A Study on Throughput and Delay Performance Analysis in Two-Hop Relay Mobile Ad Hoc Networks

著者	Liu Jiajia
学位授与機関	Tohoku University
URL	http://hdl.handle.net/10097/54572

A Study on Throughput and Delay Performance Analysis in Two-Hop Relay Mobile Ad Hoc Networks

by

Jiajia Liu

Submitted to
Tohoku University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Graduate School of Information Sciences
Tohoku University

September, 2012

To My Family

Abstract

With the rapid development of communication technologies, the wireless networks nowadays are increasingly becoming irreplaceable communication techniques for the next generation networks. Due to its distinctive features of robustness, self-organization, quick deployment and reconfiguration, mobile ad hoc network (MANET) appears attractive for lots of application scenarios, such as disaster relief, emergency response, daily information exchange, military troop communication, pedestrian network and vehicular ad hoc network, etc.

However, there is still a long way to go before MANETs could be widely commercialized and implemented. The very roadblock that has been stunting the application of MANETs is the lack of a general network information theory, which is expected to establish a thorough understanding on the fundamental performances in such networks, like the throughput capacity, delivery delay, end-to-end delay, etc. The available works on this line mainly focus on deriving order sense results, which are helpful for us to explore the general scaling laws of throughput capacity and delay but tell us little about the exact throughput and delay performances there. Towards such a target, we develop theoretical frameworks to analytically study the MANET throughput and delay performances in this thesis. Specifically, we focus on an important class of MANETs—the two-hop relay MANETs, i.e., the MANETs adopting the popular and efficient two-hop relay algorithms for packet routing.

Based on the concept of automatic feedback control and the Markov chain model, we first develop a general theoretical framework to characterize the throughput capacity in a two-hop relay MANET, where the transmission power of each node can be controlled to adapt to a specified transmission range and a limited number of redundant copies can be distributed for each packet. With the help of the framework, we are then able to derive the exact per node throughput capacity for a fixed setting of both the transmission range and packet redundancy limit. Based on the new throughput result, we further explore the optimality properties of the considered MANETs in terms of the per node throughput capacity, by considering all possible settings of transmission range and packet redundancy limit.

We further extend the conventional two-hop relay and propose a general group-based

two-hop relay algorithm with packet redundancy, which covers the available two-hop relay protocols with out-of-order or strictly in-order reception as special cases. A Markov chain-based theoretical framework is further developed to analyze how the mean value and variance of packet delivery delay vary with the redundancy limit and group size, where the important medium contention, interference and traffic contention issues are carefully incorporated into the analysis. Extensive simulation and theoretical results are provided to illustrate the performance of the proposed algorithm and the corresponding theoretical framework, which indicate that the theoretical framework is efficient in delay analysis and the new algorithm actually enables both the mean value and variance of packet delivery delay to be flexibly controlled in a large region for the challenging MANET.

Acknowledgments

I am truly and deeply indebted to so many people that there is no way to acknowledge them all, or even any of them properly. Without their help and encouragement, this work would not and could not have been done. I extend my deepest gratitude to all.

First and foremost, I am deeply indebted, grateful to my supervisor, Professor Nei Kato, for his continuous guidance, supervision and constant support in my scientific research and daily life. I wish to thank him for his kindness, his constant support, and his invaluable advice during the course of my Ph.D. study. In addition, he creates so nice a research environment which has made my Ph.D. period one of the best periods in my life.

I would like to give a special thanks to Professor Xiaohong Jiang, for his insightful guidance, positive encouragement and continuous support in both my research and my life. I learned a lot from the discussions with him and the comments from him. I appreciate all that he has taught me and all that he has done to aid my development–both professionally and personally. Without him, this thesis could not have been finished.

I would also like to acknowledge my thesis committee members, Professor Kazuyuki Tanaka and Professor Takuo Suganuma, for their interests and for their constructive comments that help to improve this thesis.

I would like to give my sincere thanks to other members of the Kato Lab in Tohoku University. In particular, I want to thank Associate Professor Hiroki Nishiyama whose diligence has set an example for me, and Motoko Shiraishi who has helped me a lot with my paper works. Additionally, I would also like to thank Yuta Tada, Qing Chen, Wei Liu, Zubair and Ahamed, who has helped me a lot in the past three years. It is my big pleasure to be a friend of yours. With your friendship, the time of my life became joyful and memorable.

Acknowledgments are also given to Japanese Ministry of Education, Culture, Sports, Science and Technology (Monbukagakusho) for their strong support that enabled me to pursue the challenging route during my doctoral studies.

This thesis is dedicated to my wife, my son, my parents and my mother-in-law, whose love and unconditional support provide a constant inspiration in my life. In particular, I would like to thank my wife for her support during my studies, and for her commitment to live and love along with me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background to Mobile Ad Hoc Networks

With the rapid development of communication technologies, wireless networks nowadays are increasingly becoming irreplaceable communication techniques for people's daily life. Motivated by their promising applications, lots of research efforts have been dedicated towards the designing and building of wireless networks, such as mobile ad hoc network, satellite network, cellular network, Wi-Fi (or hotspot) network, wireless sensor network, etc. Among these wireless networks, the mobile ad hoc network is of special interests to researchers from both academia and industry due to its distinctive features.

A mobile ad hoc network (MANET) is a peer-to-peer network with fully self-organized mobile nodes. In such an autonomous network system, mobile users randomly move around and freely communicate to each other via wireless links without the aid of any pre-existing infrastructure or centralized administration. Therefore, any mobile objects including the human beings, animals and vehicles, could easily form a MANET as long as each object carries a wireless communication device.

Compared with the available wireless network architectures, including the satellite network, cellular network and Wi-Fi network, the mobile ad hoc network has lots of attractive advantages. First, it incurs much lower establishment expenditure and maintenance cost than other wireless networks, as no infrastructure support or base station is required for the building of a MANET. Second, it is robust against the single point of failure, i.e., the death or diminishing of any particular network node will never affect the whole network performance. Finally, it can be rapidly deployed and flexibly reconfigured even in those geographically tough areas. Due to these specific features, the MANET holds great promise for a lot of future applications, such as the disaster relief, emergency response, daily information exchange, military troop communication, pedestrian network and vehicular ad hoc network, etc. It is believed that the MANET will become one of the most important and indispensable component among the next generation networks.

Motivated by the above great application potentials of MANETs, extensive researches from all over the world have been conducted towards a thorough understanding of the fundamental MANET performances in the last decades, such as the maximum supportable per node throughput, the expected end-to-end delay performance, the average packet delivery delay, etc. However, such important network performances remain largely unknown for the general MANETs and some research problems remain open by now, due to some challenging issues which will be introduced in the next section. Therefore, there is still a long way to go before the MANET could become as popular and prevalent as the cellular network and Wi-Fi network today.

1.2 Challenging Issues and Open Problems

As introduced in last section, since the nodes in mobile ad hoc networks (MANETs) randomly move around in network region according to their own mobility patterns, it is very difficult to predict the node movement behaviors. A significant amount of works has been done to understand the mobility patterns of human beings, where a lot of mobility trajectories have been collected and analyzed. A lot of mobility models have also been proposed to characterize the mobility pattern, such as the i.i.d. (independently and identically distributed) model, random walk model, random waypoint model, Brownian model, restricted mobility model, the Levy walk model and the Levy fly model. Although these models could provide a good approximation of the node mobility pattern to some extent, it remains unknown for the key characteristics lying behind the real mobility traces, such as the fundamental factors that will affect the node movement behavior, the inter-meeting time and the meeting duration between any node pair in a general MANET, the mobility model that matches best with the real mobility traces, etc.

Furthermore, all nodes in MANETs perform the required data delivery through node cooperations without an inherent infrastructure, where mobile nodes transmit to each other spontaneously without any centralized scheduling or administration. The wireless interference incurred by such spontaneous transmissions becomes much more challenging than that in cellular and Wi-Fi networks, in which the data transmissions are scheduled by the base station and access point, respectively. In the general MANETs, the ongoing data reception at some node could easily get corrupted by the data transmission initiated by some neighboring node, which is out of the sensing range and thus has no idea of the channel occupation status. Although some medium access control schemes have been proposed to address the interference issues, like the time division multiple access (TDMA), frequency-division multiple access (FDMA), orthogonal frequency-division multiple access (OFDMA) and carrier sense multiple access (CSMA), these schemes were proposed either for centralized networks or for static ad hoc networks, rather than for MANETs. The problem that how to characterize the interference issue and analyze the transmission opportunity and success probability in the general MANETs remains open now.

Due to the complicated node mobility and wireless interference issues in MANETs, the traditional information theory borne with the Shannon limit theory which has been vital for links and centralized networks, has not been successfully applied to decentralized wireless networks. Even if this was accomplished, due to the mobility issue, there is still a long way to go for such a theory to truly characterize the limits of deployed MANETs. The lack of a MANET capacity theory has been stunting the development and commercialization of many types of wireless networks, including emergency, military, sensor, and community mesh networks. It is expected that such a MANET capacity theory could provide a thorough understanding of the fundamental network performances, like the per node throughput capacity, i.e., the maximum supportable traffic input rate. Some problems remain challenging for the researchers and engineers, like how to determine whether a traffic input rate vector could be supported by a given MANET or not? If so, what is the corresponding routing and scheduling algorithm? Furthermore, what is the corresponding expected end-to-end delay at such input rate? By considering all possible routing and scheduling algorithms, in which range could the end-to-end delay be flexibly controlled?

1.3 Research Objectives

In this thesis, the overall aim is to develop a general Shannon-limit like network information theory so as to provide a deeper understanding on the fundamental throughput and delay performances in MANETs. Specifically, we will focus on an important class of mobile ad hoc networks, i.e., the mobile ad hoc networks adopting the two-hop relay algorithms as routing protocols (two-hop relay MANETs for short). By adopting the i.i.d. mobility model and the concurrent-set based scheme to address the node mobility issue and interference issue among simultaneous data transmissions, respectively, we will develop theoretical frameworks to analytically study the throughput and delay performances in two-hop relay MANETs.

In the study of throughput performance, we will focus on a two-hop relay MANET where the transmission power of each node can be controlled to adapt to a specified transmission range and a limited number of redundant copies can be distributed for each packet. We will first develop a general theoretical framework to characterize the throughput capacity in such a MANET by adopting the concept of automatic feedback control and the Markov chain model. With the help of the framework, we will then be able to derive the exact per node throughput capacity for a fixed setting of both the transmission range and packet redundancy limit. Based on the new throughput result, we will further explore the optimality properties of the considered MANETs in terms of the per node throughput capacity, by considering all possible settings of transmission range and packet redundancy limit.

In the delay performance study, we will focus on the average per packet delivery delay in mobile ad hoc networks. We will first extend the conventional two-hop relay and propose a general group-based two-hop relay algorithm with packet redundancy, which will cover the available two-hop relay protocols with out-of-order or strictly in-order reception as special cases. We will further develop a Markov chain-based theoretical framework to analyze how the mean value and variance of packet delivery delay vary with the redundancy limit and group size, where the important medium contention, interference and traffic contention issues will also be carefully incorporated into the analysis. We will also conduct extensive simulation studies to validate the developed theoretical framework for delivery delay analysis. Based on the developed framework, we will further explore how the proposed algorithm could enable a flexible delay control to be made to adapt to various applications with different delay requirements.

1.4 Thesis Outline

The remainder of this thesis is outlined as follows:

Chapter 2. Two-Hop Relay Mobile Ad Hoc Networks. In this chapter, we introduce the two-hop relay mobile ad hoc networks considered in this thesis. We adopt the cell partition technique to divide the network region into multiple cells, which enables a flexible control of both the node movement speed and the transmission range by adjusting the cell side length. We adopt the i.i.d. model as the mobility model and adopt the protocol interference model to address the wireless interference among simultaneous transmissions. We then introduce the permutation traffic pattern and also define the queue structures at each node. After defining the concurrent-set and introducing the transmission scheduling process, we give an overall introduction of the two-hop relay algorithm and its basic transmission modes.

Chapter 3. Exact Throughput Capacity under Power Control in MANETs. In this chapter, we focus on per node throughput capacity analysis in MANETs. We first analyze the packet delivery process of a tagged flow from the source to the destination in a two-hop relay MANET, where the transmission power of each node can be controlled to adapt to a specified transmission range and a limited number of redundant copies can be distributed for each packet. Based on the distinctive features of the delivery process, we then model the considered network into an automatic feedback control system. After introducing the concept of virtual queue at the destination, we develop Markov chain models to characterize the copy dispatching process at the source and also the packet reception process at the destination. By analyzing the properties of the two service times at the source and the destination, we derive closed-form expressions for the exact per node throughput capacity under a fixed setting of transmission range and packet redundancy limit. Then we conduct extensive simulation studies to validate the developed theoretical framework for the per node throughput capacity. Finally, we further explore the optimality properties of the considered MANETs in terms of the per node throughput capacity, and find out the optimum setting of packet redundancy limit to achieve the optimum throughput.

Chapter 4. Generalized Two-Hop Relay for Flexible Delay Control in MANETs. In this chapter, we focus on the packet delivery delay performance, which measures the time it takes a network to deliver a packet from the source to the destination. We first analyze the available delay performance studies and find out their limitations in delay control. We then accordingly propose a general group-based two-hop relay algorithm with limited packet redundancy so as to provide flexible delay control for the challenging MANETs. After analyzing the properties of the start state and the end state of the delivery process, we define general transient states to represent intermediate network transient states. By further defining the transition diagrams for a general transient state under all possible transition scenarios, we develop a Markov chain-based theoretical framework to derive closed-form results for both the mean value and the standard deviation of the packet delivery delay. Finally, we conduct extensive simulation studies to validate the developed theoretical framework and delay results.

In the last chapter, we conclude the whole thesis and discuss the future works.

1.5 Thesis Contribution

In this thesis, we studied the exact per node throughput capacity and also the packet delivery delay performance in two-hop relay mobile ad hoc networks. The main contributions of this thesis are summarized as follows:

- We first develop a general theoretical framework to depict the complicated packet delivery process in the challenging MANET environment. With the help of the theoretical framework, we then develop closed-form expressions for the exact per node throughput capacity $\mu(\upsilon, f)$ under any specified setting of transmission range υ and packet redundancy limit f. Simulation results are also provided to validate the throughput capacity result. Based on the new throughput result, we further explore the optimal capacity $\max_f\{\mu(\upsilon, f)\}$ for a fixed υ and also determine the corresponding optimum setting of f to achieve it. This result helps us to understand how such optimal capacity varies with υ and to find a suitable υ (and also f) to achieve the maximum possible throughput capacity $\max_{\upsilon,f}\{\mu(\upsilon, f)\}$ of such a network.

- We propose a new 2HR-(f, g) algorithm, where each packet is delivered to at most f distinct relay nodes and can be accepted by its destination if it is a fresh packet to the destination and also it is among g packets of the group the destination is currently requesting. This algorithm is general and covers all the available two-hop routing protocols as special cases.

- We further develop a general theoretical framework based on the multi-dimensional Markov chain, which is powerful in the sense it enables not only the mean value but also the variance of packet delivery delay to be derived analytically with a careful consideration of the important medium contention, interference and traffic contention issues. Extensive simulation and theoretical results are provided to validate the 2HR-(f, g) algorithm and the Markov chain theoretical framework. These results indicate that the theoretical framework is very efficient in packet delay analysis, and more importantly, the new 2HR-(f, g) algorithm makes it possible for us to flexibly control the packet delivery delay (and its variance) in a large region through the proper settings of f and g.

Chapter 2

Two-Hop Relay Mobile Ad Hoc Networks

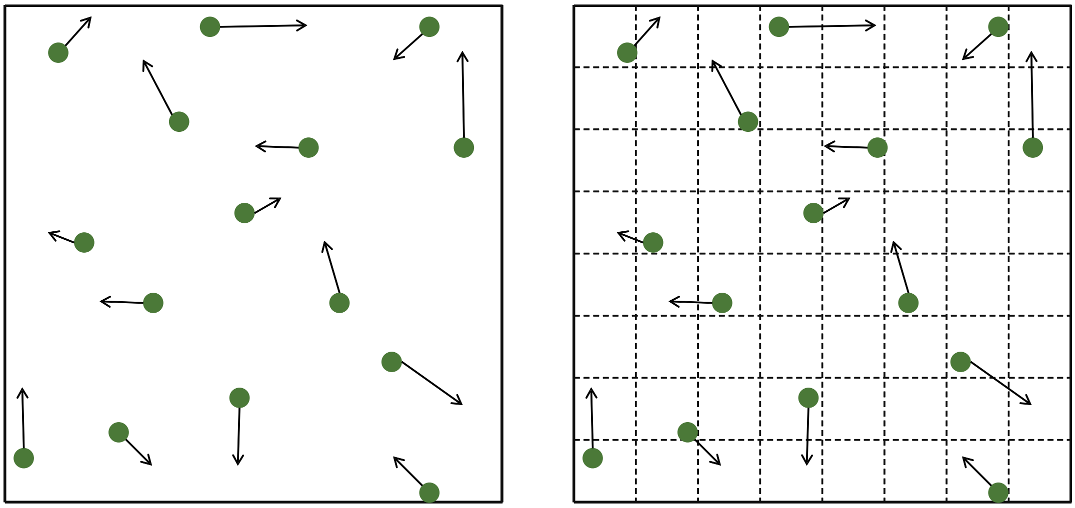
2.1 Introduction

In this thesis, we focus on an important class of mobile ad hoc networks (MANETs)—the two-hop relay mobile ad hoc networks, i.e., the MANETs adopting the popular and efficient two-hop relay algorithms for packet routing [1, 2, 3, 4, 5, 6]. In this chapter, we first introduce the system models considered in this thesis, including the network model, mobility model, interference model, traffic model and queue structure at each node. After defining the concurrent-set based transmission scheduling scheme, we then give an introduction to the two-hop relay routing protocols.

2.2 System Models

2.2.1 Network Model

Mobile ad hoc network: Consider a wireless ad hoc network with n mobile nodes moving around randomly and independently in a unit square. Similar to previous works [7, 8, 9], the unit square is assumed to be a torus, i.e., the top and bottom edges of the square touch



(a) Network before cell partition. (b) Network after cell partition

Figure 2.1: Illustration of cell partitioned network structure.

each other and the left and right ones also touch each other. As shown in Fig. 2.1(a), each node moves around in the two-dimensional unit torus according to its own mobility pattern.

Similar to previous works [9, 10, 11, 12], we assume a time slotted system, where the time is divided into consecutive time slots with equal duration, and apply the cell partition technique to divide the network area into $m \times m$ equal sized cells, each cell with a side length of $1/m$. As shown in Fig. 2.1(b), the cell partition technique maps the random node mobility in the network region into the node movement among different network cells. As a common partition technique for time slotted network, the general $m \times m$ cell partition actually enables a flexible control of both the node movement speed and the node transmission range to be made for the challenging MANET, by accordingly setting the cell side length (or parameter m). As indicated in [10, 11, 7, 8, 13, 14], the cell partition technique is also very helpful for network performance modeling and analysis.

We further assume the limit channel bandwidth scenario such that the total number of bits that can be transmitted per time slot between any node pair is fixed and normalized to one packet. We suppose that during each time slot each node has the knowledge about which cell it falls within based on its location information (For node localization, please refer to [15, 16]).

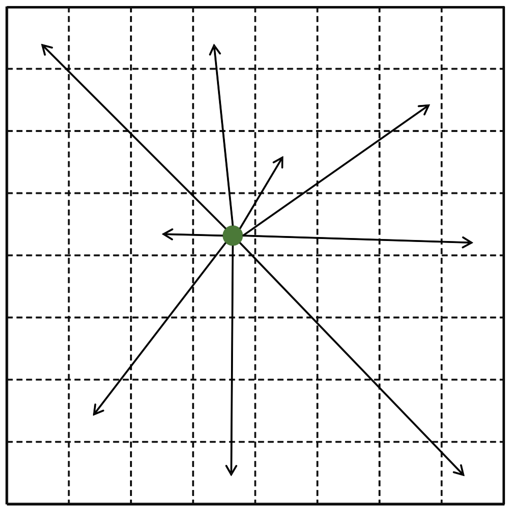


Figure 2.2: Illustration of the i.i.d. mobility model.

2.2.2 Mobility Model

i.i.d. mobility model: For the cell-partitioned network considered in this thesis, we focus on the i.i.d. node mobility model [7, 17, 18], where each node first independently and uniformly chooses a destination cell over all m^2 cells at the beginning of each time slot, and then stays within it for the whole time slot. As shown in Fig. 2.2, suppose a node is in a cell at a time slot, it may appear at any network cell in the next time slot with equal probability, i.e., each cell has the same probability of $1/m^2$ to be selected as the destination cell. Notice that under the i.i.d. mobility model, the time a node takes to move from one cell to another cell is neglected.

Indeed, the i.i.d. mobility model is popular for network performance analysis and has been widely adopted in literature [10, 11, 3, 4, 7, 13, 19]. Since the network topology varies dramatically and the network behavior can never be predicted under the bi-dimensional i.i.d. mobility model, the i.i.d. model is able to capture the node movements in the regime of infinite mobility and thus the network performance analysis derived under such mobility model provides a meaningful bound in the limit of infinite mobility. Furthermore, the results in [10, 20, 4] indicate that the network capacity derived under the i.i.d. mobility model is actually identical to the one derived under other non-i.i.d. mobility models (like the Markovian random walk model and random waypoint model) if they follow the same

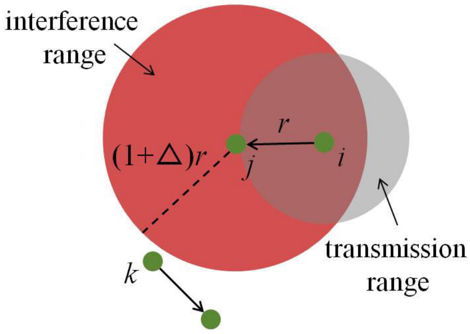


Figure 2.3: Illustration of protocol interference model.

steady state channel distribution, as later to be shown in Chapter 3.

2.2.3 Interference Model

Protocol interference model: To account for the interference issue among simultaneous transmissions, the Protocol interference model introduced in [21, 9] is adopted here. For a link x at time slot t, we use $T_x(t)$ and $R_x(t)$ to denote the positions of the corresponding transmitter and receiver, respectively. Based on the Protocol model, the transmission of the link x can be successful at the time slot t if for any other link y with simultaneous transmission we have

$$|T_y(t) - R_x(t)| \geq (1 + \Delta)|T_x(t) - R_x(t)|$$

here Δ is a protocol specified guard factor for interference control (a fixed positive constant representing the guard zone in the Protocol model).

As shown in Fig. 2.3, node i is transmitting to node j and node k is transmitting to another node at a time slot. The transmission range is defined as the circle area centered at node i with a radius of r, i.e., the shaded circle. According to the Protocol interference model, the interference range is defined as the red circle area centered at node j with a radius of $(1 + \Delta)r$, and the data reception at node j will be successful if and only if there is no any other simultaneous transmitter in the interference range. We can see from Fig. 2.3 that, the transmission of node k will not affect the data reception at node j, since node k is outside the interference range of j.

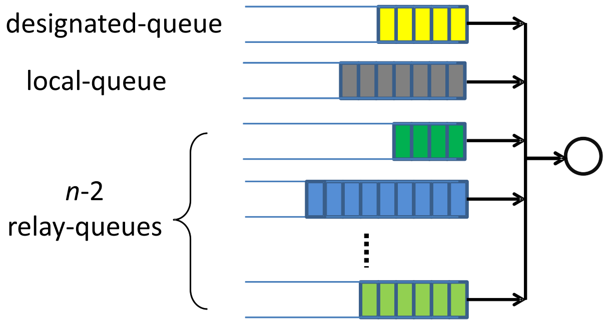


Figure 2.4: The queue structure at a node with the permutation traffic pattern.

2.2.4 Traffic Model

Permutation traffic pattern: We consider the permutation traffic pattern widely adopted in previous studies [1, 10, 13, 7, 22, 23, 11]. Under such traffic model, there will be in total n distinct uni-cast flows, where each node is the source of its locally generated traffic flow and at the same time the destination of a flow originated from another node. For the traffic flow originated at each node, we assume it has an average rate of λ (packets/slot). The packet arrival process at each node is independent of its mobility process and all packets arrive at the beginning of each time slots. For the purpose of throughput capacity and delay analysis, we simply assume that no lifetime is associated with each packet and the buffer size at each node is large enough such that the packet loss due to buffer overflow will never happen.

The permutation traffic pattern can be regarded as the worst-case uni-cast scenario, under which each node has a local outgoing traffic to deliver and also an incoming traffic to receive. According to the two-hop relay algorithm, therefore, a node will choose to forward traffic for other flows only when the node does not meet the destination of its own outgoing flow. In light of the fact that in the real-world MANETs some nodes may have no traffic to deliver or receive, i.e., may serve as pure relays, the throughput capacity derived under the permutation traffic pattern may serve as an achievable lower bound.

Since there are n distinct uni-cast flows, without loss of generality, we focus on a tagged flow and denote its source and destination by node S and node D, respectively. According to the two-hop relay [1, 10, 11, 24], S can also be a potential relay for other

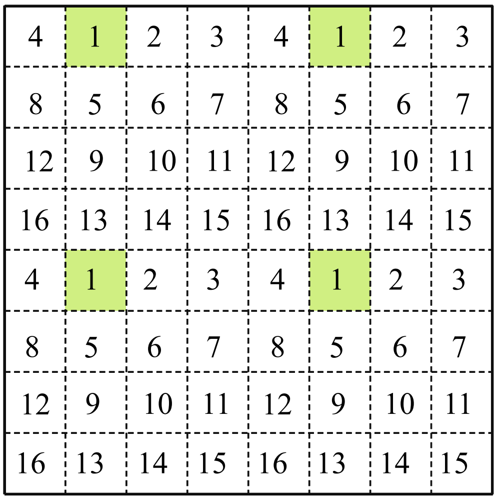


Figure 2.5: Illustration of concurrent-set in a cell partitioned network.

$n-2$ flows (except the two flows originated from and destined for itself). As illustrated in Fig. 2.4, we assume that S maintains n individual queues at its buffer, one designated-queue for storing the packets that are destined for itself, one local-queue for storing the packets that are locally generated and destined for node D, and $n-2$ parallel relay-queues for storing packets destined for other $n-2$ nodes (excluding S and D).

2.3 Overview of Concurrent-Set Based Transmission Scheduling Scheme

2.3.1 Definition of Concurrent-Set

According to the Protocol interference model, multiple links could simultaneously transmit if they are sufficiently far away from each other. To support as many simultaneous link transmissions as possible while ensuring an acceptable interference among nodes, we consider here a concurrent-set based scheduling scheme similar to [12, 11, 13, 19].

Concurrent-set: A concurrent-set is a subset of cells, where any two of them have a vertical and horizontal distance of some multiple of α cells and all of them could conduct transmission simultaneously.

An example of concurrent-set is illustrated in Fig. 2.5, where the distance parameter

$\alpha = 4$. From Fig. 2.5, we can see that the network cells are divided into 16 concurrent-sets and all the highlighted cells belong to the same concurrent-set, i.e., the concurrent-set 1, in which each cell can simultaneously support a transmitting node without interfering with each other.

2.3.2 Transmission Scheduling and Transmitting Node Selection

It is notable that for the concurrent-set based scheduling with distance parameter α, there will be in total α^2 distinct concurrent-sets, where each cell belongs to one distinct concurrent-set. If all concurrent-sets alternatively become active (i.e., get the transmission opportunity), then each concurrent-set (and thus each cell) becomes active in every α^2 time slots. For a $\sqrt{n} \times \sqrt{n}$ cell-partitioned network, the results in [11] indicate that at any time slot, the event that there are at least two nodes falling within an an active cell happens with a non-negligible probability (approaches $1 - 2e^{-1}$ as n goes to infinity). When there are more than one node in an active cell, the selection of transmitting node from the cell can be implemented by a contention window based selection mechanism similar to the DCF, which is defined as follows.

Selection of Transmitting Node: At the beginning of each time slot, each node independently judges whether it is inside an active cell or not. If not, it remains silent and will not contend for the transmitting opportunity. Otherwise, it starts its back-off counter with a seed randomly selected from $(0, CW]$ (CW represents the contention window), and then overhears the channel until its back-off counter becomes 0 or it hears a broadcasting message from a transmitter. If no broadcasting message is heard during the back-off counting process, it broadcasts out a message denoting itself as the transmitter.

Fig. 2.6 illustrates the flowchart of the contention window based transmitting node selection scheme. It is easy to see that based on the back-off counting mechanism, each node of an active cell has the same probability to win the transmission opportunity and thus becomes the transmitting node.

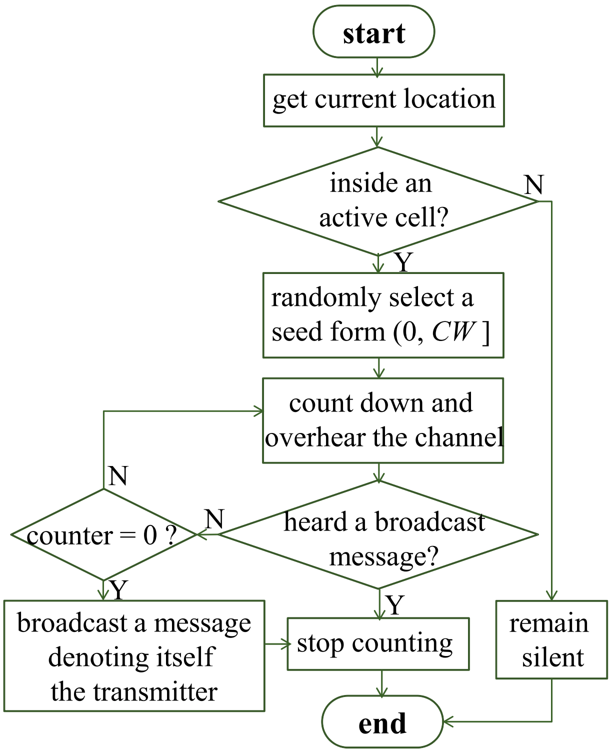


Figure 2.6: Flowchart of the contention window based transmitting node selection scheme.

2.4 Overview of Two-Hop Relay Routing Protocols

2.4.1 Elemental Operations

Due to random node mobility in MANETs, the network topology may vary dramatically and no contemporaneous end-to-end path may ever exist at any given time instant [25, 26, 27]. The traditional route-based routing protocols such as DSR [28], AODV [29], etc., fail to work properly as they require the simultaneous availability of a number of links. As a "store-carry-forward" kind of routing protocol, the two-hop relay, which relies on the mobility of nodes and sequences of their contacts to compensate for lack of continuous connectivity and thus enable packets to be delivered from end to end, becomes a promising routing protocol for the MANETs [1, 30, 11, 27, 3].

The two-hop relay routing protocol, simple yet efficient, has been proved to be able to provide a flexible control of both the throughput and packet delay for the challenging MANETs [27, 3]. Under such a routing protocol, there are three elemental operations.

- "source-to-relay" transmission: the source node replicates copies of its packets to

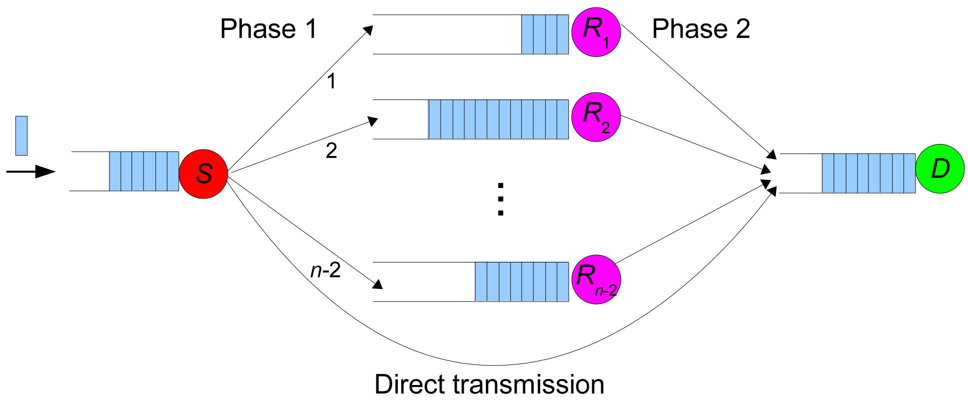


Figure 2.7: Illustration of the two-hop relay routing protocol.

other nodes (i.e., the relay nodes) it encounters.

- "relay-to-destination" transmission: an intermediate relay node carrying a copy can forward the copy only to the destination node.
- "source-to-destination" transmission: the source node directly sends a packet to the destination.

2.4.2 Basic Protocol and Various Variants

Since the seminal work of Grossglauser and Tse (2001) [1], the two-hop relay algorithm and its variants have become a class of attractive routing algorithms for mobile ad hoc networks, which enable the capacity and delay to be studied analytically. As illustrated in Fig. 2.7, the two-hop relay routing protocol defines two phases for packet transmission, where a packet is first transmitted from its source node to an intermediate node (relay node) in phase 1 (i.e., the "source-to-relay" transmission), and then in phase 2 the packet is transmitted from the relay node to its destination node (i.e., "relay-to-destination" transmission). Since the source node can directly transmit a packet to its destination node (i.e., the "source-to-destination" transmission) once such transmission opportunity arises, every packet goes through at most two hops to reach its destination in a two-hop relay network.

By now, a lot of two-hop relay variants have been proposed and explored in literature.

In the in-order two-hop relay algorithms, like the ones proposed in [10, 31, 32, 11], each packet should be received in-order at its destination. The algorithms in [1, 33, 34] can be regarded as the out-of-order two-hop without redundancy, where a packet has at most one copy and gets accepted by its destination if it is "fresh" (never received before). The out-of-order two-hop relay with redundancy has also been explored recently [27, 35, 36, 37, 38, 39, 40, 41], where each packet may have multiple copies in the transmission process. To achieve a flexible trade-off between the in-order reception and out-of-order reception, the general group-based two-hop relay algorithm has been proposed [24, 3], where the packets are delivered and received in group order. More recently, the two-hop relay algorithms with coding technique have also been explored in [42, 43, 6, 44].

2.5 Summary

In this chapter, we introduced the considered two-hop relay mobile ad hoc networks, including the cell-partitioned network model, the i.i.d. mobility model, the protocol interference model, the permutation traffic pattern and the queue structure at each node. In order to schedule as many simultaneous transmissions as possible in a cell-partitioned network, we defined the concurrent-set based scheduling scheme, where the transmission scheduling and the contention window based transmitting node selection scheme were also introduced. We also gave an overall introduction of the two-hop relay routing protocol, including its three elemental transmissions, the basic protocol and various variants.

Chapter 3

Exact Throughput Capacity under Power Control in MANETs

3.1 Introduction

The mobile ad hoc network (MANET), a very flexible and self-autonomous wireless network architecture, is very promising to find many important applications in the daily information exchange, disaster relief, military troop communication, etc. By now, the lack of a general Shannon limit-like network capacity theory is still a challenging roadblock stunting the development and commercialization of MANETs [45]. It is expected such a theory can help us to understand the basic network throughput capacity limit and thus serves as an instruction guideline for the network design, performance optimization and engineering of future MANETs [46, 47]. In this chapter, we first develop theoretical framework to characterize the packet delivery process in a two-hop relay MANET with limited packet redundancy, derive closed-form expressions for the per node throughput capacity, and then explore the maximum per node throughput capacity and the corresponding optimum packet redundancy limit to achieve it.

3.2 Related Research Works

Since the seminal work of Grossglauser and Tse (2001) [1], a lot of research efforts have been devoted to a better understanding of the MANET throughput capacity under various mobility models. Grossglauser and Tse [1] showed that under the i.i.d. mobility model, it is possible to achieve a $\Theta(1)$ per node throughput by employing a two-hop relay scheme. Following this line, it was later proved that the $\Theta(1)$ per node throughput can also be achieved under other mobility models, like the random walk model[8], the two-dimensional Brownian motions model[9] and the restricted mobility model [34]. Moraes *et al.* further showed that under uniform mobility model, we can still have the $\Theta(1)$ throughput even with a variant of the two-hop relay scheme, where each packet is only broadcasted once by its source and all nodes that receive the packet will act as its relays [48].

Recently, the trade-off between the throughput capacity and delay performances in MANETs has also been extensively explored. Perevalov *et al.* [33] studied the delay-limited throughput of MANETs and reported that under the i.i.d. mobility model, the achievable throughput is of order $\Theta(n^{-1/3})$ for a fixed delay value d and the throughput increases as $d^{2/3}$ when the delay d is a moderate value. Lin *et al.* [49] considered the Brownian motion model and showed that the two-hop relay scheme proposed by Grossglauser and Tse, while capable of achieving a per node throughput of $\Theta(1)$, incurs an expected packet delay of $\Omega(\log n/\sigma_n^2)$, where σ_n^2 is the variance parameter of the Brownian motion model. Neely *et al.* [10] proved that under the i.i.d. mobility model, it is able to achieve $O(1/\sqrt{n})$ throughput and $O(\sqrt{n})$ delay by introducing exact $\sqrt{n}$ redundancy for each packet. More recently, the per node throughput capacity and delay trade-off has also been studied under the random waypoint model [50] and the Brownian mobility model [31].

It is also noticed that some prior works has been done to analyze the exact capacity for MANETs. Neely and Modiano [10] established the network capacity for a cell partitioned MANET with i.i.d. mobility model, and showed that as the network size scales up the per node capacity there remains constant for a given node density. The capacity results in [10] was later extended to a delay tolerant MANET with general Markovian mobility models [20], where it is assumed that two nodes could transmit to each other as long as they are in the same cell or in adjacent cells.

3.2.1 Limitations of Available Works

It is notable that the above works mainly focus on deriving the order sense results of MANET throughput capacity. Although the order sense results are helpful for us to understand the general scaling law and thus the growth rate of the throughput capacity with network size n, they tell us little about the exact achievable per node throughput capacity. In practice, however, such exact network throughput capacity is of great interest for network designers. The available exact throughput capacity in [10, 20] was explored without careful consideration of the important wireless interference, medium contention and traffic contention issues in MANETs. Another limit of these works is that the impact of node transmission range on the throughput capacity of MANETs has been largely neglected. Since it is generally believed that the local transmission mode could result in the maximum per node throughput capacity, these work generally adopt the local transmission mode in their analysis, where either each node has a small transmission range of $\Theta(1/\sqrt{n})$ [1, 48, 34, 33, 49, 50, 31], or it can only transmit to some other node(s) in the same cell [8, 9, 10].

In this chapter, we study the exact per node throughput capacity of a MANET, where the transmission power (and thus transmission range) of each node can be controlled such that the impact of the transmission range on per node throughput capacity can be explored. For packet routing, we consider a generalized two-hop relay with limited packet redundancy (i.e., with a specified limit on the maximum number of distinct relays for each packet), which covers the available two-hop relay schemes [10, 50, 31] as special cases.

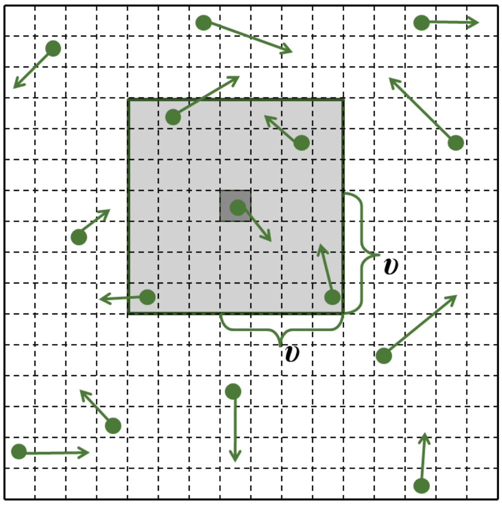


Figure 3.1: Illustration of the general transmission range.

3.3 System Assumptions and Definitions

We assume that the time is slotted and the network is evenly divided into $\sqrt{n} \times \sqrt{n}$ [1] cells with $1/n$ area each, as illustrated in Fig. 3.1. In order to explore the impact of power control on network throughput capacity, similar to [51] we assume that each node employs a power level so as to cover a set of cells with horizontal and vertical distances no more than $\upsilon - 1$ cells away from its current cell, where $1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$, and $\lfloor x \rfloor$ is the floor function. With such power control, a node could transmit to any other node in a square area centered at the cell of the node and is of side length $(2\upsilon - 1)/\sqrt{n}$, as that illustrated in Fig. 3.1.

Throughput Capacity: We call a traffic input rate λ (packets/slot) feasible or achievable if there exists a spatial and temporal scheduling algorithm such that under this input rate the queue length at each node will never increase to infinity as the time goes to infinity. The per node throughput capacity is then defined as the maximum feasible input rate λ. Without incurring any ambiguity, hereafter we call such capacity as throughput capacity for brevity.

[1] In the case that $\sqrt{n}$ is not an integer, i.e., when the network is divided into $m \times m$ equal cells where $m = \lceil n \rceil$ or $m = \lfloor n \rfloor$, the corresponding per node throughput capacity can also be easily derived based on the theoretical framework developed in this chapter.

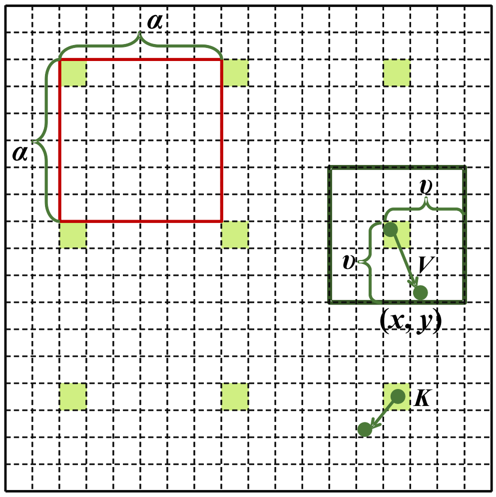


Figure 3.2: Tuning of concurrent-set under the general transmission range control.

3.4 Transmission Scheduling and Routing

3.4.1 Tuning of Concurrent-Set

Setting of Parameter α: To support as many simultaneous transmissions as possible, we need to properly set the parameter α of concurrent-set based on the parameter υ for power control and the parameter Δ for interference control. As illustrated in Fig. 3.2, suppose that the node V is scheduled to receive from some transmitting node, while the node K in another active cell of the same concurrent-set is transmitting to some other node. Notice that in this chapter we consider a network scenario where each node employs a power level so as to cover a set of cells which have a horizontal and vertical distance of no greater than $\upsilon - 1$ cells away from its current cell, $1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$. Thus, we assume that the node V is at a distance of (x, y) $(x, y \in [-\upsilon + 1, \upsilon - 1])$ cells away from its transmitting node, where the x and y denote the horizontal distance and vertical distance, respectively. It is trivial to see that we only need to consider the cases that $x \in [0, \upsilon - 1], y = \upsilon - 1$. We can easily see that the distance from node V to its transmitting node is at most $\frac{1}{\sqrt{n}}\sqrt{\upsilon^2 + (x+1)^2}$, while another simultaneous transmitting node (say the node K in Fig. 3.2) is at least $\frac{1}{\sqrt{n}}\sqrt{(\alpha - \upsilon)^2 + x^2}$ away from the node V. According to the interference model, the condition that K will not interfere with the

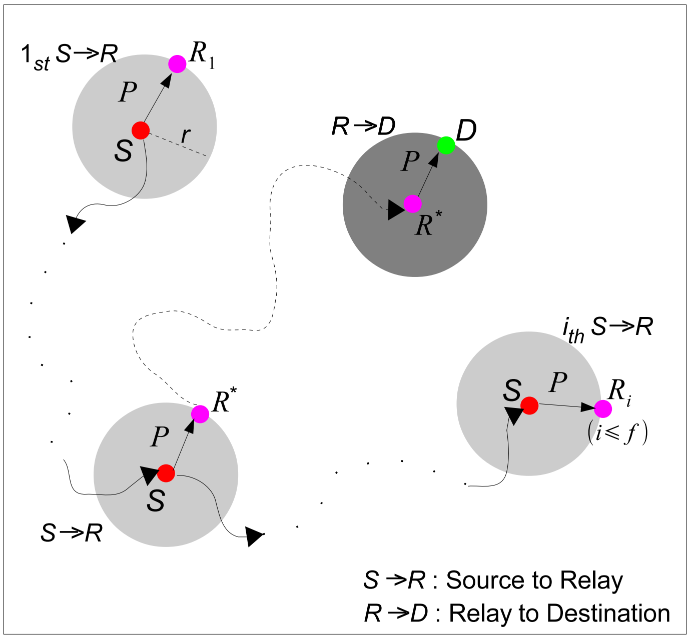


Figure 3.3: Illustration of the 2HR-f routing scheme for a tagged flow, whose source node S is transmitting packet P to the destination node D.

reception at the V is that for any $x \in [0, \upsilon - 1]$,

$$\frac{1}{\sqrt{n}}\sqrt{(\alpha-\upsilon)^2+x^2} \geq (1+\Delta)\frac{1}{\sqrt{n}}\sqrt{\upsilon^2+(x+1)^2} \tag{3.1}$$

To ensure above inequality for each $x \in [0, \upsilon - 1]$, we have

$$\alpha \geq \upsilon + \sqrt{2(\Delta+1)^2\upsilon^2 - (\upsilon-1)^2} \tag{3.2}$$

Since α is an integer and $\alpha \leq \sqrt{n}$, we can set α as follow to support as many simultaneous transmissions as possible.

$$\alpha = \min\{\upsilon + \lceil\sqrt{2(\Delta+1)^2\upsilon^2 - (\upsilon-1)^2}\rceil, \lfloor\sqrt{n}\rfloor\} \tag{3.3}$$

3.4.2 2HR-f Routing Scheme

In this chapter, we consider a generalization of the classic two-hop routing scheme with f-cast (2HR-f) [10, 27, 11], $f \in [1, n-2]$, where each packet waiting at the source is delivered to at most f distinct relay nodes (i.e., each packet has a limited redundancy f) and should be received in order at its destination, as shown in Fig. 3.3. For the permutation traffic pattern considered in this chapter, there are in total n distinct traffic flows. Without loss of generality, we focus on a tagged flow in our discussion and use the S and D to denote the source node and the destination node, respectively. The source S labels each packet P of the tagged flow with a *sequence number* $SN(P)$, while the destination D maintains a *request number* $RN(D)$ to indicate the sequence number of the packet it is currently requesting. Notice that the sequence number mechanism ensures that every packet is received in order at the destination and also helps to remove the remnant copies of those packets already received [10]. The overall 2HR-f scheme is summarized as follows.

2HR-f Routing Scheme: When the source S wins the transmission opportunity at the current time slot, S first overhears the channel for a specified interval of time to check whether the node D is inside the one-hop transmission range.

1) If S hears the reply from D within the specified time interval, it initiates a handshake with D and then transmits a packet directly to D ("Source-to-Destination" transmission);

2) If no broadcasting reply is overheard during the specified time interval, a receiving node (say R) is randomly selected among the nodes within the one-hop transmission range of S based on a mechanism similar to the selection of transmitting node. With probability $1/2$, the S and R then perform either the "Source-to-Relay" or "Relay-to-Destination" transmission:

   - Source-to-Relay: Suppose that the packet P is the packet locally generated for which S is currently delivering copies, S first initiates a handshake with R to check whether R has already received a copy of P before. If not, S delivers out a new copy of P to R if only less than f copies of P have already been delivered out from S by now; otherwise, S remains idle for this time slot.

   - Relay-to-Destination: As shown in Fig. 3.4, S initiates a handshake with R to check if S carries a packet P^* destined for node R with $SN(P^*) = RN(R)$. If

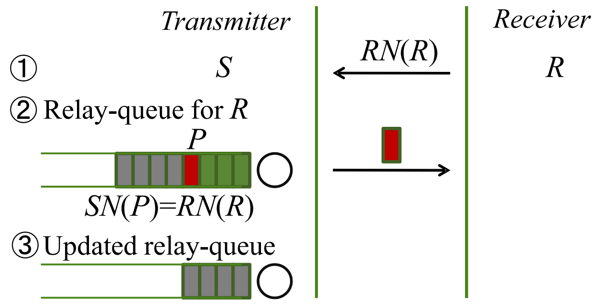


Figure 3.4: Illustration of Relay-to-Destination transmission in the 2HR-f routing scheme.

so, S delivers the packet P^* to node R; otherwise, S remains idle for this time slot.

3.5 Throughput Capacity Analysis

3.5.1 Some Basic Probabilities

Lemma 1. *For the tagged flow and a given time slot, we use p_1 and p_2 to denote the probability that S conducts a source-to-destination transmission and the probability that S conducts a source-to-relay or relay-to-destination transmission, respectively. By setting $m = (2\upsilon - 1)^2$, we have*

$$p_1 = \frac{1}{\alpha^2}\Big\{\frac{m-1}{n-1}\Big(1 - \Big(\frac{n-1}{n}\Big)^{n-1}\Big) + \frac{1}{n}\Big(\frac{n-1}{n}\Big)^{n-1}\Big\} \tag{3.4}$$

$$p_2 = \frac{1}{\alpha^2}\Big\{\frac{n-m}{n-1}\Big(1 - \Big(\frac{n-1}{n}\Big)^{n-1}\Big) - \Big(\frac{n-m}{n}\Big)^{n-1}\Big\} \tag{3.5}$$

Proof. Consider a tagged active cell, the node S can conduct a source-to-destination transmission with D only under the following two mutually exclusive cases: both S and D are in this cell; or S is in this cell while D is in one of the $4\upsilon^2 - 4\upsilon$ surrounding cells. If we further assume that aside from S and D, there are k other nodes in this cell, $k \in [0, n-2]$, then the probability that S is selected as the transmitter is $\frac{1}{k+2}$ (resp. $\frac{1}{k+1}$) under the former case (resp. under the latter case). Summing up the probabilities under these two cases, then we have

$$\begin{aligned}
p_1 &= \frac{1}{\alpha^2}\Bigg\{\sum_{k=0}^{n-2}\binom{n-2}{k}\Big(\frac{1}{n}\Big)^k\Big(\frac{n-1}{n}\Big)^{n-2-k}\frac{1}{n(k+2)} \\
&\quad+\sum_{k=0}^{n-2}\binom{n-2}{k}\Big(\frac{1}{n}\Big)^k\Big(\frac{n-1}{n}\Big)^{n-2-k}\frac{4\upsilon^2-4\upsilon}{n(k+1)}\Bigg\} \\
&= \frac{1}{\alpha^2}\Bigg\{\frac{4\upsilon^2-4\upsilon}{n-1}-\Big(\frac{4\upsilon^2-4\upsilon}{n-1}-\frac{1}{n}\Big)\Big(\frac{n-1}{n}\Big)^{n-1}\Bigg\} \\
&= \frac{1}{\alpha^2}\Bigg\{\frac{m-1}{n-1}\Big(1-\Big(\frac{n-1}{n}\Big)^{n-1}\Big)+\frac{1}{n}\Big(\frac{n-1}{n}\Big)^{n-1}\Bigg\}
\end{aligned} \tag{3.6}$$

where $m=(2\upsilon-1)^2$.

Similarly, S conducts a source-to-relay or relay-to-destination transmission iff the following four events happen simultaneously: S is in an active cell, S is selected as the transmitter, there is at least one other node (except S and D) in the same cell of S or its $4\upsilon^2-4\upsilon$ surrounding cells, and D is in one of the other $n-(2\upsilon-1)^2$ cells. Consider a tagged active cell, the probability that D is in one of the other $n-(2\upsilon-1)^2$ cells is $\frac{n-(2\upsilon-1)^2}{n}$. Further notice that S can conduct a source-to-relay or relay-to-destination transmission with some other node only under the following two mutually exclusive cases: this cell contains only S; or this cell contains at least one other node aside from S. If we assume that there are k ($k\in[1,n-2]$) other nodes in this cell (resp. in the $4\upsilon^2-4\upsilon$ surrounding cells), the other $n-2-k$ nodes can be in any cell of the other $n-1$ (resp. $n-(2\upsilon-1)^2$) cells. Summing up the probabilities under these two cases, then we have

$$\begin{aligned}
p_2 &= \frac{n-m}{n\alpha^2}\Bigg\{\sum_{k=1}^{n-2}\binom{n-2}{k}\Big(\frac{1}{n}\Big)^k\Big(\frac{n-1}{n}\Big)^{n-2-k}\frac{1}{k+1} \\
&\quad+\sum_{k=1}^{n-2}\binom{n-2}{k}\Big(\frac{m-1}{n}\Big)^k\Big(\frac{n-m}{n}\Big)^{n-2-k}\Bigg\} \\
&= \frac{1}{\alpha^2}\Bigg\{\frac{n-m}{n-1}\Big(1-\Big(\frac{n-1}{n}\Big)^{n-1}\Big)-\Big(\frac{n-m}{n}\Big)^{n-1}\Bigg\}
\end{aligned} \tag{3.7}$$

Then we finish the proof for Lemma 1. □

Lemma 2. *For the tagged flow, suppose that the source S is delivering copies for some packet P in the current time slot, the destination D is also requesting for P, i.e., $SN(P)=$*

$RN(D)$, and there are already j ($1 \leq j \leq f+1$) copies of P in the network (including the original one at S). For the next time slot, we use $P_r(j)$ to denote the probability that D will receive P, use $P_d(j)$ to denote the probability that S will successfully deliver out a copy of P to some new relay (if $j \leq f$), and use $P_s(j)$ to denote the probability of simultaneous source-to-relay and relay-to-destination transmissions. By setting $m = (2v-1)^2$, we have

$$P_r(j) = p_1 + \frac{j-1}{2(n-2)} p_2 \tag{3.8}$$

$$P_d(j) = \frac{n-j-1}{2(n-2)} p_2 \tag{3.9}$$

$$P_s(j) = \frac{(j-1)(n-j-1)(n-\alpha^2)}{4n\alpha^4} \sum_{k=0}^{n-5} \binom{n-5}{k} h(k) \cdot \Big\{ \sum_{t=0}^{n-4-k} \binom{n-4-k}{t} h(t) \Big(\frac{n-2m}{n}\Big)^{n-4-k-t} \Big\} \tag{3.10}$$

where

$$h(x) = \frac{m\big(\frac{m}{n}\big)^{x+1} - (m-1)\big(\frac{m-1}{n}\big)^{x+1}}{(x+1)(x+2)} \tag{3.11}$$

Proof. Given that there are j copies of packet P in the network, we know the source node S has already distributed $j-1$ replicas of P to $j-1$ distinct relay nodes. In the next time slot, the destination node D may either receive P directly from S or from one of the $j-1$ relay nodes, say node R. Notice that the probability that D receives P from S is p_1, and the probability that D receives P from R is $\frac{p_2}{2(n-2)}$. Since R can be any one of the available $j-1$ relay nodes, and the events that D receives P from S or the $j-1$ relay nodes are mutually exclusive and independent, by summing up the probabilities of these j events, the formula (3.8) follows.

Similarly, given that there are $j-1$ relay nodes each carrying the packet P, in the next time slot S may deliver out a new copy of P to any one of the remaining $n-j-1$ nodes. Since these $n-j-1$ events are also mutually exclusive and independent, and the probability that S delivers out a copy to a single relay node is $\frac{p_2}{2(n-2)}$, the probability that S will deliver out a new copy for P can then be determined as the formula (3.9).

To derive $P_s(j)$, let's focus on a specific node R from the $j-1$ nodes carrying packet P and a specific node V from the $n-j-1$ nodes without P, and use $P(S \to V, R \to D)$ to denote the probability that a source-to-relay transmission from S to V and a relay-to-destination transmission from R to D happen simultaneously in the next time slot. Thus, the $P_s(j)$ can be determined as

$$P_s(j) = (j-1)(n-j-1) \cdot P(S \to V, R \to D) \tag{3.12}$$

The basic idea is to treat the event $(S \to V, R \to D)$ as two simultaneous but mutually independent transmissions, i.e., the relay-to-destination transmission $R \to D$ and the source-to-relay transmission $S \to V$, then divide each transmission into multiple mutually exclusive cases, and finally represent each case with several simultaneous but independent sub-events. First consider the active cell with node R. The node R can conduct a relay-to-destination transmission with D only under the following two mutually exclusive cases: D is in this cell or D is in one of the $4\upsilon^2 - 4\upsilon$ surrounding cells. If we further assume that except the nodes S, D, R, V and the destination node of R's local traffic, there are in total k other nodes in the one-hop neighborhood of R, $k \in [0, n-5]$, among them i nodes are in the same cell as R, $i \in [0, k]$, and the other $k-i$ nodes are in the $4\upsilon^2 - 4\upsilon$ surrounding cells. Then the probability that R and D are selected as the transmitter and the receiver, respectively, is $\frac{1}{(i+2)(k+1)}$ (resp. $\frac{1}{(i+1)(k+1)}$) under the former case (resp. under the latter case). Summing up the probabilities under these two cases, then we get the corresponding probability of the relay-to-destination transmission $R \to D$. Similarly, we can also get the probability of the source-to-relay transmission $S \to V$. By the law of multiplication, then we have

$$P(S \to V, R \to D) = \frac{n - \alpha^2}{4n\alpha^4} \sum_{k=0}^{n-5} \binom{n-5}{k}$$
$$\cdot \Big(\sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \Big(\frac{1}{i+2} + \frac{m-1}{i+1} \Big) \Big(\frac{1}{n} \Big)^{i+1} \Big(\frac{m-1}{n} \Big)^{k-i} \Big)$$

$$\cdot \sum_{t=0}^{n-4-k} \binom{n-4-k}{t} \Bigg(\sum_{j=0}^{t} \binom{t}{j} \frac{1}{t+1} \Big(\frac{1}{j+2} + \frac{m-1}{j+1} \Big) \cdot \Big(\frac{1}{n}\Big)^{j+1} \Big(\frac{m-1}{n}\Big)^{t-j} \Bigg) \Big(\frac{n-2m}{n}\Big)^{n-4-k-t} \tag{3.13}$$

Notice that

$$\begin{aligned}
&\sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \Big(\frac{1}{i+2} + \frac{m-1}{i+1} \Big) \Big(\frac{1}{n}\Big)^{i+1} \Big(\frac{m-1}{n}\Big)^{k-i} \\
&= \sum_{i=0}^{k} \binom{k+1}{i+1} \frac{1}{k+1} \frac{1}{i+2} \Big(\frac{1}{n}\Big)^{i+1} \Big(\frac{m-1}{n}\Big)^{k-i} \\
&- \sum_{i=0}^{k} \binom{k}{i+1} \frac{1}{k+1} \frac{1}{i+2} \Big(\frac{1}{n}\Big)^{i+1} \Big(\frac{m-1}{n}\Big)^{k-i} \\
&+ \sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \frac{m-1}{i+1} \Big(\frac{1}{n}\Big)^{i+1} \Big(\frac{m-1}{n}\Big)^{k-i} \\
&= \frac{1}{(k+1)(k+2)} \Bigg(m\Big(\frac{m}{n}\Big)^{k+1} - (k+1+m)\Big(\frac{m-1}{n}\Big)^{k+1} \Bigg) \\
&- \frac{1}{(k+1)^2} \Bigg((m-1)\Big(\frac{m}{n}\Big)^{k+1} - (k+m)\Big(\frac{m-1}{n}\Big)^{k+1} \Bigg) \\
&+ \frac{m-1}{(k+1)^2} \Bigg(\Big(\frac{m}{n}\Big)^{k+1} - \Big(\frac{m-1}{n}\Big)^{k+1} \Bigg) \\
&= \frac{m\big(\frac{m}{n}\big)^{k+1} - (m-1)\big(\frac{m-1}{n}\big)^{k+1}}{(k+1)(k+2)}.
\end{aligned} \tag{3.14}$$

The (3.10) then follows by combining (3.12), (3.13) and (3.14). □

3.5.2 Service Time at Source S and Destination D

Before deriving the service time at S and D, we first define the following two queues. As shown in Fig. 3.5 that the first queue is a local queue at the source S, which stores the locally generated packets and operates as follows: every time a local packet arrives at S, it is put to the end of the queue; every time S finishes the copy dispatching for the head-of-line packet, S takes it out of the queue and moves ahead the remaining packets behind it. Thus, the head-of-line packet of the queue is the one for which S is currently distributing copies.

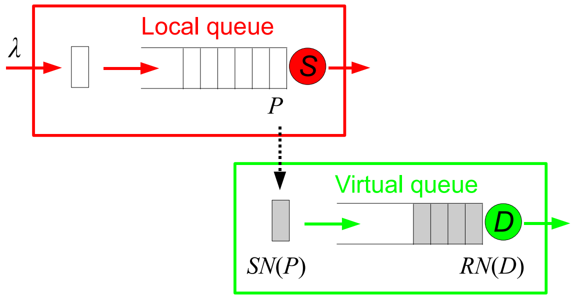


Figure 3.5: Correlationship between the local queue at the source S and the virtual queue at the destination D.

The second queue is a virtual queue defined at the destination D. As shown in Fig. 3.5, the virtual queue stores only the sequence numbers of those packets not received yet by D and operates as follows: every time a packet P is moved to the head-of-line of the local queue at S, the corresponding packet sequence number $SN(P)$ is put to the end of the virtual queue; every time D receives a packet whose sequence number equals to the head-of-line entry, D moves the head-of-line entry out of the virtual queue and moves ahead the remaining entries. Thus, the head-of-line entry of the virtual queue is the sequence number of the packet that D is currently requesting for, i.e., the $RN(D)$.

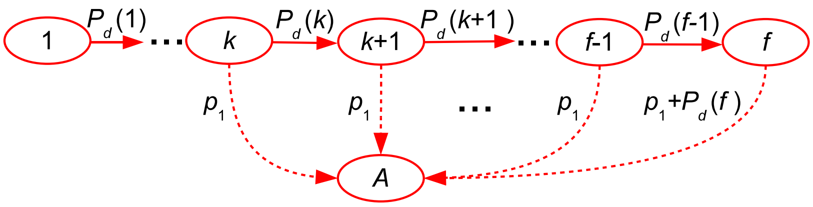
The service time at S and D can then be defined as follows:

Definition 1. *For a packet P of the tagged flow, its service time at the source S is defined as the time elapsed between the time slot when S starts to deliver copies for P and the time slot when S stops distributing copies for P.*

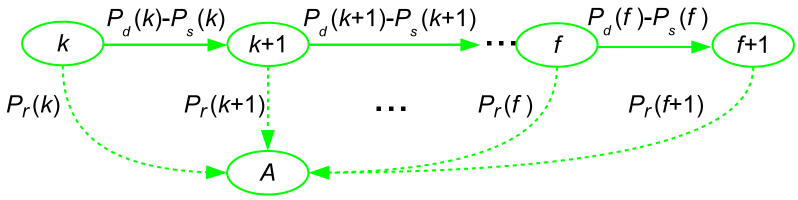
Definition 2. *For a packet P of the tagged flow, the service time at the destination D is defined as the time elapsed between the time slot when D starts to request for P and the time slot when D receives P.*

For a packet P of the tagged flow, suppose that there are k copies of P in the network when its destination D starts to request for P, $1 \leq k \leq f+1$. If we denote by state A the absorbing state (i.e., the termination of the service process) for P, then the service processes for the packet P at its source S and at its destination D can be defined by two finite-state absorbing Markov chains shown in Fig. 3.6(a) and Fig. 3.6(b), respectively.

Suppose that there are k copies of P in the network when D starts to request for the



(a) Absorbing Markov chain for the packet distribution process at the source node S.



(b) Absorbing Markov chain for the packet reception process at the destination node D.

Figure 3.6: Absorbing Markov chains for a packet P of the tagged flow.

packet, we denote by $X_S(k)$ and $X_D(k)$ the corresponding service time of the packet at S and D, respectively [2]. From the theory of Markov chain [52] we know that the $X_S(k)$ is the time the Markov chain in Fig. 3.6(a) takes to become absorbed given that the chain starts from the state 1, and the $X_D(k)$ is the time the Markov chain in Fig. 3.6(b) takes to become absorbed given that the chain starts from the state k.

Lemma 3. *For a packet P of the tagged flow, suppose that there are k copies of P in the network when the destination D starts to request for P, $1 \leq k \leq f+1$, then we have*

$$\mathbb{E}\{X_S(k)\} = \begin{cases} \sum_{i=1}^{k-1} \frac{1}{P_d(i)} + \frac{1}{p_1+P_d(k)}\left(1+\sum_{j=1}^{f-k}\phi_1(k,j)\right) & \text{if } 1 \leq k \leq f, \\ \sum_{i=1}^{f} \frac{1}{P_d(i)} & \text{if } k = f+1. \end{cases} \tag{3.15}$$

[2]The $X_S(f+1)$ corresponds to the case that D starts to request for the packet P from the state that there are $f+1$ copies in the network, i.e., the f copies of P have already been distributed.

$$\mathbb{E}\{X_D(k)\} = \begin{cases} \frac{1}{P_r(k)+P_d(k)-P_s(k)}\left(1+\sum_{j=1}^{f-k}\phi_2(k,j)+\frac{P_d(f)-P_s(f)}{P_r(f+1)}\phi_2(k,f-k)\right) \\ \qquad\qquad\qquad\qquad\qquad \textit{if } 1 \le k \le f-1, \\ \frac{1}{P_r(f)+P_d(f)-P_s(f)}\left(1+\frac{P_d(f)-P_s(f)}{P_r(f+1)}\right) \qquad \textit{if } k = f, \\ \frac{1}{P_r(f+1)} \qquad\qquad \textit{if } k = f+1. \end{cases} \tag{3.16}$$

where

$$\phi_1(k,j) = \prod_{t=1}^{j}\frac{P_d(k+t-1)}{p_1+P_d(k+t)}$$

$$\phi_2(k,j) = \prod_{t=1}^{j}\frac{P_d(k+t-1)-P_s(k+t-1)}{P_r(k+t)+P_d(k+t)-P_s(k+t)}$$

Proof. As the expected service time $\mathbb{E}\{X_S(k)\}$ can be obtained after some derivations similar to that in [11], here we only derive the expected service time $\mathbb{E}\{X_D(k)\}$. For the absorbing Markov chain in Fig. 3.6(b) and a given k there, $1 \le k \le f+1$, if we denote by a_i the mean time the Markov chain takes to become absorbed given that the chain starts from the state i, $k \le i \le f+1$, and denote by q_{ij} the transition probability from state i to state j, $i, j \in [k, f+1]$, then we have

$$\mathbb{E}\{X_D(k)\} = a_k \tag{3.17}$$

$$a_i = \frac{1+\sum_{j\in[k,f+1],j\neq i} q_{ij}\cdot a_j}{1-q_{ii}} \tag{3.18}$$

Notice that in the Markov chain of Fig. 3.6(b), except transiting back to itself and transiting to the absorbing state A, the state i can only transit to its next state, i.e., the state $i+1$. Thus, the a_i can be further determined as

$$a_i = \begin{cases} \frac{1+(P_d(i)-P_s(i))\cdot a_{i+1}}{P_r(i)+P_d(i)-P_s(i)} & \text{if } k \le i \le f, \\ \frac{1}{P_r(f+1)} & \text{if } i = f+1. \end{cases} \tag{3.19}$$

The formula (3.16) can then be derived from (3.19) recursively. □

Lemma 4. *For any* $1 \leq k \leq f$, *we have*

$$\mathbb{E}\{X_S(k)\} < \mathbb{E}\{X_S(k+1)\} \tag{3.20}$$

$$\mathbb{E}\{X_D(k)\} > \mathbb{E}\{X_D(k+1)\} \tag{3.21}$$

Proof. As the proof of (3.20) is similar to that in [11], we omit it here. Before proceeding to prove (3.21), we first employ the mathematical induction to prove the following inequality

$$P_r(k) \cdot \mathbb{E}\{X_D(k+1)\} < 1, \quad 1 \leq k \leq f \tag{3.22}$$

which will be used for the proof of (3.21).

Initial step: for $k = f$, it is easy to see that

$$P_r(f) \cdot \mathbb{E}\{X_D(f+1)\} = \frac{P_r(f)}{P_r(f+1)} < 1 \tag{3.23}$$

Inductive step: we assume that (3.22) holds for some $k = t$, $1 < t \leq f$, i.e., $P_r(t) \cdot \mathbb{E}\{X_D(t+1)\} < 1$. We need to prove (3.22) still holds for $k = t-1$.

$$\begin{aligned}
&P_r(t-1) \cdot \mathbb{E}\{X_D(t)\} \\
&= P_r(t-1) \cdot \frac{1 + (P_d(t) - P_s(t)) \cdot \mathbb{E}\{X_D(t+1)\}}{P_r(t) + P_d(t) - P_s(t)} \\
&< \frac{P_r(t-1) + P_r(t)\mathbb{E}\{X_D(t+1)\}(P_d(t) - P_s(t))}{P_r(t) + P_d(t) - P_s(t)} &(3.24)\\
&< \frac{P_r(t-1) + P_d(t) - P_s(t)}{P_r(t) + P_d(t) - P_s(t)} < 1 &(3.25)
\end{aligned}$$

where (3.24) follows after $P_r(t-1) < P_r(t)$. Combining (3.23) and (3.25), we prove (3.22).

Now we are ready to prove (3.21). For the case that $k = f$, we have

$$\begin{aligned}
&\mathbb{E}\{X_D(f)\} - \mathbb{E}\{X_D(f+1)\} \\
&= \frac{1 + \frac{P_d(f) - P_s(f)}{P_r(f+1)}}{P_r(f) + P_d(f) - P_s(f)} - \frac{1}{P_r(f+1)}
\end{aligned}$$

$$= \frac{1}{P_r(f+1)}\left(\frac{P_r(f+1)+P_d(f)-P_s(f)}{P_r(f)+P_d(f)-P_s(f)}-1\right) > 0 \tag{3.26}$$

For the case that $1 \leq k < f$, we have

$$\begin{aligned} &\mathbb{E}\{X_D(k)\} - \mathbb{E}\{X_D(k+1)\} \\ &= \frac{1+(P_d(k)-P_s(k))\mathbb{E}\{X_D(k+1)\}}{P_r(k)+P_d(k)-P_s(k)} - \mathbb{E}\{X_D(k+1)\} \\ &= \frac{1-P_r(k)\mathbb{E}\{X_D(k+1)\}}{P_r(k)+P_d(k)-P_s(k)} > 0 \end{aligned} \tag{3.27}$$

where (3.27) follows after substituting (3.22). After combining (3.26) and (3.27), we have (3.21) and then complete the proof for Lemma 4. □

For the tagged flow, if we further denote by $\overline{X}_S$ the average service time of all packets locally generated at the source S, and denote by $\overline{X}_D$ the average service time of all packets received at the destination D, then we can establish the following result based on Lemma 4.

Lemma 5. *For any given transmission range parameter υ and packet redundancy limit f, $1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$, $1 \leq f \leq n-2$, we have*

$$\mathbb{E}\{X_S(1)\} \leq \overline{X}_S \leq \mathbb{E}\{X_S(f+1)\} \tag{3.28}$$

$$\mathbb{E}\{X_D(f+1)\} \leq \overline{X}_D \leq \mathbb{E}\{X_D(1)\} \tag{3.29}$$

Proof. Since $\overline{X}_S$ is averaged among all locally generated packets at S, and $\overline{X}_D$ is also averaged over all received packets at D, combining with (3.20) and (3.21) in Lemma 4, it then follows Lemma 5. □

3.5.3 Per Node Throughput Capacity

For the tagged flow, suppose that currently packet P is the head-of-line packet at the local queue of S, D just starts to request for P and there are k $(1 \leq k \leq f)$ copies of

P now in the network. We further assume that the packet waiting right behind P in the local queue is packet P', and D will start to request for P' when there are k' copies of P' in the network. Then we have the following two cases:

- If $\mathbb{E}\{X_S(k)\} \leq \mathbb{E}\{X_D(k)\}$, then in the average case we have $k' \geq k$. According to (3.20) and (3.21), it follows that $\mathbb{E}\{X_S(k')\} \geq \mathbb{E}\{X_S(k)\}$ and $\mathbb{E}\{X_D(k')\} \leq \mathbb{E}\{X_D(k)\}$. Then we have

$$\mathbb{E}\{X_D(k') - X_S(k')\} \leq \mathbb{E}\{X_D(k) - X_S(k)\} \tag{3.30}$$

  The above condition indicates that statistically the expected gap between the service time at the destination and the service time at the source tends to reduce. Since (3.30) also holds for the packets (if any) waiting behind P', we can see that the $\overline{X}_S$ and $\overline{X}_D$ (and thus the average service rates at S and D) will gradually approach each other until a balance is achieved [3].

- If $\mathbb{E}\{X_S(k)\} > \mathbb{E}\{X_D(k)\}$, on average we then have $k' < k$. Similar to the above case, it follows that

$$\mathbb{E}\{X_S(k') - X_D(k')\} < \mathbb{E}\{X_S(k) - X_D(k)\} \tag{3.31}$$

  This condition indicates that statistically the service time at S tends to decrease while the service time at D tends to increase, thus the network system will gradually evolve towards a stable state.

The above analysis indicates that under the 2HR-f routing scheme, the parameter k is automatically updated from packet to packet to adjust to the service rates at the S and D. Based on this intrinsic feature of automatic updating for parameter k, we can model the packet delivery process of the tagged flow as an automatic feedback control system shown

[3] As we will show later that since the server for the local queue at S and the server for the virtual queue at D may have vacancy time, even in the case that the two average service rates $1/\overline{X}_S$ and $1/\overline{X}_D$ can not achieve a balance, the network system may still be stabilized as long as the input rate λ is feasible.

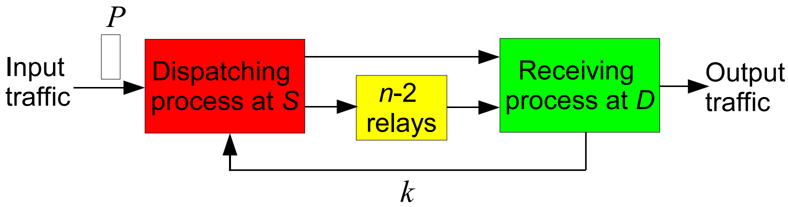


Figure 3.7: Illustration of the automatic feedback control system defined for the packet delivery process of the tagged flow.

in Fig. 3.7, where the packet dispatching process at S and the packet receiving process at D can be defined by the two absorbing Markov chains in Fig. 3.6(a) and Fig. 3.6(b), respectively[4].

Now we are ready to derive the throughput capacity for the tagged flow. We first denote by V_S the long-term average packet dispatching rate at S and denote by V_D the long-term average packet receiving rate at D, where

$$V_S = \lim_{t\to\infty} \frac{\text{the number of dispatched packets at } S}{t} \tag{3.32}$$

$$V_D = \lim_{t\to\infty} \frac{\text{the number of received packets at } D}{t} \tag{3.33}$$

We then have the following result.

Lemma 6. *For the tagged flow and any given parameters of υ and f, we have*

$$V_S \leq \frac{1}{\mathbb{E}\{X_S(1)\}} \tag{3.34}$$

$$V_D \leq \frac{1}{\mathbb{E}\{X_D(f+1)\}} \tag{3.35}$$

Proof. We first prove (3.34). For the local queue at the source S, suppose that during some time interval t node S has successfully served a total of $N_S(t)$ locally generated packets (i.e., S has distributed copies for $N_S(t)$ local packets). According to the definition in

[4]Notice that for a packet P of the tagged flow, its corresponding parameter k depends only on the delivery process of the last packet received just before the P. Thus, for the automatic feedback control system in Fig. 3.7, the parameter k will be automatically updated for each packet and there is no need for any extra transmissions to deliver the parameter k between S and D.

(3.32), we have

$$V_S = \lim_{t\to\infty} \frac{N_S(t)}{t} \tag{3.36}$$

Notice that during the time interval t, the local queue may be empty and thus the queue server at S may become idle. Denote by $I_S(t)$ the accumulated vacancy time at S during the time interval t, then we have

$$\overline{X}_S = \lim_{t\to\infty} \frac{t - I_S(t)}{N_S(t)} \tag{3.37}$$

Since $I_S(t) \geq 0$, combining the (3.36) and (3.37), we have

$$V_S \leq \frac{1}{\overline{X}_S} \tag{3.38}$$

Substituting (3.28) into (3.38), (3.34) then follows. After a similar derivation, (3.35) also follows. □

We now can establish the following main result on per node throughput capacity.

Theorem 1. *Consider a cell partitioned MANET, where nodes move according to the i.i.d. mobility model, each node could transmit to the cells with horizontal and vertical distance no more than $\upsilon - 1$ cells away from its current cell, $1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$, and each packet follows the 2HR-f routing scheme, $1 \leq f \leq n-2$. If we denote by $\mu(\upsilon, f)$ the per node (flow) throughput capacity, i.e., the network can stably support any input rate $\lambda \leq \mu(\upsilon, f)$, then we have*

$$\mu(\upsilon, f) = \min\left\{ p_1 + \frac{f}{2(n-2)} p_2, \quad \frac{p_1 + p_2/2}{1 + \sum_{j=1}^{f-1} \prod_{t=1}^{j} \frac{(n-t-1)p_2}{2(n-2)p_1 + (n-t-2)p_2}} \right\} \tag{3.39}$$

Proof. For the tagged flow, as its packet delivery process can be defined by the automatic feedback control system in Fig. 3.7, we can see that if the network is stable (i.e., the queue length at each node will not go to infinity) under the input rate λ, then we have

$$\lambda = V_S = V_D \tag{3.40}$$

This is due to the fact that in a stable control system, the long-term average rate of the input traffic is equal to that of the output one.

Based on (3.34) and (3.35), we then have

$$\lambda \leq \min\left\{\frac{1}{\mathbb{E}\{X_S(1)\}}, \frac{1}{\mathbb{E}\{X_D(f+1)\}}\right\} \tag{3.41}$$

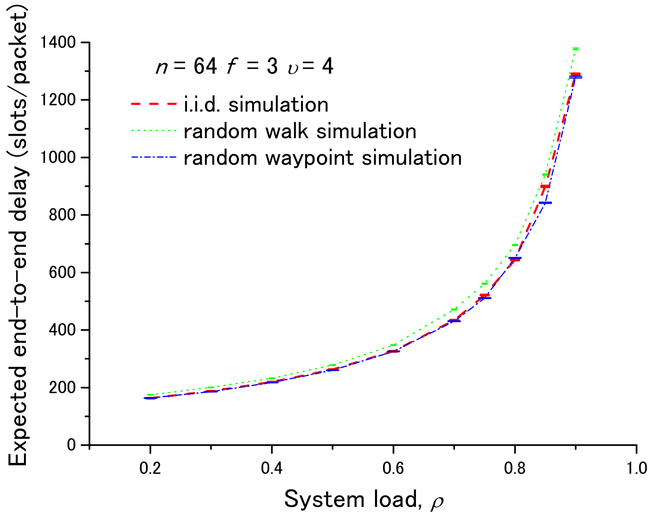
After substituting (3.15) and (3.16) into (3.41), (3.39) follows. □

3.5.4 Model Validation

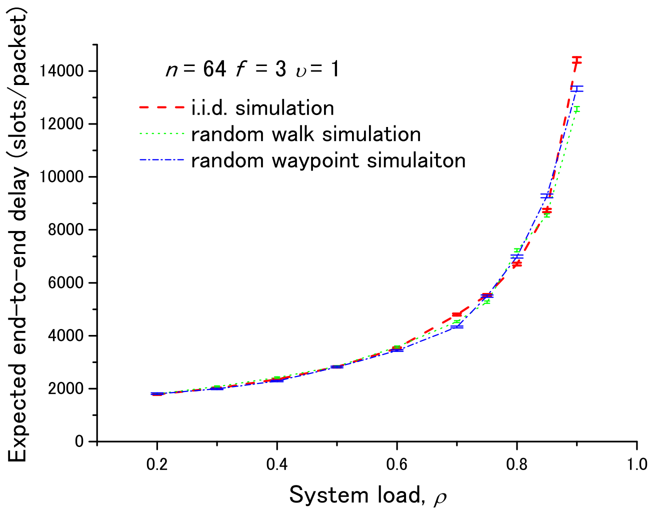
To validate our theoretical per node throughput capacity, a dedicated simulator in C++ was developed, which is now available at [53]. In our simulation, the traffic flow originated from each node is assumed to be a Poisson stream, and, similar to the settings in [54, 55], the guard factor Δ is fixed as $\Delta = 1$ here. In addition to the i.i.d. mobility model considered in this chapter, we also implemented the simulations for the popular random walk and random waypoint models:

- **Random Walk Model [8]:** At the beginning of each time slot, each node independently makes a decision regarding its mobility action, either staying inside its current cell or moving to one of its eight adjacent cells. Each action happens with the same probability of 1/9.

- **Random Waypoint Model [17]:** At the beginning of each time slot, each node independently and randomly generates a two-dimensional vector $\boldsymbol{\nu} = [\nu_x, \nu_y]$, where the values of ν_x and ν_y are uniformly drawn from $[1/\sqrt{n}, 3/\sqrt{n}]$. The node then moves a distance of ν_x along the horizontal direction and a distance of ν_y along the vertical direction.



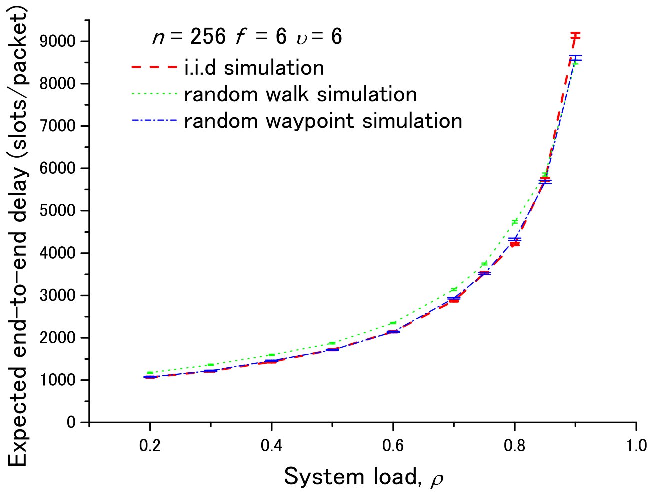
(a) Network scenario ($n = 64$, $f = 3$, $\upsilon = 4$) with $\mu(4, 3) = 7.60 \times 10^{-3}$ (packets/slot).



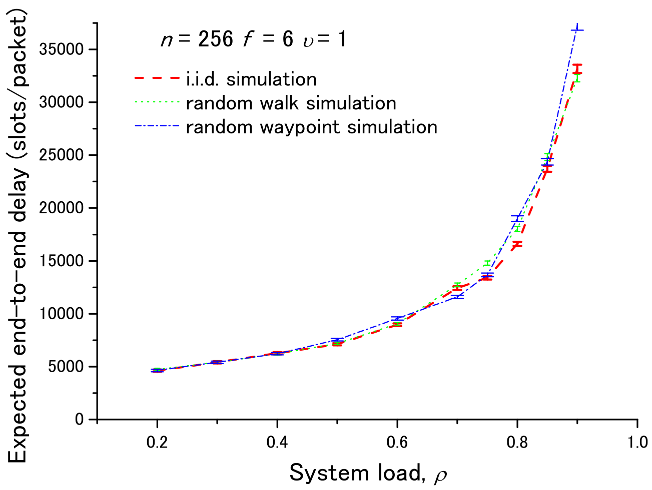
(b) Network scenario ($n = 64$, $f = 3$, $\upsilon = 1$) with $\mu(1, 3) = 7.53 \times 10^{-4}$ (packets/slot).

Figure 3.8: The expected end-to-end packet delay for $n = 64$.

In our simulation, we consider both a smaller network with $n = 64$, $f = 3$ and $\upsilon = \{1, 4\}$ and a larger network with $n = 256$, $f = 6$ and $\upsilon = \{1, 6\}$. The corresponding simulation results are summarized in Fig. 3.8 and Fig. 3.9, where all the results are reported with the 95% confidence interval. Figs. 3.8 and 3.9 indicate clearly that our theoretical models can precisely depict the throughput capacity of a MANET under the



(a) Network scenario ($n = 256,\ f = 6,\ \upsilon = 6$) with $\mu(6,6) = 1.17 \times 10^{-3}$ (packets/slot).



(b) Network scenario ($n = 256,\ f = 6,\ \upsilon = 1$) with $\mu(1,6) = 2.84 \times 10^{-4}$ (packets/slot).

Figure 3.9: The expected end-to-end packet delay for $n = 256$.

packet redundancy control and the transmission range control. As can be observed from these two figures that when the system load $\rho = \lambda/\mu$ approaches to 1 (i.e., when the traffic input rate λ approaches the throughput capacity μ), the packet delay rises up sharply and becomes extremely sensitive to the variations of ρ. Such skyrocketing behavior of packet delay as ρ approaches 1 serves as an intuitive validation for the throughput capacity determined by our theoretical model.

It is also interesting to notice from Figs. 3.8 and 3.9 that the networks we consider here actually exhibit very similar behaviors under either the i.i.d. model, the random walk model or the random waypoint model. In this sense, our theoretical throughput capacity model, although was developed under the i.i.d. mobility model, can also be used to nicely capture the network behaviors under the random walk and the random waypoint models as well.

3.6 Throughput Optimization

Based on the new theoretical capacity model, we further explore the following throughput optimization problem.

Throughput Optimization Problem: For a 2HR-f-based MANET with a fixed transmission range υ for each node, calculate its maximum per node throughput capacity for any value of f.

For a fixed transmission range υ, if we denote by μ^* the corresponding maximum per node throughput capacity, then the throughput optimization problem can be formulated as

$$\begin{aligned} \mu^* &= \max_f \{\mu(\upsilon, f)\} \\ &= \max \ \ \min \left\{ \frac{1}{\mathbb{E}\{X_S(1)\}}, \frac{1}{\mathbb{E}\{X_D(f+1)\}} \right\} \end{aligned} \tag{3.42}$$

subject to:

$$1 \leq f \leq n-2, 1 \leq \upsilon \leq \lfloor \frac{\sqrt{n}+1}{2} \rfloor$$

where the $\mathbb{E}\{X_S(1)\}$ and $\mathbb{E}\{X_D(f+1)\}$ are defined in (3.15) and (3.16), respectively.

Regarding the solution of this optimization problem, we have the following result.

Lemma 7. *For any given* $\upsilon \in [1, \lfloor \frac{\sqrt{n}+1}{2} \rfloor]$*, we have*

$$\mu^* = max\left\{ \frac{1}{\mathbb{E}\{X_D(f+1)\}\big|_{f=f_0}}, \frac{1}{\mathbb{E}\{X_S(1)\}\big|_{f=f_1}} \right\} \tag{3.43}$$

where

$$f_0 = \max \Big\{ f \mid \mathbb{E}\{X_S(1)\} \leq \mathbb{E}\{X_D(f+1)\} \Big\} \tag{3.44}$$

$$f_1 = \min \Big\{ f \mid \mathbb{E}\{X_D(f+1)\} \leq \mathbb{E}\{X_S(1)\} \Big\} \tag{3.45}$$

Proof. We first prove that f_0 and f_1 defined above do exist. According to (3.15) and (3.16), we have

$$\begin{aligned} \mathbb{E}\{X_S(1)\}|_{f=1} &= \frac{1}{p_1 + P_d(1)} \\ &\leq \frac{1}{p_1 + \frac{p_2}{2(n-2)}} = \mathbb{E}\{X_D(f+1)\}|_{f=1} \end{aligned} \tag{3.46}$$

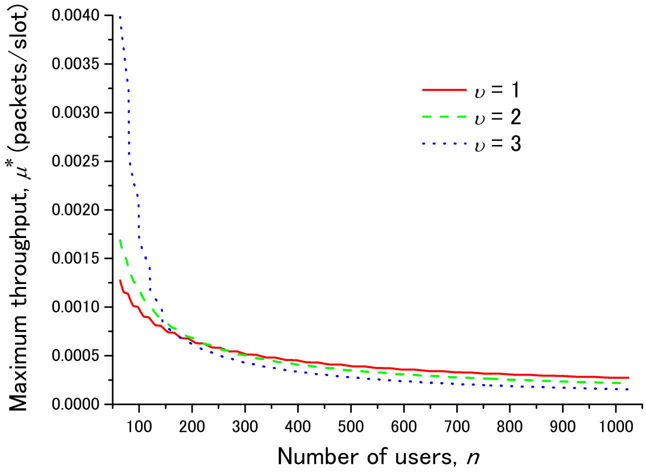
Notice that

$$\begin{aligned} &\mathbb{E}\{X_S(1)\}|_{f=n-2} \\ =& \frac{1}{p_1 + p_2/2}\Big(1 + \sum_{j=1}^{n-3} \prod_{t=1}^{j} \frac{(n-t-1)p_2}{2(n-2)p_1 + (n-t-2)p_2}\Big) \\ \geq& \frac{1}{p_1 + p_2/2} = \mathbb{E}\{X_D(f+1)\}|_{f=n-2} \end{aligned} \tag{3.47}$$

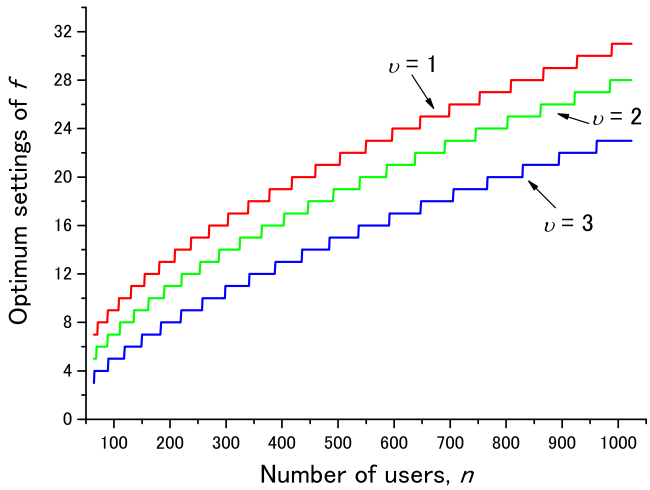
It is easy to see from (3.15) and (3.16) that as f increases, the $\mathbb{E}\{X_S(1)\}$ monotonically increases while the $\mathbb{E}\{X_D(f+1)\}$ monotonically decreases. Combining with the results in (3.46) and (3.47), we can see that the f_0 and f_1 defined above do exist.

From (3.41) we know that

$$\mu(\upsilon, f) = \begin{cases} 1/\mathbb{E}\{X_D(f+1)\} & \text{if } 1 \leq f \leq f_0, \\ 1/\mathbb{E}\{X_S(1)\} & \text{if } f_1 \leq f \leq n-2. \end{cases} \tag{3.48}$$



(a) The maximum throughput capacity μ^* vs. n



(b) The optimum setting of f vs. n

Figure 3.10: The maximum throughput capacity μ^* and the corresponding optimum setting of f for networks with n varying from 64 to 1024.

Thus, (3.43) can be derived directly based on (3.48). The above results indicate that for a MANET with a fixed υ, there exists an optimum setting of f (f_0 or f_1) to achieve the optimal per node throughput capacity μ^*. $\square$

To illustrate the optimal throughput capacity μ^*, we show in Fig. 3.10(a) and Fig. 3.10(b) that for $\upsilon = \{1, 2, 3\}$, how μ^* and the corresponding optimum setting of f vary with network size n. Fig. 3.10(a) shows clearly that for all the three settings of υ here, although

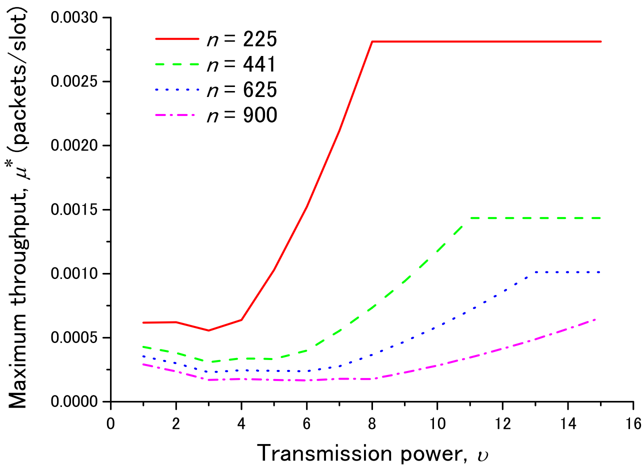


Figure 3.11: The throughput capacity μ^* vs. node transmission region υ.

the corresponding μ^* all decrease quickly as the network size increases, their varying tendencies with n are actually different. A careful observation of Fig. 3.10(a) indicates that among the three settings of υ here, the μ^* of the case $\upsilon = 3$ decreases most dramatically with n while the one of the case $\upsilon = 1$ decreases least significantly with n. It is also interesting to notice that when $n \leq 143$, the μ^* of the case $\upsilon = 3$ is always the greatest one among that of all three cases, while the μ^* of the case $\upsilon = 1$ becomes the greatest one when n is larger than 270. The results in Fig. 3.10(b) show that for a given υ, the corresponding optimum setting of f is actually a piecewise function of n. We can also see from the figure that for each network size n, the optimum setting of f of the case $\upsilon = 3$ is the smallest one among that of all three cases. This can be intuitively interpreted as follows. For one given network, if a bigger υ (and thus a larger transmission range) is adopted, a node will have higher probability to meet its destination or relay nodes and thus can deliver packets more fast, resulting in a fewer number of redundant copies for each packet before it arrives at its destination.

To further explore how υ affects μ^*, we summarize in Fig. 3.11 how μ^* varies with υ for networks of $n = \{225, 441, 625, 900\}$. It is interesting to see that, for each network scenario here, as υ increases μ^* always first decreases and then increases. This is because that the effect of increasing υ is two-fold: on one hand, it increases the probability that a node meets the destination or relay nodes, but on the other hand it decreases the number of simultaneous transmissions. As illustrated in Fig. 3.11 that when υ is small, the latter negative effect dominates, while as υ gradually increases, the former positive effect becomes the dominant one. It is further noticed that for the cases $n = 225$, $n = 441$ and $n = 625$, μ^* does not increase any more when υ increases beyond some threshold ($\upsilon = 8, 11$ and 13, respectively), where a node is able to cover the whole network region. It is notable that the results in both Fig. 3.11 and Fig. 3.10(a) actually imply a fact that for the network scenarios considered in this chapter, we may have a very significant per node throughput capacity improvement and may achieve its maximum possible value through adopting a bigger υ (and thus a larger transmission range) for each node, which is different from what is generally believed in literature that a smaller υ usually results in a higher throughput capacity.

3.7 Summary

Distinguished from the available works which mainly focus on deriving the order sense results and exploring the scaling laws of the throughput capacity in MANETs, in this chapter we addressed another basic problem: for a MANET with general node transmission range control and packet redundancy control, what is the exact achievable per node throughput capacity. We found that for the considered network scenarios, it may not be always true that adopting local transmission can achieve the maximum per node throughput capacity, as what is generally believed in literature. This finding indicates that further deliberate studies are necessary to reveal the real achievable network throughput of MANETs. Another interesting finding of our work is that the MANETs considered in this chapter actually exhibit very similar behaviors in terms of packet delay and per node throughput under different node mobility models, like the i.i.d., random walk and random waypoint.

Chapter 4

Generalized Two-Hop Relay for Flexible Delay Control in MANETs

4.1 Introduction

The available two-hop relay protocols with out-of-order or strictly in-order reception cannot provide a flexible control for the packet delivery delay, which may significantly limit their applications to the future MANETs with different delay requirements. In this chapter, we extend the conventional two-hop relay and proposes a general group-based two-hop relay algorithm with packet redundancy. A Markov chain-based theoretical framework is further developed to analyze how the mean value and variance of packet delivery delay vary with the parameters packet redundancy and group size, where the important medium contention, interference and traffic contention issues are carefully incorporated into the analysis.

4.2 Related Research Works

The available two-hop relay algorithms adopt either out-of-order or strictly in-order reception, which are the two extreme cases of reception mode. In the in-order two-hop relay algorithms, like the ones proposed in [10, 31, 32], each packet should be received in-order at its destination. The algorithms in [1, 33, 34] can be regarded as the out-of-order two-hop without redundancy, where a packet has at most one copy and gets accepted by its destination if it is "fresh" (never received before). The out-of-order two-hop relay with redundancy has also been explored recently [27, 35], where each packet may have multiple copies in the transmission process.

A significant amount of work has been done on the delay performance of the two-hop relay algorithms. These works mainly focus on closed-form analysis or order-sense scaling law study of expected packet delay in a two-hop relay network.

Closed-form Delay Analysis: Liu *et al.* [32] considered a two-hop relay algorithm with redundancy and in-order reception in a time-slotted system, and derived closed-form results for the expected end-to-end per packet delay. The expected delivery delay analysis under continuous system models was conducted in [43, 56, 57, 58], where the inter-meeting time between two nodes, i.e., the time elapsed between two consecutive encounters for a given pair of nodes, is assumed to be exponentially distributed, and it is assumed that the network has only one source-destination pair, and the source node has only one single packet to deliver to the destination.

Order-sense Delay Scaling Laws: The delay scaling law of the two-hop relay with out-of-order reception but without redundancy has been extensively examined in the regime of ad hoc mobile networks. Gamal *et al.* [8] showed that under the random walk model, the two-hop relay results in a $\Theta(n \log n)$ delay and achieves a $\Theta(1)$ throughput. Later, Mammen *et al.* [34] proved that the same delay and throughput scalings are also achievable even with a variant of the two-hop relay and a restricted mobility model. Gamal *et al.* [9] showed that under the two-dimensional Brownian motion on a torus of size $\sqrt{n} \times \sqrt{n}$, the delay scales as $\Theta(n^{1/2}/v(n))$, where $v(n)$ is the velocity of mobile nodes. Lin *et al.* [49] also considered the Brownian mobility model, and showed that the two-hop relay results in an expected delay of $\Omega(\log n/\sigma_n^2)$, where σ_n^2 is the variance parameter of the Brownian motion model. Sharma *et al.* [59] further showed that when the network is divided into $n^\alpha \times n^\alpha$ cells, the two-hop delay is $\Theta(n)$ for $0 \leq \alpha < 1/2$ and $\Theta(n \log n)$ for $\alpha = 1/2$ under a family of discrete random direction models, while the delay becomes $\Theta(n)$ for $\alpha < 1/2$ and $\Theta(n \log n)$ for $\alpha = 1/2$ when a family of hybrid random walk models are considered. Recently, the delay performance of a variant of two-hop relay has been examined under a correlated mobility model [23], where nodes are partitioned into different groups and all nodes of the same group have to reside concurrently within a circular region around the group center.

In the case of allowing packet redundancy and in-order reception, Neely and Modiano [10] considered a modified version of the two-hop relay algorithm for ad hoc mobile networks, and proved that under the i.i.d. mobility model it achieves $O(\sqrt{n})$ delay with exact $\sqrt{n}$ redundancy for each packet. Sharma and Mazumdar explored the order-sense delay results in ad hoc mobile networks with multiple redundancy for each packet, and proved that it achieves $\Theta(T_p(n)\sqrt{n})$ delay under the random way-point mobility model [50] and achieves $O(T_p(n)\sqrt{n \log n})$ delay under the Brownian mobility model [31], where $T_p(n)$ is the packet transmission time.

4.2.1 Limitation of Available Works

For the two-hop relay with in-order reception, lot of reception opportunities may be wasted as the destination only accepts packets according to their sequence orders, resulting in an increase in the packet delivery delay. The out-of-order two-hop relay, on the other hand, can take the full advantage of each reception opportunity but each mobile node there should potentially carry a very big (if not infinite) buffer to accommodate all possible arrivals, which is not really practical for the MANETs. Also, the early arrived packets there may need to wait a long time for the arrivals of other related packets, which may make the early arrived packets become expired and thus useless. The packet delay for two-hop relay MANETs has been extensively studied in the literature, in terms of its order sense scaling laws with network size or its closed-form analytical models. These delay results indicate that the available out-of-order or strictly in-order two-hop relay protocols, although simple and easy to operate, can not provide a flexible control for the packet delivery delay. The lack of a flexible delay control in available two-hop relay protocols may significantly limit their ability to support various delay sensitive applications in the future MANETs, like VoIP [60, 61, 62], video streaming [63, 64], real-time monitoring and networked control [65, 66], etc.

4.3 System Assumptions

The concerned network consists of n mobile nodes inside a unit square, which is evenly divided into $m \times m$ cells. We focus on a slotted system and a fast mobility scenario [23], where only one-hop transmissions are possible within each time slot, and the total number of bits transmitted per slot is fixed and normalized to 1 packet. The nodes independently roam from cell to cell, following the bi-dimensional i.i.d. mobility model [10]. At the beginning of each time slot, each node independently and uniformly selects a cell among all m^2 cells and stays in it for the whole time slot. The protocol model with guarding factor Δ in [21] is adopted as the interference model here. We further assume a permutation traffic pattern in the saturated case [23], where each node is a source and at the same time a destination of some other node, and each source node always has packets waiting for delivery. For a given source-destination pair, we call the traffic between them as a flow.

Notice that each node can be a potential relay for other $n-2$ flows (except the two flows originated from and destined for itself), thus, to support the operation of the 2HR-(f, g) algorithm, we assume that each node maintains $n+1$ individual queues at its buffer: one local-queue for storing the packets that are locally generated at the node and waiting for their copies (up to f copies for each packet) to be dispatched, one already-sent-queue for storing packets whose f copies have already been dispatched but their reception status are not confirmed yet (from destination node), one designated-queue for storing packets destined to this node and $n-2$ parallel relay-queues for storing packets of other flows (one queue per flow).

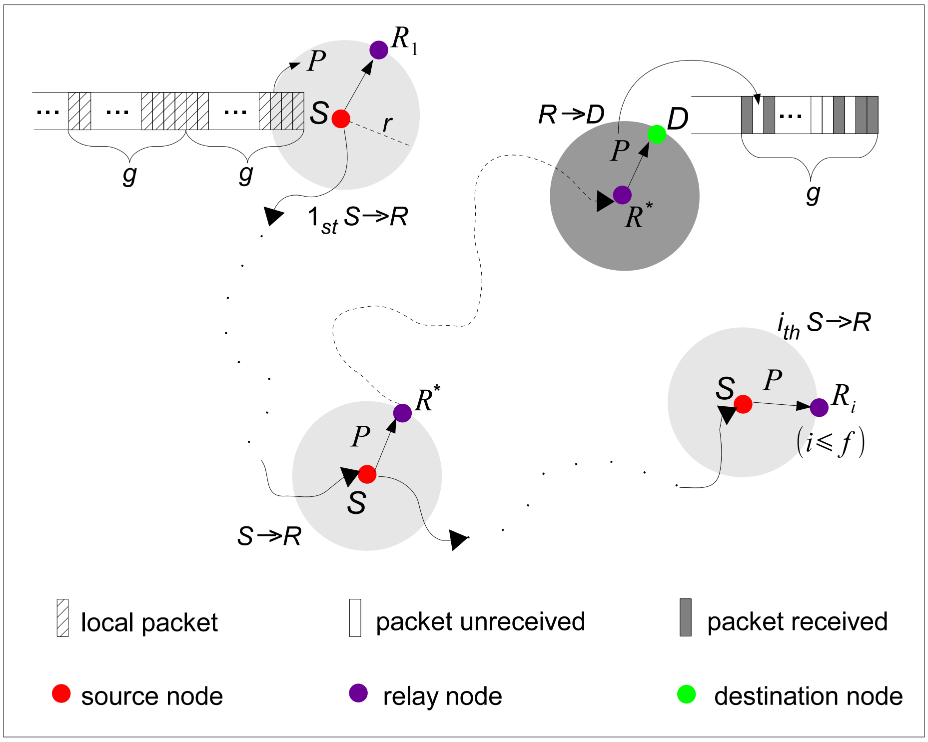


Figure 4.1: Illustration of the 2HR-(f, g) algorithm for a tagged flow.

4.4 2HR-(f, g) Algorithm and Transmission Scheduling

4.4.1 2HR-(f, g) Algorithm

Without loss of generality, we focus on a tagged flow and denote its source node and destination node as S and D, respectively. As illustrated in Fig. 4.1 that with the 2HR-(f, g) algorithm, the source node S will deliver at most f copies of a packet P to distinct relay nodes, while the destination D may finally receive the packet from one relay node R^*.

To support the group based transmission in the 2HR-(f, g) algorithm, the source node S divides packets waiting at its local-queue into consecutive groups, g packets per group (as shown in Fig. 4.1), and labels each packet P with a *send group number* $SG(P)$ and a *sequence number* $SN(P)$ $(1 \leq SN(P) \leq g)$. Similarly, the node D also maintains a *request group number* $RG(D)$ and an *indicator vector* $IN(D)$. The $IN(D)$ is a g-bit

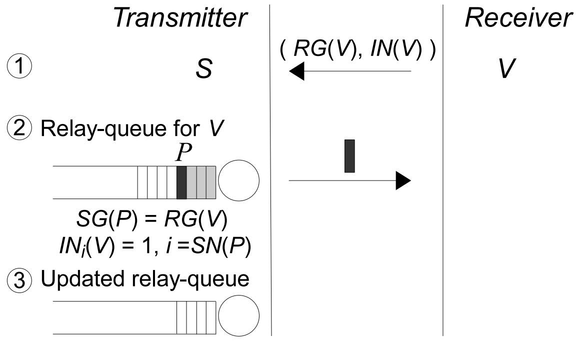


Figure 4.2: Illustration of the Relay-to-Destination mode, where the node S (acting as a relay) transmits a fresh packet P destined for the node V.

binary vector that records the reception status of current requesting group at D, where the i_{th} bit $IN_i(D)$ is set as 0 (resp. 1) if the i_{th} packet of the current requesting group has (resp. has not) been received. To simplify the analysis, we assume that each relay node will carry at most one packet for any particular group. We further introduce the following definitions:

- **Fresh packet and non-fresh packet:** A packet is called a fresh packet if it has not been received yet by its destination; a non-fresh packet, otherwise.
- **Fresh node and non-fresh node:** For a tagged packet group, a node (except the source S and the destination D) is called a fresh node if it is carrying a fresh packet for the group; otherwise, if the node is either carrying a non-fresh packet or carrying no packet for the tagged group, it is called a non-fresh node.

Based on the above definitions, the 2HR-(f, g) algorithm can be summarized as follows.

2HR-(f, g) Algorithm: For the tagged flow, every time the node S gets a transmission opportunity, it operates as follows:

Step 1: (Source-to-Destination) If the node D is among its one-hop neighbors, it initiates a handshake with D to get its $RG(D)$ and $IN(D)$. Then it tries to transmit a fresh packet directly to D, where the packet to be transmitted is selected as follows: it first checks its local-queue, starting from its head-of-line packet P_h, to find a fresh packet; if it fails, then it tries to retrieve a fresh packet from the already-sent-queue.

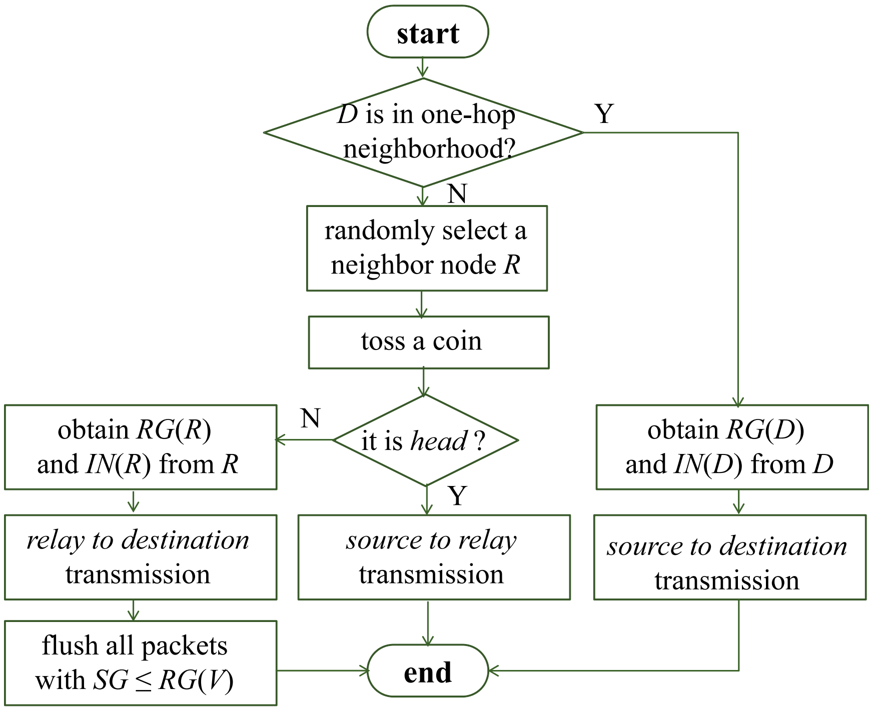


Figure 4.3: Flowchart of the 2HR-(f, g) algorithm.

Step 2: Otherwise, if the node D is not among the one-hop neighbors of S, the node S randomly chooses one of the following two operations to perform:

- (Source-to-Relay) It first randomly selects one node (say R) from its current one-hop neighbors, then initiates a handshake with R to check whether the node R is a non-fresh node. If so, it delivers a new copy of P_h to R; otherwise it remains idle for this time slot.

- (Relay-to-Destination) It acts as a relay and randomly selects one node (say V) as the receiver from its one-hop neighbors. As indicated in Fig. 4.2 that it first initiates a handshake with V to get the $RG(V)$ and $IN(V)$, then checks its relay-queue specified for V whether there exists a fresh packet of group $RG(V)$. If so, it delivers this packet to V and deletes all packets with $SG \leq RG(V)$ from its relay-queue for V; otherwise it remains idle for this time slot.

Fig. 4.3 illustrates the flowchart of the 2HR-(f, g) algorithm. Notice that in the above source-to-relay transmission, every time S sends out a copy of P_h it checks whether f copies of P_h have already been delivered. If yes, it puts P_h to the end of the already-sent-

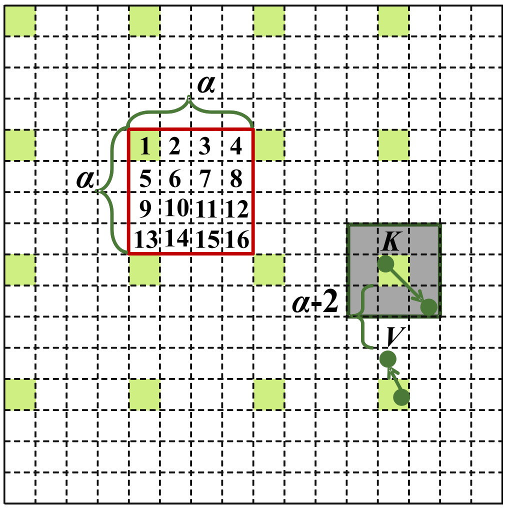


Figure 4.4: Tuning of the concurrent-set for the 2HR-(f, g) algorithm.

queue and then moves ahead the remaining packets in the local-queue. At the relay node R, P_h is put at the end of its relay-queue dedicated to the node D. Thus, each packet may have at most $f+1$ copies in the network (including the one in the already-sent-queue of its source node).

Remark 1. *In the 2HR-(f, g) algorithm, if the node D is currently requesting for packets of group i, then any fresh packet belonging to the group i is eligible for reception at the node. The node D will start to receive packets of the next group $i+1$ only after all packets of the group i have been received. Thus, the 2HR-(f, g) algorithm ensures that the inter-group packet reception is strictly in-group-order while the intra-group packet reception is totally out-of-order.*

Remark 2. *The 2HR-(f, g) algorithm is flexible and general, since its packet delivery process can be flexibly controlled by a proper setting of the redundancy f and group size g. Actually, the new algorithm covers all the available two-hop relays as special cases, like the out-of-order ones with redundancy [27, 35] ($f > 1$, $g = \infty$) or without redundancy [1, 33, 34] ($f = 1$, $g = \infty$), and the strictly in-order ones [10, 31, 32] ($f \geq 1$, $g = 1$).*

4.4.2 Tuning of Concurrent-Set

To guarantee the simultaneous transmissions in a concurrent-set without interfering with each other, the parameter α should be set properly. We consider a local transmission scenario, in which a node in some cell can only send packets to the nodes in the same cell or its eight adjacent cells. Two cells are called adjacent if they share a common point. Thus, the maximum distance between a transmitting node (transmitter) and a receiving node (receiver) is $\sqrt{8}/m$, so we set the communication range as $r = \sqrt{8}/m$. Due to the wireless interference, only cells that are sufficiently far away could simultaneously transmit without interfering with each other. As shown in Fig. 4.4, suppose that during some time slot, the node V is scheduled to receive a packet. According to the definition of "concurrent-set", we know that except the transmitting node of V, another transmitting node (say node K) in the same concurrent-set is at least $(\alpha - 2)/m$ away from V. The condition that K will not interfere with the reception at V is that, $(\alpha - 2)/m \geq (1+\Delta)\cdot r$. By substituting $r = \sqrt{8}/m$, we obtain that $\alpha \geq (1+\Delta)\sqrt{8} + 2$. As α is an integer and $\alpha \leq m$, we set

$$\alpha = \min\left\{\lceil (1+\Delta)\sqrt{8}\rceil + 2, m\right\},$$

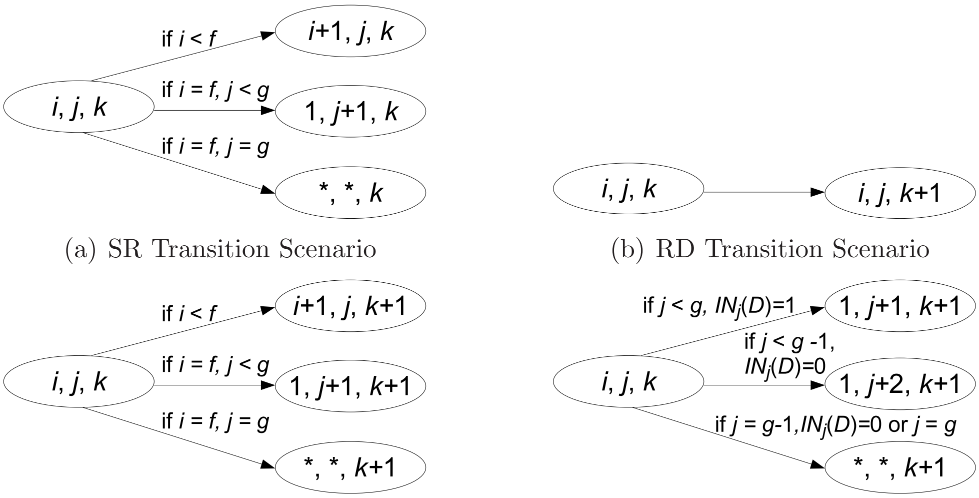
where $\lceil x \rceil$ returns the smallest integer not less than x.

Notice that each cell will become active (i.e., get transmission opportunity) in every α^2 time slots. If there are more than one nodes inside an active cell, a transmitting node is selected randomly from them, and the selected node then follows the 2HR-(f, g) algorithm for packet transmission.

4.5 Markov Chain-Based Framework

4.5.1 Theoretical Framework

For a tagged packet group at the source node S, we use a three-tuple (i, j, k) to denote the transient state that S is delivering the i_{th} $(1 \leq i \leq f)$ copy for the j_{th} $(1 \leq j \leq g)$ packet while the destination node D has already received any k $(0 \leq k < g,\ k \leq j)$ of the g



(a) SR Transition Scenario

(b) RD Transition Scenario

(c) SR+RD Transition Scenario

(d) SD Transition Scenario

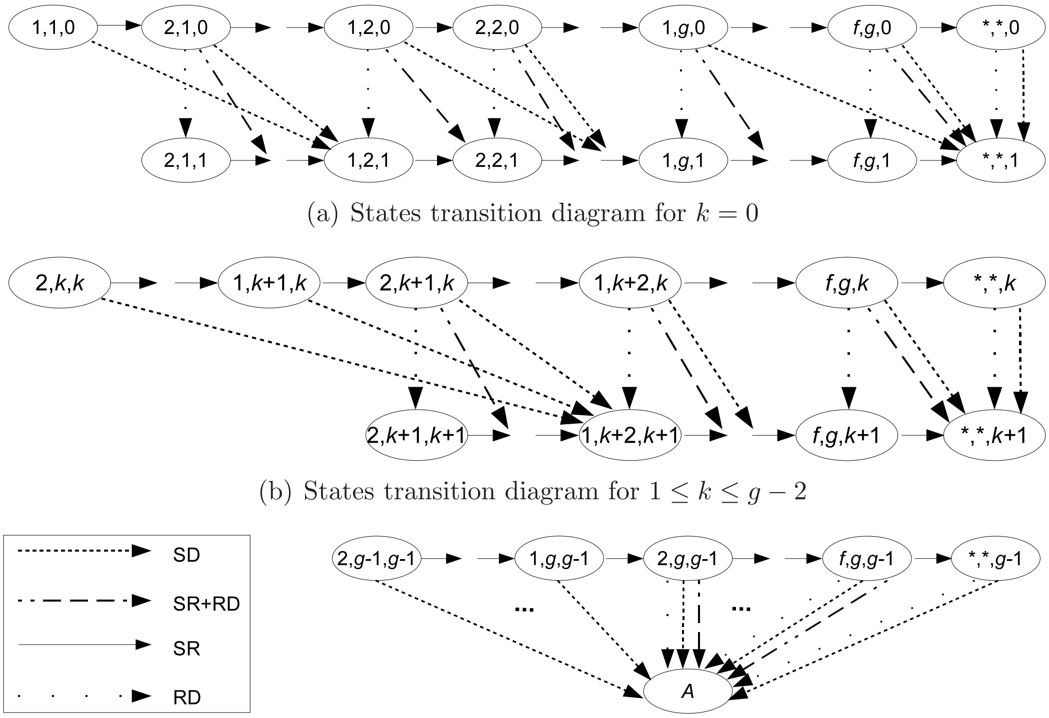
Figure 4.5: Transition scenarios of a state (i, j, k), where $1 \leq i \leq f$, $1 \leq j \leq g$ and $0 \leq k < g$, $k \leq j$.

packets. We further use $(*, *, k)$ to denote the transient state that S has already finished dispatching the copies of all packets in the tagged group while D has only received k $(0 \leq k < g)$ of them. From the operation of the 2HR-(f, g) algorithm we know that if a node pair (S, D) is in state (i, j, k) at the current time slot, then only one of the following four transmission scenarios illustrated in Fig. 4.5 may happen in the next time slot:

- SR Scenario: source-to-relay transmission only, i.e., S successfully delivers the i_{th} copy to a new relay while none of the relays delivers a fresh packet to D. As shown in Fig. 4.5(a) that under such a transition scenario, the state (i, j, k) may transit to three different neighboring states depending on the current copy index i and also the sequence number j of the current packet.

- RD Scenario: relay-to-destination transmission only, i.e., some relay node successfully delivers a fresh packet to node D while S fails to deliver out the i_{th} copy to a new relay node. As shown in Fig. 4.5(b) that there is only one target state $(i, j, k+1)$ under the RD transition scenario.

- SR+RD Scenario: both source-to-relay and relay-to-destination transmissions, i.e., these two transmissions happen simultaneously. We can see from Fig. 4.5(c) that depending on both the values of i and j, there are three possible target states under



(a) States transition diagram for $k = 0$

(b) States transition diagram for $1 \le k \le g - 2$

(c) States transition diagram for $k = g - 1$

Figure 4.6: Transition diagram of the Markov chain for the general 2HR-(f, g) algorithm.

the SR+RD transition scenario, similar to that under the SR scenario.

- SD Scenario: source-to-destination transmission, i.e., S successfully delivers out a fresh packet to D. As shown in Fig. 4.5(d) that under the SD transition scenario, the state (i, j, k) may transit to $(1, j + 1, k + 1)$, $(1, j + 2, k + 1)$ or $(*, *, k + 1)$, depending on the sequence number j of the current packet and also its reception status $IN_j(D)$.

If we use A to denote the absorbing state that the destination node D has received all the g packets of the tagged group, then the transition diagrams in Fig. 4.5 indicate that the packet delivery process in a 2HR-(f, g)-based network can be modeled as a discrete-time finite-state absorbing Markov chain illustrated in Fig. 4.6, where Figs. 4.6(a), 4.6(b) and 4.6(c) each represents some cases of the full chain. Specifically, Fig. 4.6(a) defines the transitions among neighboring states when no more than one packet is received by D, i.e., $k = 0$; Fig. 4.6(b) represents the cases that D may receive at most one more fresh packet of the tagged group given that it has already received k packets of the group, $1 \le k \le g - 2$;

Fig. 4.6(c) shows the transition diagrams of how D may receive the last packet. The transitions of SD, SR, RD and SR+RD in Fig. 4.6 correspond to the 2HR-(f, g) transmissions of source-to-destination, source-to-relay, relay-to-destination, and both source-to-relay and relay-to-destination, respectively.

Remark 3. *The Markov chain model in Fig. 4.6 covers the available models for conventional two-hop relay analysis as special cases when we set $g = 1$ there [43, 56, 57, 58].*

Although the Markov chain framework in Fig. 4.6 is general enough to model the packet delivery process in a 2HR-(f, g)-based network, it is difficult to directly apply such framework for an accurate packet delay analysis, even for the simple scenario of $f > 1$ and $g = 1$ [43, 57]. This is mainly due to the complicated transitions that may happen among transient states. As shown in Fig. 4.5 that for a state (i, j, k), its next state may vary significantly with the transition scenarios (SR, RD, SR+RD or SD), with the values of i, j and k, and also with the reception status $IN(D)$.

To simplify the analysis of packet delivery delay and enable both its mean value and variance to be derived analytically, we introduce the following assumption regarding the complex SD transition scenario:

Assumption 1. *Under the SD transition scenario in Fig. 4.5(d), the transient state (i, j, k) will always transit to state $(1, j + 1, k + 1)$ whenever $k < j < g$.*

The Assumption 1 indicates that for a transient state (i, j, k) under the SD transition scenario, if $k < j < g$, i.e., the source node S is currently delivering the j_{th} packet while only less than j packets have been received at the destination node D by now, then we assume that this j_{th} packet has not been received yet by D.

This assumption is due to the following observations: 1) To incorporate all the reception details of the j_{th} packet into analysis, a much complex Markov chain with bigger state space and more complex transitions among neighboring states should be adopted. We must be careful to avoid arriving at intractably complex models, so the case that the j_{th} packet has been received by D is neglected here; 2) Notice that the source node S

always delivers out packets sequentially, so the packet delivered out earlier will be likely received early at the destination. Thus, the j_{th} packet currently being delivered at S is very likely not received yet by D given that only $k < j$ packets have been received at D by now. Notice also that this assumption only applies to the special SD transition scenario, which in general happens with negligible probability in comparison with that of the SR or RD transition scenario in a large MANET. Therefore, the simplification introduced by the Assumption 1 will not cause a significant error in the overall packet delay analysis, as to be validated in Section V.

4.5.2 Some Basic Results

We can easily see from Fig. 4.6 that in the Markov-chain model, the total number of transient states β is determined as

$$\beta = \frac{f}{2}(g^2 + 3g - 2) + 1. \tag{4.1}$$

Actually, these β transient states are arranged into g rows, where the number of transient states L_k in the k_{th} row $(0 \le k \le g-1)$ is given by

$$L_k = \begin{cases} (g+1-k)f & \text{if } 1 \le k \le g-1, \\ g \cdot f + 1 & \text{if } k = 0. \end{cases} \tag{4.2}$$

We now establish the following lemmas regarding some basic results of the Markov-chain model in Fig. 4.6, which will help us for packet delay analysis in Section IV.

Lemma 8. *For the Markov-chain model in Fig. 4.6, the number of fresh nodes u_r and the number of non-fresh nodes u_o in the t_{th} transient state of k_{th} row, $t \in [1, L_k]$, $k \in [1, g-1]$, can be determined as*

$$u_r \approx t - f, \tag{4.3}$$

$$u_o \approx n - 2 - t + k - (k-1)f. \tag{4.4}$$

Proof. From the Assumption 1 we can easily see that for a transient state (i, j, k) in the k_{th} row of the Markov chain model in Fig. 4.6, the number of corresponding fresh nodes u_r can be approximated as

$$u_r \approx i - 1 + (j - 1 - k)f. \tag{4.5}$$

Notice that in a large MANET, the probability of direct source-to-destination transmission is negligible in comparison with that of the source-to-relay or relay-to-destination transmissions, so with high probability the destination will receive each of the k packets from a relay node rather than the source node. Thus, the number of corresponding non-fresh nodes u_o can be approximated as

$$u_o \approx n - 2 - (i - 1) - (j - 1)f + k. \tag{4.6}$$

Suppose that the state (i, j, k) is the t_{th} transient state in the k_{th} row of the Markov chain model in Fig. 4.6, $k \in [1, g - 1]$, $t < L_k$, then we have

$$t = (j - k)f + i - 1. \tag{4.7}$$

By combining (4.7) with (4.5) and (4.6), the formulas (4.3) and (4.4) then follow. □

Lemma 9. *For a given time slot and a tagged flow, we use p_1 and p_2 to denote the probability that S conducts a source-to-destination transmission and the probability that S conducts a source-to-relay or relay-to-destination transmission, respectively. Then we have*

$$p_1 = \frac{1}{\alpha^2}\left(\frac{9n - m^2}{n(n-1)} - \left(1 - \frac{1}{m^2}\right)^{n-1} \frac{8n + 1 - m^2}{n(n-1)}\right), \tag{4.8}$$

$$p_2 = \frac{1}{\alpha^2}\left(\frac{m^2 - 9}{n - 1}\left(1 - \left(1 - \frac{1}{m^2}\right)^{n-1}\right) - \left(1 - \frac{9}{m^2}\right)^{n-1}\right). \tag{4.9}$$

Proof. Consider a tagged active cell, the node S can conduct a source-to-destination transmission with D only under the following two mutually exclusive cases: both S and

D are in this cell; or S is in this cell while D is in the eight adjacent cells of this cell. If we further assume that aside from S and D, there are k other nodes in this cell, $k \in [0, n-2]$, then the probability that S is selected as the transmitter is $\frac{1}{k+2}$ (resp. $\frac{1}{k+1}$) under the former case (resp. under the latter case). Summing up the probabilities under these two cases, then we have

$$
\begin{aligned}
p_1 = & \frac{1}{\alpha^2}\Bigg(\sum_{k=0}^{n-2}\binom{n-2}{k}\Big(\frac{1}{m^2}\Big)^k\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{1}{m^2(k+2)} \\
& +\sum_{k=0}^{n-2}\binom{n-2}{k}\Big(\frac{1}{m^2}\Big)^k\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{8}{m^2(k+1)}\Bigg) \\
= & \frac{1}{\alpha^2}\Bigg(\sum_{k=0}^{n-2}\binom{n-1}{k+1}\Big(\frac{1}{m^2}\Big)^{k+1}\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{1}{k+2} \\
& -\sum_{k=0}^{n-2}\binom{n-2}{k+1}\Big(\frac{1}{m^2}\Big)^{k+1}\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{1}{k+2} \\
& +\sum_{k=0}^{n-2}\binom{n-2}{k}\Big(\frac{1}{m^2}\Big)^{k+1}\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{8}{k+1}\Bigg) \\
= & \frac{1}{\alpha^2}\Bigg(\frac{9-m^2}{n-1}+\frac{m^2}{n}-\frac{8}{n-1}\Big(\frac{m^2-1}{m^2}\Big)^{n-1} \\
& +\Big(\frac{m^2}{n-1}-\frac{m^2}{n}\Big)\Big(\frac{m^2-1}{m^2}\Big)^{n}\Bigg).
\end{aligned} \tag{4.10}
$$

The formula (4.8) can then be easily derived from (4.10) after some basic algebraic operations.

Similarly, S conducts a source-to-relay or relay-to-destination transmission iff the following four events happen simultaneously: S is in an active cell, S is selected as the transmitter, there is at least one other node (except S and D) in the same cell of S or its eight adjacent cells, and the node D is in one of the other m^2-9 cells. Thus, we have

$$
\begin{aligned}
p_2 = & \frac{m^2-9}{m^2\alpha^2}\Bigg(\sum_{k=1}^{n-2}\binom{n-2}{k}\Big(\frac{1}{m^2}\Big)^k\Big(\frac{m^2-1}{m^2}\Big)^{n-2-k}\frac{1}{k+1} \\
& +\sum_{k=1}^{n-2}\binom{n-2}{k}\Big(\frac{8}{m^2}\Big)^k\Big(\frac{m^2-9}{m^2}\Big)^{n-2-k}\Bigg) \\
= & \frac{1}{\alpha^2}\Bigg(\frac{m^2-9}{n-1}\Bigg(1-\Big(1-\frac{1}{m^2}\Big)^{n-1}\Bigg)-\Big(1-\frac{9}{m^2}\Big)^{n-1}\Bigg).
\end{aligned}
$$

$\square$

Lemma 10. *For a tagged flow, suppose that the source node S is delivering copies for some packet group i in the current time slot, the destination node D is requesting the packets of the group i, and there are currently t_1 fresh nodes and t_2 non-fresh nodes for the group i in the network. For the next time slot, we use $P_r(t_1)$ to denote the probability that D will receive a fresh packet, use $P_d(t_2)$ to denote the probability that S will successfully deliver out a copy to some new relay, and use $P_s(t_1, t_2)$ to denote the probability that both a successful source-to-relay transmission and a relay-to-destination transmission will be performed. Then we have*

$$P_r(t_1) = p_1 + \frac{t_1}{2(n-2)} p_2, \tag{4.11}$$

$$P_d(t_2) = \frac{t_2}{2(n-2)} p_2, \tag{4.12}$$

$$P_s(t_1, t_2) = \frac{t_1 t_2 (m^2 - \alpha^2)}{4m^2\alpha^4} \sum_{k=0}^{n-5} \binom{n-5}{k} h(k) \cdot \left\{ \sum_{t=0}^{n-4-k} \binom{n-4-k}{t} h(t) \Big(1 - \frac{18}{m^2}\Big)^{n-4-k-t} \right\}, \tag{4.13}$$

where

$$h(x) = \frac{9\big(\frac{9}{m^2}\big)^{x+1} - 8\big(\frac{8}{m^2}\big)^{x+1}}{(x+1)(x+2)}. \tag{4.14}$$

Proof. In the next time slot, the destination node D may receive a fresh packet either from the source node S or from one of the t_1 fresh nodes. Notice that these $t_1 + 1$ events are mutually exclusive, the probability that D receives a fresh packet from S is p_1, and the probability that D receives a fresh packet from a single fresh node is $\frac{p_2}{2(n-2)}$. By summing the probabilities of these $t_1 + 1$ events, the formula (4.11) follows.

Similarly, given that there are t_2 non-fresh nodes, in the next time slot the node S may deliver out a new copy to any one of them. Notice that these t_2 events are also exclusive, and the probability that S delivers out a copy to a single non-fresh node is $\frac{p_2}{2(n-2)}$, so the formula (4.12) follows.

To derive $P_s(t_1, t_2)$, let's focus on a specific fresh node R and a specific non-fresh node V, and use $P(S \to V, R \to D)$ to denote the probability that the a source-to-relay transmission from S to V and a relay-to-destination transmission from R to D happen simultaneously in the next time slot. Thus, the $P_s(t_1, t_2)$ can be determined as

$$P_s(t_1, t_2) = t_1 t_2 \cdot P(S \to V, R \to D). \tag{4.15}$$

The basic idea is to treat the event $(S \to V, R \to D)$ as two simultaneous but mutually independent transmissions, i.e., the relay-to-destination transmission $R \to D$ and the source-to-relay transmission $S \to V$, then divide each transmission into multiple mutually exclusive cases, and finally represent each case with several simultaneous but independent sub-events. First consider the active cell with node R. The node R can conduct a relay-to-destination transmission with D only under the following two mutually exclusive cases: D is in this cell or D is in one of the eight adjacent cells. If we further assume that except the S, D, R, V, and the destination node of R's local traffic,there are in total k other nodes in the one-hop neighborhood of R, $k \in [0, n-5]$, among them i nodes are in the same cell as R, $i \in [0, k]$, and the other $k-i$ nodes are in the eight adjacent cells. Then the probability that R and D are selected as the transmitter and the receiver, respectively, is $\frac{1}{(i+2)(k+1)}$ (resp. $\frac{1}{(i+1)(k+1)}$) under the former case (resp. under the latter case). Summing up the probabilities under these two cases, then we get the corresponding probability of the relay-to-destination transmission $R \to D$. Similarly, we can also get the probability of the source-to-relay transmission $S \to V$. By the law of multiplication, then we have

$$P(S \to V, R \to D) = \frac{m^2 - \alpha^2}{4m^2\alpha^4} \sum_{k=0}^{n-5} \binom{n-5}{k}$$
$$\cdot \left(\sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \left(\frac{1}{i+2} + \frac{8}{i+1} \right) \left(\frac{1}{m^2} \right)^{i+1} \left(\frac{8}{m^2} \right)^{k-i} \right)$$

$$\cdot \sum_{t=0}^{n-4-k} \binom{n-4-k}{t} \Bigg(\sum_{j=0}^{t} \binom{t}{j} \frac{1}{t+1} \Big(\frac{1}{j+2} + \frac{8}{j+1} \Big)$$

$$\cdot \Big(\frac{1}{m^2} \Big)^{j+1} \Big(\frac{8}{m^2} \Big)^{t-j} \Bigg) \Big(\frac{m^2-18}{m^2} \Big)^{n-4-k-t}. \tag{4.16}$$

Notice that

$$\begin{aligned}
&\sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \Big(\frac{1}{i+2} + \frac{8}{i+1} \Big) \Big(\frac{1}{m^2} \Big)^{i+1} \Big(\frac{8}{m^2} \Big)^{k-i} \\
=& \sum_{i=0}^{k} \binom{k+1}{i+1} \frac{1}{k+1} \frac{1}{i+2} \Big(\frac{1}{m^2} \Big)^{i+1} \Big(\frac{8}{m^2} \Big)^{k-i} \\
&- \sum_{i=0}^{k} \binom{k}{i+1} \frac{1}{k+1} \frac{1}{i+2} \Big(\frac{1}{m^2} \Big)^{i+1} \Big(\frac{8}{m^2} \Big)^{k-i} \\
&+ \sum_{i=0}^{k} \binom{k}{i} \frac{1}{k+1} \frac{8}{i+1} \Big(\frac{1}{m^2} \Big)^{i+1} \Big(\frac{8}{m^2} \Big)^{k-i} \\
=& \frac{1}{(k+1)(k+2)} \Bigg(9 \Big(\frac{9}{m^2} \Big)^{k+1} - (k+10) \Big(\frac{8}{m^2} \Big)^{k+1} \Bigg) \\
&- \frac{1}{(k+1)^2} \Bigg(8 \Big(\frac{9}{m^2} \Big)^{k+1} - (k+9) \Big(\frac{8}{m^2} \Big)^{k+1} \Bigg) \\
&+ \frac{8}{(k+1)^2} \Bigg(\Big(\frac{9}{m^2} \Big)^{k+1} - \Big(\frac{8}{m^2} \Big)^{k+1} \Bigg) \\
=& \frac{9 \big(\frac{9}{m^2} \big)^{k+1} - 8 \big(\frac{8}{m^2} \big)^{k+1}}{(k+1)(k+2)}.
\end{aligned} \tag{4.17}$$

The (4.13) then follows by combining (4.15), (4.16) and (4.17). □

4.6 Packet Delivery Delay Analysis

With the help of the Markov-chain framework and related basic results in Section III, this section provides the study of both expected value and standard deviation of packet delivery delay under the 2HR-(f, g) algorithm. We first introduce the following definition about the delivery delay of a packet group.

Definition 3. *For a packet group at a source node S, the delivery delay of the group is the time elapsed between the time slot S moves the first packet of the group into the head-of-line at its local-queue and the time slot when the destination node D receives the* *last packet of the group.*

For the 2HR-(f, g) relay algorithm, if we denote by $T(f, g)$ the delivery delay of a packet group and denote by T_p the average delivery delay of one packet, then we have

$$T_p = \frac{T(f,g)}{g}. \tag{4.18}$$

Remark 4. *Under the 2HR-(f, g) algorithm, the destination node D queues up the packets of a group until it receives all packets of that group, and then considers the packets of the group delivered. Thus, as the (4.18) shows that the per packet delivery delay T_p is calculated on a group basis.*

4.6.1 Expected Packet Delivery Delay

As illustrated in Fig. 4.6, all β transient states in the Markov chain model are arranged into g rows. We number these transient states sequentially as 1, 2, . . ., β, in a left-to-right and top-to-down way. For these transient states, let q_{ij} denote the transition probability from transient sate i to transient state j, then we can define a matrix $\mathbf{Q} = (q_{ij})_{\beta\times\beta}$ of transition probabilities among transient states there. From the theory of Markov chain [52] we know that the fundamental matrix $\mathbf{N}$ of the Markov chain in Fig. 4.6 is given by

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}, \tag{4.19}$$

where $\mathbf{N} = (a_{ij})_{\beta\times\beta}$ and the entry a_{ij} denotes the expected number of times in the j_{th} transient state until absorption given that the chain starts from the i_{th} transient state.

Based on the Markov chain structure in Fig. 4.6, we can actually partition the matrix $\mathbf{N}$ into g-by-g blocks as $\mathbf{N} = (\mathbf{N}_{tk})_{g\times g}$, where the block (i.e., sub-matrix) $\mathbf{N}_{tk}$ corresponds to the expected number of times in the transient states of $(k-1)_{th}$ row of the Markov chain structure given that the Markov chain starts from the transient states of $(t-1)_{th}$ row there. If we use $\mathbf{N}_{tk}(i, j)$ to denote the ij-entry of a block $\mathbf{N}_{tk}$, and further use b_i to denote the time the Markov chain takes to become absorbed given that the chain starts from the i_{th} transient state ($1 \le i \le \beta$), then we have

$$T(f, g) = b_1, \tag{4.20}$$

where the expected value $\mathbb{E}\{b_1\}$ of b_1 is given by

$$\mathbb{E}\{b_1\} = \sum_{k=1}^{g} \sum_{j=1}^{L_{k-1}} \mathbf{N}_{1k}(1, j). \tag{4.21}$$

Then the expected packet delivery delay $\mathbb{E}\{T_p\}$ can be determined as

$$\mathbb{E}\{T_p\} = \frac{1}{g} \sum_{k=1}^{g} \sum_{j=1}^{L_{k-1}} \mathbf{N}_{1k}(1, j). \tag{4.22}$$

4.6.2 Standard Deviation

From (4.18) and (4.20) we can easily see that the variance of packet delivery delay $Var\{T_p\}$ can be determined as

$$Var\{T_p\} = \frac{1}{g^2} Var\{b_1\}. \tag{4.23}$$

Since $Var\{b_1\} = \mathbb{E}\{{b_1}^2\} - (\mathbb{E}\{b_1\})^2$ and $\mathbb{E}\{b_1\}$ can be determined by (4.21), we only need to derive the $\mathbb{E}\{{b_1}^2\}$ here.

Based on the definition of b_i we can see that the expected value $\mathbb{E}\{{b_i}^2\}$ is given by

$$\begin{aligned} \mathbb{E}\{{b_i}^2\} &= \sum_{j=1}^{\beta} q_{ij} \mathbb{E}\{(1 + b_j)^2\} \\ &= 1 + 2 \sum_{j=1}^{\beta} q_{ij} \cdot \mathbb{E}\{b_j\} + \sum_{j=1}^{\beta} q_{ij} \cdot \mathbb{E}\{{b_j}^2\}. \end{aligned} \tag{4.24}$$

Let $\mathbf{b}^{(j)} = (\mathbb{E}\{{b_1}^j\}, \mathbb{E}\{{b_2}^j\}, \ldots, \mathbb{E}\{{b_\beta}^j\})^T$, then we can rearrange (4.24) as

$$\mathbf{I} \cdot \mathbf{b}^{(2)} = \mathbf{c} + 2\mathbf{Q} \cdot \mathbf{b}^{(1)} + \mathbf{Q} \cdot \mathbf{b}^{(2)}, \tag{4.25}$$

where $\mathbf{c}$ is the $\beta \times 1$ column vector with all entries being 1, i.e., $\mathbf{c} = \{1, 1, \ldots, 1\}^T$.

Then, according to [67], we have

$$\mathbf{b}^{(1)} = \mathbf{N} \cdot \mathbf{c}, \tag{4.26}$$

$$\mathbf{b}^{(2)} = \mathbf{N}(\mathbf{I} + 2\mathbf{Q} \cdot \mathbf{N})\mathbf{c}. \tag{4.27}$$

Since $\mathbb{E}\{{b_1}^2\} = \mathbf{e} \cdot \mathbf{b}^{(2)}$, where $\mathbf{e} = \{1, 0, \ldots, 0\}$, the $\mathbb{E}\{{b_1}^2\}$ and $Var\{T_p\}$ can be derived based on $\mathbf{Q}$ and $\mathbf{N}$.

The above results indicate clearly that the only remaining issue for the calculation of both $\mathbb{E}\{T_p\}$ and $Var\{T_p\}$ is the derivation of matrices $\mathbf{Q}$ and $\mathbf{N}$, as discussed in the follows.

4.6.3 Derivation of Matrix Q

Notice that for the Markov chain in Fig. 4.6, the transitions happen only among the transient states of the same row or neighboring rows, so the matrix $\mathbf{Q}$ there can be defined as

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_0 & \mathbf{Q}_0^{'} & & & & & & \\ & \mathbf{Q}_1 & \mathbf{Q}_1^{'} & & & & & \\ & & \ddots & \ddots & & & & \\ & & & \mathbf{Q}_k & \mathbf{Q}_k^{'} & & & \\ & & & & \ddots & \ddots & & \\ & & & & & \mathbf{Q}_{g-2} & \mathbf{Q}_{g-2}^{'} \\ & & & & & & \mathbf{Q}_{g-1} \end{bmatrix}, \tag{4.28}$$

here the block (sub-matrix) $\mathbf{Q}_k$ of size $L_k \times L_k$ corresponds to the transition probabilities among the transient states of k_{th} row in the Markov chain, while the block $\mathbf{Q}_k^{'}$ of size $L_k \times L_{k+1}$ corresponds to the transition probabilities from the transient states of the k_{th} row to that of the $(k+1)_{th}$ row. All the other blocks are zero matrices here and thus omitted for simplicity. The (4.28) indicates that to derive the matrix $\mathbf{Q}$, we just need to calculate the sub-matrices $\mathbf{Q}_k$ and $\mathbf{Q}_k^{'}$ there.

Calculation of $\mathbf{Q}_k$: Let $\mathbf{Q}_k(i,j)$ denote the ij-entry of the sub-matrix $\mathbf{Q}_k$, $i,j \in [1, L_k]$, then the non-zero entries of $\mathbf{Q}_k$ can be determined as

- when $1 \le k \le g-1$,

$$\mathbf{Q}_k(i,i+1) = \begin{cases} P_d(u_o) & \text{if } 1 \le i \le f, \\ P_d(u_o) - P_s(u_r,u_o) & \text{if } f+1 \le i < L_k, \end{cases} \tag{4.29}$$

$$\mathbf{Q}_k(i,i) = \begin{cases} 1 - p_1 - P_d(u_o) & \text{if } 1 \le i \le f, \\ 1 - P_d(u_o) - P_r(u_r) + P_s(u_r,u_o) & \text{if } f+1 \le i < L_k, \\ 1 - P_r(u_r) & \text{if } i = L_k. \end{cases} \tag{4.30}$$

- when $k = 0$,

$$\mathbf{Q}_0(i,i+1) = P_d(n-1-i) - P_s(i-1,n-1-i) \quad \text{if } 1 \le i < L_0, \tag{4.31}$$

$$\mathbf{Q}_0(i,i) = \begin{cases} 1 - P_d(n-1-i) - P_r(i-1) + P_s(i-1,n-1-i) & \text{if } 1 \le i < L_0, \\ 1 - P_r(g \cdot f) & \text{if } i = L_0. \end{cases} \tag{4.32}$$

Calculation of $\mathbf{Q}_k^{'}$: The sub-matrix $\mathbf{Q}_k^{'}$ is of size $L_k \times L_{k+1}$, where its non-zero ij-entry $\mathbf{Q}_k^{'}(i,j)$ is determined as

- when $1 \le k \le g-2$,

$$\mathbf{Q}_k^{'}(i, j(i) \cdot f) = p_1 \quad \text{if } 1 \le i < L_k, \tag{4.33}$$

where

$$j(i) = \begin{cases} 1 & \text{if } 1 \le i \le f, \\ \lfloor \frac{i}{f} \rfloor & \text{if } f+1 \le i < L_k, \end{cases} \tag{4.34}$$

and $\lfloor x \rfloor$ returns the largest integer not greater than x.

$$\mathbf{Q}_k^{'}(i, i-f) = \begin{cases} P_r(u_r) - p_1 - P_s(u_r, u_o) \text{ if } f+1 \le i < L_k, \\ P_r(u_r) \quad \text{if } i = L_k, \end{cases} \tag{4.35}$$

$$\mathbf{Q}_k^{'}(i, i-f+1) = P_s(u_r, u_o) \quad \text{if } f+1 \le i < L_k. \tag{4.36}$$

- when $k = 0$,

$$\mathbf{Q}_0^{'}(i, \lceil \frac{i}{f} \rceil \cdot f) = p_1 \qquad \text{if } 1 \le i < L_0, \tag{4.37}$$

and $\lceil x \rceil$ returns the smallest integer not less than x.

$$\mathbf{Q}_0^{'}(i, i-1) = \begin{cases} P_r(i-1) - p_1 - P_s(i-1, n-1-i) & \text{if } 2 \le i < L_0, \\ P_r(g \cdot f) & \text{if } i = L_0, \end{cases} \tag{4.38}$$

$$\mathbf{Q}_0^{'}(i, i) = P_s(i-1, n-1-i) \quad \text{if } 2 \le i < L_0. \tag{4.39}$$

4.6.4 Derivation of Matrix N

We denote the matrix $\mathbf{I} - \mathbf{Q}$ as $\mathbf{G}$, so $\mathbf{N} = \mathbf{G}^{-1}$. Based on the structure of $\mathbf{Q}$ we can see that the matrix $\mathbf{G}$ can also be defined in the block structure. Let $\{\mathbf{G}_k\}$ and $\{\mathbf{G}_k^{'}\}$ denote the main diagonal and upper diagonal blocks of $\mathbf{G}$, then we have

$$\mathbf{G}_k(i, j) = \begin{cases} 1 - \mathbf{Q}_k(i, j) & \text{if } i = j, \\ -\mathbf{Q}_k(i, j) & \text{otherwise}, \end{cases} \tag{4.40}$$

$$\mathbf{G}_k^{'}(i,j) = -\mathbf{Q}_k^{'}(i,j). \tag{4.41}$$

The following lemma indicates that the matrix $\mathbf{N}$ can be calculate based on $\{\mathbf{G}_k^{-1}\}$ and $\{\mathbf{G}_k^{'}\}$.

Lemma 11. *The fundamental matrix $\mathbf{N} = (\mathbf{N}_{ij})_{g\times g}$ of the Markov chain in Fig. 4.6 can be determined as*

$$\mathbf{N}_{ij} = \begin{cases} \mathbf{0} & \text{if } i > j, \\ \mathbf{G}_{i-1}^{-1} & \text{if } i = j, \\ (-1)^{j-i}\big(\prod_{k=i-1}^{j-2} \mathbf{G}_k^{-1}\mathbf{G}_k^{'}\big)\mathbf{G}_{j-1}^{-1} & \text{if } i < j, \end{cases} \tag{4.42}$$

where $i, j \in [1, g]$.

Proof. We can easily see that the matrix $\mathbf{G} = \mathbf{I} - \mathbf{Q}$ also has a block partition similar to that of $\mathbf{Q}$ in (4.28), with sub-matrices (blocks) in the main diagonal and upper diagonal, while other blocks are zero matrices. As to be proved in Lemma 12, each main diagonal block $\mathbf{G}_k$ of $\mathbf{G}$ is invertible (i.e., its inverse matrix $\mathbf{G}_k^{-1}$ exists), $k \in [0, g-1]$. Similarly, we can easily see that $\mathbf{G}$ is invertible and its inverse matrix $\mathbf{G}^{-1}$ (i.e., $\mathbf{N}$) is an upper block triangular matrix. Thus, the formula (4.42) follows after some basic row operations. □

Since $\{\mathbf{G}_k^{'}\}$ can be calculated based on (4.41), the only remaining issue for evaluating $\mathbf{N}$ is to determine $\{\mathbf{G}_k^{-1}\}$.

Lemma 12. *Each sub-matrix $\mathbf{G}_k$ of $\mathbf{G}$ has an inverse $\mathbf{G}_k^{-1}$, where ij-entry $\mathbf{G}_k^{-1}(i,j)$ of $\mathbf{G}_k^{-1}$ can be determined as*

$$\mathbf{G}_k^{-1}(i,j) = \begin{cases} 0 & \text{if } i > j, \\ \frac{1}{\mathbf{G}_k(i,i)} & \text{if } i = j, \\ (-1)^{j-i}\big(\prod_{t=i}^{j-1} \frac{\mathbf{G}_k(t,t+1)}{\mathbf{G}_k(t,t)}\big)\frac{1}{\mathbf{G}_k(j,j)} & \text{if } i < j, \end{cases} \tag{4.43}$$

where $k \in [0, g-1]$, $i, j \in [1, L_k]$.

Proof. As indicated in (4.40) that for each $\mathbf{G}_k$, we have $\mathbf{G}_k = \mathbf{I}_k - \mathbf{Q}_k$. Obviously, $\mathbf{G}_k$ is a square matrix of size $L_k \times L_k$. Combining the definitions of $\mathbf{Q}_k$ in the (4.29), (4.30), (4.31) and (4.32) with the fact that $0 < \mathbf{Q}_k(i,i) < 1$ and $\mathbf{Q}_k(i,i+1) > 0$, we know that $0 < \mathbf{G}_k(i,i) < 1$, $\mathbf{G}_k(i,i+1) < 0$ and all other off-diagonal entries are zero. Thus, $\mathbf{G}_k$ is invertible and its inverse matrix $\mathbf{G}_k^{-1}$ is an upper triangular matrix. After some basic row operations, the (4.43) follows. □

4.7 Numerical Results

In this section, we first verify the efficiency of the Markov chain-based framework through simulation, then apply it to explore how the parameters f and g would affect the packet delivery delay in a 2HR-(f,g) MANET.

4.7.1 Simulation Setting

A simulator was developed to simulate the packet delivery process in a 2HR-(f,g) MANET, which is now available at [53]. Similar to the settings adopted in [54, 55], the guard factor here is fixed as $\Delta = 1$, and hence the concurrent-set is defined with $\alpha = \min\{8, m\}$. Besides the bi-dimensional i.i.d. mobility model considered in this chapter, we also implemented the simulator for the popular random walk model and random waypoint model, which are defined as follows:

- **Random Walk Model [7]:** At the beginning of each time slot, each node independently makes a decision regarding its mobility action, either staying inside its current cell or moving to one of its eight adjacent cells. Each action happens with the same probability of 1/9.

- **Random Waypoint Model [68]:** At the beginning of each time slot, each node independently and randomly generates a two-dimensional vector $v = [v_x, v_y]$, where the values of v_x and v_y are uniformly drawn from $[1/m, 3/m]$. The node then moves

Table 4.1: Comparison between simulated and theoretical results for model validation, $m = 16$, Simulated / Theoretical

		$\mathbb{E}\{T_p\}$	$\sqrt{Var\{T_p\}}$
$n = 100$	$g = 1, f = 2$	$1850.3 \pm 3.5/1849.7$	$1805.5/1803.9$
	$g = 5, f = 4$	$1196.7 \pm 0.9/1198.8$	$456.62/456.81$
	$g = 10, f = 6$	$1086.6 \pm 0.5/1118.7$	$242.19/242.68$
$n = 250$	$g = 1, f = 2$	$2703.1 \pm 5.3/2702.5$	$2676.3/2675.1$
	$g = 5, f = 4$	$2043.1 \pm 1.7/2043.2$	$845.23/844.89$
	$g = 10, f = 6$	$1662.8 \pm 0.8/1670.7$	$424.09/423.55$
$n = 600$	$g = 1, f = 2$	$4686.2 \pm 9.2/4685$	$4671.6/4665$
	$g = 5, f = 4$	$4045.3 \pm 3.4/4047.6$	$1741.1/1739.6$
	$g = 10, f = 6$	$3444.5 \pm 1.9/3446$	$967.89/967.94$

a distance of v_x along the horizontal direction and a distance of v_y along the vertical direction.

The simulated expected delivery delay (SE) is calculated as the average value of 10^2 batches of simulation results, where each batch consists of 10^4 random and independent simulations. The simulated standard deviation (SSD) is the sample standard deviation, which is calculated as

$$SSD = \sqrt{\frac{1}{w-1}\sum_{i=1}^{w}(x_i - SE)^2}, \tag{4.44}$$

where $w = 10^6$, and x_i is the observed delivery delay in the i_{th} simulation. Notice that all the simulation results of the expected delivery delay are reported with the 95% confidence intervals.

4.7.2 Model Validation

Extensive simulations have been conducted to verify the Markov chain-based theoretical framework. For the fixed setting of $m = 16$, we considered three different network scenarios of $n = 100$, 250 and 600, which correspond to the sparse network (with node density 0.39), ordinary network (with node density 0.98) and dense network (with node density 2.34), respectively. For each network scenario, three different settings of parameters f and g have been examined, i.e., ($g = 1, f = 2$), ($g = 5, f = 4$) and ($g = 10, f = 6$). The corresponding simulation results and theoretical results are summarized in Table 4.1.

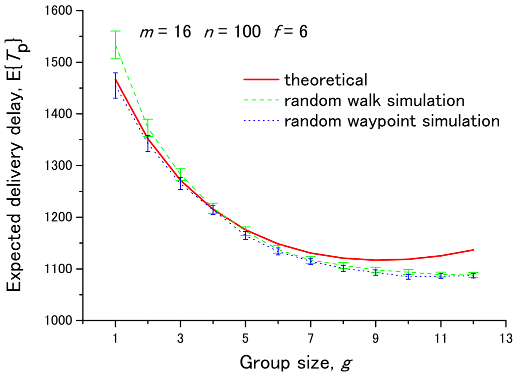
Table 4.1 indicates clearly that the simulation results match nicely with the theoretical ones for both the expected value and standard deviation of packet delivery delay, so our theoretical framework can be used to efficiently model the packet delivery process. A further careful observation of Table 4.1 shows that there is still a very small gap ($\leq 5\%$) between the simulation results and theoretical ones. For example, for the case that $n = 100$, $g = 10$ and $f = 6$, the simulated value for $\mathbb{E}\{T_p\}$ is 1086.6 while the theoretical value is 1118.7. Regarding the standard deviation, when $n = 600$, $g = 1$ and $f = 2$, the simulated and theoretical results are 4671.6 and 4665, respectively. This small gap is mainly due to the following two reasons. The first one is that the simplification adopted in the Assumption 1 slightly "slows" down the absorbing speed of the Markov chain, and thus results in a higher absorption time (i.e., delivery delay). The other reason is that we adopted approximations (4.3) and (4.4) for the fresh nodes and non-fresh nodes in the theoretical delay analysis, which made the theoretical results shift slightly from simulation ones.

To further illustrate the applicability of our theoretical framework to other mobility models, we show in Fig. 4.7 the $\mathbb{E}\{T_p\}$ and relative standard deviation δ of packet deliver delay under the i.i.d, random walk and random waypoint mobility models, where δ is defined as

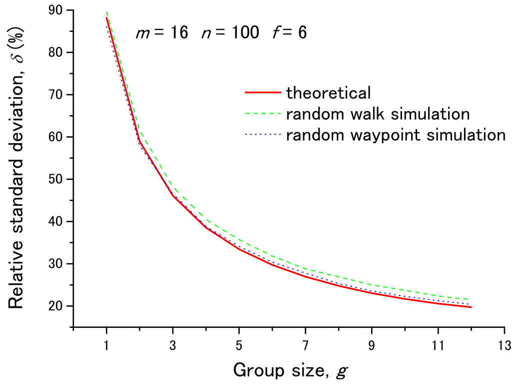
$$\delta = \frac{\sqrt{Var\{T_p\}}}{\mathbb{E}\{T_p\}}. \tag{4.45}$$

It's interesting to observe from Fig. 4.7 that the analytical models of $\mathbb{E}\{T_p\}$ and δ, although were developed under the i.i.d. mobility model, can also well approximate the general trends of $\mathbb{E}\{T_p\}$ and δ under the other two mobility models.

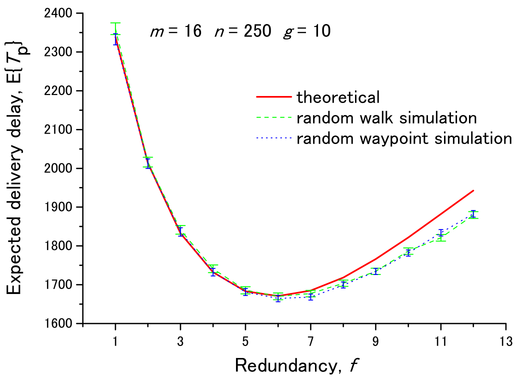
Regarding the $\mathbb{E}\{T_p\}$ performance, the results in Fig. 4.7(a) and Fig. 4.7(c) indicate that the behavior of $\mathbb{E}\{T_p\}$ vs. g under the i.i.d model is slightly different from that of under other two models, but the behavior of $\mathbb{E}\{T_p\}$ vs. f is similar for all the three mobility models. In particular, from Fig. 4.7(a) we can see that for the concerned network scenario the minimum $\mathbb{E}\{T_p\}$ of the i.i.d model is reached at $g = 9$, while the minimum $\mathbb{E}\{T_p\}$ of the random walk and random waypoint are reached at $g = 12$ and $g = 10$,



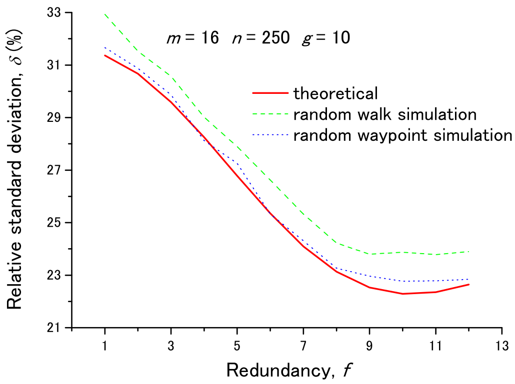
(a) $\mathbb{E}\{T_p\}$ vs. g



(b) δ vs. g



(c) $\mathbb{E}\{T_p\}$ vs. f



(d) δ vs. f

Figure 4.7: Delivery delay vs. group size g and redundancy f under random walk and random waypoint mobility models.

respectively. For the $\mathbb{E}\{T_p\}$ vs. f results in Fig. 4.7(c), however, the minimum $\mathbb{E}\{T_p\}$ is reached at the same setting of $f = 6$ for all three mobility models. Different from that of the $\mathbb{E}\{T_p\}$ performance, the results of δ in Fig. 4.7(b) and Fig. 4.7(d) show that the behavior of δ vs. g is very similar for all three mobility models, while the δ vs. f behavior of the random walk is a little different from that of other two models.

4.7.3 Achievable Delay Region

Based on the new Markov chain theoretical framework, we now explore the achievable delay performance region of the 2HR-(f, g) algorithm in terms of its $(\delta,\ \mathbb{E}\{T_p\})$. For the scenario of $m = 8$, $n = 50$ and $g = \{3, 4\}$, Fig. 4.8 shows the region of $(\delta,\ \mathbb{E}\{T_p\})$

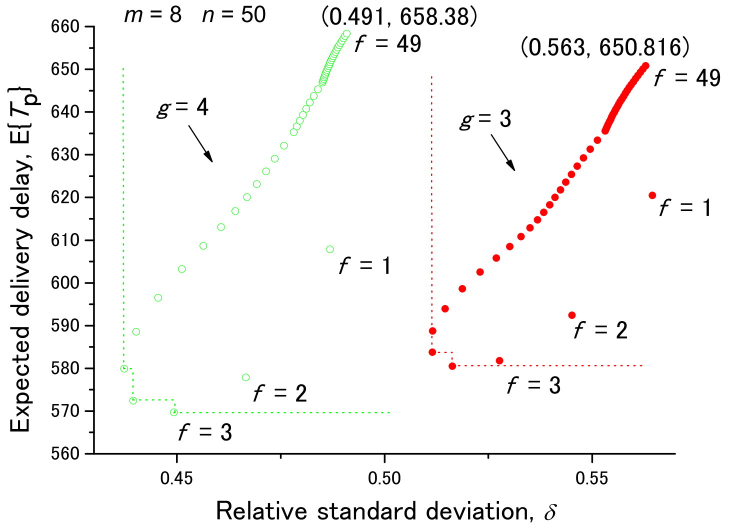
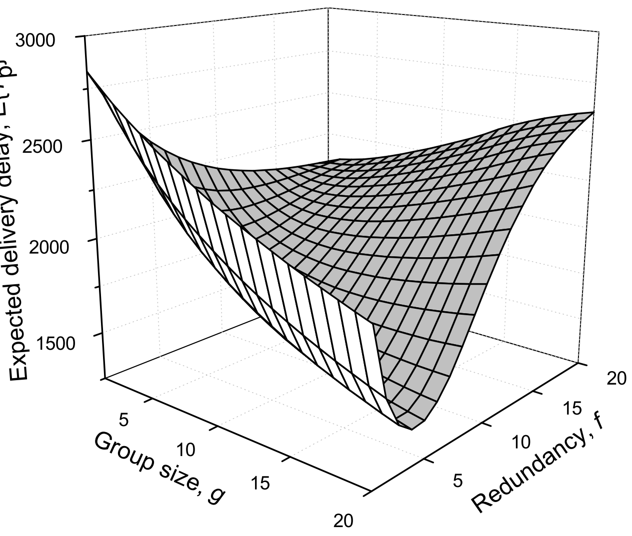


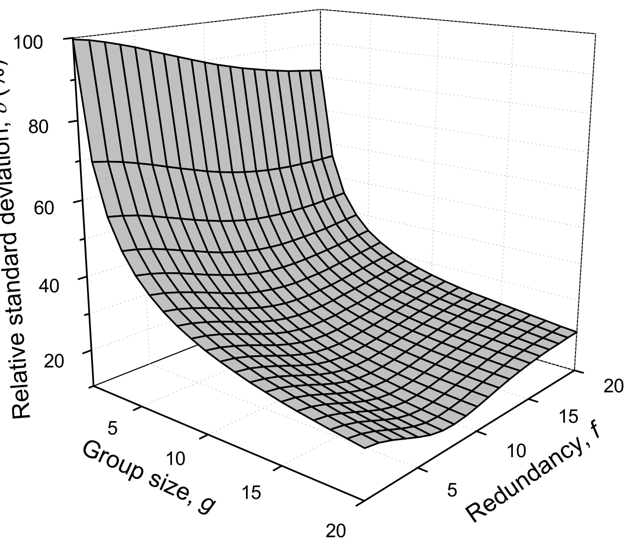
Figure 4.8: Achievable delay performance region of a 2HR-(f, g) MANET for the cases of $m = 8$, $n = 50$ and $g = \{3, 4\}$.

that the 2HR-(f, g) can achieve by varying the parameter f. Notice that each curve in Fig. 4.8 consists of multiple discrete points and each point corresponds to a specific value of f, so it may happen that two distinct settings of f achieve different $\mathbb{E}\{T_p\}$ but very similar (even the same) δ. For example, for the curve of $g = 4$, the settings of $f = 1$ and $f = 36$ achieve the same $\delta = 0.487$ but different expected delivery delay of 607.875 and 651.632, respectively. The results in Fig. 4.8 indicate that the 2HR-(f, g) algorithm actually enables the delay performance (δ, $\mathbb{E}\{T_p\}$) to be flexibly controlled in a large region to adapt to various applications with different delay (and variance) requirements.

It is interesting to notice from Fig. 4.8 that for a specified group size g, the achievable delay performance region is actually defined by some vertical and horizontal lines determined by several key points, i.e., the Pareto optimal points [69]. For example, when $g = 3$, the achievable delay performance region is determined by the point $(0.516, 580.506)$ ($f = 4$) that results in the minimum $\mathbb{E}\{T_p\}$ of 580.506 and the point $(0.511, 583.765)$ ($f = 5$) that results in the minimum δ of 0.511. For the case that $g = 4$, the achievable delay region is co-determined by three points, i.e., the point $(0.449, 569.695)$ ($f = 3$), point $(0.439, 572.447)$ ($f = 4$) and point $(0.437, 579.933)$ ($f = 5$). Thus, for a specified



(a) $\mathbb{E}\{T_p\}$ vs. (g, f)



(b) δ vs. (g, f)

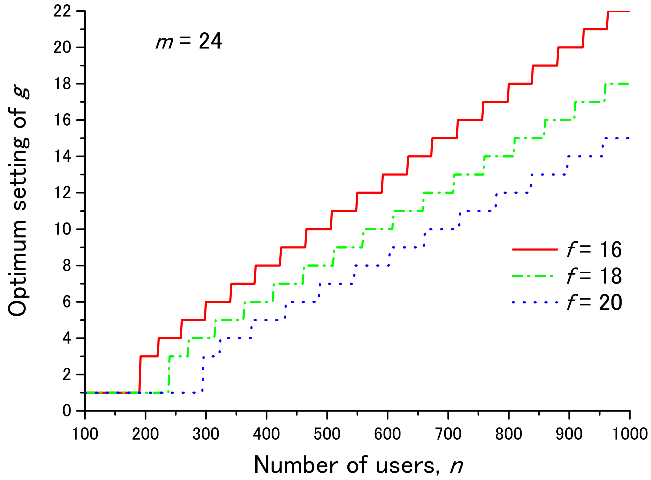
Figure 4.9: Delivery delay for a specific network $m = 16$, $n = 250$.

group size g, any delay performance requirement in terms of (δ, $\mathbb{E}\{T_p\}$) can be supported by the 2HR-(f, g) algorithm as long as the point (δ, $\mathbb{E}\{T_p\}$) falls within the corresponding performance region defined by the group size.

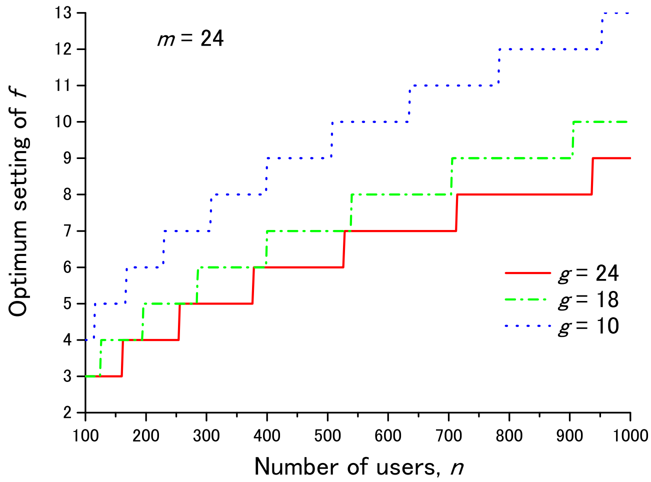
4.7.4 Delay Control

To see how the delay can be controlled according to a specified delay target, we now apply our framework to a network scenario of ($m = 16$, $n = 250$) and show in Fig. 4.9 how its delay performance ($\delta, \mathbb{E}\{T_p\}$) varies with both g and f there. As shown in Fig. 4.9(a) (resp. Fig. 4.9(b)) that for a specified target t_p of the mean delay value (resp. a target δ_0 of the relative standard deviation), we can accordingly define a target plane intersecting the z-axis orthogonally at the point $(1, 1, t_p)$ (resp. at the point $(1, 1, \delta_0)$), and thus can get a set of (g, f)-pairs corresponding to the surface below the defined target plane there. By finding the intersection of these two sets of (g, f)-pairs, we can determine the set of (g, f)-pairs to achieve the specified delay target in terms of t_p and δ_0. Fig. 4.9(a) also shows that for the network scenario there, although the $\mathbb{E}\{T_p\}$ has different varying trends with g and f but a minimum delivery delay can always be identified. For example, we can see that when $g = 1$, the $\mathbb{E}\{T_p\}$ monotonically decreases with f; for any fixed $g \geq 2$, we can find an optimum setting of f to achieve the corresponding minimum delivery delay. Similarly, when $f \leq 6$ the $\mathbb{E}\{T_p\}$ monotonically decreases with g; for any fixed $f \geq 7$, there also exists an optimum setting of g to achieve the minimum delivery delay. For the network scenario here and all $g, f \in [1, 20]$, the global minimum delivery delay of 1426.75 is achieved at the setting of ($g = 20, f = 4$).

Fig. 4.9(a) indicates that for a network scenario with fixed g (resp. f), there exists a corresponding optimum setting of f (resp. optimum setting of g) to achieve the minimal delivery delay. We show in Fig. 4.10 how such optimum setting of g (resp. f) varies under different network scenarios. One can observe from Fig. 4.10 that there does not exist a particular optimal value of g (or f) which applies to all networks of different size n. Actually, the optimal setting of g (or f) is a piecewise function of n and one optimal setting only applies to a small range of n. A further careful observation of Fig. 4.10 indicates that, as n scales up, the optimum setting of f in Fig. 4.10(b) becomes less sensitive to the variation of n (i.e., as n increases up, an optimal setting of f applies to a wider range of n), but this is not the case for the optimum setting of g in Fig. 4.10(a).



(a) Optimum setting of g vs. n



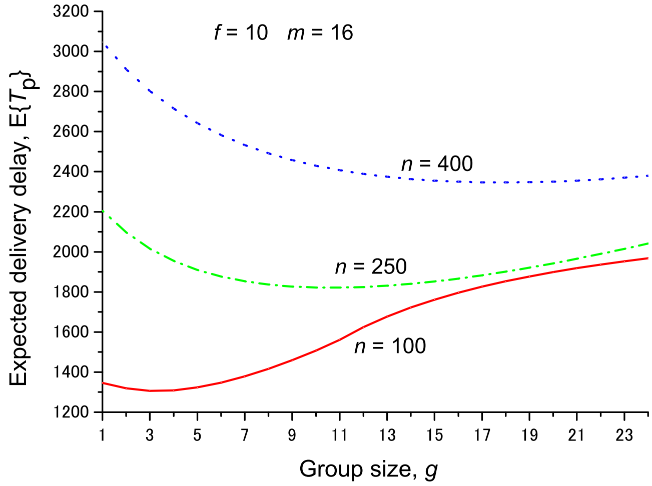
(b) Optimum setting of f vs. n

Figure 4.10: Optimum parameter settings vs. number of nodes n

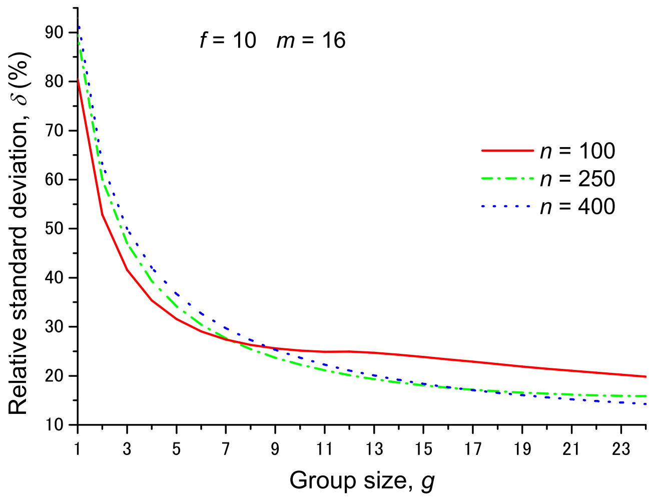
Thus, compared with the optimum setting of f (under a given g), the optimum setting of g (under a given f) depends more heavily on n.

4.7.5 Performance Analysis

We now explore how the performance (δ, $\mathbb{E}\{T_p\}$) of the 2HR-(f, g) algorithm varies with different parameters. For the scenarios of $n = \{100, 250, 400\}$ and the fixed setting of $f = 10$ and $m = 16$, Fig. 4.11 illustrates how $\mathbb{E}\{T_p\}$ and δ vary with group size g. It



(a) $\mathbb{E}\{T_p\}$ vs. g



(b) δ vs. g

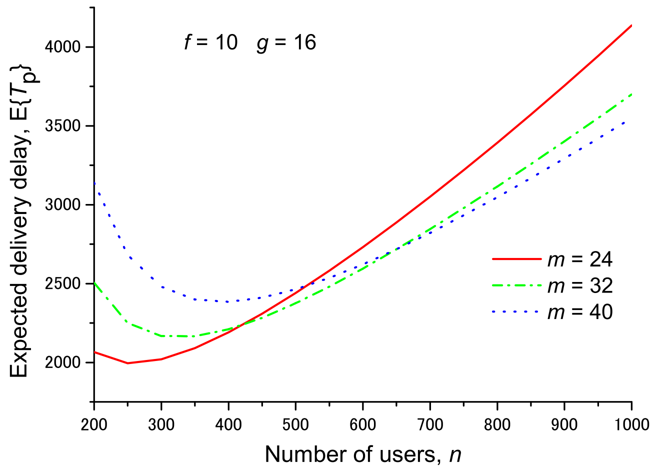
Figure 4.11: Delivery delay vs. group size g

is interesting to see from Fig. 4.11 that for a given network with a fixed value of f, the δ always monotonously decreases as g increases, but this is not the case for $\mathbb{E}\{T_p\}$. As shown in Fig. 4.11(b) that when $g \leq 2$, the δ is quite high (larger than 50%). It is notable, however, that for most of the MANET applications, the destination node will allow a certain degree of packet out of order determined by the parameter g, so a moderate value of g is usually acceptable. As we can see from Fig. 4.11(b) that as g increases beyond 2, the δ drops dramatically to a low level for all three network scenarios considered here, which indicates that our algorithm can stably control the delivery delay for most interested settings of g. Notice also that the affordable group size is limited by the buffer size at each mobile node, and a large group may unavoidably force the early arrived packets at the destination node to wait a long time for other packets (of the same group), which may make the early arrived packets become expired before the arrival of the last packet of the same group. This indicates that using a large group size in the 2HR-(f, g) algorithm may significantly limit its applications to support the delay-sensitive applications in the MANETs. Thus, the group size g should be carefully dimensioned with the considerations of $\mathbb{E}\{T_p\}$, δ and buffer limitation in each node.

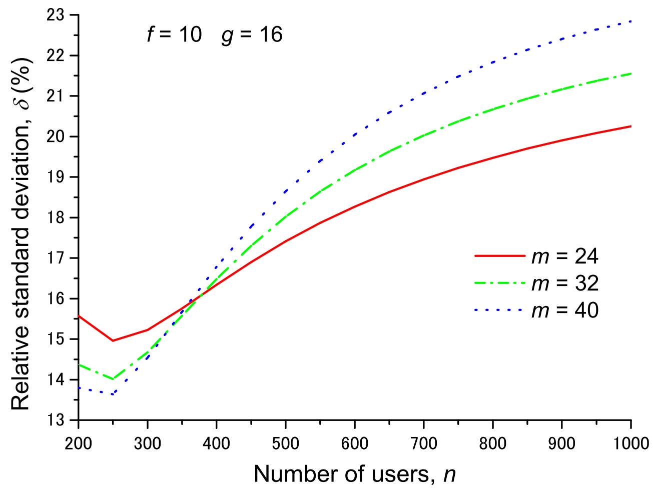
Finally, we examine in Fig. 4.12 how metrics δ and $\mathbb{E}\{T_p\}$ vary with network size n, given that $f = 10$, $g = 16$, and $m = \{24, 32, 40\}$. We can see from Fig. 4.12(a) that for a given m (determined by communication range r as $m = \sqrt{8}/r$), we can find a most suitable network size n^* to achieve the minimum expected packet delay $\mathbb{E}\{T_p\}$. A further careful observation of Fig. 4.12(a) indicates that the most suitable network size for a minimum expected packet delay varies with m and can be roughly determined as $n^* \approx m \cdot f$. For example, for the setting of $m = 24$, 32 and 40, the corresponding n^* are roughly 240, 320 and 400, respectively. Regarding the performance of δ, the results in Fig. 4.12(b) indicate that for a given m, there also exists a most suitable network size to achieve the minimum δ. However, it is interesting to see from Fig. 4.12(b) that the most suitable network size for a minimum δ is always 250 for the scenario of $f = 10$ and $g = 16$ here, which actually does not change as m varies.

4.8 Summary

In this chapter, we proposed a general 2HR-(f, g) algorithm for MANETs, and also developed a Markov chain-based theoretical framework for corresponding performance modeling. We proved that the 2HR-(f, g) algorithm has the capability of flexibly controlling packet delay and its variance in a large region, an important property for future MANETs to support various applications of different delay (and delay variance) requirements. The



(a) $\mathbb{E}\{T_p\}$ vs. n



(b) δ vs. n

Figure 4.12: Delivery delay vs. number of nodes n

results in this chapter indicate that the control parameters f and g of the 2HR-(f,g) algorithm may affect the packet delay and its variance in very different ways, and a target packet delay (and delay variance) requirement can be actually achieved through various combinations between f and g. Thus, a careful trade-off among packet delay (and delay variance) requirement, packet redundancy (f) and node buffer limitation (related to g) should be examined for the efficient support of a target application.

Chapter 5

Conclusion

5.1 Summary and Discussions

As a flexible and autonomous network architecture, the mobile ad hoc network holds great promise for lots of future application scenarios due to its distinctive features. However, the lack of a MANET capacity theory has been stunting its application and commercialization in the last decades. In this thesis, we have investigated the fundamental throughput and delay performances in the popular two-hop relay mobile ad hoc networks. The main contributions are listed as follows:

- In Chapter 2, we defined the system models adopted in the throughput and delay performance analysis, which includes the cell-partitioned network model, the i.i.d. mobility model, the protocol interference model and the permutation traffic model. After defining the system models, we also discussed the motivations behind the adopted system models and their possible applications to other network scenarios. To schedule as many simultaneous transmissions as possible, we then defined the concurrent-set for cell-partitioned network, introduced the concurrent-set based transmission scheduling, and proposed the transmitting node selection scheme for all cells of an active concurrent-set. After defining the transmission scheduling scheme, we gave an overall introduction of the basic operations in two-hop relay routing protocol and its various variants.

- In Chapter 3, we addressed a basic problem: for a MANET with general node transmission range control and packet redundancy control, what is the exact achievable per node throughput capacity. Distinguished from the available works which mainly focus on deriving the order sense results and exploring the scaling laws of the throughput capacity in MANETs, we aimed to derive the exact per node throughput capacity and thus to provide a detailed understanding. We found that for the network scenarios considered in this thesis, it may not be always true that adopting local transmission can achieve the maximum per node throughput capacity, as what is generally believed in literature. This finding indicates that further deliberate studies are necessary to reveal the real achievable network throughput of MANETs. Another interesting finding of our work is that the MANETs considered in this thesis actually exhibit very similar behaviors in terms of packet delay and per node throughput under different node mobility models, like the i.i.d., random walk and random waypoint.

- In Chapter 4, we proposed a general 2HR-(f, g) algorithm for MANETs, and also developed a Markov chain-based theoretical framework for corresponding performance modeling. We proved that the 2HR-(f, g) algorithm has the capability of flexibly controlling packet delay and its variance in a large region, an important property for future MANETs to support various applications of different delay (and delay variance) requirements. The results in this thesis indicate that the control parameters f and g of the 2HR-(f, g) algorithm may affect the packet delay and its variance in very different ways, and a target packet delay (and delay variance) requirement can be actually achieved through various combinations between f and g. Thus, a careful trade-off among packet delay (and delay variance) requirement, packet redundancy (f) and node buffer limitation (related to g) should be examined for the efficient support of a target application.

5.2 Future Work

In this thesis, we developed theoretical models for analytical study of the throughput and delay performances in two-hop relay mobile ad hoc networks. The possible future works are as follows:

- Notice that the theoretical models and closed-form results for per node throughput capacity developed in this thesis hold only for the concurrent-set based scheduling scheme, so one of our future research directions is to develop theoretical models for other more commonly used MAC schemes, like the 802.11 DCF.

- In the proposed 2HR-(f, g) algorithm, we considered a very simple scenario where only one node is randomly selected from the one-hop neighbors for the source-to-relay transmission or relay-to-destination transmission, which may cause a waste of the transmission opportunity if a wrong node is selected. Therefore, one future work is to further explore the performance of 2HR-(f, g) under a more flexible scenario, where not only one but many (even all) one-hop neighbors will be considered for the source-to-relay or relay-to-destination transmission to take the full advantage of each transmission opportunity.

- Another interesting future direction is to extend the theoretical models in this thesis to analytically determine the combinations of group size g and redundancy f for the proposed 2HR-(f, g) algorithm to satisfy a given delay requirement and further derive the optimum combination of g and f to achieve the minimum delivery delay under a specific network scenario.

- Since all theoretical models in this thesis were developed for throughput capacity and delay analysis in the two-hop relay mobile ad hoc networks, it would be interesting to further extend the developed theoretical models to analyze the throughput and delay performances in the general k-hop ($k \geq 2$) mobile ad hoc networks.

Bibliography

[1] M. Grossglauser and D. N. Tse, "Mobility increases the capacity of ad hoc wireless networks," in *INFOCOM*, 2001.

[2] E. Altman, T. Basar, and F. D. Pellegrini, "Optimal monotone forwarding policies in delay tolerant mobile ad-hoc networks," in *Inter-Perf*, 2008.

[3] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Generalized two-hop relay for flexible delay control in manets," IEEE/ACM Transactions on Networking, 2012, to appear.

[4] ——, "Exact throughput capacity under power control in mobile ad hoc networks," in *INFOCOM*, 2012.

[5] E. Altman, T. Basar, and F. D. Pellegrini, "Optimal control in two-hop relay routing," *IEEE Transactions on Automatic Control*, vol. 56, no. 3, pp. 670–675, March 2011.

[6] E. Altman and F. D. Pellegrini, "Forward correction and fountain codes in delay tolerant networks," in *INFOCOM*, 2009.

[7] L. Ying, S. Yang, and R. Srikant, "Optimal delay-throughput trade-offs in mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4119–4143, September 2008.

[8] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Optimal throughput-delay scaling in wireless networks-part i: The fluid model," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2568–2592, June 2006.

[9] ——, "Throughput-delay trade-off in wireless networks," in *INFOCOM*, 2004.

[10] M. J. Neely and E. Modiano, "Capacity and delay tradeoffs for ad-hoc mobile networks," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1917–1936, June 2005.

[11] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Delay and capacity in ad hoc mobile networks with f-cast relay algorithms," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2738–2751, August 2011.

[12] S. R. Kulkarni and P. Viswanath, "A deterministic approach to throughput scaling in wireless networks," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1041–1049, June 2004.

[13] P. Li, Y. Fang, and J. Li, "Throughput, delay, and mobility in wireless ad-hoc networks," in *INFOCOM*, 2010.

[14] R. J. La, "Distributional convergence of inter-meeting times under generalized hybrid random walk mobility model," *IEEE Transactions on Mobile Computing*, vol. 9, no. 9, pp. 1201–1211, September 2010.

[15] M. Li and Y. Liu, "Rendered path: Range-free localization in anisotropic sensor networks with holes," in *MobiCom*, 2007.

[16] Z. Zhu, A. M.-C. So, and Y. Ye, "Universal rigidity: Towards accurate and efficient localization of wireless networks," in *INFOCOM*, 2010.

[17] S. Zhou and L. Ying, "On delay constrained multicast capacity of large-scale mobile ad-hoc networks," in *INFOCOM*, 2010.

[18] Y. Wang, X. Chu, X. Wang, and Y. Cheng, "Optimal multicast capacity and delay tradeoffs in manets: A global perspective," in *INFOCOM*, 2011.

[19] C. Zhang, Y. Fang, and X. Zhu, "Throughput-delay tradeoffs in large-scale manets with network coding," in *INFOCOM*, 2009.

[20] R. Urgaonkar and M. J. Neely, "Network capacity region and minimum energy function for a delay-tolerant mobile ad hoc network," *IEEE/ACM Transactions on Networking*, vol. 19, no. 4, pp. 1137–1150, August 2011.

[21] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[22] M. Garetto, P. Giaccone, and E. Leonardi, "Capacity scaling in ad hoc networks with heterogeneous mobile nodes: The subcritical regime," *IEEE/ACM Transactions on Networking*, vol. 17, no. 6, pp. 1888–1901, December 2009.

[23] D. Ciullo, V. Martina, M. Garetto, and E. Leonardi, "Impact of correlated mobility on delay-throughput performance in mobile ad-hoc networks," in *INFOCOM*, 2010.

[24] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Group-based two-hop relay with redundancy in manets," in *HPSR*, 2011.

[25] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, "Delay-tolerant networking: an approach to interplanetary internet," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 128–136, June 2003.

[26] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, June 2007.

[27] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 77–90, February 2008.

[28] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, 1996.

[29] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *WMCSA*, 1999.

[30] A. A. Hanbali, M. Ibrahim, V. Simon, E. Varga, and I. Carreras, "A survey of message diffusion protocols in mobile ad hoc networks," in *Inter-Perf*, 2008.

[31] G. Sharma and R. Mazumdar, "On achievable delay/capacity trade-offs in mobile ad hoc networks," in *Wiopt*, 2004.

[32] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Delay and capacity in ad hoc mobile networks with f-cast relay algorithms," in *ICC*, 2011.

[33] E. Perevalov and R. S. Blum, "Delay-limited throughput of ad hoc networks," *IEEE Transactions on Communications*, vol. 52, no. 11, pp. 1957–1968, November 2004.

[34] J. Mammen and D. Shah, "Throughput and delay in random wireless networks with restricted mobility," *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1108–1116, 2007.

[35] T. Small and Z. Hass, "Resource and performance tradeoffs in delay-tolerant wireless networks," in *ACM SIGCOMM Workshop on Delay Tolerant Networks (WDTN)*, 2005.

[36] E. Altman, F. D. Pellegrini, and L. Sassatelli, "Dynamic control of coding in delay tolerant networks," in *INFOCOM*, 2010.

[37] E. Altman, G. Neglia, F. D. Pellegrini, and D. Miorandi, "Decentralized stochastic control of delay tolerant networks," in *INFOCOM*, 2009.

[38] E. Altman, A. P. Azad, T. Basar, and F. D. Pellegrini, "Optimal activation and transmission control in delay tolerant networks," in *INFOCOM*, 2010.

[39] F. D. P. Wissam Chahin and, R. EI-Azouzi, and A. P. Azad, "Blind online optimal forwarding in heterogeneous dtns," in *IFIP Wireless Days*, October 2011.

[40] C. Singh, A. Kumar, and R. Sundaresan, "Delay and energy optimal two-hop relaying in delay tolerant networks," in *WiOpt*, May 2010.

[41] R. EI-Azouzi, F. D. Pellegrini, and V. Kamble, "Evolutionary forwarding games in delay tolerant networks," in *WiOpt*, May 2010.

[42] Y. Wang, S. Jain, M. Martonosi, and K. Fall, "Erasure-coding based routing for opportunistic networks," in *WDTN*, 2005.

[43] A. A. Hanbali, A. A. Kherani, and P. Nain, "Simple models for the performance evaluation of a class of two-hop relay protocols," in *Proc. IFIP Networking*, 2007.

[44] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Performance modeling for two-hop relay with erasure coding in manets," in *GLOBECOM*, 2011.

[45] J. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, "Rethinking information theory for mobile ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, December 2008.

[46] A. Goldsmith, M. Effros, R. Koetter, M. Medard, A. Ozdaglar, and L. Zheng, "Beyond shannon: The quest for fundamental performance limits of wireless ad hoc networks," *IEEE Communications Magazine*, vol. 49, no. 5, pp. 195–205, May 2011.

[47] L. X. Cai, X. S. Shen, J. W. Mark, and L. Cai, "Capacity analysis and mac enhancement for uwb broadband wireless access networks," *Elsevier Computer Networks*, vol. 51, no. 11, pp. 3265–3277, August 2007.

[48] R. M. de Moraes, H. R. Sadjadpour, and J. Garcia-Luna-Aceves, "Throughput-delay analysis of mobile ad-hoc networks with a multi-copy relaying strategy," in *SECON*, 2004.

[49] X. Lin, G. Sharma, R. R. Mazumdar, and N. B. Shroff, "Degenerate delay-capacity tradeoffs in ad-hoc networks with brownian mobility," *IEEE/ACM Transactions on Networking, Special Issue on Networking and Information Theory*, vol. 52, no. 6, pp. 2777–2784, June 2006.

[50] G. Sharma and R. Mazumdar, "Delay and capacity trade-off in wireless ad hoc networks with random way-point mobility," in *Dept. Elect. Comput. Eng., Purdue Univ.,West Lafayette, IN*, 2005. [Online]. Available: https://ece.uwaterloo.ca/~mazum/adhoc_mobility.pdf

[51] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming, W. M. McEneaney, G. Yin, and Q. Zhang, pp. 547–566, 1998.

[52] C. M. Grinstead and J. L. Snell, *Introduction to Probability: Second Revised Edition*. American Mathematical Society, 1997.

[53] C++ programs for the 2hr-(f, g) manet. [Online]. Available: http://distplat.blogspot.com

[54] The network simulator ns-2. [Online]. Available: http://www.isi.edu/nsnam/ns/

[55] Qualnet. [Online]. Available: http://www.scalable-networks.com/products/qualnet/

[56] A. A. Hanbali, P. Nain, and E. Altman, "Performance of ad hoc networks with two-hop relay routing and limited packet lifetime-extended version," *Performance Evaluation*, vol. 65, no. 6-7, pp. 463–483, June 2008.

[57] A. Panagakis, A. Vaios, and I. Stavrakakis, "Study of two-hop message spreading in dtns," in *WiOpt*, April 2007.

[58] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, no. 1-4, pp. 210–228, October 2005.

[59] G. Sharma, R. Mazumdar, and N. B. Shroff, "Delay and capacity trade-offs for mobile ad hoc networks: A global perspective," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 981–992, October 2007.

[60] A. Kashyap, S. Ganguly, S. R. Das, and S. Banerjee, "Voip on wireless meshes: models, algorithms and evaluation," in *INFOCOM*, 2007.

[61] S. Ganguly, V. Navda, K. Kim, A. Kashyap, D. Niculescu, R. Izmailov, S. Hong, and S. R. Das, "Performance optimizations for deploying voip services in mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 2147–2158, November 2006.

[62] J. Cano, J.-C. Cano, C. K. Toh, C. T. Calafate, and P. Manzoni, "Easymanet: an extensible and configurable platform for service provisioning in manet environments," *IEEE Communications Magazine*, vol. 48, no. 12, pp. 159–167, December 2010.

[63] S. Khan, Y. Peng, and E. Steinbach, "Application-driven cross-layer optimization for video streaming over wireless networks," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 122–130, January 2006.

[64] L. Zhou, X. Wang, W. Tu, G.-M. Muntean, and B. Geller, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 409–420, April 2010.

[65] I.-H. Hou, V. Borkar, and P. R. Kumar, "A theory of qos for wireless," in *INFOCOM*, 2009.

[66] I.-H. Hou and P. R. Kumar, "Scheduling heterogeneous real-time traffic over fading wireless channels," in *INFOCOM*, 2010.

[67] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. D. Van Nostrand, 1963.

[68] S. Zhou and L. Ying, "On delay constrained multicast capacity of large-scale mobile ad-hoc networks," in *INFOCOM*, 2010.

[69] H. Boche, S. Naik, and M. Schubert, "Pareto boundary of utility sets for multiuser wireless systems," *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, pp. 589–601, April 2011.

Publications

Journal Articles

[1] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Generalized Two-hop Relay for Flexible Delay Control in MANETs," *IEEE/ACM Transactions on Networking*, to be published.

[2] J. Liu, X. Jiang, H. Nishiyama, R. Miura, Nei Kato and N. Kadowaki, "Optimal Forwarding Games in Mobile Ad Hoc Networks with Two-Hop f-cast Relay," *IEEE Journal of Selected Areas in Communications*, to be published.

[3] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Delay and Capacity in Ad Hoc Mobile Networks with f-cast Relay Algorithms," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2738-2751, August 2011.

[4] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Reliability Assessment for Wireless Mesh Networks Under Probabilistic Region Failure Model," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2253-2264, June 2011.

[5] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Performance Modeling for Relay Cooperation in Delay Tolerant Networks," *Mobile Networks and Applications*, to be published.

[6] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "On the Delivery Probability of Two-Hop Relay MANETs with Erasure Coding," *IEEE Transactions on Communication*, under revision.

[7] J. Liu, J. Gao, X. Jiang, H. Nishiyama and N. Kato,, "Capacity and Delay of Probing-Based Two-Hop Relay in MANETs," *IEEE Transactions on Wireless Communications*, under revision.

Conference Papers

[8] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Throughput Capacity of the Group-Based Two-Hop Relay Algorithm in MANETs," in *2012 IEEE Global Communications Conference (GLOBECOM)*, Anaheim, California, USA, 3-7 December 2012.

[9] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Message Delivery Probability of Two-Hop Relay with Erasure Coding in MANETs," in *7th International ICST Conference*

on Communications and Networking in China (CHINACOM), Kunming, China, 8-10 August, 2012.

[10] J. Liu, J. Gao, X. Jiang, H. Nishiyama and N. Kato, "Probing-Based Two-Hop Relay with Limited Packet Redundancy," in *IEEE International Conference on High Performance Switching and Routing (HPSR)*, Belgrade, Serbia, 24-27 June 2012.

[11] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Delivery Ratio in Two-Hop Relay MANETs with Limited Message Lifetime and Redundancy," in *IEEE International Conference on Communications (ICC)*, Ottawa, Canada, 10-15 June 2012.

[12] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Capacity vs. Delivery Delay in MANETs with Power Control and f-cast Relay," in *IEEE International Conference on Communications (ICC)*, Ottawa, Canada, 10-15 June 2012.

[13] J. Liu, X. Jiang, H. Nishiyama, N. Kato and X. Shen, "End-to-End Delay in Mobile Ad Hoc Networks with Generalized Transmission Range and Limited Packet Redundancy," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, 1-4 April 2012.

[14] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Multicast Capacity, Delay and Delay Jitter in Intermittently Connected Mobile Networks," in *the 31st Annual IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, Florida, 25-30 March 2012.

[15] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Exact Throughput Capacity under Power Control in Mobile Ad Hoc Networks," in *the 31st Annual IEEE International Conference on Computer Communications (INFOCOM)*, Orlando, Florida, 25-30 March 2012.

[16] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Performance Modeling for Two-Hop Relay with Erasure Coding in MANETs," in *2011 IEEE Global Communications Conference (GLOBECOM)*, Houston, Texas, USA, 5-9 December 2011.

[17] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Performance Modeling for Two-Hop Relay with Node Selfishness in Delay Tolerant Networks," in *2011 IEEE Online Conference on Green Communications (GREENCOM)*, 26-29 September 2011.

[18] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "A General Model for Store-Carry-Forward Routing Schemes with Multicast in Delay Tolerant Networks," in *6th International ICST Conference on Communications and Networking in China (CHINACOM)*, Harbin, China, 17-19 August, 2011 **(Invited Paper)**.

[19] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Group-based Two-Hop Relay with Redundancy in MANETs," in *IEEE International Conference on High Performance Switching and Routing (HPSR)*, Cartagena, Spain, 4-6 July 2011.

[20] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Delay and Capacity in Ad Hoc Mobile Networks with f-cast Relay Algorithms," in *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, 5-9 June 2011.

[21] J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Throughput-Delay Tradeoffs in Mobile Ad Hoc Networks with Correlated Mobility," in *INFOCOM 2013*, submitted.

Awards

- Japan Society for the Promotion of Science (JSPS) Special Research Fellow DC2 (Apr. 2012~Mar. 2014)
- 2011 Chinese Government Award for Outstanding Self-Financed Ph.D. Students Abroad
- Japanese Government (MONBUKAGAKUSHO) Scholarship (Oct. 2009~Mar. 2012)
- Best Paper Award: J. Liu, X. Jiang, H. Nishiyama, N. Kato and X. Shen, "End-to-End Delay in Mobile Ad Hoc Networks with Generalized Transmission Range and Limited Packet Redundancy," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, 1-4 April 2012.
- Student Travel Grant: J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Performance Modeling for Two-Hop Relay with Erasure Coding in MANETs," in *2011 IEEE Global Communications Conference (GLOBECOM)*, Houston, Texas, USA, 5-9 December 2011.
- Student Travel Grant: J. Liu, X. Jiang, H. Nishiyama and N. Kato, "Delay and Capacity in Ad Hoc Mobile Networks with f-cast Relay Algorithms," in *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, 5-9 June 2011.