates for the Maximum Wait method when $\lambda = 1$

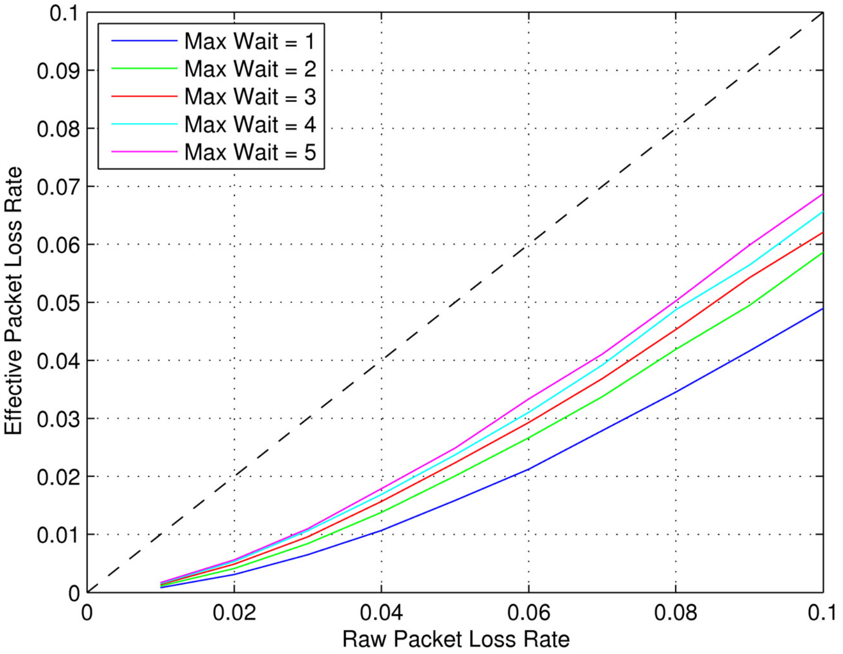

Figure 3.17: Effective packet loss rates for the Maximum Wait method when $\lambda = 3$

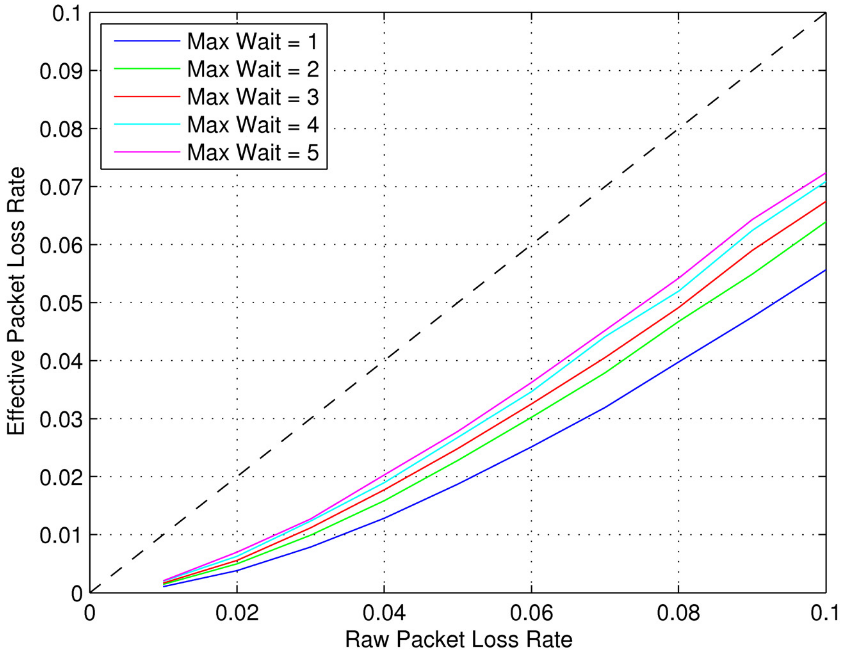

Figure 3.18: Effective packet loss rates for the Maximum Wait method when $\lambda = 5$

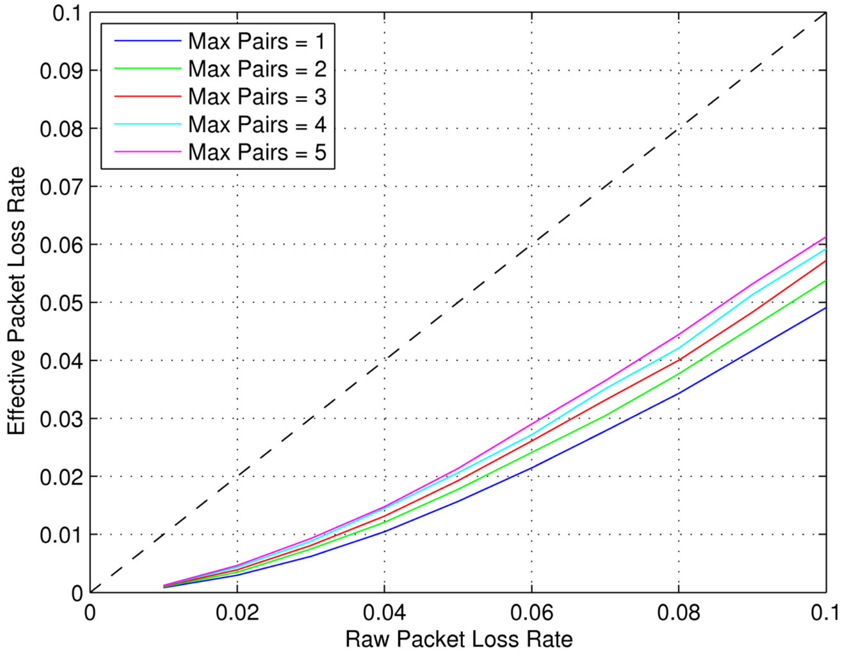

Figure 3.19: Effective packet loss rates for the Maximum Pairs method when $\lambda = 3$

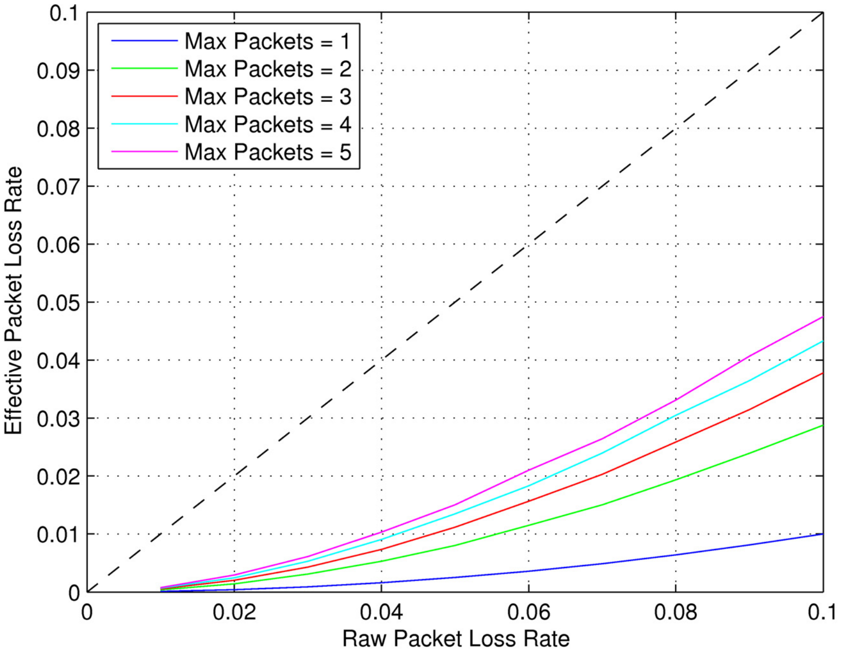

Figure 3.20: Effective packet loss rates for the Maximum Packets method when $\lambda = 3$

to the increase in the number of opportunities to introduce redundancy. For $N_P = 1$ there is one redundancy packet for every data packet, providing a high level of data protection and thus a very low PLR.

The Maximum Packet method provides the largest improvement over the raw packet loss rate. The Maximum Wait method provides some of the lowest improvements and as previously noted, these improvements decrease as λ increases. The third performance metric discussed in this section, the average increase in the number of broadcasts made by the relay node compared to the Infinite Wait method, will be looked at next.

3.5.3 Excess Broadcasts

This performance metric looks at the efficiency of the relay node and relates to the power consumption of the node (or equivalently to a reduction in the effective throughput). It is calculated as the number of transmissions made by node R for the given method, divided by the number of transmission that would have been needed if the Infinite Wait method was used minus 1, expressed as a percentage. In a mathematical formula, this calculation is given by Eqn. (3.1).

$$100 \cdot \left(\left(\frac{\sharp\ transmissions\ current\ method}{\sharp\ transmissions\ Infitie\ Wait} \right) - 1 \right) \tag{3.1}$$

For these simulations, $\sharp\ transmissions\ Infitie\ Wait = 100{,}000$ as every packet from node A would be paired with a packet from node B before it is transmitted.

For the Maximum Wait method, the average percentage of excess transmissions decreases as the rate parameter λ increases as seen in Fig. 3.21. This is because as the rate at which packets arrive at the relay node increases, the probability that a matching pair of packets (one from each source node) is formed before the maximum wait time elapses also increases. This reduces the number of transmissions that occur at the end of the waiting time that only contain a packet from one source and redundant information.

Turning to the Maximum Pairs and Maximum Packets methods, it can be seen in Figures 3.22 and 3.23 that changes in λ have little impact on the average percentage of excess transmissions. This is because these two methods utilize the number of received packets to determine when excess packet transmissions are necessary and therefore the rate at which the packets arrive has no impact, at least in the scenario examined when both sources are generating packets at the same rate. The Maximum Pairs method provides better performance over the Maximum Packet method and has the smallest variance in performance as the control parameter, N_{NC} is varied.

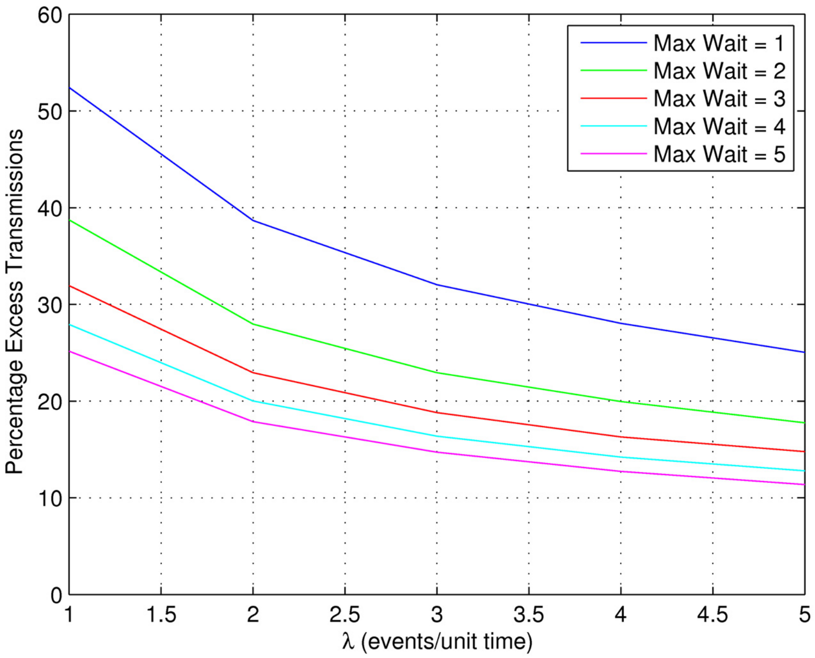

Figure 3.21: Percentage of excess broadcasts over Infinite Wait method for the Maximum Wait method

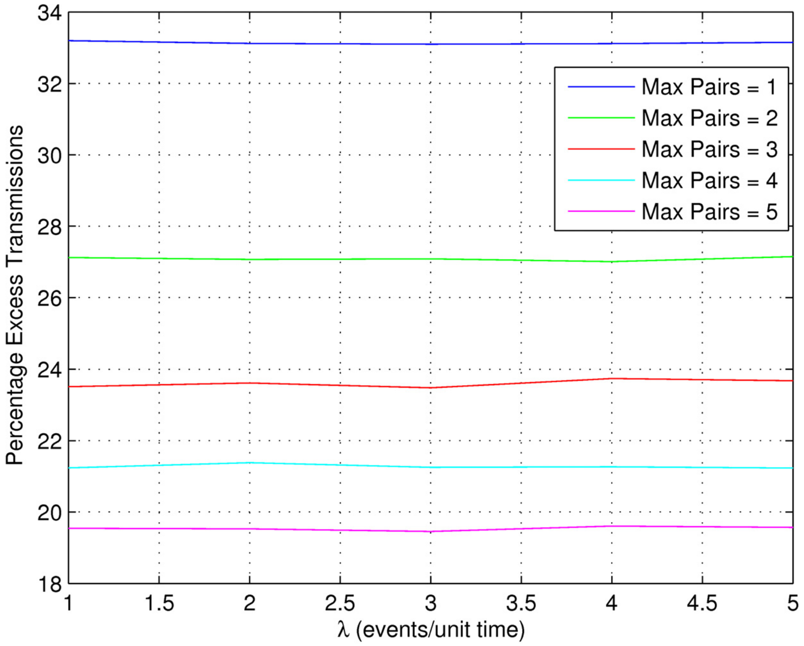

Figure 3.22: Percentage of excess broadcasts over Infinite Wait method for the Maximum Pairs method

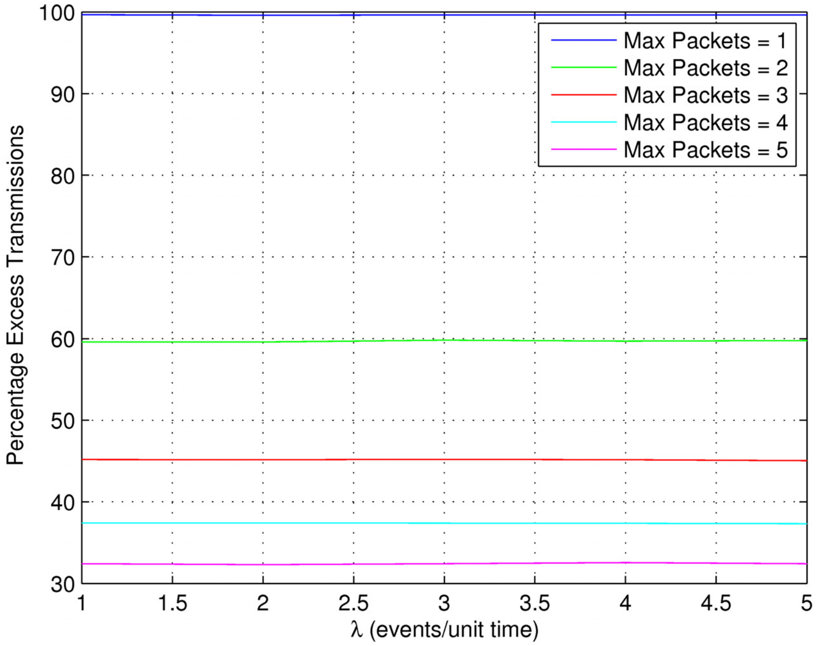

Figure 3.23: Percentage of excess broadcasts over Infinite Wait method for the Maximum Packets method

3.5.4 Performance Trade-offs

Having presented in Sections 3.5.1, 3.5.2 and 3.5.3 the results for how each of the proposed methods performed in terms of the three performance metrics used in this chapter, it is clear that trade-offs must be made as no one method had the best performance for all three metrics. When doing this comparison one must also consider the values of the control parameter for each of the methods. Namely the parameters Δ, N_{NC} and N_P must be examined to determine the appropriate values to provide the optimum performance desired for the particular network scenario. The previous plots and discussion below are based on a range of these parameters to provide a more complete overview of the performance of the outlined methods.

This section will compare two of the proposed methods using the three performance metrics in an attempt to demonstrate the trade-offs between them and determine under what conditions one would select one over the other. For this comparison the Maximum Pairs and Maximum Packets methods will be chosen.

First a comparison will be made in terms of the average packet delay. Referring back to Figs. 3.12 and 3.14 and choosing a value for λ of 2, the Maximum Pairs method has average delays (expressed in unit time) in the range of (0.3,0.575) while the Maximum Packets method has average delays in the range of (0,0.275). If λ is increased to 4, the the range of average delays for the Maximum Pairs method changes to (0.175,0.3) and the Maximum Packets method's average delay range changes to (0,0.15). For both methods their average delay drops as λ is increased but clearly the Maximum Packets method provides lower average delay.

Turning next to a comparison based on the effective PLR due to the EC, the reader will be referred back to Figs. 3.19 and 3.20 for the values presented here. For a raw PLR of 0.04, the Maximum Pairs method has effective PLRs in the range of (0.01,0.015) while the Maximum Pairs method has effective PLRs in the range of (0.001,0.011). If the raw PLR is increased to 0.08, the range of effective PLRs for the Maximum Pairs method changes to (0.035,0.045) and for the Maximum Packets method it changes to (0.0075,0.0325). For both values of raw PLR, the Maximum Packets method provides better performance then the Maximum Pairs method but variation in values for the Maximum Pairs method stay more consistent.

Lastly looking at the percentage of excess transmissions (which essentially translates in power consumption) for the two methods shown in Figs. 3.22 and 3.23, the following comparison can be made. For the Maximum Pairs method the percentage excess is in the range of 19% to 33% for all values of λ. For the Maximum Packets Method, this range is between 33% to 100% and again is constant for all values of λ. Based on this metric, the Maximum Pairs method provides the best performance.

So the decision on which of these two methods to use depends on which of these metrics is most important to the network. If delay is the critical factor then the Maximum Packets method provides the best performance while delivering a reasonable effective PLR but potentially sacrificing on the power consumption (excess transmissions). If effective PLR is the primary goal then the Maximum Packets method again is the best choice with the same benefits and drawbacks as before. If power reduction is the goal then the Maximum Pairs method is the best choice but may not provide the best delay reduction or effective PLR. So this is the three way trade-off that must be evaluated before one can choose which method is best and this decision must also take into account the data rate λ as it has an impact on some of the metrics for some of the methods.

3.6 Summary

This chapter presented a new methodology to integrate network and erasure coding in wireless networking for asynchronous two-way data exchange between terminal nodes via a relay. To limit the delay associated with conventional wireless NC, which requires prohibitive buffering for real-time applications, three forms of time limits were imposed to restrict the wait time at the relay to pair packets from end terminals for NC. This limiting theoretically forces some broadcasts from the relay to have underutilized capacity, which in the proposed methodology is used in an opportunistic fashion to send erasure coded packets, improving the relay broadcast reliability.

The benefits of time limiting the network coded broadcasts at the relay are documented for different methods in terms of erasure based improvements in PLRs, delays, energy conservation and throughput for traffic with Poisson arrival times. The trade-offs between these quality of service (QoS) performance metrics are documented in different network configurations.

Chapter 4

Optimum Selection of Relays via Power Control in Network Coded Relay Networks

The conservation of energy is an important factor to be considered when designing communication networks, particularly when designing wireless networks where network elements are powered from finite energy sources, such as in WSN or ad-hoc networks. Designers exploit energy savings opportunities at all system levels, from the application layer down to the hardware [53]. In this chapter we describe novel distance-based routing protocols optimized for achieving minimum energy communications for randomly deployed nodes in wireless relay networks with cooperative communications. The DF and NC paradigms are integrated with dynamic power allocation techniques to achieve this goal [26], [27].

Specifically, this chapter develops computationally efficient algorithms for the selection of relay nodes and the assignment of transmit power levels. The primary objective is to minimize the total transmit power consumed along the relay path chosen to permit the exchange of data between a specific pair of nodes. This work is not limited solely to the design of routing protocols for relay paths with a single relaying node, as was the case in Chapter 3. Protocols are also developed for the case where two relay nodes are involved, and extensions have been made to demonstrate that the algorithms presented here are also valid for the case of L relay nodes.

The protocols developed are based on the knowledge of interdistances between nodes, and with this knowledge, nodes are permitted to adapt their transmission power to control the network topology to achieve the minimized total energy in networks with mesh type connectivity. Geometric ideas and abstractions are used to demonstrate that nodes which are closest to equi-spaced points distributed along the line connecting a source and destination are the most energy efficient choices for relays.

A theoretical analysis and simulations of the energy efficiency of the DF and NC based relaying schemes under different propagation conditions and for various node densities are provided to demonstrate the performance of the algorithms developed in this work. The potential theoretical improvements are also provided to corroborate the results.

Section 4.1 provides some additional background information on the determination of distances and on relay strategies when more than one relay node is involved. Proof that equi-spacing of relay nodes provides the greatest power savings and a computational efficient means of determining which nodes should act as relays are presented in Section 4.2. The computationally amendable relay selection and power allocation algorithm developed in this chapter is outlined in Section 4.3. An analysis of the energy efficiency of the DF and NC relay strategies used in this algorithm, under different propagation conditions and for various node densities, are presented in Section 4.4. The chapter ends with concluding remarks in Section 4.5.

4.1 Background

This section provides additional information on how the distances between nodes are determined, along with a brief discussion of relay strategies when two or more relay nodes are employed. This will provide the groundwork for the presentation of the algorithms developed later in this chapter.

4.1.1 Determining Distances

In the initial presentation of the algorithms developed in this chapter, decisions as to which nodes are involved in forwarding the data between the source and the destination pairs are based solely on the distances between the nodes and the path loss exponent α (i.e., based on the mean channel gains). As elaborated later, in the practical implementations of the algorithms, the decisions will be based on the transmit power required to provide reliable communication between the source and the destination (as discussed in Section 1.4). Estimates of the distances can be obtained using the received signal strength indicator (RSSI) values, which are linked to the transmit power used to reach the desired nodes in the network [54]. By using the transmit power and predicting the propagation characteristics of the wireless signals, nodes are able to determine their distance from the transmitter.

The RSSI values have been used for estimating distances between nodes in a number of applications, and have been shown to have a localization accuracy of a few meters [55], [56]. To achieve this, nodes in such systems periodically exchange beacon packets with a known transmit power, thus allowing all node that receive such packets to update their distance to their neighbours. In this chapter, it will be assumed that the nodes will use a similar method to obtain the interdistance information needed to perform the routing protocols developed.

4.1.2 Relay Strategies and Power Allocation

In Section 1.2.1 the three node wireless relay network was introduced in Fig. 1.1, along with the DF and NC relay strategies. It was shown that NC could provide a 25% savings in terms of the total number of transmissions made by the network to exchange a pair of packets between nodes A and B. When all nodes were using the same transmit power, this translated into a 25% savings when looking at the total network power consumption and a 50% savings when looking solely at node R's power consumption. As the work in this chapter expands beyond a single relay node network, this section will examine the relaying strategies and power savings for a two relay node network and generalize this to an L relay node network.

To demonstrate the power saving with NC in the case of using two relay nodes, we provide in Fig. 4.1 a review of two relay communications. Specifically, when deploying the DF relay strategy as shown in Fig. 4.1 a), six packet transmissions are required to exchange the data between A and B via the relays R and T. When the NC relay strategy is deployed as in Fig. 4.1 b), only four packet transmissions are required. Using the same assumption as before, namely that all nodes broadcast with the same power P_{packet}, then there is a 33% power savings for the entire network when NC is used over DF. From the relay node perspective, it is still a 50% savings for each relay. It should be noted that the wireless NC strategy with two relays requires some additional offsetting of packets, which introduces delay and some marginal overhead [57].

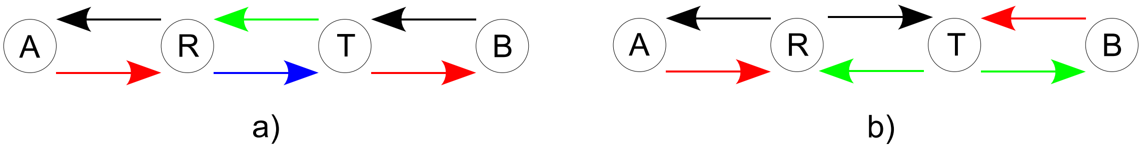

Figure 4.1: Color-coded transmissions in two-relay networks: a) with DF relaying strategy; b) with NC relay strategy

Extending the relay strategies shown in Fig. 4.1 to L relay nodes, it can be observed that for the DF relay strategy a total of $2 \cdot (L + 1)$ broadcasts must be made. For the NC relay strategy, that number drops to $L + 2$. Assuming all nodes broadcast with the same power, the percentage power savings of NC over DF can be expressed mathematically as:

$$100 \cdot \left[1 - \left(\frac{L+2}{2 \cdot (L+1)}\right)\right] . \tag{4.1}$$

Equation (4.1) indicates that as L increases, the percentage saving of NC increases, but the percentage improvement with each successive relay node decreases. Taking the limit of Eqn. (4.1) as $L \to \infty$ shows that the maximum achievable power savings of the NC strategy is 50%.

4.2 Equi-spaced Relay Positions

In this section we demonstrate through the use of geometric principles that working with relay nodes positioned close to the equi-spaced points distributed along a line connecting a pair of communicating nodes provides the largest power savings in relay networks. Results of this geometric abstraction are used as part of a computationally efficient algorithm developed in this work to select the relay nodes that result in the greatest power savings. In general, topology control problems in mesh networks, with the minimization objective being total power, tend to be computationally intractable [53]. It is the geometrical derivations using the Euclidean distance and then the statistical generalization of these observations to the L_P type norms that lead to the computationally amenable routing algorithms developed in this work.

4.2.1 Single Relay Node Forwarding

We start with the case of a single relay node by re-plotting in Fig. 4.2 the communication scenario originally shown in Fig. 1.5 b) with some additional information displayed. To exchange packets between the nodes A and B using DF via the relay node R, the total power invested in free space propagation conditions is given by Eqn. (1.10) which was:

$$P_{AB}^{\mathrm{DF}} \sim 2 \cdot (r_{AR}^2 + r_{RB}^2) \; . \tag{4.2}$$

Using the law-of-cosines and the distance from the mid-point of the line between nodes A and B and the relay node, here labelled as ρ, we can derive that:

$$r_{AR}^2 = \left(\frac{r_{AB}}{2}\right)^2 + \rho^2 - 2\frac{r_{AB}}{2}\rho\cos(\beta) \tag{4.3}$$

and

$$r_{RB}^2 = \left(\frac{r_{AB}}{2}\right)^2 + \rho^2 - 2\frac{r_{AB}}{2}\rho\cos(\pi - \beta) \; . \tag{4.4}$$

With these, Eqn. (4.2) is re-written as:

$$P_{AB}^{\mathrm{DF}} \sim 4 \cdot \left[\left(\frac{r_{AB}}{2}\right)^2 + \rho^2\right] \; , \tag{4.5}$$

which demonstrates that the choice of the relay which minimizes the total power to exchange the packets between nodes A and B is the node that is the closest to the mid-point on a line connecting the two nodes [58].

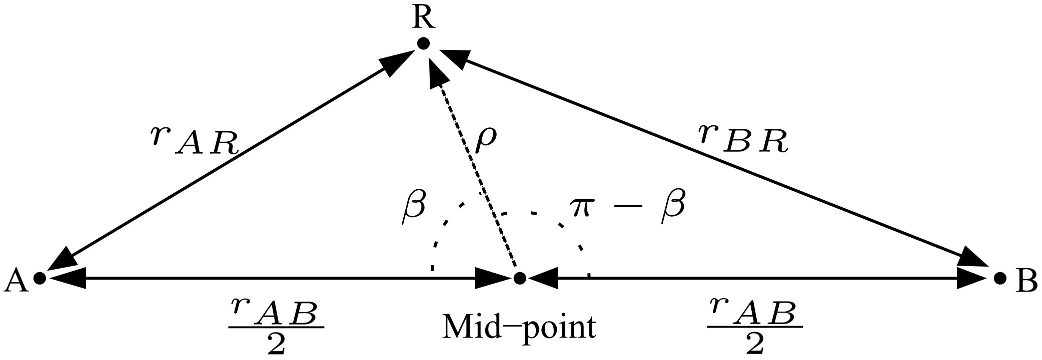

Figure 4.2: Single relay deployment in data exchange between nodes A and B

In our routing protocol, we assume that every node in the network maintains a database of the distances between all of the nodes. With this, the distance of the relay R_i ($i \in \{1, 2, \cdots, N\} - \{A, B\}$) to the mid-point of a given communication between A and B is calculated as:

$$\rho_i^2 = \frac{1}{2}(r_{AR_i}^2 + r_{R_iB}^2 - 0.5r_{AB}^2) \ . \tag{4.6}$$

Finding the minimum power path via a single node R_i using either the minimum for ρ ($\min_i(\rho_i)$) or the minimum power determined by $r_{AR}^2 + r_{RB}^2$ ($\min_i(r_{AR_i}^2 + r_{R_iB}^2)$) are comparable in terms of their computational complexity (both are $\mathcal{O}(N)$). However, it is the interpretation of forwarding via the relay node closest to the mid-point which motivates the further developments in this work.

Focusing now on the exchange of packets between the nodes A and B using NC via the relay node R, the total power invested is given by Eqn. (1.12) which was:

$$P_{AB}^{\mathrm{NC}} \sim r_{AR}^2 + r_{BR}^2 + \max\{r_{AR}^2, r_{BR}^2\} \ . \tag{4.7}$$

While minimizing the energy for the NC strategy with a single relay (when exchanging the data between A and B) using the expression in Eqn. (4.7) by finding which of the $N-2$ relays R_i offers the minimum power, we need to first calculate $\max\{r_{AR_i}^2, r_{BR_i}^2\}$ $N-2$ times. We assert that the relay with the minimum ρ_i^2 will offer a comparable energy efficient solution. This is because:

$$\begin{aligned} P_{AB}^{\mathrm{NC}} &\sim r_{AR_i}^2 + r_{BR_i}^2 + \max\{r_{AR_i}^2, r_{BR_i}^2\} \\ &= r_{AR_i}^2 + r_{BR_i}^2 + 0.5 \cdot (r_{AR_i}^2 + r_{BR_i}^2) + 0.5 \cdot |r_{AR_i}^2 - r_{BR_i}^2| \\ &= 1.5 \cdot (r_{AR_i}^2 + r_{BR_i}^2) + 0.5 \cdot |r_{AR_i}^2 - r_{BR_i}^2| \end{aligned} \tag{4.8}$$

Using expressions for $r_{AR_i}^2$ and $r_{BR_i}^2$ based on Eqns. (4.3) and (4.4) respectively, the term:

$$1.5 \cdot (r_{AR_i}^2 + r_{BR_i}^2) = 3 \cdot \left[\left(\frac{r_{AB}}{2}\right)^2 + \rho^2\right] \tag{4.9}$$

and the term:

$$0.5 \cdot \left|r_{\mathrm{AR_i}}^2 - r_{\mathrm{BR_i}}^2\right| = 0.5 \cdot \left| -2r_{\mathrm{AB}}\rho\cos\left(\frac{\beta}{2}\right)\right| = r_{\mathrm{AB}} \cdot \rho \cdot \left|\cos\left(\frac{\beta}{2}\right)\right| \ . \tag{4.10}$$

When $\rho = 0$, the expression in Eqn. (4.9) is minimized, and the expression in Eqn. (4.10) is zero. Therefor as $\rho \to 0$, the power required for NC shown in Eqn. (4.7) is minimized.

4.2.2 Double Relay Node Forwarding

Motivated by the energy benefits of using a relay close to the mid-point when relaying with a single node, this section elaborates on the positions of two relays which minimizes the total power consumed in the case of free space propagation when $\alpha = 2$.

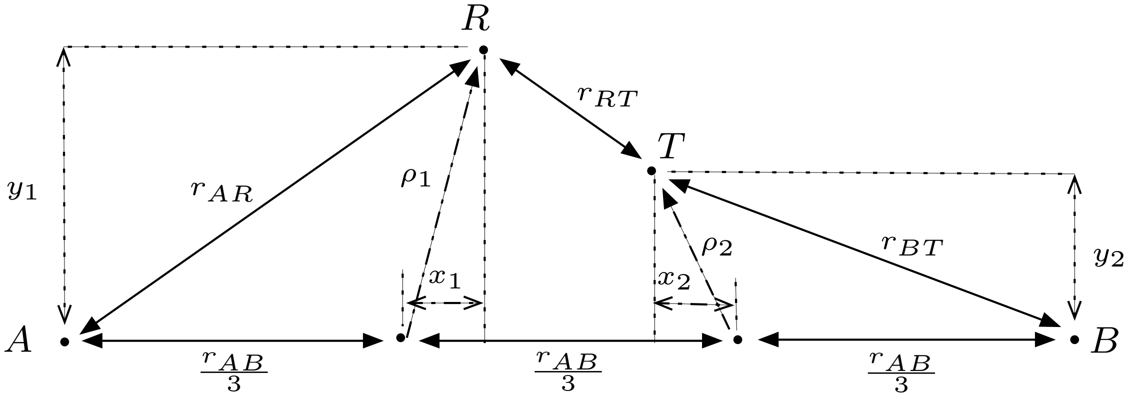

Figure 4.3: Two relay deployment in data exchange between nodes A and B

Using the distances in Fig. 4.3 documenting the relations between node interdistances $(r_{AB}, r_{AR}, r_{RT}, r_{TB})$ and the distances ρ_j $(j = 1, 2)$ from the relay nodes to two mid-points separated by $\frac{r_{AR}}{3}$ from nodes A and B as well as themselves, to exchange packets between nodes A and B via the relay nodes R and T using DF, the total power invested is:

$$P_{AB}^{\mathrm{DF}} \sim 2 \cdot \left(r_{AR}^2 + r_{RT}^2 + r_{TB}^2\right) . \tag{4.11}$$

Using coordinates of the relays (x_j, y_j) and the distances ρ_j, Eqn. (4.11) can be re-written as follows:

$$\begin{aligned} P_{AB}^{\mathrm{DF}} &\sim 2 \cdot \Big(y_1^2 + \left(\tfrac{r_{AB}}{3} + x_1\right)^2 + (y_1 - y_2)^2 + \left(\tfrac{r_{AB}}{3} - x_1 - x_2\right)^2 \\ &\quad + y_2^2 + \left(\tfrac{r_{AB}}{3} + x_2\right)^2 \Big) \\ &= 2 \cdot \left(3 \cdot \left(\tfrac{r_{AB}}{3}\right)^2 + 2 \cdot (\rho_1^2 + \rho_2^2) + 2 \cdot (x_1 x_2 - y_1 y_2) \right) . \end{aligned} \tag{4.12}$$

From the first line in Eqn. (4.12), and noting that squares involving y_i are always non-negative, P_{AB}^{DF} is minimized when $y_1 = y_2 = 0$. If this is the case, Eqn. (4.12) becomes:

$$P_{AB}^{\mathrm{DF}} \sim 2 \cdot \left(3 \cdot \left(\frac{r_{AB}}{3}\right)^2 + x_1^2 + x_2^2 + (x_1 + x_2)^2 \right) \tag{4.13}$$

and with the specific geometry considered herein, and noting that squares involving x_i are always non-negative, this is minimized when $x_1 = x_2 = 0$, meaning the relays minimizing the path energy are at the two equi-spaced points between a source-destination pair. If the relays cannot be positioned at these equi-spaced points, we observed experimentally that the relays closest in terms of ρ_1 and ρ_2 to the equi-spaced points are most likely to minimize the total transmit power over any of the $\binom{N-2}{2} = (N-2) \cdot (N-3)/2$ possible choices of two relays in our network with N nodes. The analytical justification for this fact is from the Shwatrz's inequality $\left|x_1 x_2 - y_1 y_2\right| \leq \rho_1 \cdot \rho_2$ and noting that in Eqn. (4.12), when we minimize ρ_1^2 and ρ_2^2 to minimize P_{AB}^{DF} we are also minimizing the impact of $(x_1 x_2 - y_1 y_2)$. However, since the latter term may be negative, our claim is only statistical in nature.

Finding the node R among $N - 2$ nodes that has the smallest ρ_1 and the node T that has the smallest ρ_2 offers significant computational saving over calculating $r_{AR}^2 + r_{RT}^2 + r_{TB}^2$ for $\binom{N-2}{2}$ possible pairs of two relays. The first is of order $\mathcal{O}(2N)$ while the latter is of order $\mathcal{O}(N^2)$. Additionally, if there are two node such that $\rho_1 = \rho_2 = 0$, the two relay DF will offer its greatest power saving over direct forwarding. Noting from before that the direct forward requires $P_{AB}^{\mathrm{D}} \sim 2 r_{AB}^2$ and from Eqn. (4.11) with all node interdistances being $\frac{r_{AB}}{3}$, then DF with two relays requires $P_{AB}^{\mathrm{DF}} \sim 6(\frac{r_{AB}}{3})^2$ so that the percentage energy savings in the case of free space propagation ($\alpha = 2$) is:

$$100 \cdot \frac{P_{AB}^{\mathrm{D}} - P_{AB}^{\mathrm{DF}}}{P_{AB}^{\mathrm{D}}} = 100 \cdot \frac{2r_{AB}^2 - 6(\frac{r_{AB}}{3})^2}{2r_{AB}^2} = 100 \cdot \Big(1 - \frac{1}{3}\Big) = 66.66\% \ , \tag{4.14}$$

which happens for densely distributed nodes. In case of a single relay, the comparable optimum saving is 50%. It is worth noting that since nodes only discover their interdistances from the control plain information (i.e. $r_{AB}, r_{AR}, r_{RT}, r_{TB}$), the distances ρ_1 and ρ_2 need to be calculated using the cevian triangle theorem credited to Giovanni Ceva [59].

When exchanging data between nodes A and B using NC with two relays, the total required power is:

$$P_{AB}^{\mathrm{NC}} \sim r_{AR}^2 + r_{TB}^2 + \max\{r_{AR}^2, r_{RT}^2\} + \max\{r_{TB}^2, r_{RT}^2\} \ . \tag{4.15}$$

With the proper interpretation of $\max\{\cdot\}$, Eqn. (4.15) is re-written as:

$$\begin{aligned} P_{AB}^{\mathrm{NC}} &\sim r_{AR}^2 + r_{TB}^2 + \tfrac{r_{AR}^2 + r_{RT}^2}{2} + \tfrac{r_{TB}^2 + r_{RT}^2}{2} + \tfrac{|r_{AR}^2 - r_{RT}^2|}{2} + \tfrac{|r_{TB}^2 - r_{RT}^2|}{2} \\ &= \tfrac{3}{2}\Big(r_{AR}^2 + r_{RT}^2 + r_{TB}^2\Big) - \tfrac{r_{RT}^2}{2} + \tfrac{|r_{AR}^2 - r_{RT}^2|}{2} + \tfrac{|r_{TB}^2 - r_{RT}^2|}{2} \ . \end{aligned} \tag{4.16}$$

If the relays could be placed at the equi-spaced points, this would minimize the expression in parenthesis in Eqn. (4.16) the same way it did for P_{AB}^{DF} shown earlier in Eqn. (4.11). This will also minimize the expression involving the magnitudes of the differences. Increasing r_{RT}^2 to maximize the negative term would be offset by the increases in parenthesis, which is why we conclude that the equi-spaced position of relays is also optimal in the case of NC with two relays. Even though derivations were limited to the case of $\alpha = 2$, it is also demonstrated that the path loss exponent and node density minimally affect the power saving in the proposed routing algorithms based on near to equi-spaced distributed relays.

We conclude with the observation that following the derivations for the position of two relays minimizing the total transmission power to exchange the data between nodes A and B, it is possible to use the same arguments to demonstrate that in the case of L relays, the best positions are also equi-spaced with the separation of $\frac{r_{AB}}{L}$ offering the saving of $100 \cdot \left(1 - \frac{1}{L+1}\right)$ over the direct transmission. From this, the more relaying nodes used, the greater the energy saving obtained when DF is used over direct transmission strategy. However, these improvements also follow the law of diminishing returns, so for this work only $L = 1$ or 2 were considered.

4.3 Selection Algorithm

With the results from Section 4.2 in mind, this section outlines the computationally efficient relay selection and power allocation algorithm developed as part of this work. This algorithm first determines a candidate set of potential relay nodes before applying the near equi-spaced concept detailed in Section 4.2.

The first stage of the algorithm developed here, regardless of which relaying strategy is being employed or the number of relay nodes, is to determine a subset of the nodes in the network which would be suitable candidates to act as relays, thus reducing the number of computational operations needed in the following stages. For a given pair of communicating nodes A and B, the nodes are first sorted in order of increasing distance from node A and all nodes that have a distance greater than r_{AB} are discarded. This procedure is then repeated with node B as the starting point and any nodes with a distance greater than r_{BA} are discarded. The common elements of these two lists form the candidate nodes and will be all the nodes that exist in the overlapping region of two circles of radius r_{AB}, one centered at node A and the other at node B. Calculation of the distance to the equi-spaced point or points will then be performed on this candidate subset of nodes using the cevian triangle theorem as in Eqn. (4.6), and those with the minimum distance will be chosen as the relay node(s).

Starting with the single relay case using the DF strategy, the second stage of the algorithm calculates the power required to allow both nodes A and B to communicate using Eqn. (4.2). This value is then compared to the power needed for direct communication and the lesser of the two is chosen. This comparison is done to deal with the case where nodes A and B are extremely close together. Switching to the case of the single relay using the NC strategy, the same relay node as chosen for the DF case is used, but the power to communicate is calculated using Eqn. (4.7). This value is then compared to those of the single relay using DF and the direct method and the lowest of the three is chosen.

For the two relay case using the DF strategy, the second stage of the algorithm calculates the power using Eqn. (4.11). There is, however, the potential that during the selection process the same node was chosen as both relays, meaning that this node was the closest to both of the equi-spaced points. For this case the algorithm treats this scenario as a single relay and calculates the power as before using Eqn. (4.2). For whichever case occurs, the power calculated is compared to the power found using the single relay with DF and the direct method and the lowest of the three is chosen.

Finally, for the two relay case using the NC strategy, the potential for the same node to be chosen as both relays remains. In such an instance the algorithm will revert to the method used in the single relay case and calculate the power using Eqn. (4.7). If the two relays are unique, the power is calculated using Eqn. (4.15). The resultant power in either case is compared with the powers calculated using the two relay with DF, one relay with NC, one relay with DF and direct methods and the lowest of the five is chosen.

Simulations were run to both examine the performance of the above four schemes and compare their performance to that of direct communication. These results are presented in the next section of this chapter.

4.4 Results

To verify the analytical results presented in Section 4.2, simulations were performed using the classic Monte Carlo method in MATLAB®. By generating randomized node positions according to a Poisson point process with different densities, the average energy performance was calculated for the proposed routing protocols where, for every communicating pair, the most energy efficient relays were selected. The results provided are for different numbers of nodes N and for different area sizes.

4.4.1 Validity of Equi-spaced Relay Positions

This section first examines the validity of the choice of the relay nodes that are closest to the equi-spaced points as outlined in Section 4.3. For this simulation, an area of $50m \times 50m$ was chosen along with three different values of the path loss exponent and the node density was varied. The pair of source nodes were placed at opposite diagonal corners of the area under consideration and the remaining nodes were randomly distributed. For both the one and two relay node cases using the DF strategy, both an exhaustive search of the total transmit power and the algorithms developed were used to select the relay nodes. The resulting choices for relays were then compared to determine if the same nodes were chosen as relays and what, if any, difference in power was needed to complete the transmission. This was done for $100,000$ random networks. A summary of the simulation parameters is given in Table 4.1 and the results are displayed in Tables 4.2 and 4.3.

Table 4.1: Parameters for the first relay selection validation simulation

Area used	$50m \times 50m$
Number of random networks created	$100,000$
Values of the path loss exponent α	$2, 4, 6$
Values of the number of nodes in area	$5, 10, 15, 20, 30, 40, 50$

Table 4.2: Probability that nodes chosen closest to equi-spaced points are the same as nodes chosen by an exhaustive power search

	1 Relay			2 Relays		
N	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$
5	100.0%	83.1%	77.2%	56.4%	57.8%	56.6%
10	100.0%	80.7%	73.6%	55.7%	49.3%	45.5%
15	100.0%	80.3%	72.6%	54.7%	47.4%	42.7%
20	100.0%	79.9%	72.1%	54.5%	46.8%	41.9%
30	100.0%	79.6%	71.9%	54.3%	46.7%	41.1%
40	100.0%	79.5%	71.4%	54.4%	46.6%	40.8%
50	100.0%	79.5%	71.1%	54.4%	46.4%	40.7%

Table 4.2 shows the probability that the node(s) closest to the equi-spaced points between two communicating nodes A and B is(are) the same node(s) which result in the minimum transmit power. For the single relay case, there is 100% agreement in the choice of the node when $\alpha = 2$. As α increases, this agreement drops to between 78% and 83% for $\alpha = 4$ and between 71% and 77% for $\alpha = 6$ depending on the node density. Turning to the two relay case, this agreement drops to around 54% for $\alpha = 2$, 46% for $\alpha = 4$ and 41% for $\alpha = 6$, with the node density having a minimal impact on this agreement once a density of 20 nodes is reached.

The more important figure of merit is how much more power is consumed by the routing algorithm developed using the node(s) closest to the equi-spaced points compared to using an exhaustive search algorithm to select the node(s) that would give the minimum power consumption. As noted earlier, the computational complexities of these methods per communication pair are $\mathcal{O}(N)$ for the routing algorithm and $\mathcal{O}(N^2)$ for the exhaustive search. These excess power results are shown in Table 4.3.

In the case of a single relay node, there is no difference in the power consumed to communicate when $\alpha = 2$. With $\alpha = 4$, the percentage of excess power consumed over the minimum value ranges from 4.1% to 0.6%. The same trend exists when $\alpha = 6$, with the excess power starting at 14.5% and dropping to 2.7%. In terms of the

Table 4.3: Percentage of excess power consumed by using nodes chosen closest to equi-spaced points over the minimum transmit power from exhaustive search

	1 Relay			2 Relays		
N	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$
5	0%	4.1%	14.5%	4.2%	10.9%	20.5%
10	0%	2.7%	10.4%	3.5%	16.5%	44.1%
15	0%	2.0%	7.8%	2.4%	12.8%	36.5%
20	0%	1.5%	6.1%	1.8%	10.1%	29.3%
30	0%	1.0%	4.3%	1.3%	7.1%	20.8%
40	0%	0.8%	3.4%	0.9%	5.4%	16.0%
50	0%	0.6%	2.7%	0.8%	4.4%	12.9%

two relay case, a similar trend to that in the one relay case exists, with the percentage of excess power dropping as node density increases indicating improved performance of the algorithms. For $\alpha = 2$ the range is from 4.2% to 0.8%, for $\alpha = 4$ the range is from 10.9% to 4.4% and for $\alpha = 6$ the range is from 44.1% to 12.9%.

As the results presented in Tables 4.2 and 4.3 are for a single average over a group of $100,000$ random networks, they represent the mean or average performance of the method developed here. To provide more statistical data, a second simulation was performed. The parameters of this simulation are the same as that of the previous one with the exception that now only 1000 random networks are generated and the results averaged together. But this process is repeated 1000 times and from this a mean and standard deviation is derived. The parameters are outlined in detail in Table 4.4. The results of this simulation are shown in Tables 4.5 and 4.6.

Table 4.4: Parameters for the second relay selection validation simulation

Area used	$50m \times 50m$
Number of random networks averaged	1000
Number of iterations run	1000
Values of the path loss exponent α	$2, 4, 6$
Values of the number of nodes in area	$5, 10, 15, 20, 30, 40, 50$

Table 4.5 shows the statistics for the probability that the algorithms developed here chooses the same node(s) as compared to an exhaustive search. For each value of α the mean and standard deviation are presented. Comparing the means in this table to the values presented in Table 4.2 one can see that the agreement between the values is very good and in the worst case scenario there is less then a half a percentage point difference. Turning to the values of the standard deviation, for the single relay case where $\alpha = 2$ there is no deviation. For all other cases the standard deviation is around 1.5% indicating that their not a lot of variation in the performance results of the proposed algorithms.

Table 4.5: Statistical results for probability of choosing the same node(s)

	1 Relay					
	$\alpha = 2$		$\alpha = 4$		$\alpha = 6$	
N	Mean	Std	Mean	Std	Mean	Std
5	100.00%	0.00%	83.02%	1.14%	77.43%	1.29%
10	100.00%	0.00%	80.72%	1.19%	73.58%	1.38%
15	100.00%	0.00%	80.09%	1.17%	72.61%	1.39%
20	100.00%	0.00%	79.94%	1.28%	72.14%	1.45%
30	100.00%	0.00%	79.67%	1.33%	71.53%	1.43%
40	100.00%	0.00%	79.57%	1.31%	71.36%	1.41%
50	100.00%	0.00%	79.47%	1.26%	71.26%	1.41%
	2 Relays					
	$\alpha = 2$		$\alpha = 4$		$\alpha = 6$	
N	Mean	Std	Mean	Std	Mean	Std
5	56.62%	1.61%	57.60%	1.53%	56.62%	1.56%
10	55.78%	1.54%	49.26%	1.61%	45.40%	1.59%
15	54.84%	1.56%	47.60%	1.57%	42.71%	1.52%
20	54.44%	1.57%	47.14%	1.55%	41.82%	1.57%
30	54.42%	1.58%	46.74%	1.56%	41.17%	1.51%
40	54.39%	1.56%	46.68%	1.52%	40.82%	1.52%
50	54.37%	1.51%	46.61%	1.57%	40.68%	1.60%

Turning to the results for the percentage of excess power consumed, looking at Table 4.6 and comparing the mean values to the results in Table 4.3, one again sees that there is very good agreement. For the worst case scenario there is just about 0.3% difference and in most cases it is significantly less than that. Looking at the values of the standard deviation for the single relay case there is again no deviation when $\alpha = 2$ and for the remainder the deviation is generally less than half a percent. For the two relay cases the standard deviation does grow as α is increased but still is very low for higher densities.

Table 4.6: Statistical results for percentage of excess power consumed

	1 Relay					
	$\alpha = 2$		$\alpha = 4$		$\alpha = 6$	
N	Mean	Std	Mean	Std	Mean	Std
5	0.00%	0.00%	4.14%	0.41%	14.58%	1.30%
10	0.00%	0.00%	2.69%	0.26%	10.60%	0.89%
15	0.00%	0.00%	1.96%	0.19%	7.83%	0.61%
20	0.00%	0.00%	1.51%	0.15%	6.18%	0.52%
30	0.00%	0.00%	1.05%	0.11%	4.37%	0.36%
40	0.00%	0.00%	0.81%	0.08%	3.35%	0.28%
50	0.00%	0.00%	0.65%	0.07%	2.72%	0.22%
	2 Relays					
	$\alpha = 2$		$\alpha = 4$		$\alpha = 6$	
N	Mean	Std	Mean	Std	Mean	Std
5	4.20%	0.29%	10.89%	0.75%	20.40%	1.84%
10	3.50%	0.19%	16.65%	0.97%	44.08%	3.09%
15	2.42%	0.13%	12.75%	0.76%	36.21%	2.40%
20	1.84%	0.11%	10.08%	0.59%	29.13%	1.88%
30	1.24%	0.07%	7.06%	0.41%	20.66%	1.29%
40	0.94%	0.05%	5.44%	0.31%	15.89%	0.94%
50	0.76%	0.04%	4.40%	0.27%	12.96%	0.76%

4.4.2 Average Power

Simulations were run to both examine the performance of the four schemes developed as part of the routing algorithms presented in this work and to compare their performance to that of direct communication. The total power required to provide full communication between every node (a full mesh network) for each of the five communication methods was calculated. To provide more realistic results, it was assumed that the power required at the receiver to have reliable communication (a BER of 10^{-5}) was $-20dBm$. Therefore, Eqn. (1.8) becomes $P_{\text{Tx}}(r) \sim \frac{r^{\alpha}}{10^{5}}W$. The parameters are outlines in Table 4.7.

Table 4.7: Parameters for the simulations comparing the four schemes

Areas used	$25m \times 25m$, $50m \times 50m$, $100m \times 100m$
Number of random networks averaged	$100,000$
Values of the path loss exponent α	$2, 4, 6$
Values of the number of nodes in area	$5, 10, 15, 20, 40, 60, 80, 100$
Required received power	$-20dBm$

This section reports on the results for the average power per transmission. First an examination of the effect that the area the network covers will be presented. Figures 4.4, 4.5, and 4.6 plot the average power per transmission versus number of nodes for network area sizes of 25m × 25m, 50m × 50m, and 100m × 100m respectively. For these plots the path loss exponent is fixed at $\alpha = 2$.

Looking specifically at Fig. 4.5 it can be seen that all four methods generate a lower average transmission power, compared to the direct communication method which remains relatively constant over all node densities. The NC strategies outperform their DF counterparts for all node densities and both of the the two relay methods outperform the single relay methods when 20 or more nodes are present.

Turning to the matter of the impact network area has on the average power per transmission, comparing these three figures reveals that the curves appear to be identical in shape. The only difference is the values on the scale for the power. As the area increases, naturally the distances between the nodes will increase and so will the power required to communicate between nodes. But this is simply a matter of

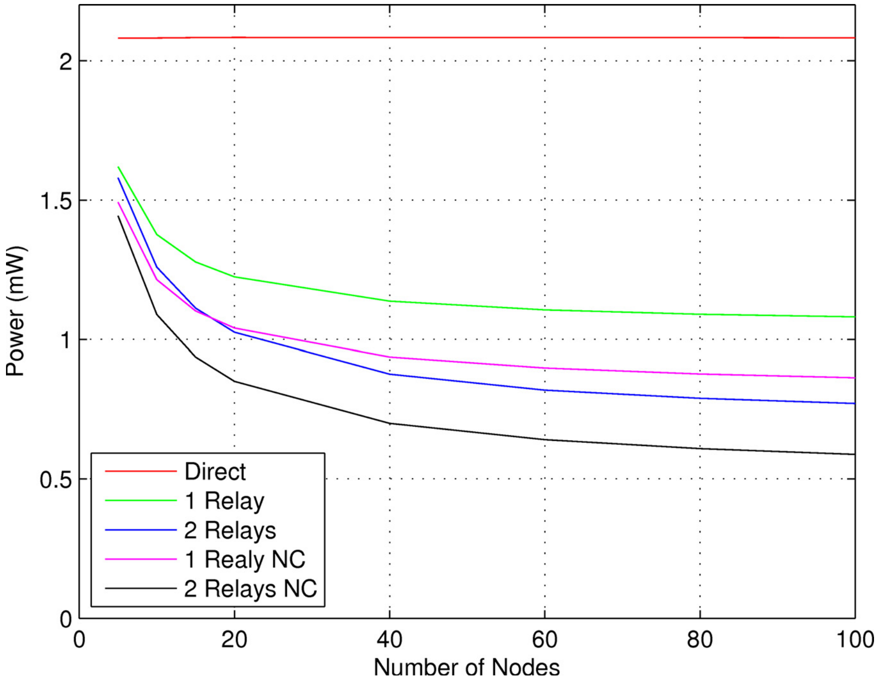

Figure 4.4: Average power per transmission, area 25m × 25m, $\alpha = 2$

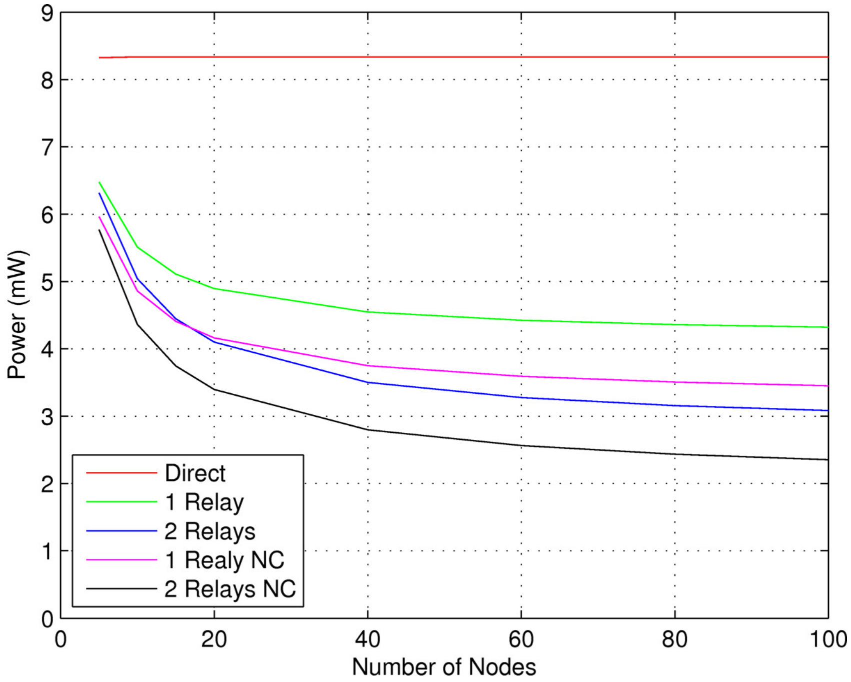

Figure 4.5: Average power per transmission, area 50m × 50m, $\alpha = 2$

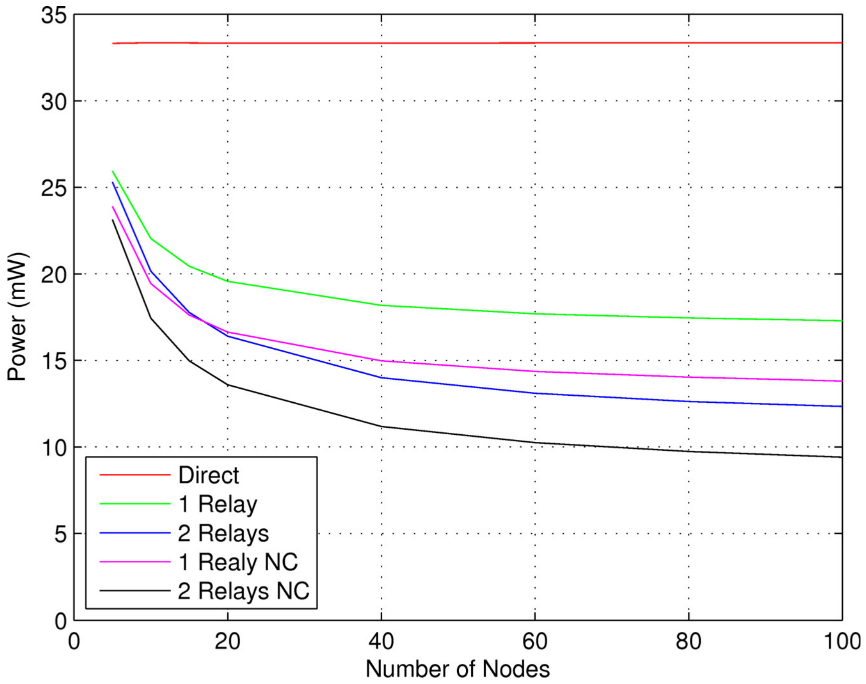

Figure 4.6: Average power per transmission, area 100m × 100m, $\alpha = 2$

scaling the power values to adjust for the changes in area, so the overall performance increase does not change. As such, area has no impact on the performance of the routing algorithms presented.

Concerning the matter of the effect the path loss parameter α has on the average power per transmission, Figs. 4.7 and 4.8 plot the average power for an area of 50m × 50m and α values of 4 and 6 respectively. Comparing these two plots with Fig. 4.5 it can be observed that as α increases, the average power values increase dramatically. This is to be expected though as power is calculated based on Eqn. (1.8) which is distance to the power α.

Looking at the shapes of the curves, the same performance improvements can be seen among the three graphs as was the case with the variation in network area. The NC strategies outperform their DF counterparts for all node densities, but now both of the two relay methods outperform the single relay methods for all node densities.

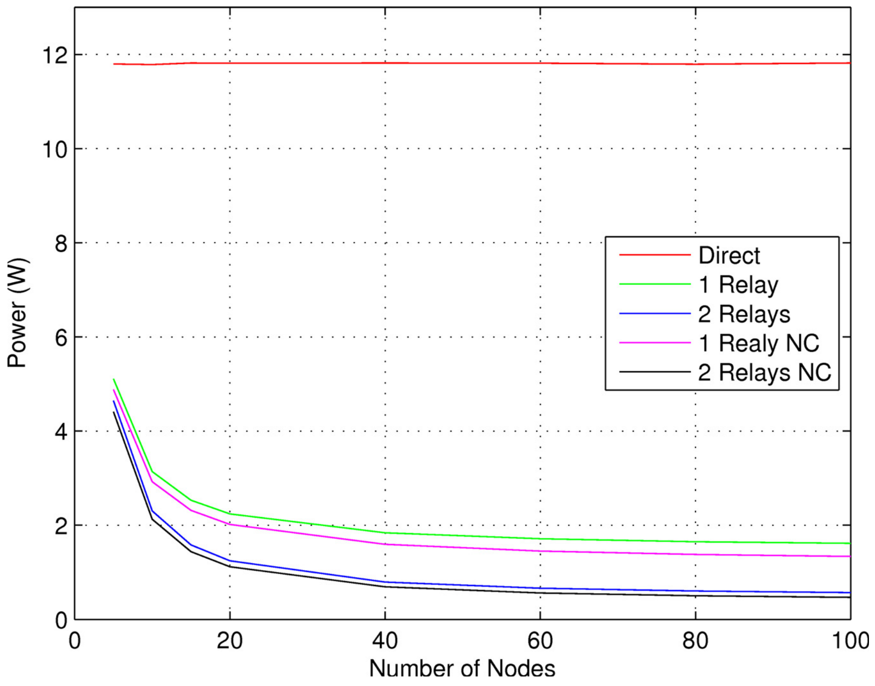

Figure 4.7: Average power per transmission, area 50m × 50m, $\alpha = 4$

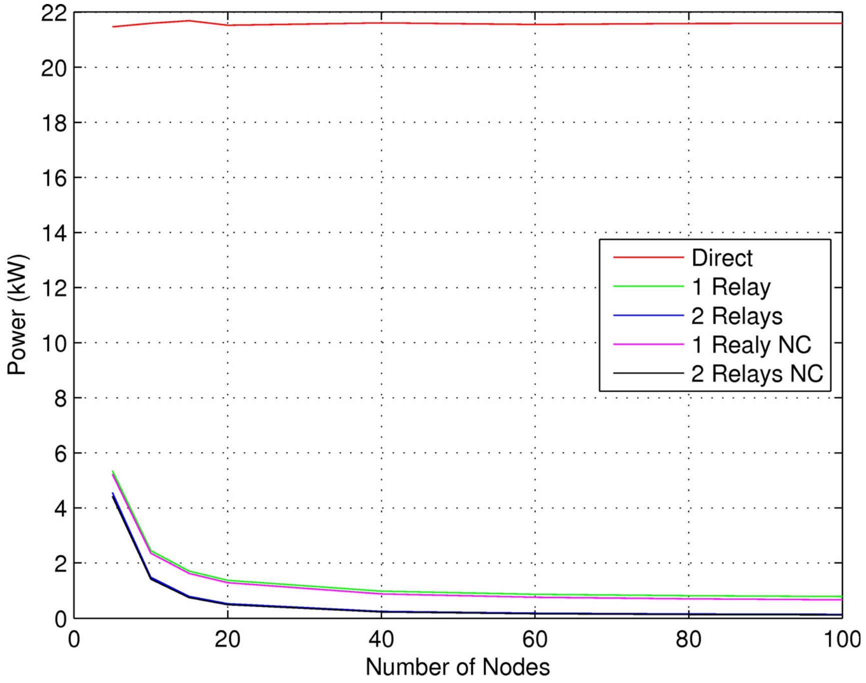

Figure 4.8: Average power per transmission, area 50m × 50m, $\alpha = 6$

As α increases, the relative performance gain of the NC strategy over the DF strategy decreases. While this is not ideal, the four methods presented still outperform the direct method by a significant margin.

All four methods demonstrate a plateau effect as node density increases. This is due to the fact that as the node density increases, the probability of a node or nodes lying very close to the equi-spaced point(s) increases. Another way to interpret these results, which would allow the removal of the effect the size of the area has on the average power, is to compare the performance of each method to that of the direct method and calculate the percentage savings. This will be presented next.

4.4.3 Power Savings

Figure 4.9 shows the percentage savings for each of the four methods over the direct method for the same results plotted in Fig. 4.5. One can observe that for low node densities, all four methods offer in the range of a 25% savings over direct communication. As the node density increases, these savings increase significantly, with the two relay with NC strategy offering the largest savings at around 72%. The power savings between the four methods also varies significantly for higher node densities, with the single relay method using the DF strategy offering the lowest savings at approximately 48%.

This graph was, of course, for an α value of 2. To see the effect that α has on the performance, simulations were run for α values of 4 and 6. Figure 4.10 shows the power savings for the four methods when $\alpha = 4$ and Fig. 4.11 for when $\alpha = 6$.

It can be seen that the performance gains grow as α increases and the plateau effect occurs at lower node densities. The gain in performance of the NC strategy is diminished somewhat, but they do still provide some performance gain over their DF counterparts. To see how these performance graphs compare to the maximum power savings that could be achieved if one was able to place the relay node(s) at the

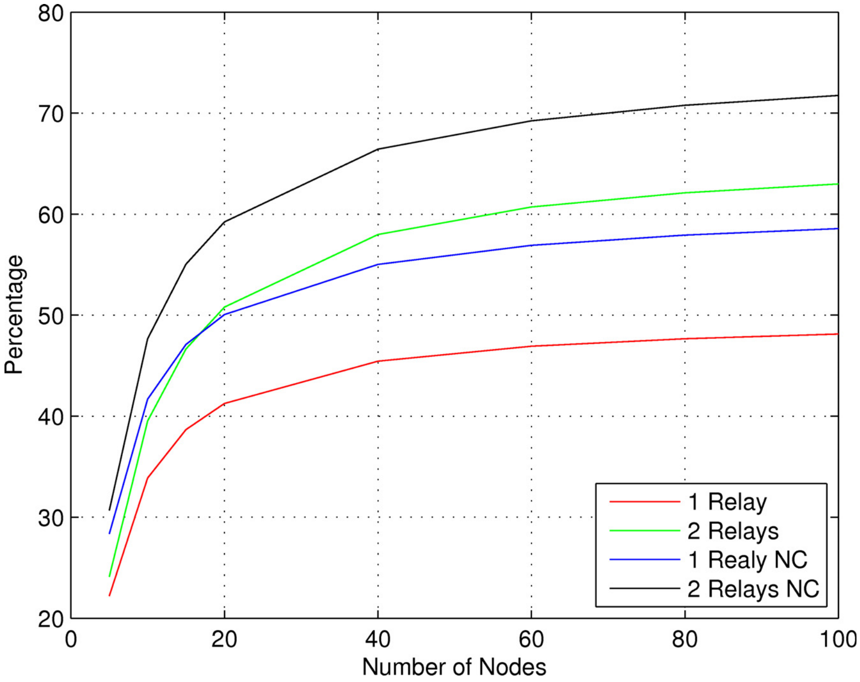

Figure 4.9: Percentage power savings per transmission, area 50m × 50m, $\alpha = 2$

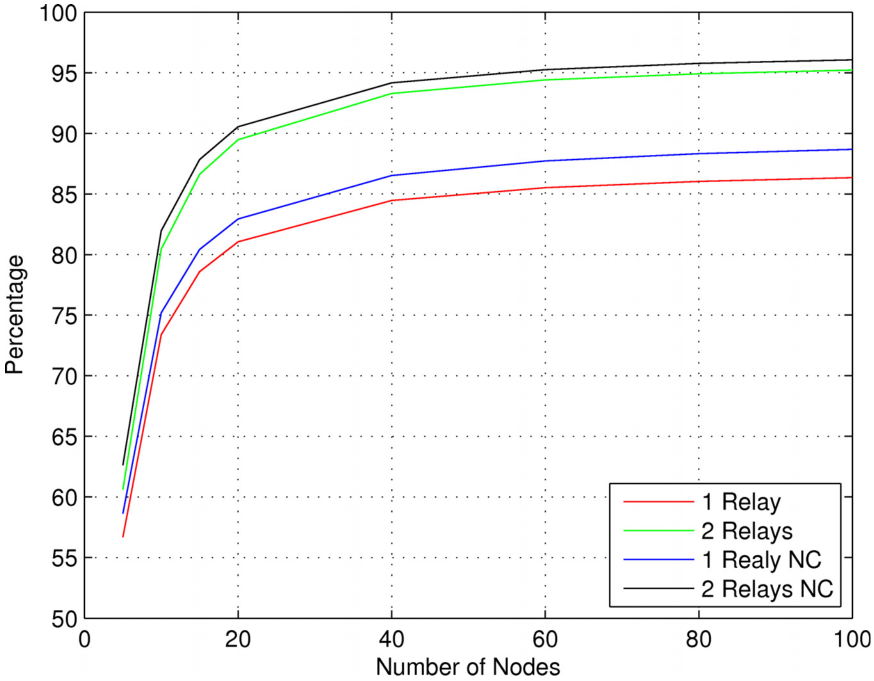

Figure 4.10: Percentage power savings per transmission, area 50m × 50m, $\alpha = 4$

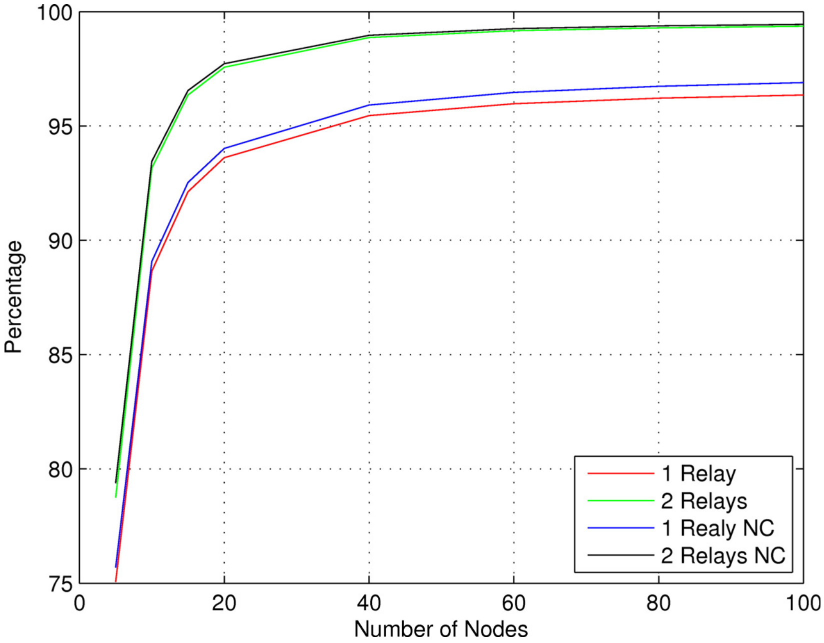

Figure 4.11: Percentage power savings per transmission, area 50m × 50m, $\alpha = 6$

equi-spaced point, the maximum savings over direct communication is derived next.

To calculate the maximum performance gain possible using the various methods shown in the plots above, it is assumed that the source and destination nodes are separated by a distance r_{AB} and the scaling factor for the power will be ignored so that the power needed to reach the destination successfully will be represented by r^{α}.

Based on this, the power required for bi-directional communication using the direct method would be $2 * r_{AB}^{\alpha}$, using a single relay node at the mid-point with DF would be $4 * (\frac{r_{AB}}{2})^{\alpha}$ and using two relay nodes evenly spaced with DF would be $6 * (\frac{r_{AB}}{3})^{\alpha}$. Applying the network coding methods outlined in sections 1.2.1 and 4.1.2, the power for one relay drops to $3 * (\frac{r_{AB}}{2})^{\alpha}$ and for two relays it is $4 * (\frac{r_{AB}}{3})^{\alpha}$. The results of comparing the four relay methods to the direct transmission in terms of their maximum power savings are shown in Table 4.8.

Recalling the performance plots in Figs. 4.9, 4.10, and 4.11 and comparing the values the various methods obtained to the values in Table 4.8, it can be observed

Table 4.8: Maximum power savings per transmission over direct communications

	$\alpha = 2$	$\alpha = 4$	$\alpha = 6$
1 Relay	50%	87.5%	96.88%
1 Relay NC	62.5%	90.63%	97.66%
2 Relays	66.67%	96.30%	99.59%
2 Relays NC	77.78%	97.53%	99.73%

that the proposed methods achieve near optimal performance for reasonable node densities.

4.5 Summary

This work developed an algorithm with a low computational complexity for interactions between topology control and routing layers using distance information between nodes. The proposed algorithm selects relays close to equi-spaced points on a line connecting the two nodes exchanging data for DF and NC relaying strategies. The objective to minimize the total transmit power for data dissemination in wireless networks with full mesh connectivity and random position of nodes has been achieved with a small energy penalty as compared to the exhaustive search routing algorithm utilizing dynamic power allocation. The savings over direct transmissions have been verified in different propagation conditions and node densities.

Chapter 5

Conclusions and Future Work

In this chapter, the contributions and conclusions of this dissertation are presented, along with suggestions for future work related to this field. After a brief overview of the research scope and methodologies within this dissertation, Section 5.1 details the major contributions and conclusions of this dissertation, while Section 5.2 suggests potential future work.

Network coding can improve the throughput, reduce delay, and increase the power efficiency of modern communication systems in ways that have not yet been fully realized. Specifically, its ability to combine information packets from different, previously independent data flows, offers optimization of networking functions which outperform traditional routing approaches. Notwithstanding this potential, that has only recently been recognized, there are still some limitations that must be addressed before practical implementations can be developed. Additionally there are a lot of opportunities that have been underutilized when it comes to integrating NC with other communication techniques and this is a motivation behind the research in this dissertation. This is especially evident when considering communication networks carrying real-time application data, as such networks are sensitive to delay and packet losses, to the extent that they can result in unacceptable QoS or even the inability to support some applications.

The notion of coding at the packet level has attracted significant interest since the publication of [4], which showed the utility of NC for multicast in wired packet networks. But the utility of NC reaches much further as demonstrated in this dissertation. In particular, the work in this dissertation developed novel approaches to applying NC to much wider networking scenarios and the performance gains of these approaches were demonstrated using both analytical methods and simulations.

Packet level NC was used to improve the throughput and delay for relay communication networks. Innovative solutions to enhance the performance of NC by designing cross-layer protocols incorporating EC and adaptive power allocation were also explored. The results presented form a framework for the development of future networking advancements, and are applicable to both the wired and wireless networking environments. The author believes that the exploration into integration of NC with other networking techniques within the scope of cross-layer design and cooperative communications done in this dissertation could shed light on future applications of NC.

5.1 Dissertation Contributions and Conclusions

The primary contributions of this dissertation are summarized as follows:

1. A framework for the combining of network coding and erasure coding to improve network throughput and data reliability in wired networks was developed. A scheme using hop-by-hop erasure recovery was implemented and a performance analysis was presented. Several sub-optimal schemes were also developed which provided some improvement over an unprotected network while successfully balancing the amount of network resources needed to implement erasure coding.

2. Two methods to reduce or limit the delay experienced by packets when network coding is applied in wireless relay networks where asynchronous traffic is present were developed. Several schemes were fashioned based on these two methods and their performances analyzed. A novel opportunistic method for applying erasure coding to provide improved data reliability was also produced.

3. The development and analysis of an efficient, near optimal algorithm for the selection of relay nodes and adaptive broadcast power level allocations to minimize the total transmit power in network coded wireless relay networks was presented.

The content of the individual chapters of this dissertation is summarized as follows:

Chapter 2 first examines the impact that PLR has on the reliability and effectiveness of NC when applied in wired networks. To combat the degradation introduced by the PLR, the addition of EC to the NC framework is discussed. This work extends the traditional use of EC from end-to-end data protection to hop-by-hop protection. A systematic parity check erasure code which generates $r = 1$ redundancy packets from k information packets is used, as it is observed that this code is highly compatible with the operations used in packet level NC. This observation allowed the development of a method to combine NC and EC which requires minimal computational effort from the relaying nodes in the network.

An optimal method where all nodes performed erasure recovery and several suboptimal methods where a subset of the network nodes perform erasure recovery were developed. Analytical results were obtained for the application of all the developed methods based on their application within the accepted standard butterfly network. Simulation results using the standard Monte Carlo approach are also presented to verify the analytical results. All methods provide an improved reliability in information packet delivery compared to an unprotected network coded network.

Chapter 3 develops two approaches to restrict the delay experienced by packets when NC is applied to wireless relay networks with asynchronous traffic. One approach is to limit by time, while the other is to limit by number of packets. From these two approaches, three methods are outlined which are applied at the relaying nodes to reduce the amount of delay experienced by packets and increase the predictability of the delay.

A novel approach to utilizing EC is also developed in this chapter. The methods developed to control the delay introduced unused capacity into the NC scheme by forcing the relay nodes to flush packets before they could be network coded. This unused capacity is utilized to carry EC recovery information, allowing the receiver nodes to recover some lost packets, thus reducing the need for retransmissions. Simulation results that demonstrate both the reduction in the delay experienced by packets and the improved packet reliability are presented.

Chapter 4 develops a computationally efficient algorithm for the selection of relay nodes in a wireless relay network. It utilizes adaptive power control to adjust the transmit radii of the network nodes to control the topology and minimize the total transmit power of the network. Geometric analysis is done to show that the optimum position for relay nodes is at equi-spaced points along the line of transmission between a source and a destination.

Knowing the power needed to reach a nodes, and thus also knowing an estimate of the distance to that node, cevian triangle theorem is used to develop a computationally efficient algorithm to determine which nodes are closest to these equi-spaced points, and thus should act as relays for that particular source-destination pair. The algorithm is fully developed for both the single relay and double relay cases and is shown to be expandable to the L relay case. Both analytical and simulation results are presented to compare the developed algorithm's performance to the exhaustive search method (which guaranties the optimal power efficient solution will be found).

5.2 Suggested Future Work

In this section, related topics for future research are suggested.

1. Applying the Developed Algorithms and Methods in a Real Network Environment

In this dissertation, all algorithms and methods were analyzed first using mathematical tools and then verified through computer simulations. Further work could be the deployment of these algorithms and methods in a real network environment, using real software and hardware implementations. For example the NC delay limiting methods developed in Chapter 3 could be implemented on a WSN platform like the one offered by Crossbow [60]. Evaluation of the scalability of the algorithms and methods should also be conducted.

2. Using More Powerful Erasure Codes

The erasure code used in this dissertation had an erasure correction capability of one ($e = 1$). This decision was based on the observation that the mechanism to implement this parity check erasure code was compatible with that of the linear packet level network coding algorithm used and is amenable to computationally efficient implementations. Many erasure codes exist which possess higher erasure correction capabilities and the integration of such codes into the framework developed here should be explored. The use of random networking coding as opposed to linear network coding should also be investigated.

3. Effects of Imperfect Channel Information

An assumption made in Chapter 4 of this dissertation was the availability at all nodes of accurate information regarding the transmission power required for every node to reach every other node in a network. A further assumption was made in this dissertation that this information was readily available and up to date. Investigations into the updating and sharing of this link state information between nodes, including time between updates and overhead associated with the information exchange, should be conducted. The statistical characteristics of the wireless channel environment for a realistic networking scenario should also be evaluated to determine their impact on the accuracy of the power (and therefore distance) values obtained compared to those based on the assumption of a homogenous wireless medium. In considering the complexity of the statistical variations in the radio signals, such research would need to be conducted using a real-world WSN prototype testbed.

Bibliography

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows - Theory, Algorithms, and Applications.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1993.

[2] J. F. Kurose and K. W. Ross, *Computer Networking - A Top-Down Approach Featuring the Internet.* Boston, MA: Addison Wesley Longman, Inc., 2001.

[3] W. Stallings, *Data and Computer Communications*, 9th ed. Upper Saddle River, NJ: Prentice Hall, Inc., 2011.

[4] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, iss. 4, pp. 1204 – 1216, July 2000.

[5] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory*, ser. Foundations and Trends in Communications and Information Theory. Boston, MA: Now Publishers Inc., 2006.

[6] C. Fragouli and E. Soljanin, *Network Coding Fundamentals*, ser. Foundations and Trends in Networking. Boston, MA: Now Publishers Inc., 2007.

[7] R. W. Yeung, "Network coding: A historical perspective," *Proc. IEEE*, vol. 99, iss. 3, pp. 366 – 371, March 2011.

[8] S.-Y. R. Li, Q. T. Sun, and Z. Shao, "Linear network coding: Theory and algorithms," *Proc. IEEE*, vol. 99, iss. 3, pp. 372 – 387, March 2011.

[9] K. Sayood, *Introduction to Data Compression*, 2nd ed. San Francisco, CA: Morgan Kaufmann Publishers, 2000.

[10] W. A. Pearlman and A. Said, *Digital Signal Compression : Principles and Practice.* New York, NY: Cambridge University Press, 2011.

[11] D. Petrovic, K. Ramchandran, and J. Rabaey, "Overcoming untuned radios in wireless networks with network coding," *IEEE Trans. Inform. Theory*, vol. 52, iss. 6, pp. 2649 – 2657, June 2006.

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

[13] M. Yang and Y. Yang, "Peer-to-peer file sharing based on network coding," in *Proceedings of the 28th International Conference on Distributed Computing Systems (ICDCS)*, Beijing, China, June 2008, pp. 168 – 175.

[14] K. Bhattad and K. R. Nayayanan, "Weakly secure network coding," in *The First Workshop on Network Coding, Theory, and Applications (NetCod 2005)*, Riva del Garda, Italy, April 2005, pp. 1 – 6.

[15] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. Karger, "Byzantine modification detection in multicast networks using randomized network coding," in *Proceedings of the International Symposium on Information Theory (ISIT)*, Chicago, IL, June 2004, p. 143.

[16] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," *IEEE/ACM Trans. Networking*, vol. 16, iss. 3, pp. 497 – 510, June 2008.

[17] M. Ghaderi, D. Towsley, and J. Kurose, "Reliability gain of network coding in lossy wireless networks," in *Proceedings of the 27th IEEE Conference on Computer Communications (INFOCOM)*, Phoenix, AZ, April 2008, pp. 2171 – 2179.

[18] D. J. MacKay, "Fountain codes," *IEE Proceedings - Communications*, vol. 152, iss. 6, pp. 1062 – 1068, December 2005.

[19] M. Luby, "LT codes," in *Proceedings of the 43rd Annual IEEE Symposium on the Foundations of Computer Science*, Vancouver, BC, Canada, November 2002, pp. 271 – 280.

[20] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inform. Theory*, vol. 52, iss. 6, pp. 2551 – 2567, June 2006.

[21] R. Ma and J. Ilow, "Regenerating nodes for real-time transmissions in multi-hop wireless networks," in *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN)*, Tampa, FL, November 2004, pp. 378 – 384.

[22] R. Coolen, "Enhancing the performance of wireless networks with packet level forward error correction," Master's thesis, Elec. and Comp. Eng. Dept., Dalhousie University, Halifax, NS, Canada, December 2006.

[23] M. Medard and A. Sprintson, Eds., *Network Coding - Fundamentals and Applications.* Boston, MA: Elsevier Inc., 2012.

[24] H. Zhao and J. Ilow, "Adaptive resource allocation in OFDMA relay networks with network coding," in *Proceedings of the 2011 IEEE International Conference on Communications (ICC)*, Kyoto, Japan, June 2011, pp. 1 – 5.

[25] S. H. Melvin and J. Ilow, "Opportunistic network and erasure coding for asynchronous two-way relay networks," in *Proceedings of the IEEE Global Communications Conference 2012 (GLOBECOM)*, Anaheim, CA, December 2012.

[26] S. H. Melvin and J. Ilow, "Relay selection in network coded communications with power control," in *Proceedings of the 2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, June 2012, pp. 697 – 702.

[27] S. H. Melvin and J. Ilow, "Equi-spaced relay selection in cooperative communication networks with power control," in *Proceedings of the IFIP IEEE Wireless Days 2011*, Niagara Falls, ON, Canada, October 2011, pp. 1 – 3.

[28] S. H. Melvin and J. Ilow, "Network and erasure coding for improved packet delivery," in *Proceedings of the IEEE 4th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Gold Coast, Australia, December 2010, pp. 1 – 5.

[29] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: Analog network coding," in *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (ACM SIGCOMM)*, Kyoto, Japan, August 2007, pp. 397 – 408.

[30] H. Wang, Y. Wang, and W. Wang, "A relay-based retransmission scheme via analog network coding in wireless cooperative communication systems," in *Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM)*, Taiyuan, China, October 2010, pp. V8–595 – V8–598.

[31] M. N. Khormuji and M. Skoglund, "Noisy analog network coding for the two-way relay channel," in *Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT)*, Saint-Petersburg, Russia, August 2011, pp. 2065 – 2069.

[32] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, 2nd ed. Boston, MA: McGraw-Hill, 2004.

[33] H.-C. Le, H. Guyennet, and N. Zerhouni, “A new contention access method for collision avoidance in wireless sensor networks,” in *Proceedings of the 6th International Conference on Networking (ICN)*, Sainte-Luce, Martinique, April 2007, pp. 1 – 7.

[34] C. Jandaeng, W. Suntiamontut, and N. Elz, “Throughput improvement of collision avoidance in wireless sensor networks,” in *Proceedings of the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, Chengdu, China, September 2010, pp. 1 – 5.

[35] F. Rossetto and M. Zorzi, “On the design of practical asynchronous physical layer network coding,” in *Proceedings of the IEEE 10th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Perugia, Italy, June 2009, pp. 469 – 473.

[36] H.-M. Wang, X.-G. Xia, and Q. Yin, “A linear analog network coding for asynchronous two-way relay networks,” *IEEE Trans. Wireless Commun.*, vol. 9, iss. 12, pp. 3630 – 3637, December 2010.

[37] M. Yang and Y. Yang, “A hypergraph approach to linear network coding in multicast networks,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, iss. 7, pp. 968 – 982, July 2010.

[38] S.-Y. R. Li, Q. T. Sun, and Z. Shao, “Linear network coding: Theory and algorithms,” *Proc. IEEE*, vol. 99, iss. 3, pp. 372 – 387, March 2011.

[39] N. Cai and R. W. Yeung, “Network coding and error correction,” in *Proceedings of the 2002 IEEE Information Theory Workshop*, Bangalore, India, July 2002, pp. 119 – 122.

[40] D. E. Comer, *Computer Networks and Internets with Internet Applications*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 2001.

[41] S. Lin and D. J. Costello, *Error Control Coding - Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1983.

[42] F. Ayres, *Theory and Problems of Matrices*, ser. Schaum’s Outline Series in Mathematics. New York, NY: Schaum Publishing Co., 1962.

[43] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed. New York, NY: Springer-Verlag New York, Inc., 2003.

[44] D. S. Dummit and R. M. Foote, *Abstract Algebra*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1998.

[45] A. F. Molisch, *Wireless Communications*, 2nd ed. West Sussex, UK: Wiley, 2011.

[46] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Upper Saddle River, NJ: Prentice-Hall, Inc., 1996.

[47] P. Santi, *Topology Control in Wireless Ad Hoc and Sensor Networks.* West Sussex, UK: John Wiley & Sons, Ltd, 2005.

[48] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proceedings of the 41st Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, October 2003, pp. 40 – 49.

[49] S. M. Ross, *Introduction to Probability Models*, 9th ed. Orlando, FL: Academic Press, Inc., 2006.

[50] H. C. Tijms, *Stochastic Models: An Algorithmic Approach.* West Sussex, UK: John Wiley & Sons, Ltd, 1994.

[51] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Hoboken, NJ: John Wiley & Sons, Ltd, 2005.

[52] J. Rudnick and G. Gaspari, *Elements of the Random Walk.* Cambridge, UK: Cambridge University Press, 2004.

[53] L. Gavrilovska and R. Prasad, *Ad Hoc Networking Towards Seamless Communications.* Dordrecht, Netherlands: Springer, 2006.

[54] F. Fitzek and M. Katz, Eds., *Cooperation in Wireless Networks: Principles and Applications.* Dordrecht, Netherlands: Springer, 2006.

[55] M. L. Sichitiu and V. Ramadurai, "Localization of wireless sensor networks with a mobile beacon," in *Proceedings of the 2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS)*, Fort Lauderdale, FL, October 2004, pp. 174 – 183.

[56] E. Goldoni, A. Savioli, M. Risi, and P. Gamba, "Experimental analysis of RSSI-based indoor localization with IEEE 802.15.4," in *Proceedings of the 2010 European Wireless Conference (EW)*, Lucca, Italy, April 2010, pp. 71 – 77.

[57] H. Zhao and J. Ilow, "Network coding in two-relay OFDMA networks using re-initialized transmission scheme," in *Proceedings of the 8th Annual Communication Networks and Services Research Conference (CNSR2010)*, Montreal, QC, Canada, 2010, pp. 23 – 28.

[58] S. Rangappa, “Cocasting and power control for energy efficient information dissemination in WSNs,” Master’s thesis, Elec. and Comp. Eng. Dept., Dalhousie University, Halifax, NS, Canada, July 2010.

[59] R. Honsberger, *Episodes in Nineteenth and Twentieth Century Euclidean Geometry.* Washington, DC: The Mathematical Association of America, 1995.

[60] Crossbow. (2012, June 1) WSN classroom kit - crossbow technology. [Online]. Available: http://bullseye.xbow.com:81/Products/productdetails.aspx?sid=233