OPTIMAL MOBILITY PATTERNS IN EPIDEMIC NETWORKS

by

SUPRIYA NIRKHIWALE

B.E., Shri Govindram Sekseria Institute of Technology and Sciences, 2006

A THESIS

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2009

Approved by:

Major Professor
Caterina Scoglio

Abstract

Disruption Tolerant Networks or opportunistic networks represent a class of networks where there is no contemporaneous path from source to destination. In other words, these are networks with intermittent connections. These networks are generally sparse or highly mobile wireless networks. Each node has a limited radio range and the connections between nodes may be disrupted due to node movement, hostile environments or power sleep schedules, etc. A common example of such networks is a sensor network monitoring nature or military field or a herd of animals under study.

Epidemic routing is a widely proposed routing mechanism for data propagation in these type of networks. According to this mechanism, the source copies its packets to all the nodes it meets in its radio range. These nodes in turn copy the received packets to the other nodes they meet and so on. The data to be transmitted travels in a way analogous to the spread of an infection in a biological network. The destination finally receives the packet and measures are taken to eradicate the packet from the network.

The task of routing in epidemic networks faces certain difficulties involving minimizing the delivery delay with a reduced consumption of resources. Every node has severe power constraints and the network is also susceptible to temporary but random failure of nodes. In the previous work, the parameter of mobility has been considered a constant for a certain setting. In our setting, we consider a varying parameter of mobility. In this framework, we determine the optimal mobility pattern and a forwarding policy that a network should follow in order to meet the trade-off between delivery delay and power consumption. In addition, the mobility pattern should be such that it can be practically incorporated. In our work,

we formulate an optimization problem which is solved by using the principles of dynamic programming. We have tested the optimal algorithm through extensive simulations and they show that this optimization problem has a global solution.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This work could not have been possible without the able guidance of my advisor Dr Caterina Scoglio. She is the person who encouraged me and inspired me to achieve my goals. I am inspired by her dedication towards her goals and kindness and compassion for her students and their success.

I would also like to thank Dr Don Gruenbacher, for guiding me throughout my Masters program. His selfless attitude and consideration for his students is greatly appreciated.

I am also thankful to Dr Sanjoy Das for his able guidance in the field of numerical computations. His zealous attitude towards his work is greatly appreciated.

Dedication

To my family, Most notably my husband, my parents and in-laws, who have set wonderful examples to me. You are the light of my life.

To my advisor, whose support and confidence in me gave me the courage to face difficult times.

To my friends, whose good wishes and prayers are always with me.

To God, who acted through all of them.

Chapter 1

Epidemic Routing

The discipline of epidemiology does not only investigate the outbreak of communicable diseases in a biological network, but makes a significant contribution towards efficient routing of data packets in a wireless ad hoc network. This work is based on a class of wireless ad-hoc routing protocols called epidemic routing protocols. The main focus of this thesis is to emphasize the effect of mobility and an optimal forwarding strategy on the performance of mobile ad hoc networks.

What is an Ad-Hoc Network?

The definition of ad hoc as given by the Webster dictionary[1] is

Ad-hoc (adjective)

1.a : concerned with a particular end or purpose <an ad hoc investigation committee>

1.b : formed or used for specific or immediate problems or needs <ad hoc solutions>

2 : fashioned from whatever is immediately available : improvised <large ad hoc parades and demonstrations>

An ad-hoc network is a local area network (LAN) that is built spontaneously as devices connect. Instead of relying on a base station to coordinate the flow of messages to each node in the network, the individual network nodes forward packets to and from each other. In other words, an ad-hoc network can establish itself despite the lack of a fixed infrastructure.

The strong increase in the popularity of ad hoc networks in recent years is due to their envisioned ease of deployment, financial benefits, recent technological advancements, terrorist attacks, and widespread use of electronic devices. Because of their wide applicability and cheap installation, other uses of ad hoc network are continuously being found and therefore this field of technology and research will continue to grow for quite some time[2].

1.1 Mobile Ad Hoc Networks

A MANET is an autonomous collection of mobile users that communicate over relatively bandwidth constrained wireless links. Since the nodes are mobile, the network topology changes rapidly and unpredictably over time. The network is decentralized, where all network activity including discovering the topology and delivering messages must be executed by the nodes themselves, i.e., routing functionality is incorporated into mobile nodes.

The nodes in mobile ad hoc networks (MANET) are mobile and form the main focus of this thesis. In these networks, mobility can easily be exploited to ensure the transfer of data from one node to another, especially if the nodes in the network are not well connected to each other. In an ad-hoc network, the connectivity of two nodes is generally determined by the physical distance between them. A particular downside of mobility is that connections between nodes are continuously set up and broken down. Also known as disruption tolerant networks or opportunistic networks, MANETs represent a class of networks where there is no contemporaneous path from source to destination. In other words, these are networks with intermittent connections. These networks are generally sparse or highly mobile wireless networks. Each node has a limited radio range and the connections between nodes may be disrupted due to node movement, hostile environments or power sleep schedules, etc.

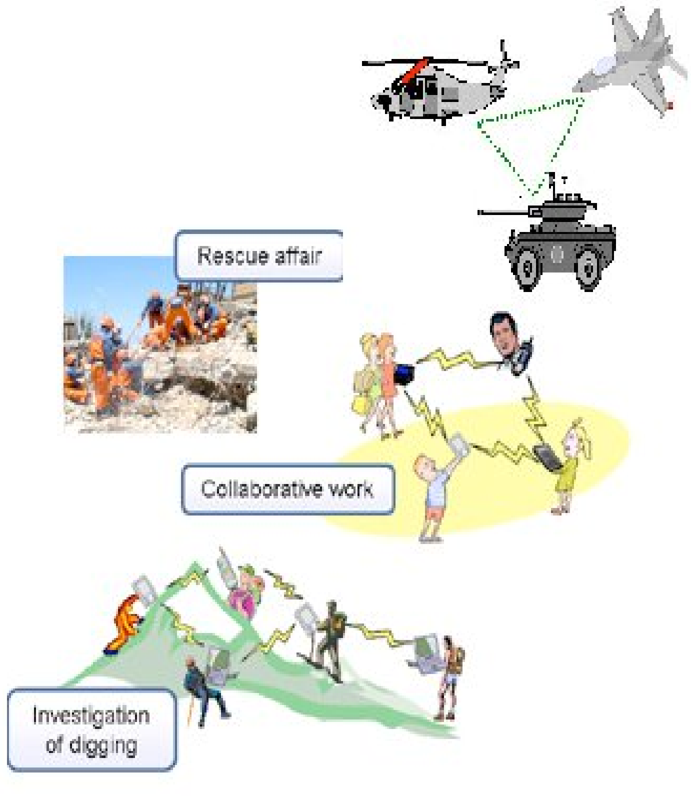

Figure 1.1: *Examples of MANETS: Military sensor networks, collaborative networks, rescue team networks, etc*

At any time, the network can be represented by an overlay graph showing the nodes which are connected to each other at that moment in time. Closely related to this are small-world graphs and peer-to-peer networks. In these networks, connections between nodes come and go and there is data that is passed from one node to another node, possibly by making use of intermediary nodes. There are however a number of distinctions which make peer-to-peer networks different from ad hoc networks and are therefore worth pointing out. In small-world and peer-to-peer networks there is large emphasis on the number of hops, the number of nodes reached per hop, search methods, and the scalability of the network. In mobile ad hoc networks however, the physical distances between the nodes play a role and therefore the performance of the network is instead characterized through the connectivity, throughput, and message delay (in terms of time instead of the number of hops).

In the coming generation of wireless communication systems, there will be a need for the rapid deployment of independent mobile users. Significant examples include establishing survivable, efficient, dynamic communication for emergency/rescue operations, disaster relief efforts, and military networks. Such network scenarios cannot rely on centralized and organized connectivity, and can be conceived as applications of mobile ad hoc networks. The set of applications for MANETs is diverse, ranging from small, static networks that are constrained by power sources, to large-scale, mobile, highly dynamic networks.

A good example of a mobile ad hoc network would be a mobile wireless sensor network. A wireless ad hoc sensor network consists of a number of sensors spread across a geographical area. Each sensor has wireless communication capability and some level of intelligence for signal processing and networking of the data. Some examples of wireless ad hoc sensor networks are the following:

1. Military sensor networks to detect and gain as much information as possible about enemy movements, explosions, and other phenomena of interest[3,4].
2. Sensor networks to detect and characterize Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) attacks and material.
3. Sensor networks to detect and monitor environmental changes in plains, forests, oceans, etc.
4. Wireless traffic sensor networks to monitor vehicle traffic on highways or in congested parts of a city.

5. Wireless surveillance sensor networks for providing security in shopping malls, parking garages, and other facilities[5,6].

6. Wireless parking lot sensor networks to determine which spots are occupied and which are free.

The above list suggests that wireless ad hoc sensor networks offer certain capabilities and enhancements in operational efficiency in civilian applications as well as assist in the national effort to increase alertness to potential terrorist threats.

The basic goals of a wireless ad hoc sensor network generally depend upon the application, but the following tasks are common to many networks:

1. Determine the value of some parameter at a given location: In an environmental network, one might one to know the temperature, atmospheric pressure, amount of sunlight, and the relative humidity at a number of locations. This example shows that a given sensor node may be connected to different types of sensors, each with a different sampling rate and range of allowed values.

2. Detect the occurrence of events of interest and estimate parameters of the detected event or events: In the traffic sensor network, one would like to detect a vehicle moving through an intersection and estimate the speed and direction of the vehicle.

3. Classify a detected object: Is a vehicle in a traffic sensor network a car, a mini-van, a light truck, a bus, etc.

4. Track an object: In a military sensor network, track an enemy tank as it moves through the geographic area covered by the network.

In these four tasks, an important requirement of the sensor network is that the required data be disseminated to the proper end users. In some cases, there are fairly strict time requirements on this communication. For example, the detection of an intruder in a surveillance network should be immediately communicated to the police so that action can be taken.

1.1.1 Routing in MANETs

The design of network protocols for these networks is a complex issue. Regardless of the application, MANETs need efficient distributed algorithms to determine network organization, link scheduling, and routing. However, determining viable routing paths and delivering messages in a decentralized environment where network topology fluctuates is not a well-defined problem. While the shortest path (based on a given cost function) from a source to a destination in a static network is usually the optimal route, this idea is not easily extended to MANETs. Factors such as variable wireless link quality, propagation path loss, fading, multiuser interference, power expended, and topological changes, become relevant issues. The network should be able to adaptively alter the routing paths to alleviate any of these effects.

Moreover, in a military environment, preservation of security, latency, reliability, intentional jamming, and recovery from failure are significant concerns. Military networks are designed to maintain a low probability of intercept and/or a low probability of detection. Hence, nodes prefer to radiate as little power as necessary and transmit as infrequently as possible, thus decreasing the probability of detection or interception. A lapse in any of these requirements may degrade the performance and dependability of the network.

Hence, the requirements of a mobile ad hoc network for efficient routing can be summarized as:

1. Large number of (mostly stationary) sensors: Aside from the deployment of sensors on the ocean surface or the use of mobile, unmanned, robotic sensors in military operations, most nodes in a smart sensor network are stationary. Networks of 10,000 or even 100,000 nodes are envisioned, so scalability is a major issue.

2. Low energy use: Since in many applications the sensor nodes will be placed in a remote area, service of a node may not be possible. In this case, the lifetime of a node may be determined by the battery life, thereby requiring the minimization of energy expenditure.

3. Network self-organization: Given the large number of nodes and their potential placement in hostile locations, it is essential that the network be able to self-organize; manual configuration is not feasible. Moreover, nodes may fail (either from lack of energy or from physical destruction), and new nodes may join the network. Therefore, the network must be able to periodically reconfigure itself so that it can continue to function. Individual nodes may become disconnected from the rest of the network, but a high degree of connectivity must be maintained.

4. Collaborative signal processing: Yet another factor that distinguishes these networks from MANETs is that the end goal is detection/estimation of some events of interest, and not just communications. To improve the detection/estimation performance, it is often quite useful to fuse data from multiple sensors. This data fusion requires the transmission of data and control messages, and so it may put constraints on the network architecture.

5. Querying ability: A user may want to query an individual node or a group of nodes for information collected in the region. Depending on the amount of data fusion performed, it may not be feasible to transmit a large amount of the data across the network. Instead, various local sink nodes will collect the data from a given area and create summary messages. A query may be directed to the sink node nearest to the desired location.

With the coming availability of low cost, short range radios along with advances in wireless networking, it is expected that wireless ad hoc sensor networks will become commonly deployed. In these networks, each node may be equipped with a variety of sensors, such as acoustic, seismic, infrared, still/motion video camera, etc. These nodes may be organized in clusters such that a locally occurring event can be detected by most of, if not all, the nodes in a cluster. Each node may have sufficient processing power to make a decision, and it will be able to broadcast this decision to the other nodes in the cluster. One node may act as the cluster master, and it may also contain a longer range radio using a protocol such as IEEE 802.11 or Bluetooth.

1.1.2 Existing Protocols

This section describes provides a classification of the existing protocols used in mobile ad-hoc network routing[7]. Existing protocols can be classified according to many criteria described as follows:

1. Proactive, reactive and hybrid routing

   One of the most popular method to distinguish mobile ad hoc network routing protocols is based on how routing information is acquired and maintained by mobile nodes. Using this method, mobile ad hoc network routing protocols can be divided into proactive routing, reactive routing and hybrid routing.

   A proactive routing protocol is also called "table driven" routing protocol. Using a proactive routing protocol, nodes in a mobile ad hoc network continuously evaluate routes to all reachable nodes and attempt to maintain consistent, up-to-date routing information. Therefore, a source node can get a routing path immediately if it needs one. In these routing protocols, all nodes need to maintain a consistent view of the network topology. When a network topology change occurs, respective updates must be propagated throughout the network to notify the change. Some examples are Wireless Routing Protocol (WRP), the Destination Sequence Distance Vector (DSDV) and the Fisheye State Routing (FSR).

Reactive routing protocols for mobile ad hoc networks are also called "on-demand" routing protocols. In a reactive routing protocol, routing paths are searched only when needed. A route discovery operation invokes a route-determination procedure. The discovery procedure terminates either when a route has been found or no route available after examination for all route permutations. The Dynamic Source Routing (DSR) and Ad hoc On- demand Distance Vector routing (AODV) are examples for reactive routing protocols for mobile ad hoc networks.

Hybrid routing protocols are proposed to combine the merits of both proactive and reactive routing protocols and overcome their shortcomings. Normally, hybrid routing protocols for mobile ad hoc networks exploit hierarchical network architectures. Proper proactive routing approach and reactive routing approach are exploited in different hierarchical levels, respectively. Some examples of hybrid routing protocols for mobile ad hoc networks are Zone Routing Protocol (ZRP), Zone-based Hierarchical Link State routing (ZHLS) and Hybrid Ad hoc Routing Protocol (HARP).

2. Structuring and delegating the routing task

   Another classification method is based on the roles which nodes may have in a routing scheme. In a uniform routing protocol, all mobile nodes have same role, importance and functionality. Examples of uniform routing protocols include Wireless Routing Protocol (WRP), Dynamic Source Routing (DSR), Ad hoc On-demand Distance Vector routing (AODV) and Destination Sequence Distance Vector (DSDV) routing protocol. Uniform routing protocols normally assume a flat network structure.

   In a non-uniform routing protocol for mobile ad hoc networks, some nodes carry out distinct management and/or routing functions. Normally, distributed algorithms are exploited to select those special nodes. In some cases, non-uniform routing approaches are related to hierarchical network structures to facilitate node organization and management. Non-uniform routing protocols further can be divided according to the organization of mobile nodes and how management and routing functions are performed. Following these criteria, non-uniform routing protocols for mobile ad hoc networks are divided into zone based hierarchical routing e.g. Zone Routing Protocol (ZRP) and Zone-based Hierarchical Link State routing (ZHLS), cluster-based hierarchical routing e.g. Clusterhead Gateway Switch Routing (CGSR), Hierarchical State Routing (HSR) and core-node based routing e.g. Core-Extraction Distributed Ad Hoc Routing (CEDAR).

3. Exploiting network metrics for routing

   Metrics used for routing path construction can be used as criteria for mobile ad hoc network routing protocol classification. Most routing protocols for mobile ad hoc networks use hop number as a metric. If there are multiple routing paths available, the path with the minimum hop number will be selected. If all wireless links in the network have the same failure probability, short routing paths are more stable than the long ones and can obviously decrease traffic overhead and reduce packet collisions. However, the assumption of the same failure properties may not be true in mobile ad hoc networks. Therefore, the stability of a link has to be considered in the route construction phase. For example, routing approaches such as Associativity Based Routing (ABR) and Signal Stability based Routing (SSR) are proposed that use link stability or signal strength as metric for routing.

4. Evaluating topology, destination and location for routing

   The existing protocols can also be divided according to the information used in deriving the best path to destination. In a topology based routing protocol for mobile ad hoc networks, nodes collect network topology information for making routing decisions. Other than topology based routing protocols, there are some destination-based routing protocols proposed in mobile ad hoc networks. In a destinationbased routing protocol a node only needs to know the next hop along the routing path when forwarding a packet to the destination. For example, DSR is a topology based routing protocol. AODV and DSDV are destination based routing protocols. In location-based routing protocols, the position relationship between a packet forwarding node and the destination, together with the node mobility can be used in both route discovery and packet forwarding. Location Aided Routing (LAR) and Distance Routing Effect Algorithm for Mobility (DREAM) are typical location-based routing protocols proposed for mobile ad hoc networks.

5. Multicast routing protocols

   Most classification methods used for unicast routing protocols for mobile ad hoc networks are also applicable for existing multicast routing protocols. For example, multicast routing algorithms for mobile ad hoc networks can be classified into reactive routing and proactive routing. The Ad-hoc Multicast Routing (AMRoute) and Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS) belong to category of proactive multicast routing and the On-Demand Multicast Routing Protocol (ODMRP) and Multicast Ad hoc On-demand Distance Vector (MAODV) are reactive multicast routing protocols.

   There is a classification method particularly used for multicast routing protocols for mobile ad hoc networks. This method is based on how distribution paths among group members are constructed. According to this method, existing multicast routing approaches for mobile ad hoc networks can be divided into tree based multicast routing, mesh based multicast routing, core based multicast routing and group forwarding based multicast.

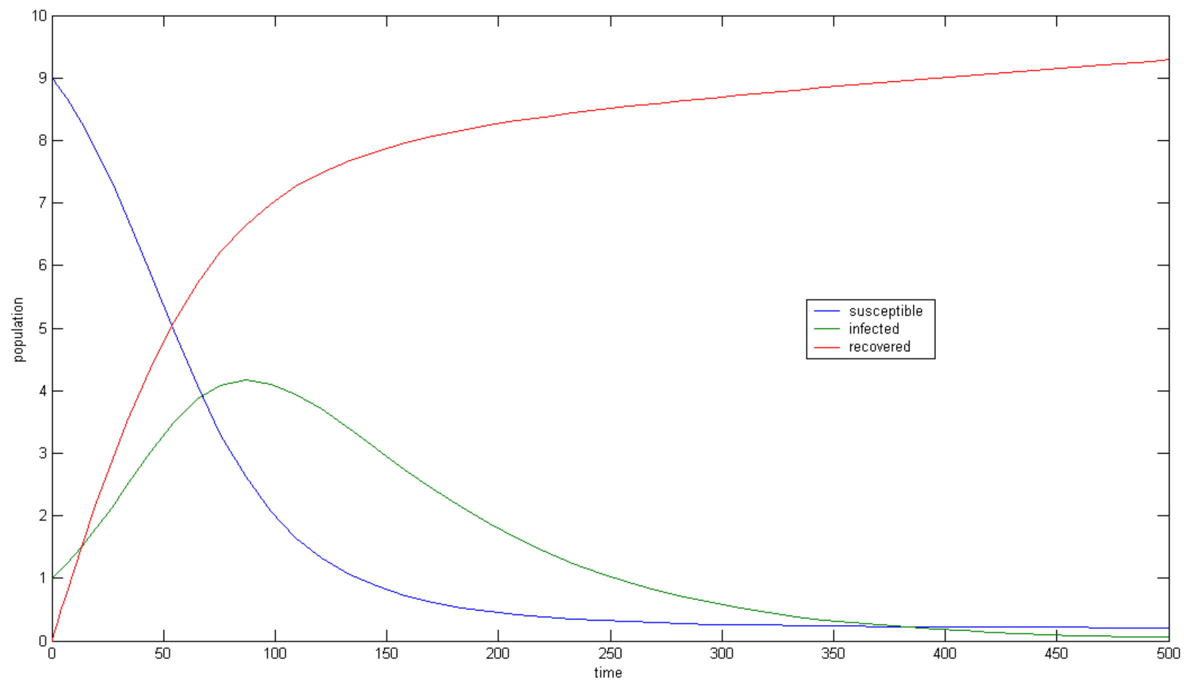

Figure 1.2: *SIR Model: Sample run for* $\beta = 0.005, \gamma = 0.013$*, population size = 10*

1.2 The SIR Model

An SIR model is an epidemiological model that computes the theoretical number of people infected with a contagious illness in a closed population over time. It was proposed by Kermack and McKendrick[8] to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865). It assumes that the population size is fixed (i.e., no births, deaths due to disease, or deaths by natural causes), incubation period of the infectious agent is instantaneous, and duration of infection is same as length of the disease. It also assumes a completely homogeneous population with no age, spatial, or social structure.

The model consists of a system of three coupled nonlinear ordinary differential equations,

$$\frac{dS}{dt} = -\beta SI \tag{1.1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I \tag{1.2}$$

$$\frac{dR}{dt} = \gamma I \tag{1.3}$$

where t is time, S is the number of susceptible people, I is the number of people infected, R is the number of people who have recovered and developed immunity to the infection, β is the infection rate, and γ is the recovery rate. A sample run for $\beta = 0.005$ and $\gamma = 0.013$ is shown in Fig.1.2.

1.3 Epidemic Routing

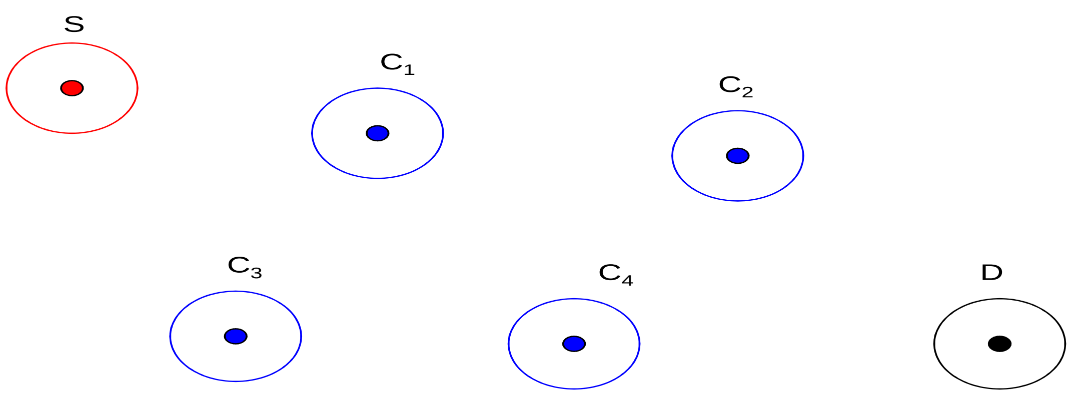

Figure 1.3: *Epidemic Routing: A random network topology at time $t = t_0$*

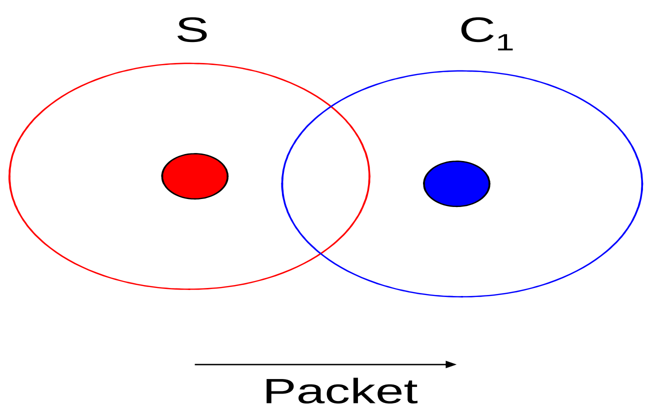

Figure 1.4: *Epidemic Routing: Packet forwarding from source to intermediate node at* $t = t_1$

Epidemic routing[9] is a widely proposed routing mechanism for data propagation in mobile ad hoc networks. The goal of this mechanism to develop techniques for delivering application data with high probability even when there is never a fully connected path between source and destination. Application messages are distributed to hosts, called carriers, within connected portions of ad hoc networks. In this way, messages are quickly distributed through connected portions of the network.

In other words, the source copies its packets to all the nodes it meets in its radio range. These nodes in turn copy the received packets to the other nodes they meet and so on. The data to be transmitted travels in a way analogous to the spread of an infection in a biological network. The destination finally receives the packet and measures are taken to eradicate the packet from the network.

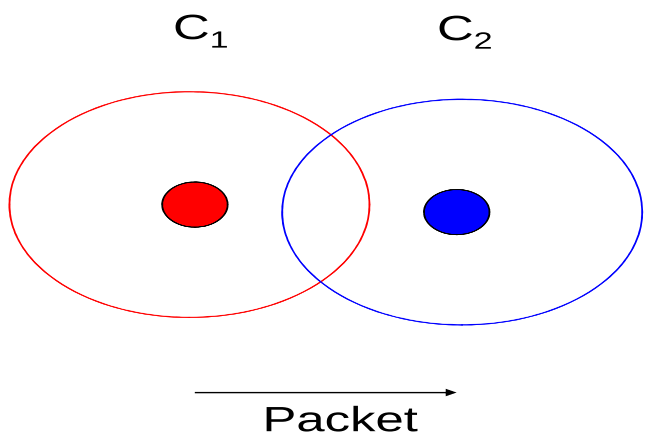

Figure 1.5: *Epidemic Routing: Packet forwarding among intermediate nodes at time* $t = t_n$

The general working of this protocol is shown in Fig1.3 to Fig.1.7. It shows a mobile ad hoc network where all nodes are moving in random directions with random speeds. The source node S wishes to send an application data packet to the destination node D. All other nodes in the network can be termed as carrier nodes C_1, C_2, C_3, etc. at this time, only the source has the data packet. This situation can be seen upon as the source node S being infected and all the other nodes in the network as susceptible to the infection.

At time $t = t_1$, S comes in contact with a carrier node say C_1. Two nodes are said to come in contact with each other when they are in each other's transmission range. At this time, the two nodes exchange information about what data packets they have, which are to be forwarded. Node S queries node C_1 and finds out that this node is susceptible. Hence, it forwards the data packet to C_1. Now C_1 is also infected. In this manner the infected nodes

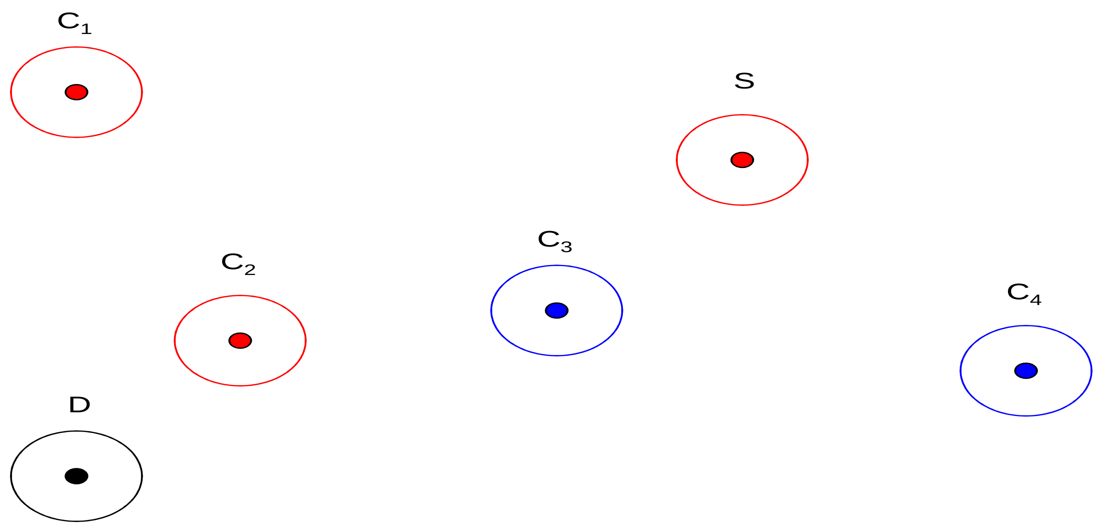

Figure 1.6: *Epidemic Routing: Random network topology at the start*

meet the susceptible nodes in the network and transfer the data packet as if spreading an infection.

Eventually, the destination receives the packet from one of the infected nodes. Let us call this time as $t = t_d$. This is the time of delivery the destination and is an important performance metric of the epidemic routing protocol. After this time, it is important that the packet be eliminated from all the infected nodes in the network. Also, further spread of this packet to the remaining susceptible nodes should be stopped. This task can be accomplished by several mechanisms known as recovery mechanisms.

Recovery mechanisms consist of spreading information about a delivered packet in the network, so that the nodes in the network adjust themselves accordingly and are able to organize and prepare themselves for further infections. In particular, as soon as an infected node re-

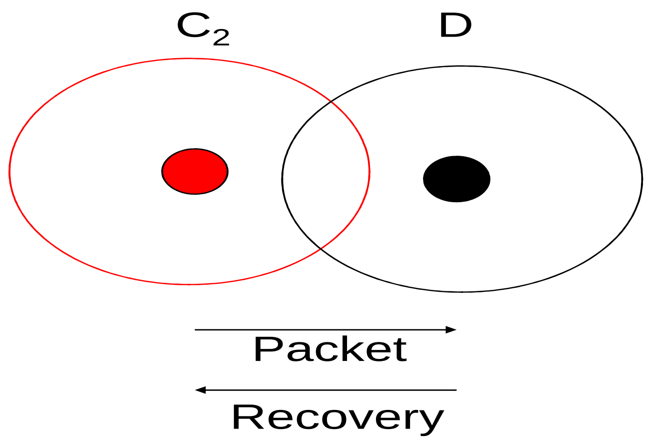

Figure 1.7: *Epidemic Routing: Packet delivery to destination and beggining of recovery at time* $t = t_d$

ceives information about a delivered packet, it deletes that packet from its own buffer. This enables it to handle a future infection and prevents further transfer of the delivered packet to other susceptible nodes. Likewise, a susceptible node can be configured to receive a packet delivery information as a way to prevent itself from getting this packet.

1.3.1 The power-delay trade-off issue

The task of routing in epidemic networks faces certain difficulties involving minimizing the delivery delay with a reduced consumption of resources. Every node has severe power constraints and the network is also susceptible to temporary but random failure of nodes. Every time a packet is copied from one node to another, power is consumed. Hence, the more the number of packet copies in the network, more is the power consumed. Similarly, the faster the nodes move, more is the power consumed. However, faster node movement and higher number of packet copies leads to a lesser delivery delay and vice versa.

Some metrics that help to evaluate the trade-off between power consumption and delivery delay are as follows:

1. The average lifetime L, of a packet is the time from when the packet is generated at the source node to the time when all copies of the packets are removed (i.e, no more infected nodes for this packet in the network).

2. The delivery delay of a packet, T_d, is the duration of the time from when the packet is generated at the source to the time the packet is first delivered to the destination.

3. For the case where nodes have a limited amount of buffer, a packet might be dropped from the network before it is delivered. The loss probability is the probability of a packet being dropped from the network before delivery.

4. There can be more than one metric related to the power consumption. The number of times a packet is copied in its entire lifetime, G; the number of times a packet is copied at the time of delivery, C being some of them. In our work, we consider that a network can control its own mobility, or the nodes can control their transmission ranges. The faster the nodes move, the more energy they use. Likewise, the more the transmission range of nodes, the higher the energy expenditure. Hence, we introduce the cost of mobility, C_m as another cost related to mobility.

1.4 Related Work

It is evident that epidemic routing aims to achieve minimum delivery delay at the cost of increased power requirements, buffer space and bandwidth. Various modifications have been suggested to bring out a balance between transmission delay and required resources. One class of suggestions involves variations in forwarding strategies. Some of them are listed as follows:

1. Limited Time Forwarding: In this setting, a certain time limit is set so that packets can be forwarded for upto that time in the network[10]. After this limited time, there is a constant number of infected nodes in the network. Packet delivery to the destination is through one of these infected nodes.

2. Probabilistic forwarding: In this setting, every time an infected node meets a susceptible node, it forwards the packet to the susceptible node with a certain probability. This parameter of probability is constant for a certain setting. In this way the total number of packet copies in kept limited[11,12].

3. K-hop forwarding: In this setting, each infected node can copy the packet only a certain number of times. Thus a packet can only travel a certain number of hops before it meets the destination[13,14]. Spray and wait routing[15] involves copying a certain number of packets in the network and waiting for delivery to destination.

Another class of strategies, involves different ways of recovery from infection. Once a node delivers a packet to the destination, all infected nodes should delete the copy from their buffers to conserve storage space and to prevent further infection. They should also store some information about this packet so that they do not get re-infected. This information is called anti-packet. Various schemes have been proposed for faster recovery from infection[16,17].

1. According to the IMMUNE scheme, a node stores a copy in the buffer until it meets the destination. Thus the infection may remain in the network long after the destination has received it.

2. The IMMUNETX scheme proposes a rather aggressive approach. In this scheme, anti-packets are propagated to all the infected nodes.

3. VACCINE, a stronger approach suggests sending the anti-packets to the susceptible nodes also so that newer susceptible nodes do not get infected if the packet delivery has occurred.

Mundur, Seligman and Lee[18] analyze and evaluate the network performance using an immunity scheme in the context of epidemic routing and its variants. The algorithm includes immunity-based information disseminated in the reverse once messages get delivered to their destination. The reverse dissemination of such information requires minimal resources and the trade-off in timely purging of delivered messages can be significant. Tower, Little and Thomas[19] introduce a scheme called SERAC, that increases the rate at which cure messages are propagated in a fragmented network for the purpose of reducing the overhead of outstanding yet incompletely disseminated messages.

The power and memory/storage constraints of miniaturized network nodes reduce the throughput capacity and increase the network latency. Sometimes, the required performance of such networks does not need to adhere to the level of services that would be required for performance-critical applications. For example, for some applications of sensor networks, minimal latency is not a critical factor and it could be traded off for a more limited resource, such as energy or throughput. Such networks are termed delay-tolerant networks.

To reduce the energy expenditure, transmission range of such sensor nodes would be quite short, leading to network topologies in which the average number of neighbors of the network nodes is very small. Here the classical networking approach of store-and-forward would not work well, as there is nearly never an intact path between a source and a destination. Several routing protocols have been proposed for this type of networking environment, one example is the Shared Wireless Infostation Model (SWIM)[20], where a packet propagates through the network by being copied (rather than forwarded) from a node to a node, as links are sporadically created. The goal is that one of the copies of the packet reaches the destination. SWIM is an example of the way that non-critical performance could be traded off for insufficient resources, such as the trade-offs between energy, delay, storage, capacity, and processing complexity.

Routing in opportunistic networks is usually based on some form of controlled flooding, which often this results in very high resource consumption and network congestion. Boldrini, Chiara, Passarella and Andrea advocate a context-based routing for opportunistic networks. This protocol is called (HiBOp)[21] and uses context information for forwarding decisions. Another approach called Context-Aware Routing (CAR) algorithm[22] provides of asynchronous communication in partially-connected mobile ad hoc networks, based on the intelligent placement of messages. This work also demonstrates that it is possible for nodes to exploit context information in making local decisions that lead to good delivery ratios and latencies with small overheads.

Many approaches rely on the use of either long range communication which leads to rapid draining of nodes' limited batteries, or existing node mobility which results in low data delivery rates and large delays. MF[23] is a mobility-assisted approach which utilizes a set of special mobile nodes called message ferries to provide communication service for nodes in the deployment area. The main idea behind the MF approach is to introduce non-randomness in the movement of nodes and exploit such non-randomness to help deliver data. The MF design exploits mobility to improve data delivery performance and reduce energy consumption in nodes.

Mobile wireless ad hoc and sensor networks can be permanently partitioned in many interesting scenarios. This implies that instantaneous end-to-end routes may not exist. Nevertheless, when nodes are mobile, it is possible to forward messages to their destinations through mobility. In many practical settings, spatial node distributions are very heterogeneous and possess concentration points of high node density. The locations of these concentration points and the flow of nodes between them tend to be stable over time. This motivates a novel mobility model, where nodes move randomly between stable islands of connectivity, where they are likely to encounter other nodes, while connectivity is very limited outside these islands.

This property has been exploited by developing algorithms that allow nodes to collaborate to discover such islands and to use them for efficient mobility forwarding. This is achieved by relying only on the evolution of the set of neighbors of each node. Sarafijanovic-Djukic, Pidrkowski and Grossglauser[24] propose an algorithm for this collaborative graph discovery problem and show that the inferred topology can greatly improve the efficiency of mobility forwarding.

PRoPHET[25] is proposed as an enhancement to the basic functionality of epidemic routing in which connectivity/position history of each node can be leveraged to make forwarding decisions, thus reducing traffic and increasing throughput with reduced latency. The Island Hopping scheme suggests that nodes often move in and out of areas with concentration points. The authors propose to move data from among concentration points via node mobility. This scheme requires that nodes form groups of close proximity and that destination geographical locations are known a priori.

Similarly, routing benefits considerably if one can take advantage of knowledge concerning node mobility. A high-dimensional euclidean space is constructed upon nodes' mobility patterns[26]. This space called MobySpace is based on the frequency of visits of nodes to each possible location. This work also determines that the degree of homogeneity of node mobility patterns has a high impact on routing. Grossglauser, Matthias and Tse[27] show that mobility has a great effect on increasing the capacity of the network.

PRioritized EPidemic (PREP)[28] for routing in opportunistic networks. PREP prioritizes bundles based on costs to destination, source, and expiry time. Costs are derived from per-link average availability information that is disseminated in an epidemic manner. PREP maintains a gradient of replication density that decreases with increasing distance from the destination.

Most forwarding algorithms aim at decreasing costs (relative to flooding the network) by forwarding only to nodes which are likely to be good relays. While it is non-trivial to decide if an encountered node is a good relay or not at the moment of encounter, it is harder still to prioritize which messages to transmit under the presence of short contact durations and which messages to drop when buffers become full. Different message prioritizing schemes have been studied by Erramilli, Vijay and Crovella[29].

Many concepts of anti-entropy, rumor mongering, and death certificates, were discussed in prior works[30] and used later (often under different names) in many of the later epidemic routing papers. Also under the assumption of no contention, epidemic routing has the minimum end-to-end delay amongst all the routing schemes proposed for such networks. The assumption of no contention was justified by arguing that since the network is sparse, there will be very few simultaneous transmissions.

Some recent papers have shown that this argument is not correct and that contention cannot be ignored while analyzing the performance of routing schemes, even in sparse networks[31]. There can be many manifestations of contention, such as the finite bandwidth of the link which limits the number of packets two nodes can exchange, the scheduling of transmissions between nearby nodes which is needed to avoid excessive interference, and the interference from transmissions outside the scheduling area.

Jain, Patra and Fall[32] formulate the delay-tolerant networking routing problem, where messages are to be moved end-to-end across a connectivity graph that is time-varying but whose dynamics may be known in advance. The problem has the added constraints of finite buffers at each node and the general property that no contemporaneous end-to-end path may ever exist. This work shows that with limited additional knowledge, far less than complete global knowledge, efficient algorithms can be constructed for routing in such environments.

Jain, Demmer, Patra and Fall[33] account for message losses due to link failures, buffer overruns, path selection errors, unscheduled delays, or other problems. This paper shows how to split, replicate, and erasure code message fragments over multiple delivery paths to optimize the probability of successful message delivery. Ideas from the modern portfolio theory literature are borrowed to solve the underlying optimization problem.

Some studies[34] suggest the use of network coding instead of replicating the packets in an epidemic network. When two nodes meet, they transmit coded packets to each other. A coded packet x is a linear combination of the K source packets.

1.5 Our Contribution

We notice that all the above schemes strive to deliver the packets (data and anti-packets) as soon as possible. In this paper, we derive an optimal mobility pattern among the nodes along with an optimal forwarding policy so that these packets are delivered with a minimum delay. Whenever two nodes meet and transmission is said to occur between them, some energy is spent. At the same time the expected delivery delay decreases. Giudici, Pagani and Rossi discuss the impact of mobility on epidemic networks[35]. It is pretty straightforward that the faster the nodes move, the quicker the packets will be delivered. However, more number of copies will be generated and power consumed. Thus this optimal mobility pattern should attain a minimum delivery delay along with minimum generation of packet copies and power consumption.

Chapter 2

The Mobility Model

To study a mobile ad hoc network protocol, it is important to simulate it and analyze it's performance. Simulations provide researchers with a number significant benefits, including repeatable scenarios, isolation of parameters and exploration of a variety of metrics. Some factors that play a key role in protocol simulation are movement pattern of nodes, communicating traffic patterns, topology, etc.

A mobility model represents the of movement of nodes, and how their location, velocity and acceleration change over time. Once the nodes have been initially distributed, the mobility model controls the movement of nodes within the network. In mobile ad hoc network research, it is very important to construct the simulation models as close to real circumstances, as possible. Wireless channels experience high variability in channel quality due to a variety of phenomenon, including multi-path, fading, atmospheric effects, and obstacles.

Previous research[36] has shown that the mobility model that we use can significantly impact the performance of ad-hoc routing protocols. A particular choice of a mobility model affects the packet delivery ratio, the control overhead, the data packet delay, etc. Hence it is important to use mobility models that accurately represent the intended scenarios in which the protocol is likely to be utilized. Unrealistic simulations may be misleading instead of being explanatory.

The mobility model that we use has been described in this chapter. A general classification of mobility models is provided in Section 2.1. Our model is then introduced in Section 2.2. Section 2.2 also describes some important properties of our model and justifies it's use for epidemic routing.

2.1 Classification of mobility models

There exists a wide variety of mobility models that have been postulated from both analytic and simulation-based studies on mobile systems. A concise categorization can be found in [37], while a simulation based comparison of a variety of mobility models can be found in [36]. This section describes a few of these models that have been designed specifically for ad hoc networks. These mobility patterns can be classified according to the following criteria:

2.1.1 Generation of mobility

Traces [38,39] are the predetermined mobility patterns that are observed in real life systems [40,41]. They provide accurate information, especially when they involve a large number of participants and an appropriately long observation period. However, new network environments (e.g. ad hoc networks) are not easily modeled if traces have not yet been created. For ad hoc networks, tracing the actual behavior of mobile nodes is a hard process. In this type of situation it is necessary to use synthetic models. Synthetic models attempt to realistically represent the behaviors of mobile nodes without the use of traces. Traces hardly let researchers to change simulation parameters, which can be a disadvantage for performance analysis of ad hoc networks.

2.1.2 Social behavior of nodes

Another way of classification of mobility models is the social behavior of nodes. The social behavior of mobile nodes can be identified by the dependence of mobile nodes among each other. In the entity mobility models, a mobile node is considered an entity that moves independently of other nodes. Examples of such models are Random Walk Mobility Model (including its many derivatives), which is simple mobility model based on random directions and speeds, Random Direction Mobility Model, a model that forces mobile nodes to travel to the edge of the simulation area before changing direction and speed, Random Waypoint Mobility Model, a model that includes pause times between changes in destination and speed, etc.

However, in some scenarios including battlefield communication and museum touring, the movement pattern of a mobile node may be influenced by certain specific 'leader' node in its neighborhood. Hence, the mobility of various nodes is correlated. The size and the movements of these groups (and within the group) vary from scenario to scenario, but several characteristics are common to all these scenarios: the nodes are split in several smaller groups, and each group acts seemingly independently of the other groups. Also, within each group, each user has its own liberty to move with respect to the center of the group or with respect to the other members of the group. In this respect, the group mobility models have two sub-models the group model describing the movements of the groups and the individual model describing the movement of a node within the group. Since the velocities of different nodes are correlated in space, this characteristic is called spatial dependency of velocity. Some group mobility models are as follows:

Reference Point Group Mobility Model

In line with the observation that the mobile nodes in mobile ad-hoc network tend to coordinate their movement, the Reference Point Group Mobility Model is proposed in[42]. One example of such mobility is that a number of soldiers may move together in a group. Another example is during disaster relief where various rescue crews (e.g., firemen, policemen and medical assistants) form different groups and work cooperatively.

In the Reference Point Group Mobility Model model, each group has a center, which is either a logical center or a group leader node. For the sake of simplicity, we assume that the center is the group leader. Thus, each group is composed of one leader and a number of members. The movement of the group leader determines the mobility behavior of the entire group. Both the movement of the logical center for each group, and the random motion of each individual mobile node within the group, are implemented via any of the stochastic mobility models.

The movement of group leader at any time not only does it define the motion of group leader itself, but also it provides the general motion trend of the whole group. Each member of this group deviates from this general motion vector by some degree. This motion vector can be randomly chosen or carefully designed based on certain predefined paths. For each node, mobility is assigned with a reference point that follows the group movement. Upon this predefined reference point, each mobile node could be randomly placed in the neighborhood.

The RPGM model was designed to depict scenarios such as an avalanche rescue. During an avalanche rescue, the responding team consisting of human and canine members work cooperatively. The human guides tend to set a general path for the dogs to follow, since they usually know the approximate location of victims. The dogs each create their own 'random' paths around the general area chosen by their human counterparts.

There can be many variations to the RPGM model. If appropriate group paths are chosen, along with proper initial locations for various groups, many different mobility applications may be represented with the RPGM model. The In-place Mobility Model partitions a given geographical area such that each subset of the original area is assigned to a specific group; the specified group then operates only within that geographic subset. The Overlap Mobility Model simulates several different groups, each of which has a different purpose, working in the same geographic region; each group within this model may have different characteristics than other groups within the same geographical boundary. For example, in disaster recovery of a geographical area, one might encounter a rescue personnel team, a medical team, and a psychologist team, each of which have unique traveling patterns, speeds, and behaviors. Lastly, the Convention Mobility Model divides a given area into smaller subsets and allows the groups to move in a similar pattern throughout each subset. Similar to the Overlap Mobility Model, some groups in the Convention Mobility Model may travel faster than others.

Other Spatially Correlated Models

Sanchez and Manzoni[43] propose a set of mobility models in which the mobile nodes travel in a co-operative manner. This set of mobility models, including Column Mobility Model, Pursue Mobility Model and Nomadic Mobility Model, are expected to exhibit strong spatial dependency between nearby nodes.

The Column Mobility Model represents a set of mobile nodes (e.g., robots) that move in a certain fixed direction. This mobility model can be used in searching and scanning activity, such as destroying mines by military robots. At a time slot t, the mobile node i is to update its reference point by adding an advance vector to its previous reference point, where the advance vector is the predefined offset used to move the reference grid of node i at time t. After the reference point is updated, the new position of mobile node i is to randomly deviate from the updated reference point by a random vector. When the mobile node is about to travel beyond the boundary of a simulation field, the movement direction is then flipped 180 degree. Thus, the mobile node is able to move towards the center of simulation field in the new direction.

The Pursue Mobility Model emulates scenarios where several nodes attempt to capture single mobile node ahead. This mobility model could be used in target tracking and law enforcement. The node being pursued (i.e., target node) moves freely according to the Random Waypoint model. By directing the velocity towards the position of the targeted node, the pursuer nodes (i.e., seeker nodes) try to intercept the target node.

The Nomadic Mobility Model is to represent the mobility scenarios where a group of nodes move together. This model could be applied in mobile communication in a conference or military application. The whole group of mobile nodes moves randomly from one location to another. Then, the reference point of each node is determined based on the general movement of this group. Inside of this group, each node can offset some random vector to its predefined reference point.

Compared to the Column Mobility Model which also relies on the reference grid, it is observed that the Nomadic Community Mobility Model shares the same reference grid while in Column Mobility Model each column has its own reference point. Moreover, the movement in the Nomadic Community Model is sporadic while the movement is more or less constant in Column Mobility Model. This set of mobility models has been utilized to analyze the protocol performance. Both Hu and Johnson[44] and Camp, Boleng and Davies[36] report that this set of mobility models behaves different than Random Waypoint model.

2.1.3 Temporal Dependencies

Mobility of a node may be constrained and limited by velocity and rate of change of direction. Hence, the current velocity of a mobile node may depend on its previous velocity. Thus the velocities of single node at different time slots are correlated'. We call this mobility characteristic the Temporal Dependency of velocity. In this section, some mobility models considering temporal dependency are discussed.

Gauss-Markov Mobility Model

The Gauss-Markov Mobility Model was first introduced by Liang and Haas[45] and widely utilized[36,45,46]. In this model, the velocity of mobile node is assumed to be correlated over time and modeled as a Gauss-Markov stochastic process. When the node is going to travel beyond the boundaries of the simulation field, the direction of movement is forced to flip 180 degree. This way, the nodes remain away from the boundary of simulation field.

In the Gauss-Markov model, the degree of dependency is determined by the memory level parameter α. α is a parameter to reflect the randomness of Gauss-Markov process. By tuning this parameter, it is capable of duplicating different kinds of mobility behaviors. At $\alpha = 0$, the model is memoryless. For $\alpha = 1$, the model has a very strong memory.

Smooth Random Mobility Model

Another mobility model considering the temporal dependency of velocity over various time slots is the Smooth Random Mobility Model. Instead of the sharp turn and sudden acceleration or deceleration, Bettstetter also proposes to change the speed and direction of node movement incrementally and smoothly. It is observed that mobile nodes in real life tend to move at certain preferred speeds, rather than at speeds purely uniformly distributed in the range $[0, V]$. Therefore, in Smooth Random Mobility model, the probability distribution of node velocity is as follows: the speed within the set of preferred speed values has a high probability, while a uniform distribution is assumed on the remaining part of entire interval

$[0, V]$.

In Smooth Random Mobility Model, the frequency of speed change is assumed to be a Poisson process. Upon an event of speed change, a new target speed is chosen according to the probability distribution function. Then, the speed of mobile node is changed incrementally from the current speed v to the targeted new speed by acceleration speed or deceleration speed $a(t)$. Thus, the speed may be controlled to increase or decrease continuously and incrementally. If $a(t)$ is a small value, then the speed is changed slowly and the degree of temporal correlation is expected to be strong. Otherwise, the speed can be changed quickly and the temporal correlation is small.

Unlike speed, the movement direction is assumed to be purely uniformly distributed. Once a movement direction is chosen, the node moves in a straight line until the direction changes. The frequency of direction change is assumed to have an exponential distribution. The change of movement direction is also be smooth and incremental. Therefore, the change is achieved in more than one time slots.

2.1.4 Geographic Restriction

In many mobility models, the nodes are allowed to move freely and randomly anywhere in the simulation field. However, in most real life applications, we observe that a nodes movement is subject to the environment. The motions of vehicles are bounded to the freeways or local streets in the urban area, and on campus the pedestrians may be blocked by the buildings and other obstacles. Therefore, the nodes may move in a pseudo-random way on predefined pathways in the simulation field. Some recent works address this characteristic and integrate the paths and obstacles into mobility models. This kind of mobility model is called a mobility model with geographic restriction.

Pathway Mobility Model

One simple way to integrate geographic constraints into the mobility model is to restrict the node movement to the pathways in the map. The map is predefined in the simulation field. Tian, Hahner and Becker[47] utilize a random graph to model the map of city. This graph can be either randomly generated or carefully defined based on certain map of a real city. The vertices of the graph represent the buildings of the city, and the edges model the streets and freeways between those buildings.

Initially, the nodes are placed randomly on the edges of the graph. Then for each node a destination is randomly chosen and the node moves towards this destination through the shortest path along the edges. Upon arrival, the node pauses for T time and again chooses a new destination for the next movement. This procedure is repeated until the end of simulation. Hence, in this graph based mobility model, the nodes are traveling in a pseudo-random fashion on the pathways. Similarly, in the Freeway mobility model and Manhattan mobility model[48], the movement of mobile node is also restricted to the pathway in the simulation field.

Obstacle Mobility Model

Another geographic constraint playing an important role in mobility modeling includes the obstacles in the simulation field[49]. To avoid the obstacles on the way, the mobile node is required to change its trajectory. Therefore, obstacles do affect the movement behavior of mobile nodes. Moreover, the obstacles also impact the way the transmission waves propagate. For example, for the indoor environment, typically, the node could not propagate the signal through obstacles without severe attenuation. When the radio propagates through an obstacle, the signal is assumed to be fully absorbed by the obstacle. More specifically, if an obstacle is in-between two nodes, the link between these nodes is considered broken until one moves out of the shadowed area of the other.

2.1.5 Random Based Mobility Models

In random-based mobility models, the mobile nodes move randomly and freely without restrictions. The destination, speed and direction are all chosen randomly and independently of other nodes. This kind of a model has been used in many simulation studies. The Random Walk model, Random Waypoint model and Random Direction model are examples of such models.

The Random Walk Model is also referred to as the Brownian Motion. This model mimics the behavior of nodes that move in an unexpected way. The nodes change their speed and direction at each time interval. The Random Walk model is a memoryless mobility process where the information about the previous status is not used for the future decision. That is to say, the current velocity is independent with its previous velocity and the future velocity is also independent with its current velocity.

The Random Waypoint Model is the model that has been using in our setting. It is explained in detail in the next section. The Random Direction model based on similar intuition is proposed by Royer, Melliar-Smith and Moser[50]. This model is able to overcome the non-uniform spatial distribution and density wave problems. Instead of selecting a random destination within the simulation field, in the Random Direction model the node randomly and uniformly chooses a direction by which to move along until it reaches the boundary. After the node reaches the boundary of the simulation field and stops with a pause time T, it then randomly and uniformly chooses another direction to travel. This way, the nodes are uniformly distributed within the simulation field. The Modified Random Direction model that allows a node to stop and choose another new direction before it reaches the boundary of the simulation field.

2.2 The Random Waypoint Model

In this section we describe the mobility model that was used in our work. The Random Waypoint Model was first proposed by Johnson and Maltz[51]. Soon, it became a benchmark mobility model to evaluate the MANET routing protocols, because of its simplicity and wide availability.

At the beginning, each mobile node randomly selects one location in the simulation field as the destination. It then travels towards this destination with constant velocity chosen uniformly and randomly from $[0, V]$, where the parameter V is the maximum allowable velocity for every mobile node. The velocity and direction of a node are chosen independently of other nodes. Upon reaching the destination, the node stops for a duration defined by the 'pause time' parameter, T . If $T = 0$, it signifies continuous mobility. After this duration, the mobile node again chooses another random destination in the simulation field and moves towards it. The whole process is repeated again and again until the simulation ends.

In the Random Waypoint model, V and T are the two key parameters that determine the mobility behavior of nodes. If the V is small and the pause time T is long, the topology of the network becomes relatively stable. On the other hand, if the node moves fast and the pause time T is small, the topology is expected to be highly dynamic. Varying these two parameters, especially V, the Random Waypoint model can generate various mobility scenarios with different levels of nodal speed.

Bettstetter, Hartenstein and Perez-Costa[52] describe Random Waypoint model as a discrete time stochastic process. The transition length or the epoch length is defined as the distance that any arbitrary node moves from one waypoint to another during the ith epoch. The average transition length in a single epoch i over all the nodes (i.e., ensemble average) is equal to the average of the transition length of a single node

The Random Waypoint model has several variations. Examples are Random Walk model, Random Direction model, etc. The Random Walk model was originally proposed to emulate the unpredictable movement of particles in physics. It is also referred to as the Brownian Motion. The Random Walk model has similarities with the Random Waypoint model because the node movement has strong randomness in both models. We can think the Random Walk model as the specific Random Waypoint model with zero pause time.

In the Random Walk model, the nodes change their speed and direction at each time interval. For every new interval, each node randomly and uniformly chooses its new direction θ from $(0, 2\pi]$. In similar way, the new speed follows a uniform distribution or a Gaussian distribution from $[0, V]$. If the node moves according to the above rules and reaches the boundary of simulation field, the leaving node is bounced back to the simulation field with the angle of $\pi - \theta$. This effect is called border effect.

Bettstetter[52,53] observes that the spatial node distribution of Random Waypoint model is transformed from uniform distribution to non-uniform distribution after the simulation starts. As the simulation time elapses, the unbalanced spatial node distribution becomes even worse. Finally, it reaches a steady state. In this state, the node density is maximum at the center region, whereas the node density is almost zero around the boundary of simulation area. This phenomenon is called non-uniform spatial distribution.

Another property of Random Waypoint model called density wave phenomenon (i.e., the average number of neighbors for a particular node periodically fluctuates along with time) is observed by Royer, Melliar-Smith and Moser[50]. This phenomenon results from the certain mobility behavior of Random Waypoint model. In Random Waypoint model, since the nodes are likely to either move towards the center of simulation field or choose a destination that requires movement through the middle, the nodes tend to cluster near the center region of simulation field and move away from the boundaries. Therefore, a non-uniform distribution is formed. At the same time, the nodes appear to converge, disperse and converge at center region periodically, resulting in the fluctuation of the node density of neighbors (i.e., density wave).

Lassila, Hyytia and Koskinen[54] focus on estimating two quantities: The probability that the network is connected and the mean duration of the connectivity periods. They also show that in sparse networks, mobility has a positive effect on connectivity, whereas in dense networks, the situation becomes the opposite. A network is said to be connected if there exists a path between all node pairs, and k connected if for each node pair, at least k node disjoint paths exist. The movement is restricted to a unit disk, but the domain of movement can be any convex region. Let $Q_{n,k}(d)$ denote the probability that an arbitrary node in the network has at least k neighbors.

$$Q_{n,k}(d) = 2 * pi * \int_0^1 rf(r) \left(1 - \sum_{i=0}^{k-1} \binom{n-1}{i} p(r,d)^i (1-p(r,d))^{n-1-i}\right) dr,$$

where $B_d(r)$ is the coverage area of any node, with a transmission range d. $p(r,d)$ is the probability that a given node is located within $B_d(r)$ and can be expressed as

$$p(r,d) = \int_{x \in B_d(r)} f(|x|) dA,$$

where $f(|x|)$ is the stationary node distribution of the random waypoint mobility model. Hence, connectivity is defined as $C_{n,k}(d)$ where a sub-network of n nodes is k connected and is defined by:

$$C_{n,k}(d) = P\{nnodesarek - connected\} \approx (Q_{n,k}(d))^n$$

Let $\bar{T_d}$ be the mean dis-connectivity time

$$\bar{T_c} = \frac{C_{n,1}(d)}{1 - C_{n,1}(d)} \bar{T_d}.$$

Similarly, a lot of work has been done to evaluate the steady state or the stationary distribution of nodes in a Random Waypoint Model[55]. The stationary node distribution can also be evaluated as a function of distance of the node from the boundaries of the simulation area[56]. Hyyti and Virtamo[57] evaluate the cell change rate and the spatial distribution of Random Waypoint mobility model in cellular networks. Bettstetter[53] evaluates the spatial distribution of nodes as composed of three distinct components: the static, pause, and mobility component.

The static component accounts for the fact that a node can remain static for the entire network operational time. The pause component accounts for the time that a mobile node rests before starting a new movement period. Finally, the mobility component accounts for the time that a mobile node is actually moving.

Rojas, Branch, and Armitage[58] validates the use of Random Wapoint Mobility Model to represent mobility patterns in large geographical areas. They use real traces of human movement, i.e destinations, pause times, direction, velocity and length of movement. The spatial distribution is fairly similar to the random waypoint model since the travels to the city center are generally more frequent and populated.

Studies show that, under the random waypoint mobility regime, average node speed tends to zero in steady state[59]. They also show that average node speed varies considerably from the expected average value for the time scales under consideration in most simulation analysis. They provides an accurate estimate of the warm-up period required by simulations using the random waypoint mobility model. Simulation data uptill the warm-up period can then be discarded to obtain accurate protocol performance results. Given that random waypoint mobility is still, by far, the most widely used mobility model in the evaluation of MANETs, the contribution of this work is potentially significant as it allows network protocol designers to continue to use the original random waypoint mobility model and yet obtain accurate results characterizing MANET protocol performance.

Chapter 3

The DP problem

This chapter deals with the mathematical formulation of the optimization problem. The problem is formulated as a Markov Decision problem and it can be solved using the principles of dynamic programming. Section3.1 introduces the basic principles of dynamic programming. Since the problem is stochastic in nature, the extension from a deterministic problem to a stochastic problem has been described in Section??. Finally, in Section 3.3 the mathematical problem is described and a solution is derived.

3.1 Dynamic programming

A dynamic programming problem deals with situations where decisions are made in stages. The outcome of each decision may not be fully predictable but it can be anticipated to some extent before the next decision is made. The objective is to minimize a certain cost a mathematical expression of what is considered an undesirable outcome.

A key aspect of such situations is that decisions cannot be viewed in isolation since one must balance the desire for low present cost with the undesirability of high future costs. The dynamic programming technique captures this trade-off. At each stage, it ranks decisions based on the sum of the present cost and the expected future cost, assuming optimal decision making for subsequent stages.

There is a very broad variety of practical problems that can be treated by dynamic programming. In this section, a broadly applicable model of optimal control of a dynamic system over a finite number of stages (a finite horizon) is discussed. A basic model has two principal features:

1. An underlying discrete time dynamic system, and
2. A cost function that is additive over time.

The dynamic system expresses the evolution of the system's state ,under the influence of decisions made at discrete instances of time. The system has the form

$$k = 0, 1, ..., N - 1,$$

where

k is the index of discrete time,

x_k is the state of the system and summarizes past information that is relevant for future optimization,

u_k is the control or decision variable to be selected at time k,

w_k is a random parameter (also called disturbance or noise depending on the context),

N is the horizon or number of times control is applied,

and f_k is a function that describes the system and in particular the mechanism by which the state is updated.

The cost function is additive in the sense that the cost incurred at time k, denoted by

$$g_k(x_k, u_k, w_k),$$

accumulates over time. The total cost is

$$g_N(X_N) + \sum_{k=0}^{N-1} g_k(X_k, U_k, W_k),$$

where $g_N(X_N)$ is a terminal cost incurred at the end of the process. However, because of the presence of W_k, the cost is generally a random variable and cannot be meaningfully optimized. We therefore formulate the problem as an optimization of the expected cost where the expectation is with respect to the joint distribution of the random variables involved. Therefore the problem is formulated as an optimization problem of the expected cost given by

$$E\left\{g_N(X_N) + \sum_{k=0}^{N-1} g_k(X_k, U_k, W_k)\right\}$$

The optimization is over the controls $u_0, u_1, ..., u_{N-1}$, but some qualification is needed here; each control U_k is selected with some knowledge of the current state X_k (either its exact value or some other related information).

3.2 The Markov Decision Problem

We consider a discrete-time dynamic system where the state X_k is an element of a space S_k, the control u_k is an element of a space O_k, and the random disturbance w_k is an element of a space D_k.

The control u_k is constrained to take values in a given nonempty subset $U_k(x_k) \subset O_k$, which depends on the current state x_k, i.e., $u_k \in U_k(x_k)$ for all $x_k \in S_k$ and k. The random disturbance w_k is characterized by a probability distribution $P_k(\cdot \mid x_k, u_k)$ that may depend explicitly on x_k and u_k but not on values of prior disturbances $w_{k-l}, .., w_0$.

We consider the class of policies (also called control laws) that consist of a sequence of functions

$$\pi = \{\mu_0, ..., \mu_{N-1}\},$$

where μ_k maps states x_k into controls $u_k = \mu_k(x_k)$ and is such that $\mu_k(x_k) \in U_k(X_k)$ for all $x_k \in S_k$. Such policies will be called admissible.

Given an initial state x_0 and an admissible policy $\{\mu_0, \mu_1, ..., \mu_{N-1}\}$, the states x_k and disturbances w_k are random variables with distributions defined through the system equation

$$x_k = f_k(x_k, \mu_k(x_k), w_k),\ k = 0, 1, ..., N-1.$$

Thus, for given functions $g_k, k = 0, 1, ..., N$, the expected cost of π starting at X_0 is

$$J_\pi(x_0) = E\left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

where the expectation is taken over the random variables w_k and X_k. An optimal policy π^* is one that minimizes this cost. i.e.,

$$J_{\pi^*}(x_0) = \min_{\pi \varepsilon \Pi} J_\pi(x_0)$$

where Π is the set of all admissible policies.

3.3 Model

We present our work as an insightful extension to the model described in[60]. The authors derive a forwarding limit which optimizes the delivery delay and the cost related to the packet copies. In our model, we consider node movement as an additional factor. The faster the nodes move, the lesser the delivery delay but higher node movement involves higher power consumption. We consider the movement to be modeled by random waypoint mobility. Such a model is characterized by the mobility parameter, β. Higher values of β denote more frequent meetings between nodes. This may be achieved by increasing their speeds or transmission ranges. Hence, there is a certain cost associated with maintaining a certain value of β for a certain time. We denote this cost by α. Figure 1 shows a three dimensional plot of the cost which is the sum of delivery delay, number of infected nodes and weighted cost of mobility, the number of infected nodes and the mobility parameter. This figure shows the cost incurred by keeping a certain value of β and infected copies till delivery occurs. We see that mobility also plays a role in achieving a trade-off between delivery delay and power consumption. Assuming that we can vary this factor of mobility, it is possible to further optimize the related cost.

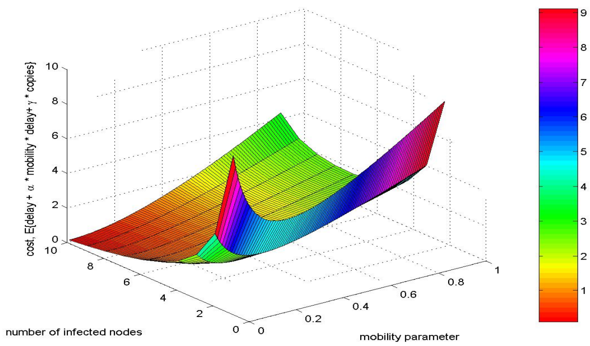

Figure 3.1: *Plot of cost vs. infected nodes vs. mobility*

In Section 3.3, we provide a description of the model and obtain analytical solutions. In the next chapter 4, we provide numerical results to compare the solution with existing strategies. Section 4.6 contains conclusions and future directions.

We assume a model of $N+1$ nodes which move according to the random waypoint mobility model in a well defined, enclosed area A. The inter-nodal meeting times are assumed to be exponentially distributed with rate β, which will be discussed in further detail. Two nodes are said to meet when they come in the transmission range of each other and can exchange packets. These nodes have a small transmission range d compared to A at all times. The node density is assumed to be low and hence the interference among the nodes can be easily ignored. We also assume that transmission of packets occurs instantaneously when the nodes meet. The node velocities are high. We also assume that the bandwidth of the network and the buffer sizes is large enough. There can be multiple source-destination pairs, but we consider only one pair to study the packet spreading through the network. Our assumption regarding large bandwidth and buffer sizes allows us to assure that different infections (i.e. packet-spreadings) will be independent.

The model for transmission of the data packets closely resembles the SIR model for the spread of infections in biological networks[61,62]. The source can be seen as the first infected node. Prior to receiving the data, the nodes can be termed as susceptible. When any susceptible nodes come in the range of infected nodes, the infected node decide to pass the copy of the packet or not. If the packet is copied, the infection spreads and the count of the total number of infected nodes increases by one. Eventually, the destination comes in contact with one of the infected nodes and the packet is delivered. As soon as the destination receives the data, the recovery process starts. This is the process of clearing the packets from the infected nodes. There are various schemes for recovery. We do not consider the process of recovery in our model as the recovery process is entirely dependent upon the total number of nodes that are infected. In our solution, we seek to optimize the total number of infected packets.

The inter-nodal meeting times are assumed to be exponentially distributed with rate β. In earlier models, β was considered to be constant for a specific model. However, in our model, we assume β to be a time varying entity. In this manner, we can study the impact of the mobility parameter on the cost. The authors in[63] confirmed that the inter-meeting time of nodes is exponential when all the nodes have a transmission range d. The rate β is given by:

$$\beta = \frac{2wdE\{V*\}}{A} \tag{3.1}$$

where w is a constant, $E[V*]$ is the average relative speed between two nodes, d is the transmission range and A is the area of the topology grid.

Changes in β can be made by many factors. For example, adjustments in node power affect the transmission range and hence cause a variation in β. Another factor might be changes in node velocities or acceleration which would vary the average relative speed between a pair of nodes. Based on this, our model assumes that these factors would have to be changed instantaneously for all nodes. There is also a possibility of considering heterogeneous nodes (i.e nodes having different radii of transmission, different average velocities, etc). Some research has been done to identify the inter-meeting patterns for such scenarios[64]. Defining equivalent mobility parameters for these cases is a matter of further research. Our present model can work upon any pattern thus obtained.

In this paper, we wish to derive the optimal policy of controlling two factors, namely packet forwarding and the mobility parameter, β so that the delivery delay, total number of packets generated and the power requirements are optimized. In particular, the infected nodes decide whether or not to forward packets when they meet other susceptible nodes and what change should be made to β in order to reach an optimal cost. Hence our cost is:

$$J = E\{T_d + \gamma M_c + C_\beta\}, \tag{3.2}$$

where T_d is the time gap between origination of packet from source and the delivery to the destination, γ is the cost related to copying a packet, M_c is the total number of packet copies generated in the process and C_β is the cost associated with maintaining β. Since higher values of β suggest high power consumption, C_β has to be proportional to the magnitude of β. However, higher values of β would also contribute to lesser delivery delay. Hence, we take this cost to be proportional to β^2. Another reason why C_β is chosen to be proportional to β^2 is the following. We consider varying β by varying the transmission radius d, and a linear change in d implies a quadratic change in the power. Hence, $C_\beta = \alpha\beta^2$. The values of α and γ are a matter of design choice. The higher the values, the more importance we give to the respective costs.

We now describe the process. At each time t, the state of the process is the pair $x_t = (n_I(t), \beta(t))$, where $n_I(t)$ denotes the current number of infected nodes and $\beta(t)$ denotes the current value of the mobility parameter. Since we do not consider the recovery process, the number of infected nodes is set to 0 when the destination receives the packet, i.e. $n_I(T_d) = 0$.

At the beginning of the process $t_0 = 0, n_I(0) = 1$. The state of the process can assume values $0, 1, \cdots, N$, where N is the number of nodes in the system (except the destination). The state changes whenever an infected node meets a susceptible node or the destination. Hence, if t_k is the k^{th} inter-meeting time between an infected node and a non-infected node, then $n_I(t_k) = n_I(t_{k-1}) + 1$ (meets susceptible node) or $n_I(t_k) = 0$ (meets destination node). We also define our decision at any time t_k, which is denoted by $u_{t_k} = \{p, q\}$. In this tuple, $p = \{c, \bar{c}\}$ denotes the forwarding strategy. $p = c$ implies that the packet should be copied, whereas $p = \bar{c}$ implies that the packet should not be copied to the susceptible node. Also, $q = \{\lambda_+, \lambda_-, \lambda_0\}$ denotes strategy of changing β. $q = \lambda_+$, implies that β should be increased by the magnitude λ, $y = \lambda_-$, implies that β should be decreased by the magnitude λ, $q = \lambda_0$, implies that β should not be changed. Here, λ is a predefined step-size parameter of design choice. In our model, β can be varied as long as it has a positive value between 0 and 1. Thus, if at a particular instant the decision is such that it would take β out of these bounds, then the system would be compelled to choose between other decision choices.

We also assume perfect state information at each node. This means that all nodes know the exact number of infected nodes in system at all times. They also maintain the same value of d, the transmission radius at all times. Due to this assumption, the distributed nature of our system is lost. Even though, the actuator of a certain decision is a single infected node which meets a susceptible node, the system assumes a central controller which causes all the nodes to be aware of the decision and execute it. In addition, we observe that there is no decision to be made when an infected node comes in contact with the destination. The packet is simply copied to the destination. Since, this is the absorbing state of the model, there is no need to decide about any further change to be made to β.

Hence we are able to define an optimal policy as the infinite set of functions $\pi = \{\mu_1, \mu_2, \cdots\}$, where μ_k maps the state x_{t_k} into control $u_k = \mu_k(x_{t_k})$. This problem can be studied as a stochastic shortest path with finite state and exponential transition time. The initial state of the problem is $x_0 = (1, \beta(0))$. The cost function of an admissible policy starting at time t_1 from state $x_{t_1} = (n_I(t_1), \beta(t_1))$ is the limit of the cost from t_1 to t_k as $k \to \infty$.

$$\begin{aligned} & J_\pi(n_I(t_1), \beta(t_1)) \\ = \; & \lim_{k\to\infty} \sum_{u=1}^{k-1} E\left\{\widehat{g}(x_{t_k}, \mu_k(x_{t_k}))|x_{t_1}\right\} + \\ & E\left\{\int_{t_k}^{t_{k+1}} g(x_{t_k}, \mu_k(x_{t_k}))dt|x_{t_1}\right\}, \end{aligned} \tag{3.3}$$

where $\widehat{g}(x_{t_k}, \mu_k(x_{t_k}))$ is the finite cost of the decision taken at time t_k, while $g(x_{t_k}, \mu_k(x_{t_k}))$ is the cost per unit time. Comparing (3.3) to (3.2), we see that

$$g(x_{t_k}, \mu_k(x_{t_k})) = 1 + \alpha\beta(t_k)^2,$$

and

$$\widehat{g}(x_{t_k}, \mu_k(x_{t_k}) = \begin{cases} 0 & \text{if } u = (\bar{c}, \cdot) \\ \gamma & \text{if } u = (c, \cdot). \end{cases}$$

The total cost also includes the time from t_0 to t_1, which is the first inter-meeting time, and the cost associated with copying the packet to the destination, are unavoidable and do not depend on the policy. Hence, these are kept out of the optimization framework. (3.3) can be written as

$$\begin{aligned} & J_\pi(n_I t_1, \beta(t_1)) \\ = \; & \widehat{g}(x_{t_1}, \mu_1(x_{t_1})) + (1 + \alpha\beta(t_1)^2)G(x_{t_1}, \mu_1(x_{t_1})) + \\ & \sum_{j=0}^{N} m_{n_I(t_1),j}(\mu_1(x_{t_1})J_{\pi_2}(j, \beta(t_2)) \end{aligned} \tag{3.4}$$

where J_{π_2} is the cost-to-go of the policy $\pi_2 = \{\mu_2, \mu_3, \cdots\}$ that is used from the second meeting time, $m_{i,j}(u)$ is the probability of transition from a state with i infected nodes to a state with j infected nodes under the decision u (it is independent of the β value), and $G(i, u)$ is the average transition time from state i to another state when the decision is u.

The expressions for $G(i,u)$ and $m_{i,j}$ for $i,j \neq 0$ are as follows:

$$G(i,\beta,u) = \begin{cases} \frac{1}{(\beta+\lambda)i(N-i+1)} & \text{if } u = (\bar{c},\lambda_+), \\ \frac{1}{\beta i(N-i+1)} & \text{if } u = (\bar{c},\lambda_0), \\ \frac{1}{(\beta-\lambda)i(N-i+1)} & \text{if } u = (\bar{c},\lambda_-), \\ \frac{1}{(\beta+\lambda)(i+1)(N-i)} & \text{if } u = (c,\lambda_+) \\ \frac{1}{\beta(i+1)(N-i)} & \text{if } u = (c,\lambda_0) \\ \frac{1}{(\beta-\lambda)(i+1)(N-i)} & \text{if } u = (c,\lambda_-) \end{cases}$$

$$m_{i,j} = \begin{cases} \frac{N-i}{N-i+1} & \text{if } j = i \text{ and } u = (\bar{c},\cdot), \\ \frac{N-i-1}{N-i} & \text{if } j = i+1 \text{ and } u = (c,\cdot), \\ 0 & \text{otherwise.} \end{cases}$$

The explanation for these equations is as follows. If $u = (\bar{c},\cdot)$, then the number of infected nodes does not change, and the transition rate is $\beta i(N-i+1)$. The number of infected nodes at the next meeting time is still i with the probability $\frac{N-i}{N-i+1}$, which is the probability that an infected node meets another susceptible node before the destination. Otherwise, if $u = (c,\cdot)$, the number of infected nodes increases to $i+1$ and the transition rate is $\beta(i+1)(N-i)$. The number of infected nodes at the next meeting time is still $i+1$ with probability $\frac{N-i-1}{N-i}$. We consider $i,j \neq 0$ because we assume zero cost in the final absorbing state. Hence, the transitions to this state do not appear in (3.4), and we do not evaluate transition probabilities $m_{i,0}$.

We evaluate the Bellman equation

$$\begin{aligned} & J^*(i,\beta) \\ = \; & \min_u \left\{ h(i,\beta,u) + \sum_{j=1}^{N} m_{i,j}(u) J^*(j,\widetilde{\beta}) \right\}. \end{aligned} \tag{3.5}$$

where $\widetilde{\beta}$ is the value of β at the next meeting instant and

$$h(i,\beta,u) = \widehat{g}(i,\beta,u) + (1+\alpha\beta^2)G(i,\beta,u)$$

In other words, $J^*(i,\beta)$ is the minimum of the following six equations.

$$\begin{aligned} J_{c,\lambda_+}(i,\beta) = \; & \gamma + \frac{(1+\alpha(\beta+\lambda)^2)}{(\beta+\lambda)(i+1)(N-i)} + \\ & \frac{(N-i-1)}{(N-i)} J^*(i+1,\beta+\lambda). \end{aligned} \tag{3.6}$$

$$
\begin{aligned}
J_{c,\lambda_0}(i,\beta) &= \gamma + \frac{(1+\alpha\beta^2)}{\beta(i+1)(N-i)} + \\
&\quad \frac{(N-i-1)}{(N-i)} J^*(i+1,\beta).
\end{aligned} \tag{3.7}
$$

$$
\begin{aligned}
J_{c,\lambda_-}(i,\beta) &= \gamma + \frac{(1+\alpha(\beta-\lambda)^2)}{(\beta-\lambda)(i+1)(N-i)} + \\
&\quad \frac{(N-i-1)}{(N-i)} J^*(i+1,\beta-\lambda).
\end{aligned} \tag{3.8}
$$

$$
\begin{aligned}
J_{\bar{c},\lambda_+}(i,\beta) &= \gamma + \frac{(1+\alpha(\beta+\lambda)^2)}{(\beta+\lambda)i(N-i+1)} + \\
&\quad \frac{(N-i)}{(N-i+1)} J^*(i,\beta+\lambda).
\end{aligned} \tag{3.9}
$$

$$
\begin{aligned}
J_{\bar{c},\lambda_0}(i,\beta) &= \gamma + \frac{(1+\alpha\beta^2)}{\beta i(N-i+1)} + \\
&\quad \frac{(N-i)}{(N-i+1)} J^*(i,\beta).
\end{aligned} \tag{3.10}
$$

$$
\begin{aligned}
J_{\bar{c},\lambda_-}(i,\beta) &= \gamma + \frac{(1+\alpha(\beta-\lambda)^2)}{(\beta-\lambda)i(N-i+1)} + \\
&\quad \frac{(N-i)}{(N-i+1)} J^*(i,\beta-\lambda).
\end{aligned} \tag{3.11}
$$

From these equations, we observe that under the optimal scheme if the best decision with l infected nodes is not to copy, then there will be atmost l copies in the system. The proof of this result is provided in Appendix 5. Consider $\beta = \beta_{l_1}$ at that time. The expected cost from this state is equal to the expected time to meet the destination from this state, i.e., $\frac{1+\alpha\beta_{l_1}^2}{l\beta_{l_1}}$. After this point of time any decision can only be made regarding β. We consider the term $\frac{1+\alpha\beta_{l_1}^2}{l\beta_{l_1}}$, and note that it has a global minimum at $\frac{1}{\sqrt{\alpha}}$. Thus, the decision about

β is made such that it moves closer to $\frac{1}{\sqrt{\alpha}}$. Comparing among (3.6), (3.7), (3.8) and (3.9), (3.10), (3.11), we see the same pattern as above. From an analysis similiar to that done in[60], we see that the maximum number of copies in the network will be equal to h, where

$$h(h-1) < \frac{2}{\sqrt{\alpha}\gamma} \leq h(h+1). \tag{3.12}$$

The optimal algorithm thus derived is as follows. Evidently, the algorithm depends on the step size λ. We prove the dependence rigorously in Appendix 6.

1. Calculate h from the (4.1).
2. If the number of infected nodes in the system is less than h, then the packet can be copied further. If it is equal to h, then the packet should not be copied any further.
3. The decision about β depends upon the step size λ.
4. If $\beta - \lambda \geq \frac{1}{\sqrt{\alpha}}$, then λ_- is the best choice.
5. If $\beta + \lambda \leq \frac{1}{\sqrt{\alpha}}$, then λ_+ is the best choice.
6. If $\beta = \frac{1}{\sqrt{\alpha}}$, then λ_0 is the best choice.
7. If $\beta > \frac{1}{\sqrt{\alpha}}$ and $\beta - \lambda < \frac{1}{\sqrt{\alpha}}$, then
   (a) If $\beta - \lambda \leq \frac{1}{\alpha\beta}$, then λ_0 is the best choice.
   (b) If $\beta - \lambda > \frac{1}{\alpha\beta}$, then λ_- is the best choice.
8. If $\beta < \frac{1}{\sqrt{\alpha}}$ and $\beta + \lambda > \frac{1}{\sqrt{\alpha}}$, then
   (a) If $\beta + \lambda \leq \frac{1}{\alpha\beta}$, then λ_0 is the best choice.
   (b) If $\beta + \lambda > \frac{1}{\alpha\beta}$, then λ_+ is the best choice.

Chapter 4

Results

This chapter deals with the results derived in the previous chapter. The analytical solution has been incorporated in simulations executed on Matlab. Section 4.1 describes our setting for simulations. Sections 4.2, 4.3 and 4.4 deal with the analysis of individual costs. Section 4.5 describes the cumulative cost comparison. In the end, we present a brief analysis of the solution and conclusion of our work.

4.1 Simulation Setting

We carried the simulations in Matlab version 6.0.0.88 R12. For the sake of comparison, we considered two other forwarding protocols: Spray and Wait forwarding and Probabilistic forwarding.

In Spray and Wait forwarding scheme[65], the power-delay trade-off issue is handled by controlling the total number of packet copied in the network. The parameter of the protocol is K, which is the maximum allowable number of copies. For every message originating at the source node, K message copies are initially spread, either by the source itself or by other infected nodes. This is called the Spray phase. Then comes the Wait phase in which, the network has to wait for one of these K infected nodes to contact the destination.

In Probabilistic forwarding scheme, the power-delay trade-off issue is handled by controlling

γ	α	$\beta(0)$
0.5	100	0.01
0.5	100	0.1
0.5	100	0.2
0.5	500	0.01
0.5	500	0.1
0.5	500	0.2
4	100	0.01
4	100	0.1
4	100	0.2
4	500	0.01
4	500	0.1
4	500	0.2

Table 4.1: *Test parameter combinations for Probabilistic and Spray and Wait forwarding strategies.*

a probability parameter p. Each time a node receives a packet either from the source or from some other infected node, it is entitled to forward the packet further by a probability p. Eventually one of these infected nodes meets the destination and the packet is forwarded to the destination. Hence, it this protocol there is no Wait phase.

In our setting, we consider $N + 1 = 101$ nodes with a single source-destination pair. Here N includes the source node and the intermediate nodes. The additional 1 denotes the destination. For the two schemes mentioned above, we consider three values of β, $\beta_{01} = 0.01$, $\beta_{02} = 0.1$, $\beta_{03} = 0.2$. These schemes consider a constant mobility pattern throughout a run. For the cost of packet copies, we compare the effect of two values of γ, $\gamma_1 = 0.5$, $\gamma_2 = 4$. For the mobility cost, we consider $\alpha_1 = 100$, $\alpha_2 = 500$. According to the analytical results, the optimal values of β should be 0.1 and 0.0447 respectively. For the optimal policy, we consider two step sizes to vary β, $\lambda_1 = 0.001$ and $\lambda_2 = 0.009$.

The various parameters for evaluation have been summarized in Table 4.1 and Table 4.2.

γ	α	$\beta(0)$	λ
0.5	100	0.01	0.001
0.5	100	0.01	0.009
0.5	100	0.1	0.001
0.5	100	0.2	0.001
0.5	100	0.2	0.009
0.5	500	0.01	0.001
0.5	500	0.01	0.009
0.5	500	0.1	0.001
0.5	500	0.2	0.001
0.5	500	0.2	0.009
4	100	0.01	0.001
4	100	0.01	0.009
4	100	0.1	0.001
4	100	0.2	0.001
4	100	0.2	0.009
4	500	0.01	0.001
4	500	0.01	0.009
4	500	0.1	0.001
4	500	0.2	0.001
4	500	0.2	0.009

Table 4.2: *Test parameter combinations for Optimal forwarding strategy.*

4.2 The cost of Delay

The comparison of the delay cost derived under the optimal scheme has been compared with that of probabilistic forwarding in Fig.4.1, Fig.4.2, Fig.4.3 and Fig.4.4. Similarly, the comparison of the delay cost derived under the optimal scheme has been compared with that of Spray and Wait forwarding in Fig.4.1, Fig.4.2, Fig.4.3 and Fig.4.4. In both, probabilistic forwarding and Spray and Wait forwarding, the delivery delay decreases as β_0 increases. Also, for lower values of probability and spray value, there are a fewer number of copies in the network. Hence there is a significant delay at these values.

For the optimal scheme, Fig.4.1, Fig.4.2, Fig.4.5, Fig.4.6 refer to cases where $\alpha = 100$. For these settings, the value to β converges to 0.1. Thus, for cases with $\beta_0 = 0.01$, the mobility parameter is lesser than optimal and increases gradually. Similarly, for cases with $\beta_0 = 0.2$, the mobility parameter is greater and decreases gradually. Hence the delivery delay for cases starting with higher mobility is greater.

Comparing the optimal scheme with other two schemes, we find that delivery delay for lower values of probability and spray is greater than that of the optimal scheme. As probability and spray values increase, the delivery delay of these two schemes becomes lesser than that of the optimal scheme. The reason for this is there is a hard limit to the number of packet copies in the optimal scheme. Larger values of probability and spray denote more number of copies than the limit of the optimal scheme. As a result the delay in optimal scheme is greater.

This section compares the delivery delay from the source to the destination in the three forwarding schemes.

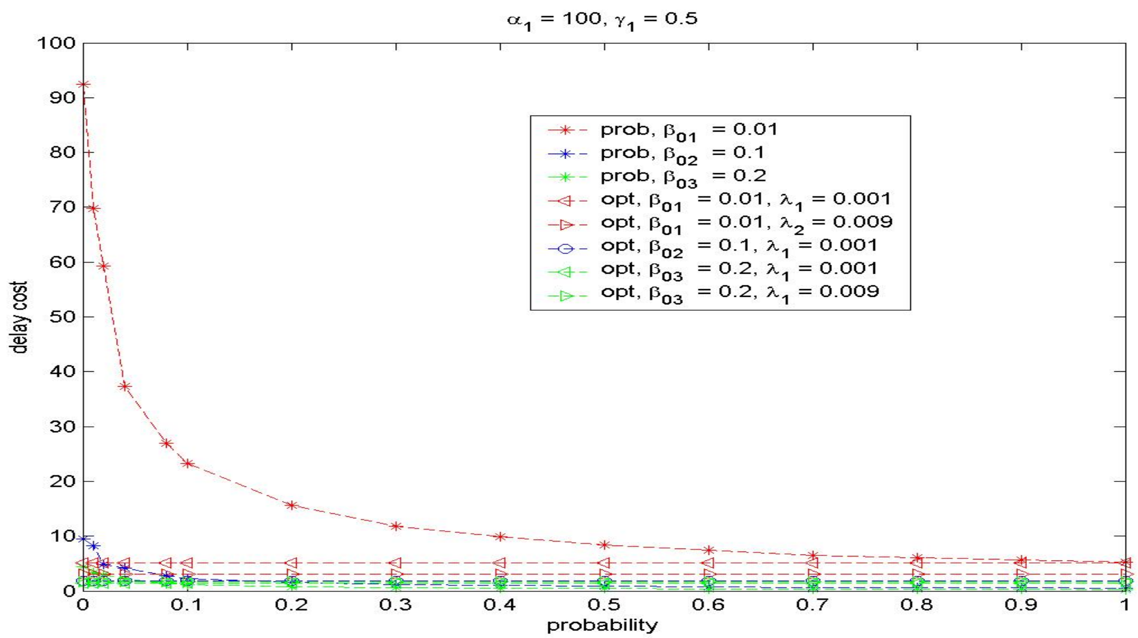

Figure 4.1: *Delay cost: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

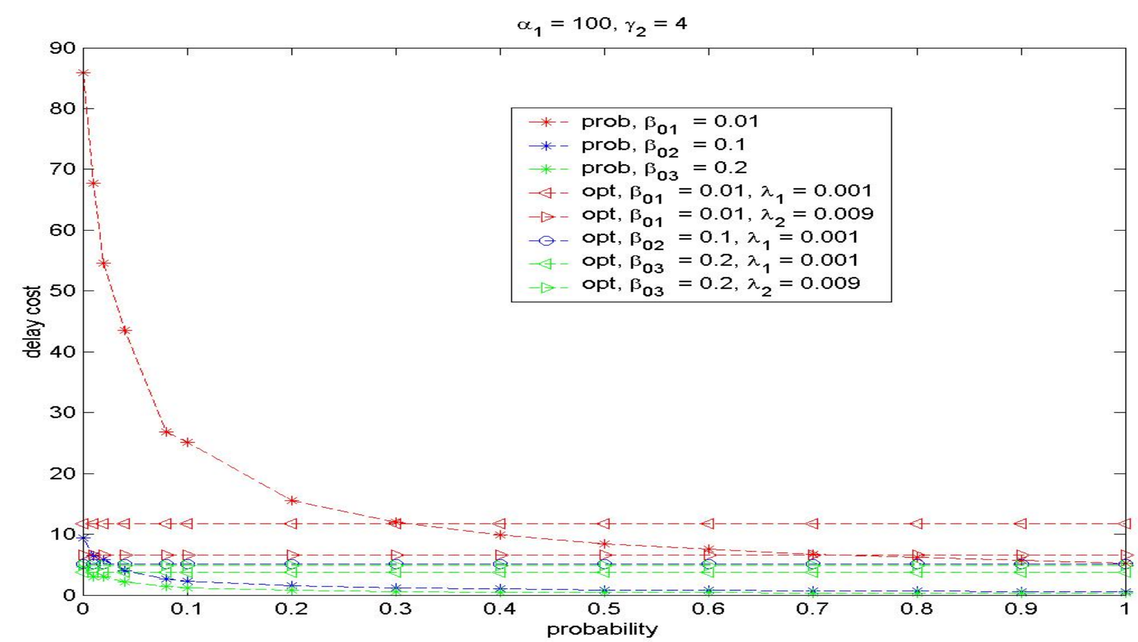

Figure 4.2: *Delay cost: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

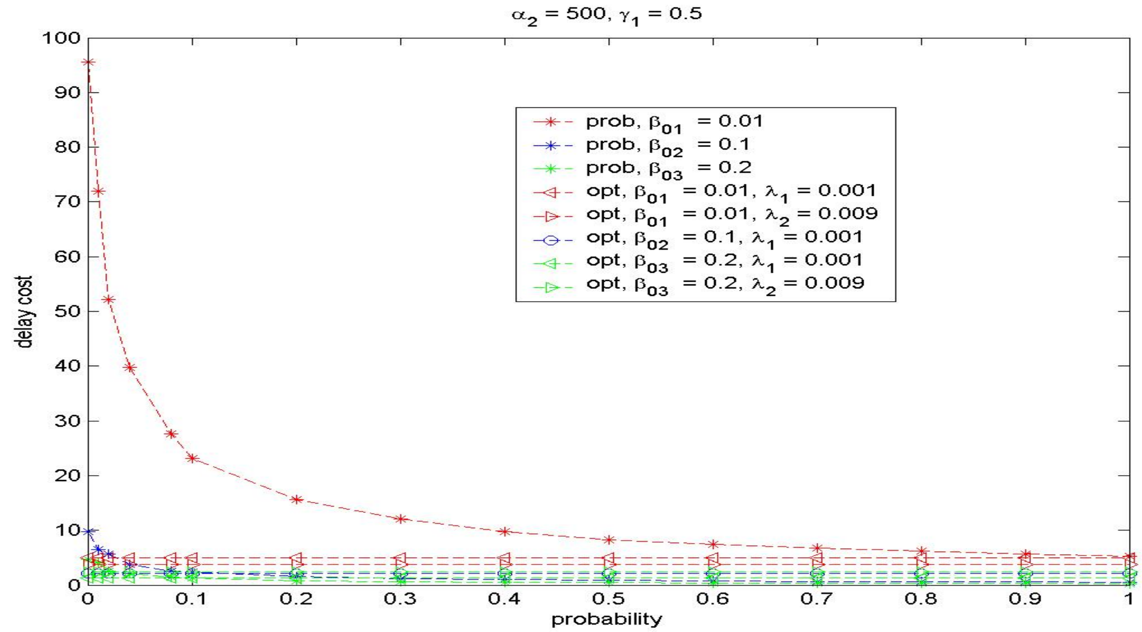

Figure 4.3: *Delay cost: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

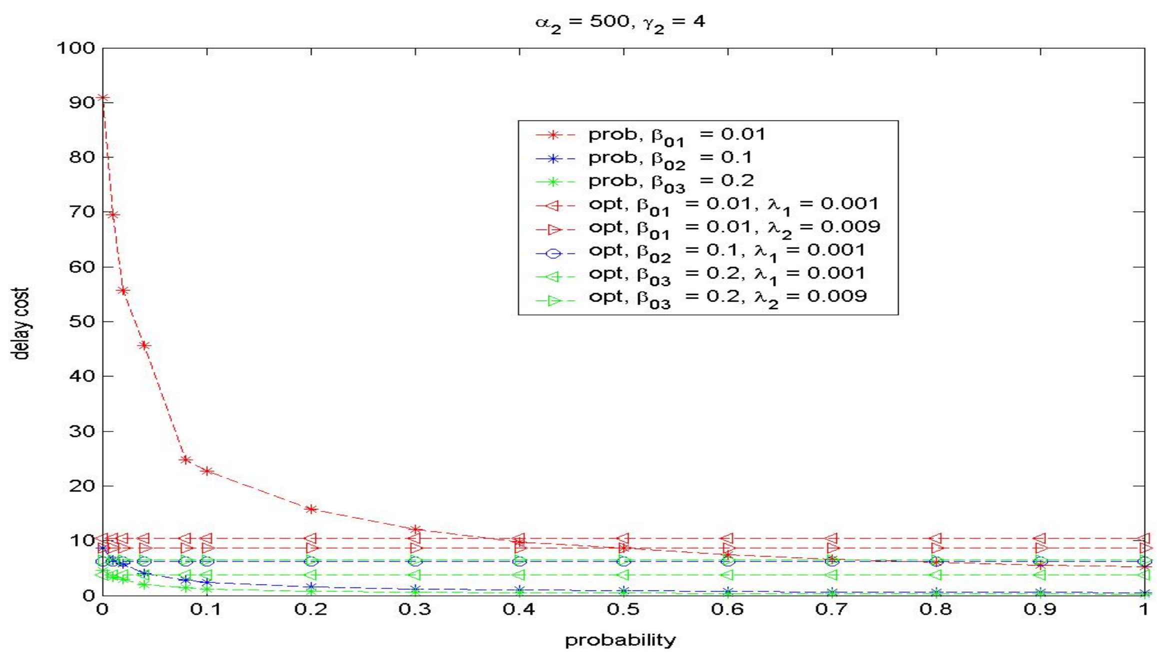

Figure 4.4: *Delay cost: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

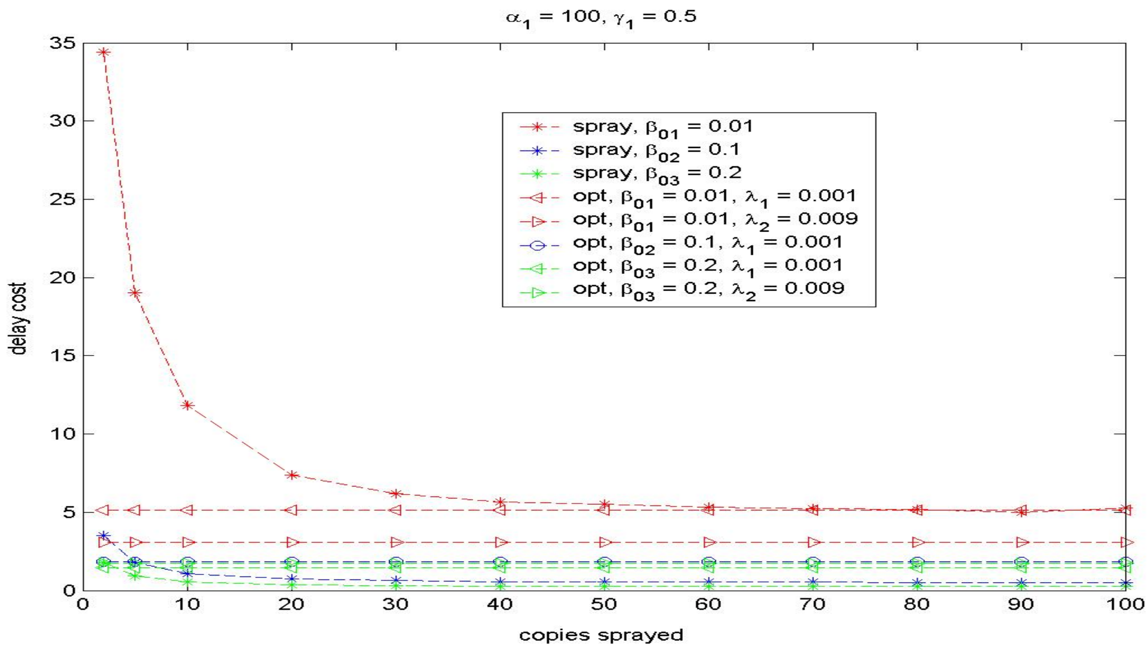

Figure 4.5: *Delay cost: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

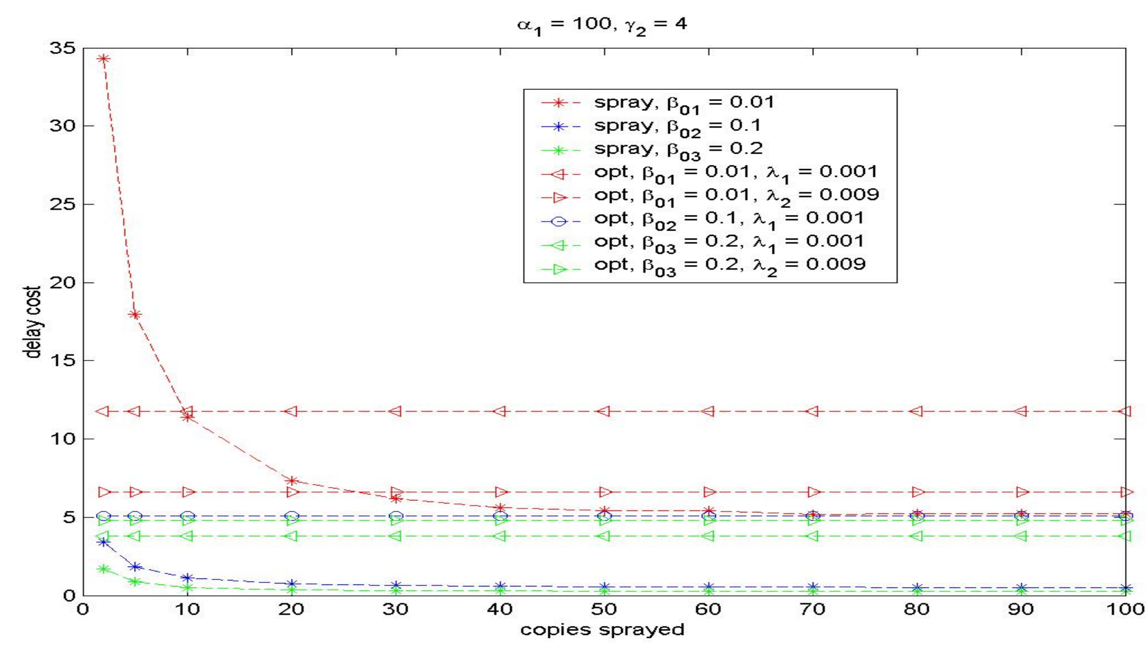

Figure 4.6: *Delay cost: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

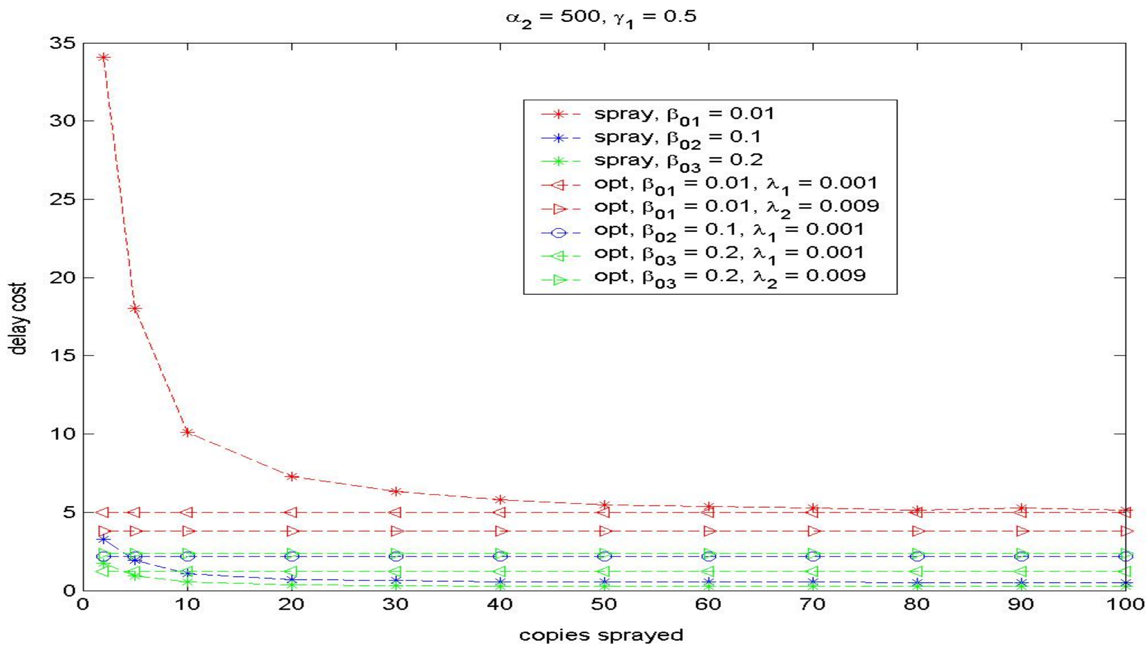

Figure 4.7: *Delay cost: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

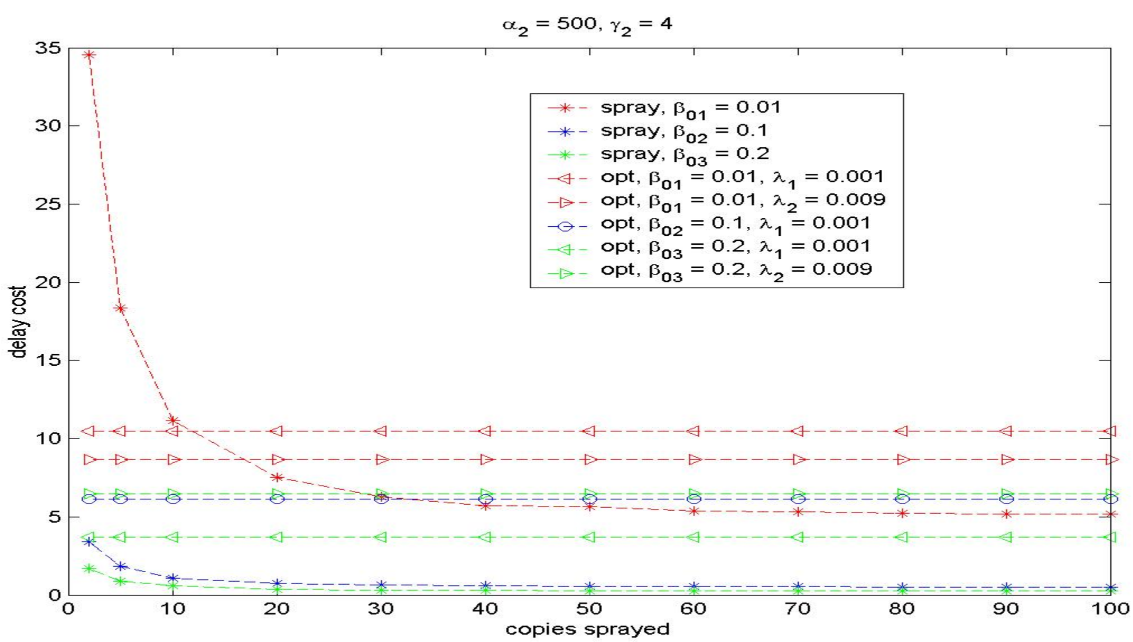

Figure 4.8: *Delay cost: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

4.3 Cost of Packet copies

The comparison of the cost of packet copies derived under the optimal scheme has been compared with that of probabilistic forwarding in Fig.4.9, Fig.4.10, Fig.4.11 and Fig.4.12. Similarly, the comparison of the cost of packet copies derived under the optimal scheme has been compared with that of Spray and Wait forwarding in Fig.4.13, Fig.4.14, Fig.4.15 and Fig.4.16. We see that as the values of probability and spray value, increase, there are more number of copies in the network. Hence the cost of packet copies increases.

For the optimal scheme, the cost of packet copies is larger for larger values of γ. The limit of packet copies in the system is governed by the equation

$$h(h-1) < \frac{2}{\sqrt{\alpha}\gamma} \leq h(h+1). \tag{4.1}$$

As β converges to the optimal value the limit of the packet copies decreases according to the equation4.1. The limit of packet copies at the optimal value of beta minimum.

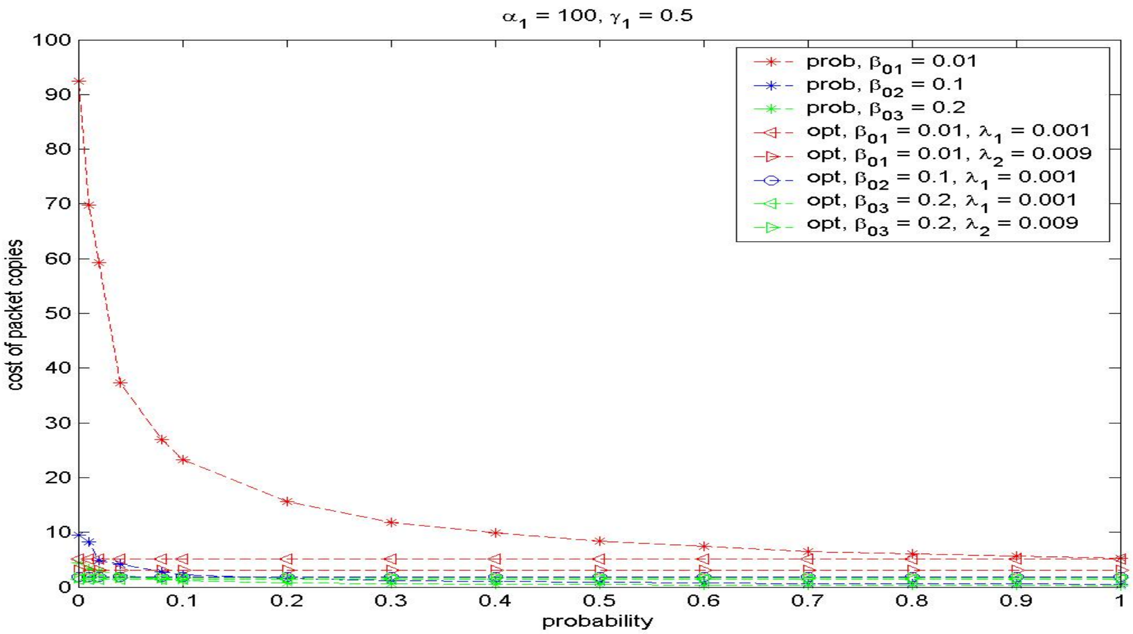

Figure 4.9: *Cost of packet copies: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

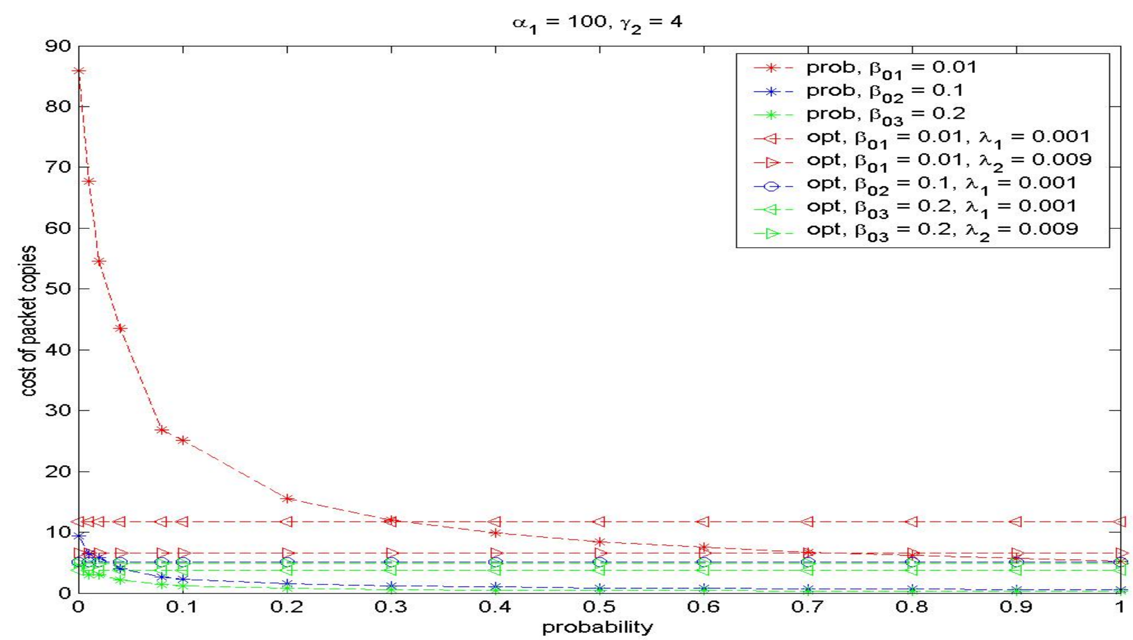

Figure 4.10: *Cost of packet copies: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

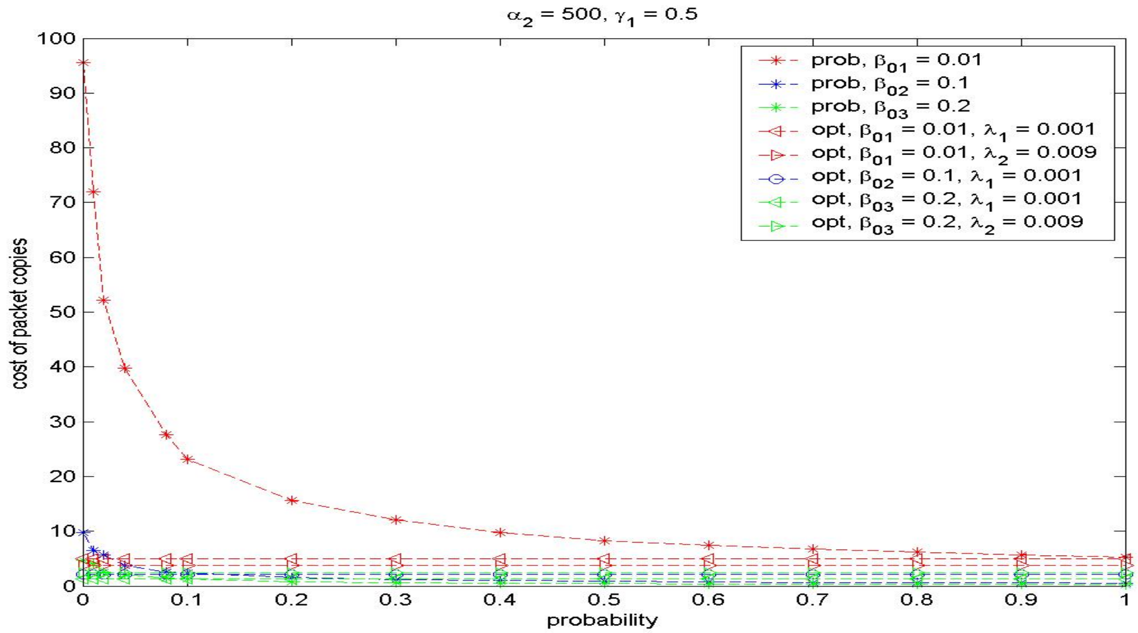

Figure 4.11: *Cost of packet copies: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

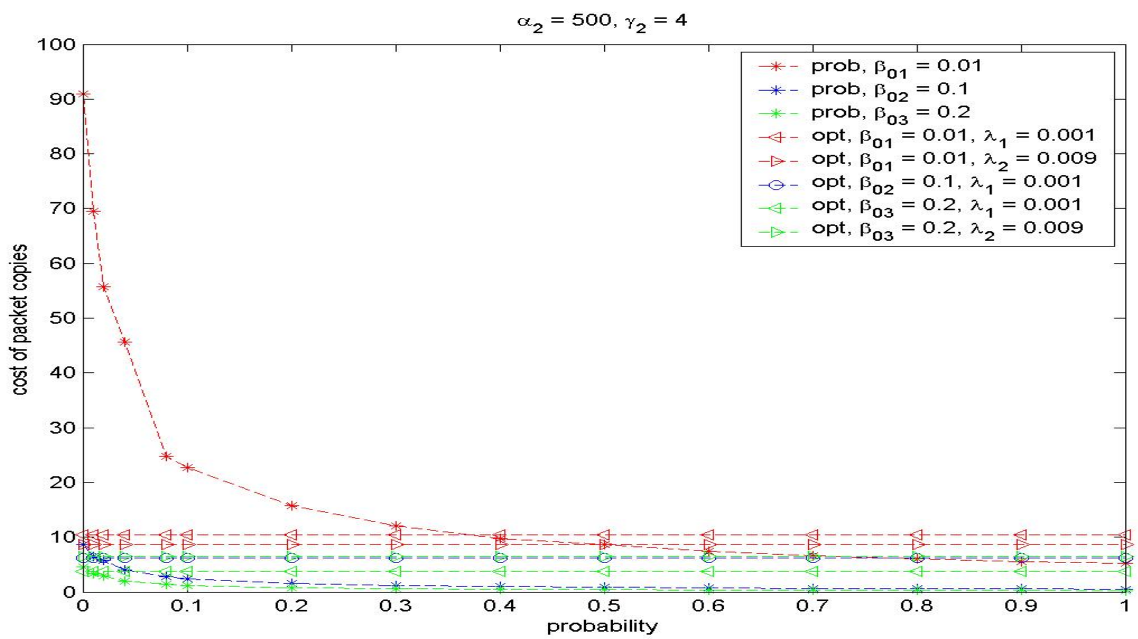

Figure 4.12: *Cost of packet copies: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

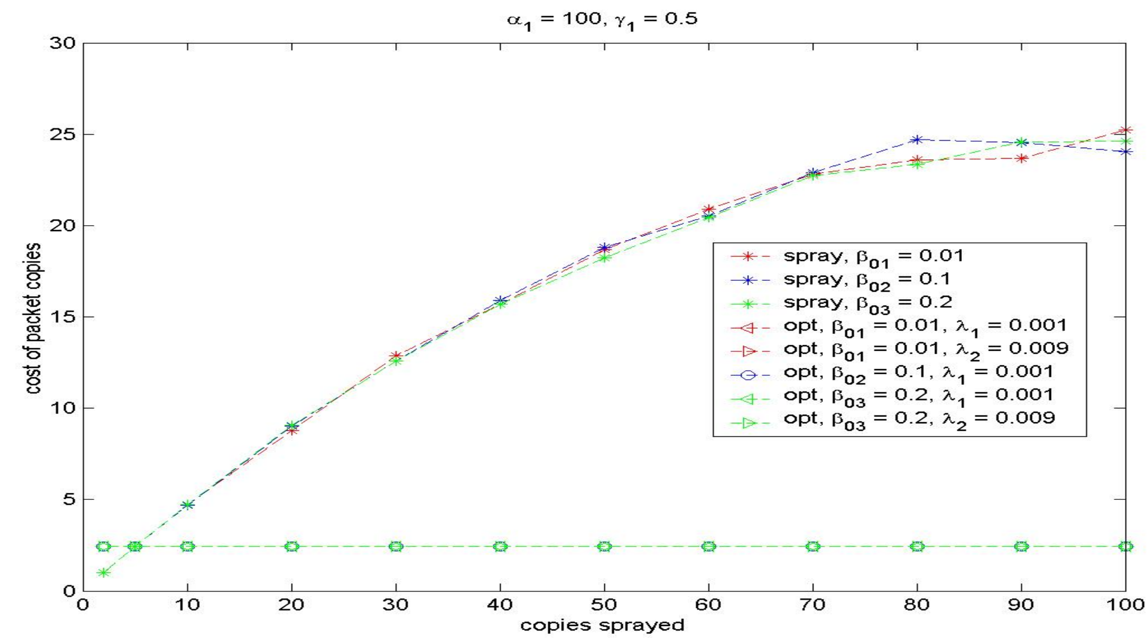

Figure 4.13: *Cost of packet copies: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

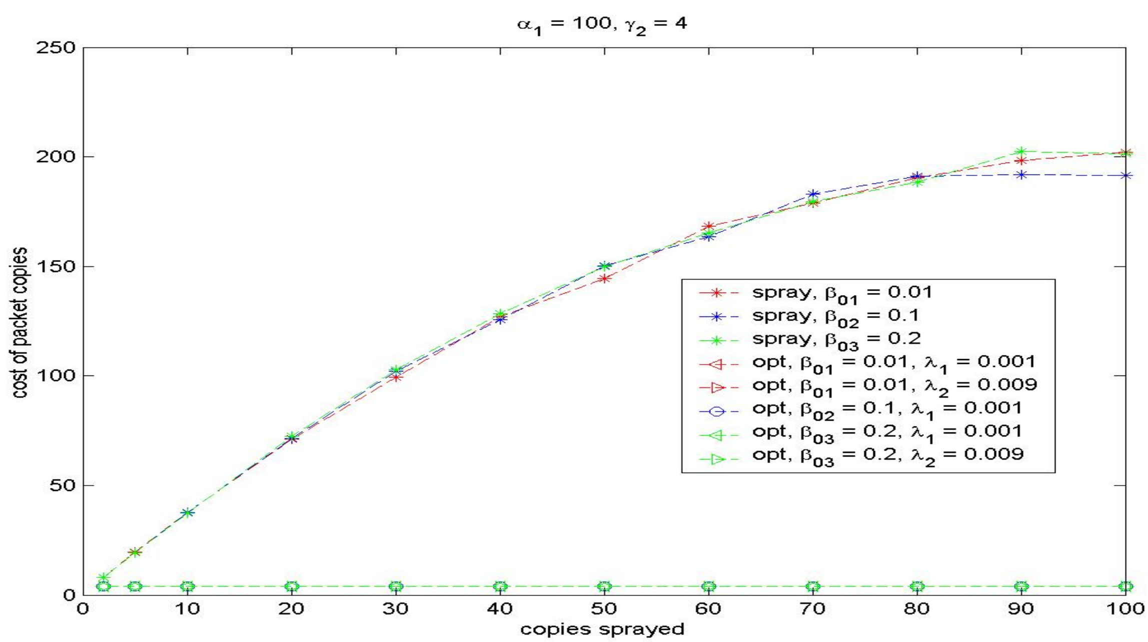

Figure 4.14: *Cost of packet copies: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

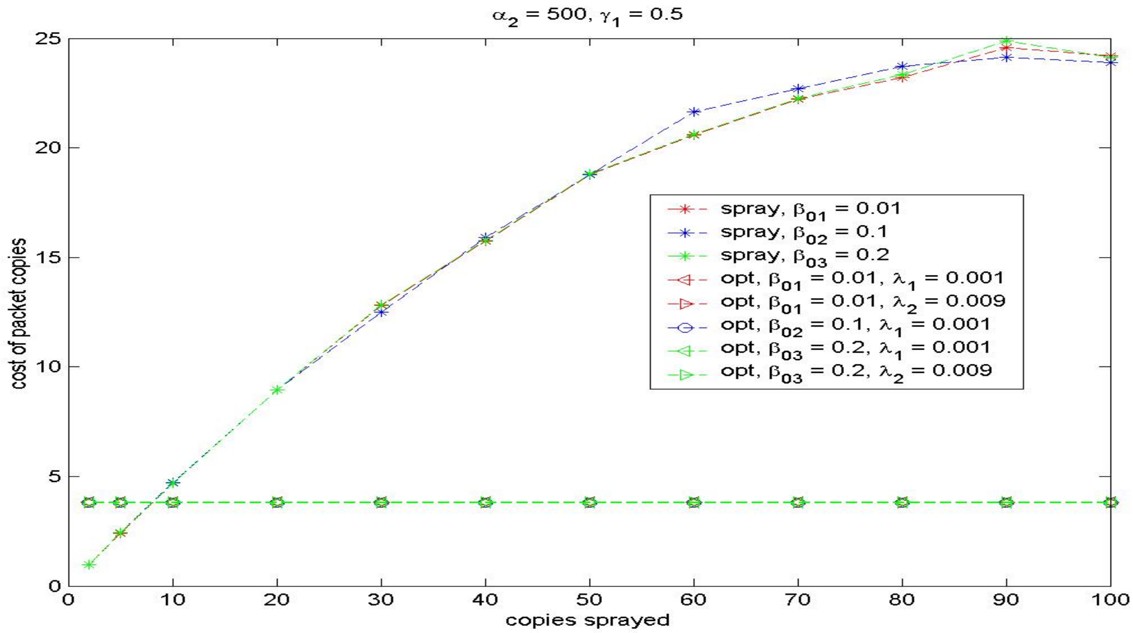

Figure 4.15: *Cost of packet copies: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

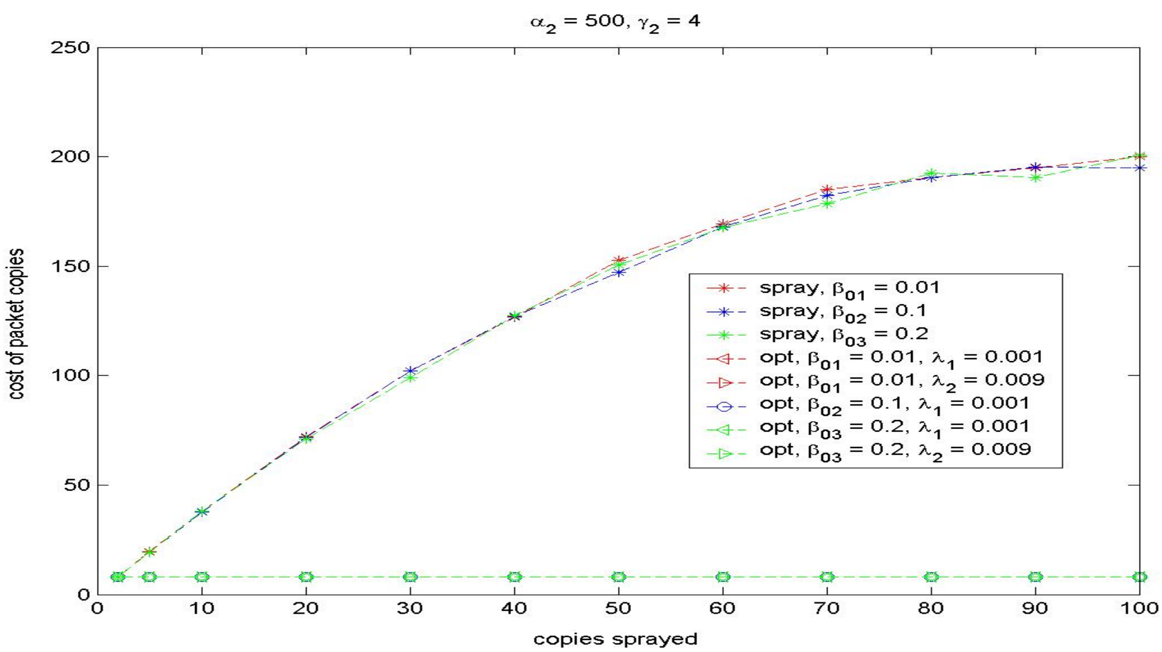

Figure 4.16: *Cost of packet copies: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

4.4 Cost of mobility

The comparison of the cost of mobility derived under the optimal scheme has been compared with that of probabilistic forwarding in Fig.4.17, Fig.4.18, Fig.4.19 and Fig.4.20. Similarly, the comparison of the cost of mobility derived under the optimal scheme has been compared with that of Spray and Wait forwarding in Fig.4.21, Fig.4.22, Fig.4.23 and Fig.4.24.

For both probabilistic forwarding and Spray and Wait forwarding, we see that starting with higher mobility renders more mobility cost. For the optimal scheme the cost of mobility is also governed by the value to which β converges. For cases with $\alpha = 100$, that value is 0.1, whereas for $\alpha = 500$, it is 0.0447. Thus, for cases where the step size λ is more suitable to reach these values, render lower cost of mobility.

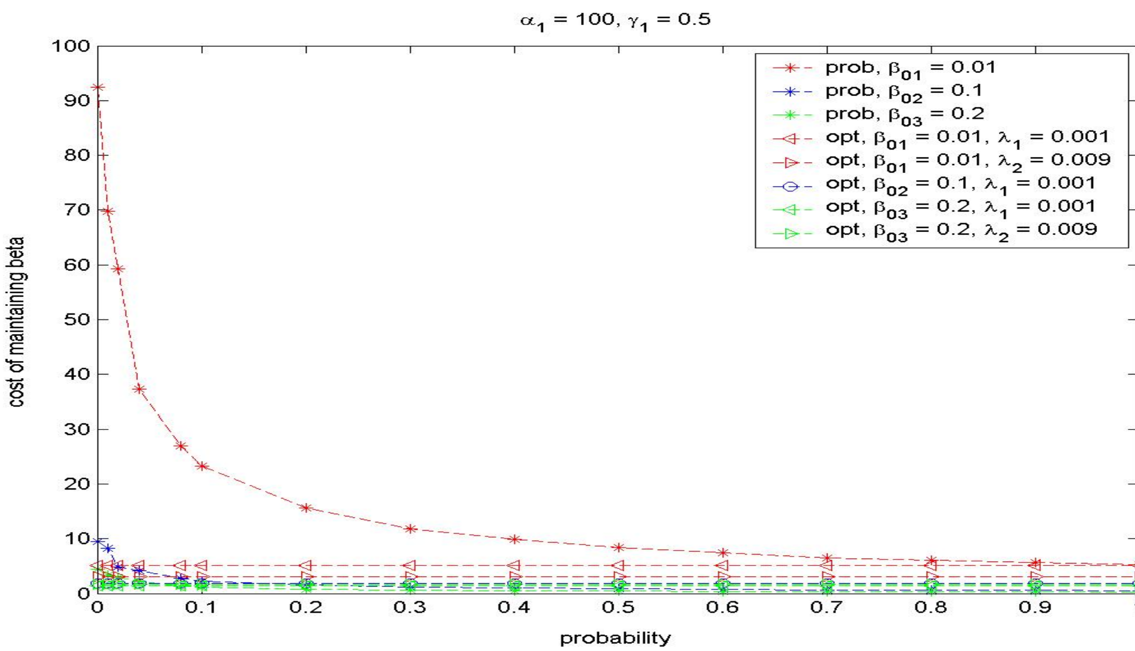

Figure 4.17: *Cost of mobility: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

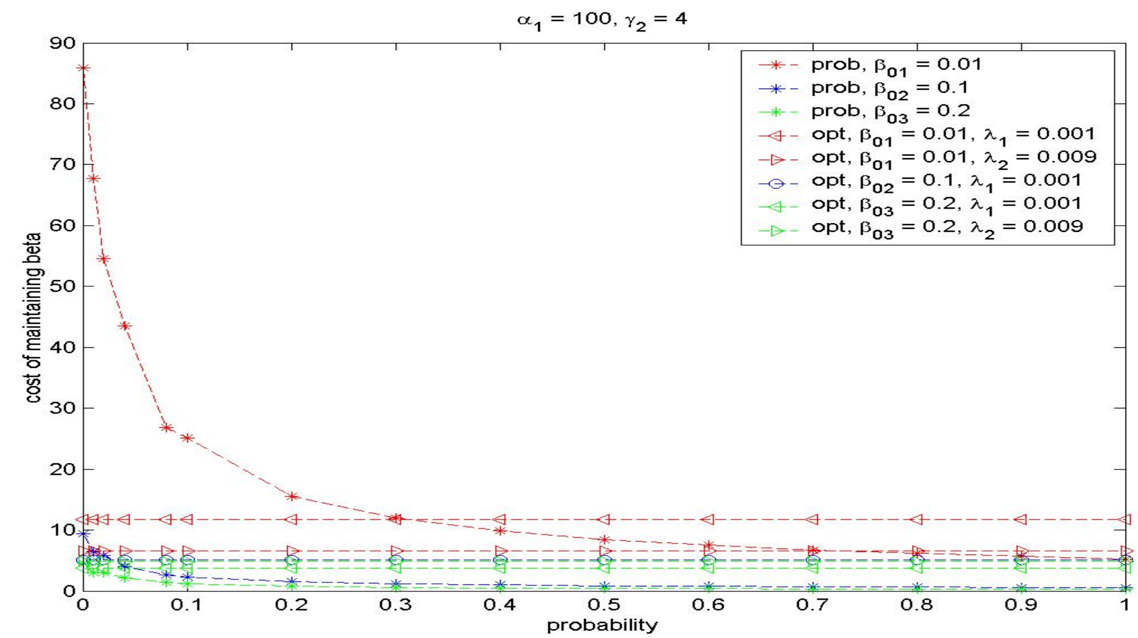

Figure 4.18: *Cost of mobility: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

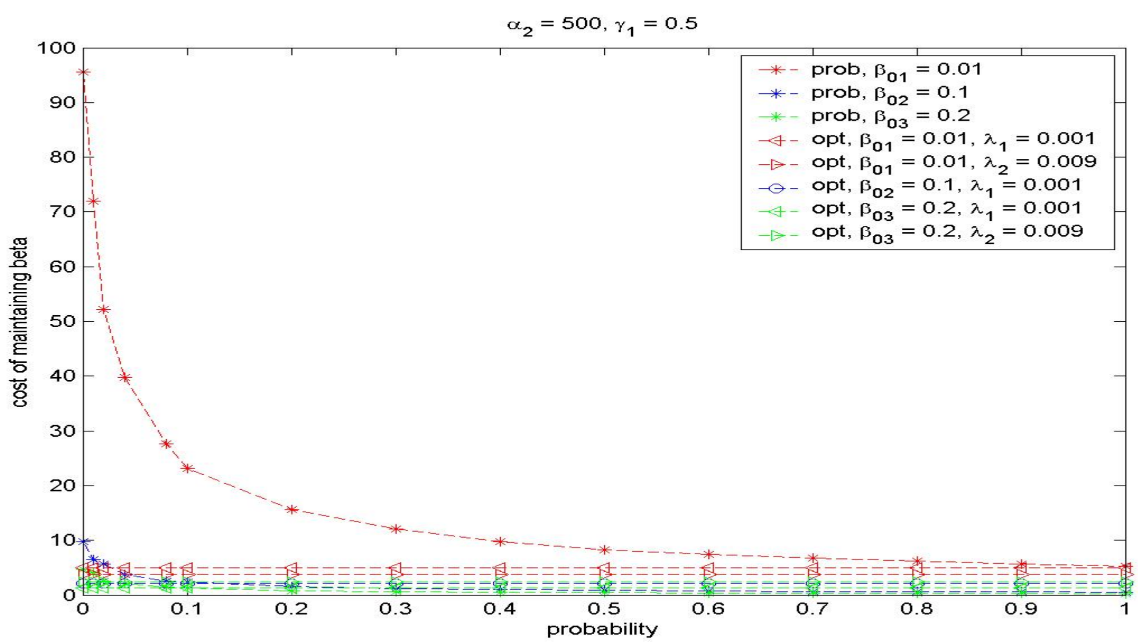

Figure 4.19: *Cost of mobility: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

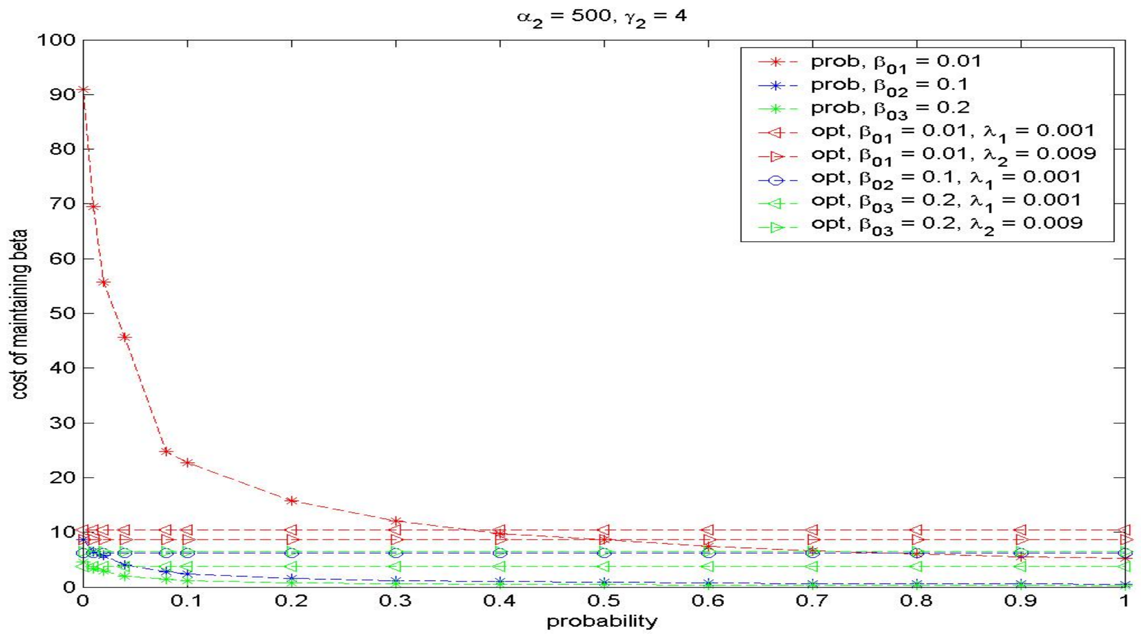

Figure 4.20: *Cost of mobility: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

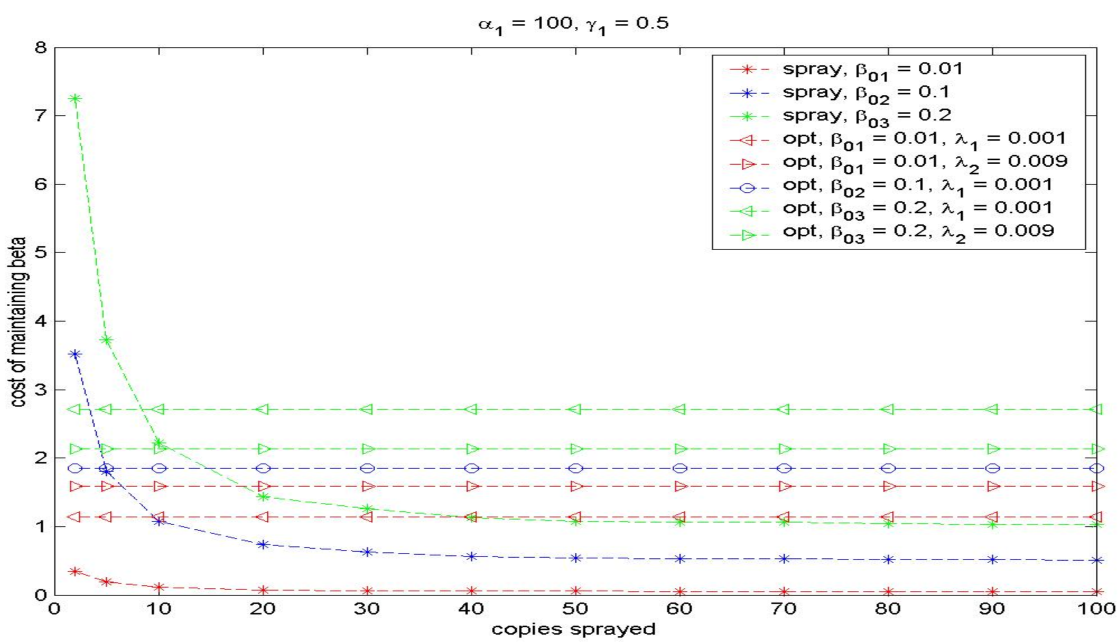

Figure 4.21: *Cost of mobility: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

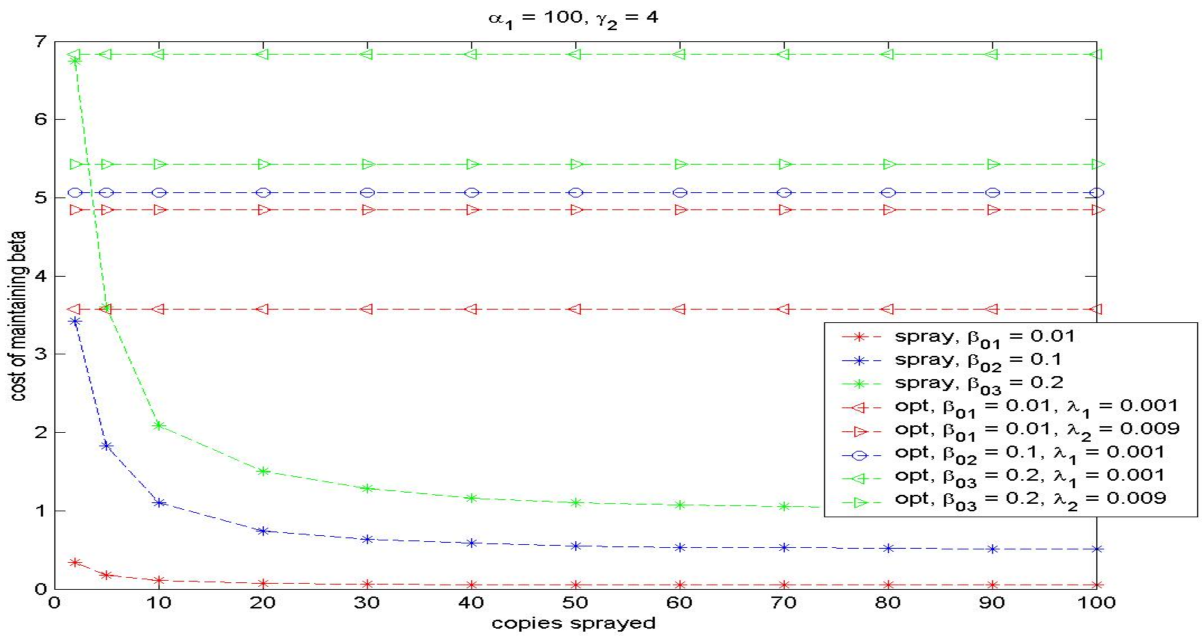

Figure 4.22: *Cost of mobility: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

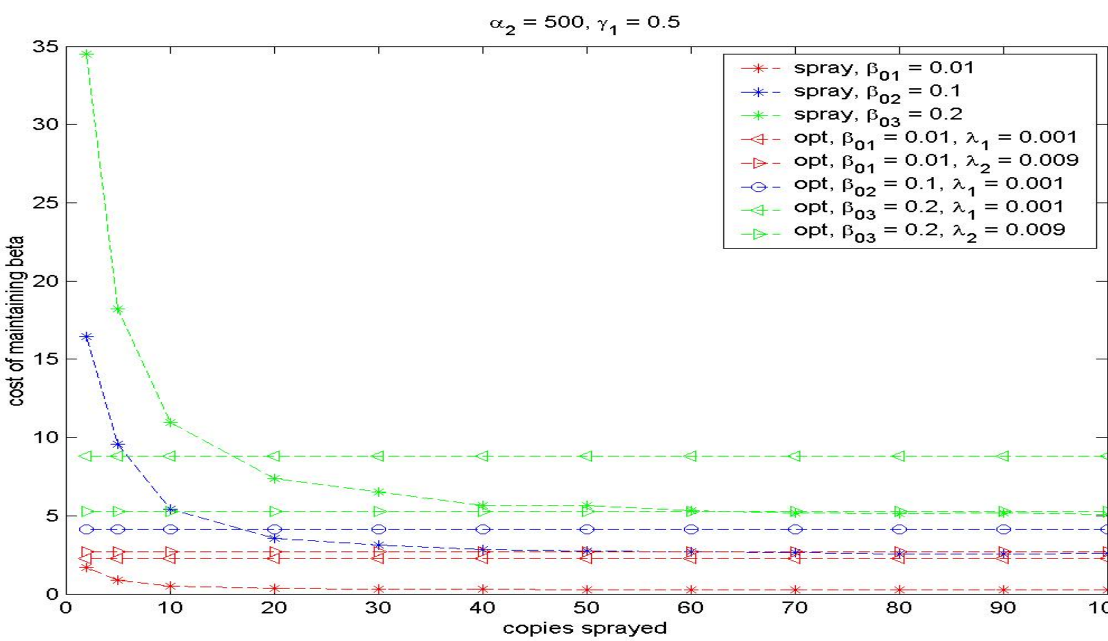

Figure 4.23: *Cost of mobility: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

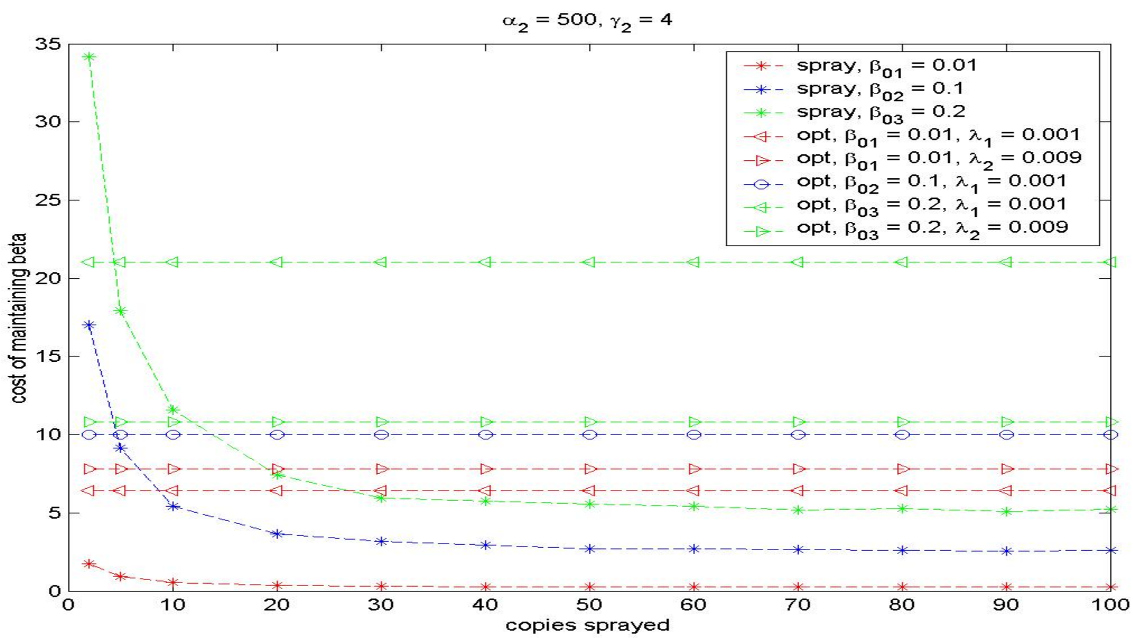

Figure 4.24: *Cost of mobility: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

4.5 Total Cost

The comparison of the total cost derived under the optimal scheme has been compared with that of probabilistic forwarding in Fig.4.25, Fig.4.26, Fig.4.27 and Fig.4.28. Similarly, the comparison of the total cost derived under the optimal scheme has been compared with that of Spray and Wait forwarding in Fig.4.29, Fig.4.30, Fig.4.31 and Fig.4.32.

It is evident from these figures that that for smaller values of probability and spray, the high cost is due to the delivery delay. Similarly, assuming a significant cost for packet copies ($\gamma_2 = 4$), the higher cost for larger values of probability and number of copies is due to the number of infected nodes. The curves for β_{01} and β_{02} are associated with a higher cost than that of β_{02}, which is the optimal value for $\alpha = 100$.

In all cases however, the optimal algorithm incurs the least cost. It considers β_{01}, β_{02}, β_{03} as starting values of the mobility parameter. The plots for optimal algorithm starting with β_{01} and β_{03} show higher values than the ones starting with β_{02}.

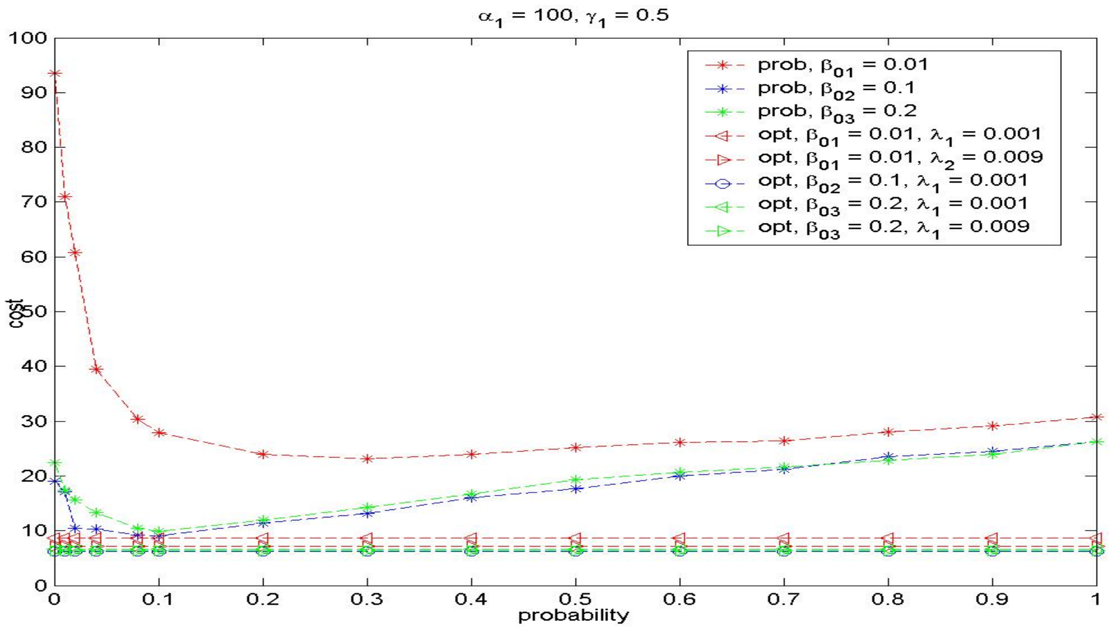

Figure 4.25: *Total cost: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

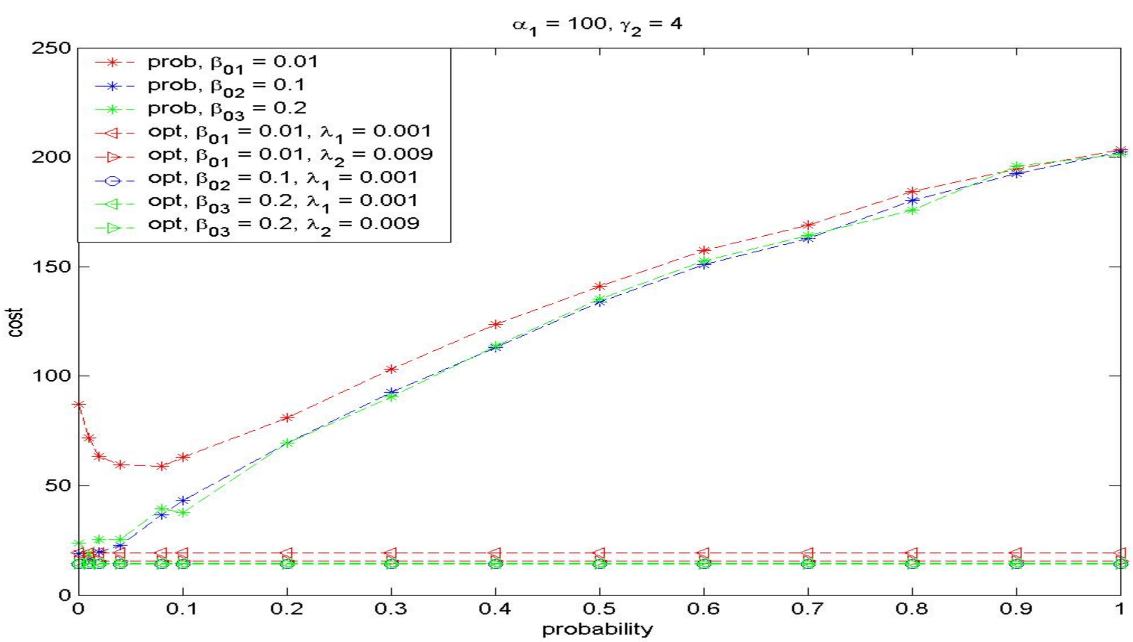

Figure 4.26: *Total cost: Comparison with probabilistic forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

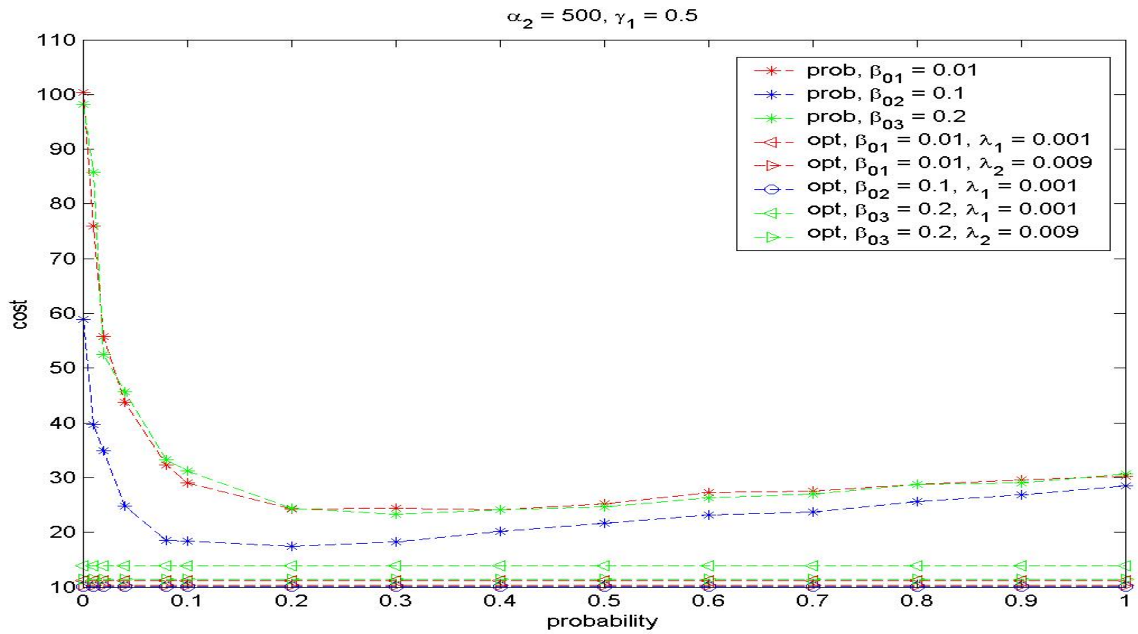

Figure 4.27: *Total cost: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

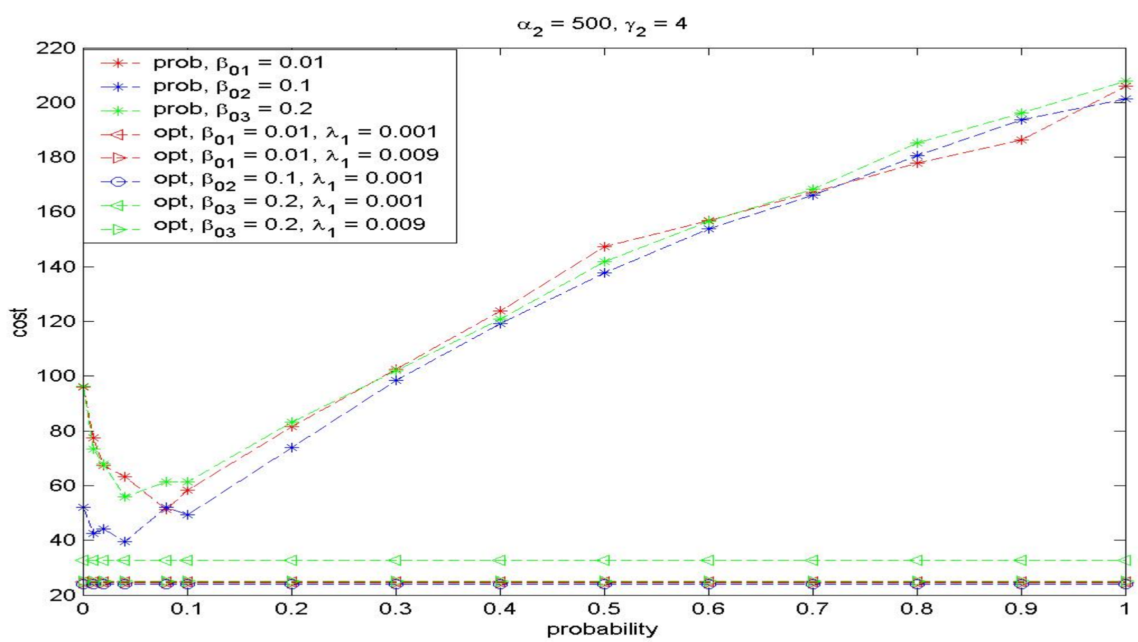

Figure 4.28: *Total cost: Comparison with probabilistic forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

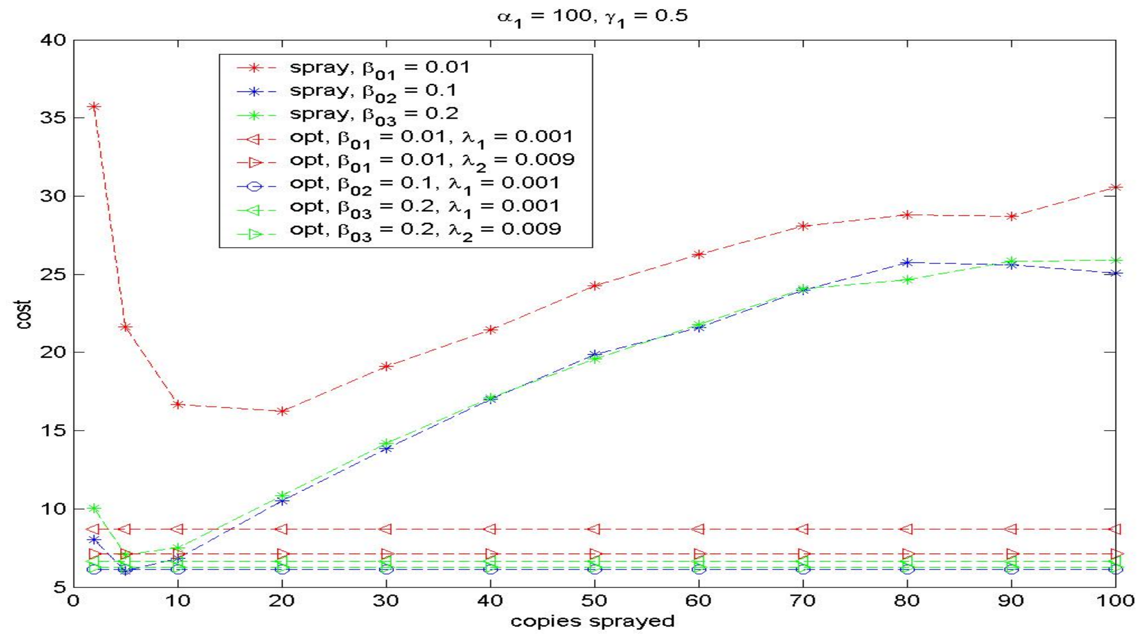

Figure 4.29: *Total cost: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_1 = 0.5$

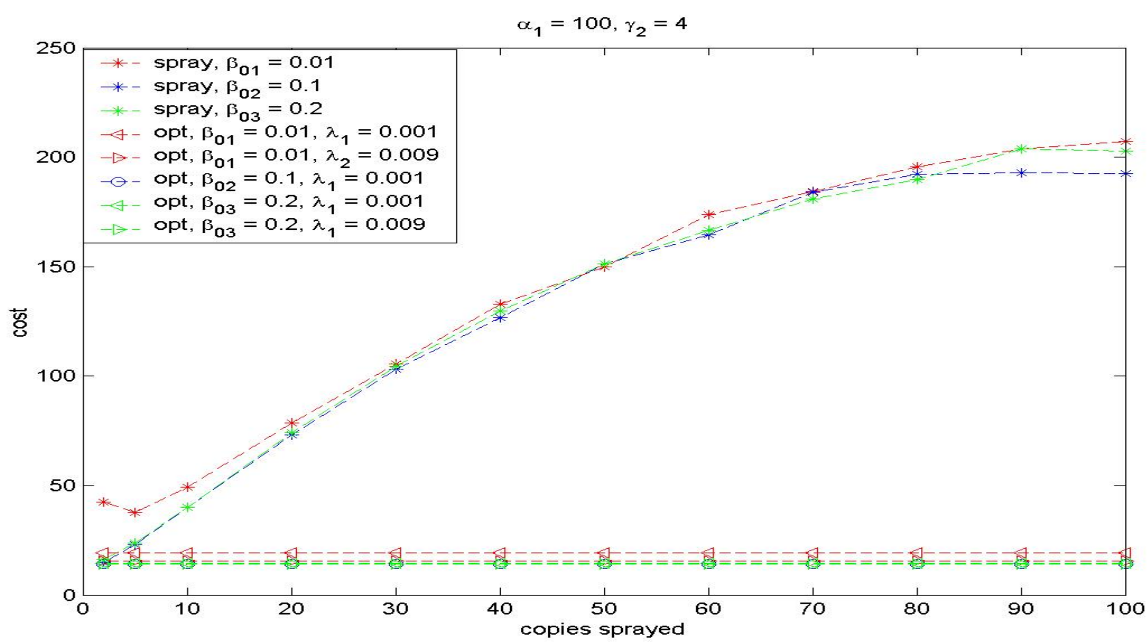

Figure 4.30: *Total cost: Comparison with Spray and Wait forwarding,* $\alpha_1 = 100, \gamma_2 = 4$

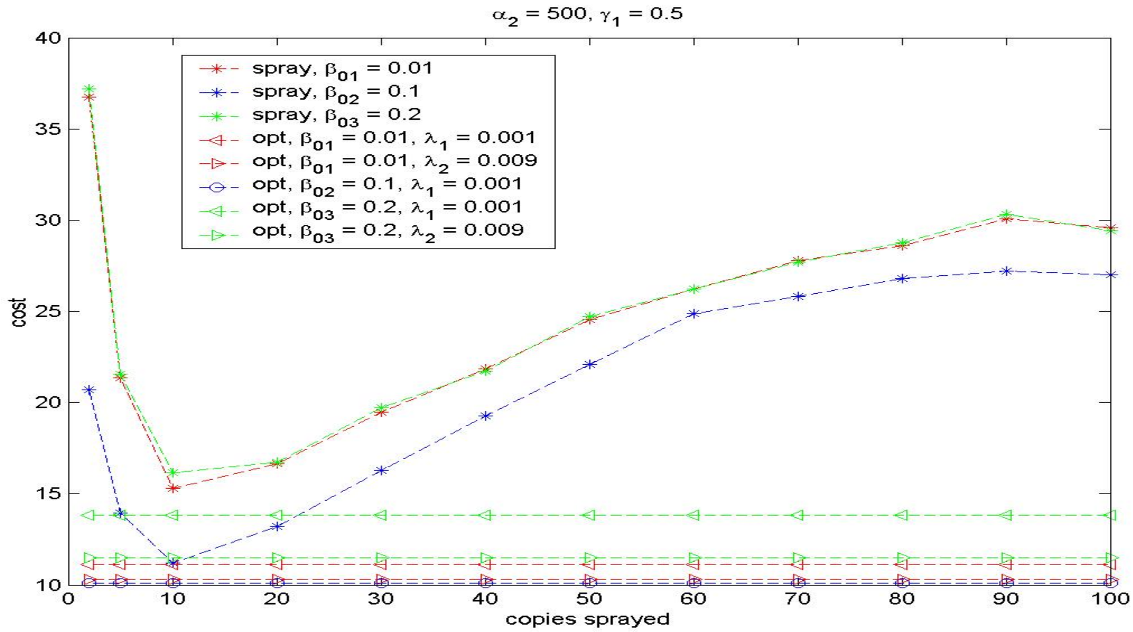

Figure 4.31: *Total cost: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_1 = 0.5$

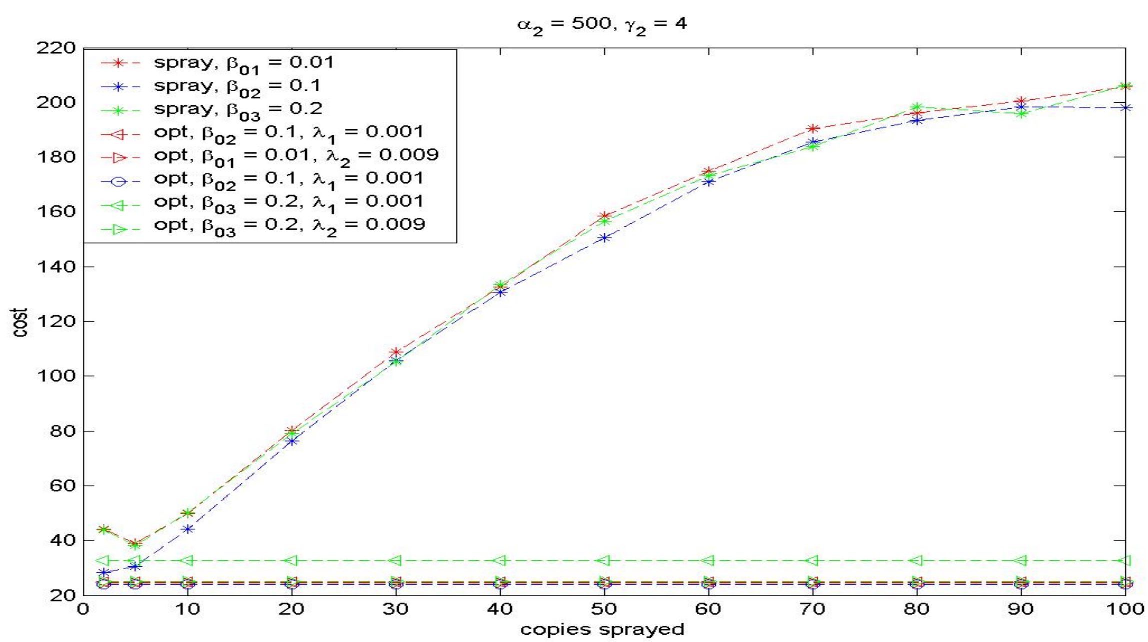

Figure 4.32: *Total cost: Comparison with Spray and Wait forwarding,* $\alpha_2 = 500, \gamma_2 = 4$

4.6 Conclusions

This work aims to emphasize on the effect of mobility on the power cost in delay tolerant networks. Higher mobility among nodes contributes to lower delivery delay, but at the same time adds to the power consumption. Thus, this work highlights the important role that mobility plays in the issue of trade-off between delivery delay and resource consumption in epidemic networks. For the sake of simplicity, we considered an ideal scenario wherein all nodes have perfect state information at all times. Under this assumption, we have been able to derive an optimal mobility pattern along with an optimal forwarding policy. This confirmed that the mobility parameter is optimal at a certain value which depends on the cost of mobility. Starting from any other value, the trajectory of β tries to reach this optimal value. Two heuristics have been compared with the optimal algorithm: Probabilistic forwarding and Spray and Wait.The preliminary results suggest that the packet forwarding limit can be studied as a configuration criterion for Spray and Wait forwarding. However, mobility is an important factor for any forwarding scheme for further optimization. Current mobility models consider that all nodes have the same properties (transmission radius, etc.). The next steps in this area are to consider mobility models for heterogeneous nodes, and work with imperfect state assumptions.

Bibliography

[1] Merriam-webster dictionary.

[2] J. King, Computerworld (2005).

[3] J. Caruso, Network World (2008).

[4] M. Hamdi, N. Boudriga, and M. S. Obaidat, Int. J. Commun. Syst. **21**, 277 (2008).

[5] B. Brewin, Computerworld (2003).

[6] B. Feder, The New York Times (2004).

[7] C. Liu and J. Kaiser, A survey of mobile ad hoc network routing protocols, 2005.

[8] W. O. Kermack and A. G. McKendrick, Royal Society of London Proceedings Series A **115**, 700 (1927).

[9] A. Vahdat and D. Becker, Epidemic routing for partially connected ad hoc networks, 2000.

[10] T. Small and Z. Haas, The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way), in *MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 233–244, New York, NY, USA, 2003, ACM.

[11] Z. J. Haas and T. Small, IEEE/ACM Trans. Netw. **14**, 27 (2006).

[12] A. Lindgren, A. Doria, and O. Schelén, SIGMOBILE Mob. Comput. Commun. Rev. **7**, 19 (2003).

[13] R. Groenevelt, P. Nain, and G. Koole, Perform. Eval. **62**, 210 (2005).

[14] G. Sharma, R. Mazumdar, and N. B. Shroff, IEEE/ACM Trans. Netw. **15**, 981 (2007).

[15] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, Spray and wait: an efficient routing scheme for intermittently connected mobile networks, in *WDTN '05: Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 252–259, New York, NY, USA, 2005, ACM Press.

[16] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, Comput. Netw. **51**, 2867 (2007).

[17] P. Mundur, M. Seligman, and J. N. Lee, Immunity-based epidemic routing in intermittent networks, in *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on*, pages 609–611, 2008.

[18] P. Mundur, M. Seligman, and J. N. Lee, Immunity-based epidemic routing in intermittent networks., in *SECON*, pages 609–611, IEEE, 2008.

[19] J. P. Tower and T. D. C. Little, A proposed scheme for epidemic routing with active curing for opportunistic networks, in *AINAW '08: Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops*, pages 1696–1701, Washington, DC, USA, 2008, IEEE Computer Society.

[20] T. Small and Z. J. Haas, Resource and performance tradeoffs in delay-tolerant wireless networks, in *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 260–267, New York, NY, USA, 2005, ACM.

[21] C. Boldrini, M. Conti, J. Jacopini, and A. Passarella, Hibop: a history based routing protocol for opportunistic networks, in *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pages 1–12, 2007.

[22] M. Musolesi, S. Hailes, and C. Mascolo, Adaptive routing for intermittently connected mobile ad hoc networks, in *WOWMOM '05: Proceedings of the Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks*, pages 183–189, Washington, DC, USA, 2005, IEEE Computer Society.

[23] W. Zhao, M. Ammar, and E. Zegura, A message ferrying approach for data delivery in sparse mobile ad hoc networks, in *MobiHoc '04: Proceedings of the 5th ACM international symposium on Mobile ad hoc networking and computing*, pages 187–198, New York, NY, USA, 2004, ACM.

[24] N. Sarafijanovic-Djukic, M. Pidrkowski, and M. Grossglauser, Island hopping: Efficient mobility-assisted forwarding in partitioned networks, in *Sensor and Ad Hoc Communications and Networks, 2006. SECON '06. 2006 3rd Annual IEEE Communications Society on*, volume 1, pages 226–235, 2006.

[25] A. Lindgren, A. Doria, and O. Schelén, SIGMOBILE Mob. Comput. Commun. Rev. **7**, 19 (2003).

[26] J. Leguay, T. Friedman, and V. Conan, Evaluating mobility pattern space routing for dtns, in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–10, 2006.

[27] M. Grossglauser and D. N. C. Tse, IEEE/ACM Trans. Netw. **10**, 477 (2002).

[28] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, Prioritized epidemic routing for opportunistic networks, in *MobiOpp '07: Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking*, pages 62–66, New York, NY, USA, 2007, ACM.

[29] V. Erramilli and M. Crovella, Forwarding in opportunistic networks with resource constraints, in *CHANTS '08: Proceedings of the third ACM workshop on Challenged networks*, pages 41–48, New York, NY, USA, 2008, ACM.

[30] A. Demers et al., SIGOPS Oper. Syst. Rev. **22**, 8 (1988).

[31] A. Jindal and K. Psounis, Performance analysis of epidemic routing under contention, in *IWCMC '06: Proceedings of the 2006 international conference on Wireless communications and mobile computing*, pages 539–544, New York, NY, USA, 2006, ACM.

[32] S. Jain, K. Fall, and R. Patra, Routing in a delay tolerant network, in *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 145–158, New York, NY, USA, 2004, ACM.

[33] S. Jain, M. Demmer, R. Patra, and K. Fall, SIGCOMM Comput. Commun. Rev. **35**, 109 (2005).

[34] Y. Lin, B. Li, and B. Liang, Selected Areas in Communications, IEEE Journal on **26**, 794 (2008).

[35] F. Giudici, E. Pagani, and G. Rossi, *Impact of Mobility on Epidemic Broadcast in DTNs*, volume 284/2008, Springer Boston, 2008.

[36] T. Camp, J. Boleng, and V. Davies, Wireless Communications and Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications **2**, 483 (2002).

[37] C. Bettstetter, Smooth is better than sharp: a random mobility model for simulation of wireless networks, in *MSWIM '01: Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, pages 19–27, New York, NY, USA, 2001, ACM.

[38] Sumatra: Stanford university mobile activity traces.

[39] Mobilib: Community-wide library of mobility and wireless networks networks measurements.

[40] A. Heinemann, J. Kangasharju, and M. Muehlhaeuser, Opportunistic data dissemination using real-world user mobility traces, in *AINAW '08: Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops*, pages 1715–1720, Washington, DC, USA, 2008, IEEE Computer Society.

[41] M. Kim and D. Kotz, Extracting a mobility model from real user traces, in *In Proceedings of IEEE INFOCOM*, 2006.

[42] X. Hong, M. Gerla, G. Pei, and C. Chiang, A group mobility model for ad hoc wireless networks, 1999.

[43] M. Snchez and P. Manzoni, Future Generation Computer Systems **17**, 573 (2001).

[44] Y.-C. Hu and D. B. Johnson, Caching strategies in on-demand routing protocols for wireless ad hoc networks, in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 231–242, New York, NY, USA, 2000, ACM.

[45] B. Liang and Z. J. Haas, IEEE/ACM Trans. Netw. **11**, 718 (2003).

[46] F. Bai and A. Helmy, Chapter 1 a survey of mobility models in wireless adhoc networks.

[47] J. Tian, J. Haehner, C. Becker, I. Stepanov, and K. Rothermel, Graph-based mobility model for mobile ad hoc network simulation, in *SS '02: Proceedings of the 35th Annual Simulation Symposium*, page 337, Washington, DC, USA, 2002, IEEE Computer Society.

[48] B. Zhou, K. Xu, and M. Gerla, Group and swarm mobility models for ad hoc network scenarios using virtual tracks, in *Military Communications Conference, 2004. MILCOM 2004. IEEE*, volume 1, pages 289–294 Vol. 1, 2004.

[49] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, Towards realistic mobility models for mobile ad hoc networks, in *MobiCom '03: Proceedings of the 9th*

annual international conference on Mobile computing and networking, pages 217–229, New York, NY, USA, 2003, ACM.

[50] E. Royer, P. Melliar-Smith, and L. Moser, An analysis of the optimum node density for ad hoc mobile networks, in *Communications, 2001. ICC 2001. IEEE International Conference on*, volume 3, pages 857–861 vol.3, 2001.

[51] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu, and J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, in *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, pages 85–97, ACM Press, 1998.

[52] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, Stochastic properties of the random waypoint mobility model: epoch length, direction distribution, and cell change rate, in *MSWiM '02: Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, pages 7–14, New York, NY, USA, 2002, ACM.

[53] C. Bettstetter, C. Wagner, and T. U. Mnchen, IEEE Transactions on Mobile Computing **2**, 257 (2003).

[54] P. Lassila, E. Hyyti, and H. Koskinen, Connectivity properties of random waypoint mobility model for ad hoc networks, in *in Proc. MedHoc-Net, le de Porquerolles*, pages 159–168, 2005.

[55] W. Navidi and T. Camp, IEEE Transactions on Mobile Computing **3**, 99 (2004).

[56] P. Lassila, IEEE Transactions on Mobile Computing **5**, 680 (2006), Member-Hyytia,, Esa and Member-Virtamo,, Jorma.

[57] E. Hyytiä and J. Virtamo, Wirel. Netw. **13**, 177 (2007).

[58] A. Rojas, P. Branch, and G. Armitage, Validation of the random waypoint mobility model through a real world mobility trace, in *TENCON 2005 2005 IEEE Region 10*, pages 1–6, 2005.

[59] K. Viswanath, K. Obraczka, A. Kottas, and B. Sansó, Simulation **83**, 157 (2007).

[60] G. Neglia and X. Zhang, Optimal delay-power tradeoff in sparse delay tolerant networks: a preliminary study, in *CHANTS '06: Proceedings of the 2006 SIGCOMM workshop on Challenged networks*, pages 237–244, New York, NY, USA, 2006, ACM.

[61] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, Comput. Netw. **51**, 2867 (2007).

[62] C. Bettstetter, SIGMOBILE Mob. Comput. Commun. Rev. **5**, 55 (2001).

[63] R. Groenevelt, E. Altman, and P. Nain, Wirel. Netw. **12**, 561 (2006).

[64] Y.-K. Ip, W.-C. Lau, and O.-C. Yue, Performance modeling of epidemic routing with heterogeneous node types, in *Communications, 2008. ICC '08. IEEE International Conference on*, pages 219–224, 2008.

[65] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, Spray and wait: an efficient routing scheme for intermittently connected mobile networks, in *WDTN '05: Proceeding of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 252–259, New York, NY, USA, 2005, ACM Press.

Chapter 5

Appendix 1

Let us define $f(\beta) := \frac{(1+\alpha\beta^2)}{\beta}$ for $\beta > 0$. Differentiating wrt β and equating to 0, we get

$$f'(\beta) = -\frac{1}{\beta^2} + \alpha = 0 \;\Leftrightarrow\; \beta = \frac{1}{\sqrt{\alpha}}. \tag{5.1}$$

Also,

$$f''\left(\frac{1}{\sqrt{\alpha}}\right) = 2\beta^{\frac{3}{2}} > 0. \tag{5.2}$$

Hence $\beta = \frac{1}{\sqrt[2]{\alpha}}$ is the global minimum for the function f with $\beta > 0$.

Theorem 1. *Under an optimal scheme, if the best decision in state $x = l, \beta$ is 'not-copy', then there will be at most l copies in the system, and the system will never reach a state where the number of copies is $l + 1$.*

Proof We consider three cases.

Case $1: \;\beta < \frac{1}{\sqrt{\alpha}}$.

We get from (3.6) to (3.11) and the definition of the function f that,

$$J_{c,\lambda_+}(l, \beta) = \min\{J_{c,\lambda_+}(l, \beta), J_{c,0}(l, \beta), J_{c,\lambda_-}(l, \beta)\},$$

and

$$J_{\bar{c},\lambda_+}(l, \beta) = \min\{J_{\bar{c},\lambda_+}(l, \beta), J_{\bar{c},0}(l, \beta), J_{\bar{c},\lambda_-}(l, \beta)\}.$$

If the best decision in state $x = l, \beta$ is 'not-copy', then $J_{\bar{c},\lambda_+}(l, \beta) < J_{c,\lambda_+}(i, \beta)$. Note that,

$$J_{\bar{c},\lambda_+}(l, \beta) < J_{c,\lambda_+}(i, \beta) \tag{5.3}$$

$$\Leftrightarrow \frac{f(\beta+\lambda)}{l(N-l+1)} + \frac{(N-l)(f(\beta+\lambda))}{l(N-l+1)} < \gamma + \frac{f(\beta+\lambda)}{(l+1)(N-l)} + \frac{(N-l-1)f(\beta+\lambda)}{(N-l)(l+1)} \tag{5.4}$$

$$\Leftrightarrow f(\beta+\lambda)\{\frac{1}{l(N-l+1)} + \frac{(N-l)}{l(N-l+1)}\} - f(\beta+\lambda)\{\frac{1}{(l+1)(N-l)} + \frac{(N-l-1)}{(l+1)(N-l)}\} < \gamma \tag{5.5}$$

$$\Leftrightarrow \frac{f(\beta+\lambda)}{l} - \frac{f(\beta+\lambda)}{(l+1)} < \gamma \tag{5.6}$$

$$\Leftrightarrow f(\beta+\lambda) < \gamma l(l+1). \tag{5.7}$$

It follows that the best decision in state $x = l, \beta$ is 'not-copy' if and only if (5.7) holds, in which case the next state is $(l, \beta+\lambda)$. However, $f(\beta)$ is a decreasing function of β for $\beta < \frac{1}{\sqrt{\alpha}}$. This means that,

$$f(\beta+2\lambda) < f(\beta+\lambda) < \gamma l(l+1),$$

and hence the system will always have l copies and will converge towards $\beta = \frac{1}{\sqrt{\alpha}}$ in succesive steps.

Case $2 : \beta > \frac{1}{\sqrt{\alpha}}$.

We get from (3.6) to (3.11) and the definition of the function f that,

$$J_{c,\lambda_-}(l,\beta) = \min\{J_{c,\lambda_+}(l,\beta), J_{c,0}(l,\beta), J_{c,\lambda_-}(l,\beta)\},$$

and

$$J_{\bar{c},\lambda_-}(l,\beta) = \min\{J_{\bar{c},\lambda_+}(l,\beta), J_{\bar{c},0}(l,\beta), J_{\bar{c},\lambda_-}(l,\beta)\}.$$

If the best decision in state $x = l, \beta$ is 'not-copy', then $J_{\bar{c},\lambda_-}(l,\beta) < J_{c,\lambda_-}(i,\beta)$. Note that,

$$J_{\bar{c},\lambda_-}(l,\beta) < J_{c,\lambda_-}(i,\beta) \tag{5.8}$$

$$\Leftrightarrow \frac{f(\beta-\lambda)}{l(N-l+1)} + \frac{(N-l)(f(\beta+\lambda))}{l(N-l+1)} < \gamma + \frac{f(\beta-\lambda)}{(l+1)(N-l)} + \frac{(N-l-1)f(\beta-\lambda)}{(N-l)(l+1)} \quad (5.9)$$

$$\Leftrightarrow f(\beta-\lambda)\{\frac{1}{l(N-l+1)} + \frac{(N-l)}{l(N-l+1)}\} - f(\beta-\lambda)\{\frac{1}{(l+1)(N-l)} + \frac{(N-l-1)}{(l+1)(N-l)}\} < \gamma \quad (5.10)$$

$$\Leftrightarrow \frac{f(\beta-\lambda)}{l} - \frac{f(\beta-\lambda)}{(l+1)} < \gamma \quad (5.11)$$

$$\Leftrightarrow f(\beta-\lambda) < \gamma l(l+1). \quad (5.12)$$

It follows that the best decision in state $x = l, \beta$ is 'not-copy' if and only if (5.7) holds, in which case the next state is $(l, \beta-\lambda)$. However, $f(\beta)$ is an increasing function of β for $\beta > \frac{1}{\sqrt{\alpha}}$. This means that,

$$f(\beta-2\lambda) < f(\beta-\lambda) < \gamma l(l+1),$$

and hence the system will always have l copies and will converge towards $\beta = \frac{1}{\sqrt{\alpha}}$ in succesive steps.

Case $3:\ \beta = \frac{1}{\sqrt{\alpha}}$.

We get from (3.6) to (3.11) and the definition of the function f that,

$$J_{c,0}(l,\beta) = \min\{J_{c,\lambda_+}(l,\beta), J_{c,0}(l,\beta), J_{c,\lambda_-}(l,\beta)\},$$

and

$$J_{\bar{c},0}(l,\beta) = \min\{J_{\bar{c},\lambda_+}(l,\beta), J_{\bar{c},0}(l,\beta), J_{\bar{c},\lambda_-}(l,\beta)\}.$$

If the best decision in state $x = l, \beta$ is 'not-copy', then $J_{\bar{c},0}(l,\beta) < J_{c,0}(i,\beta)$. This means that the system will stay in the state (l,β) till the destination node is infected. □

Theorem 2. *Suppose* $0 \leq i \leq N-1$ *and* $\beta > 0$ *are such that*

1. $J_{c,\cdot}(i+1,\beta) > J_{\bar{c},\cdot}(i+1,\beta)$,

2. $J_{\bar{c},\cdot}(i,\beta) > J_{c,\cdot}(i,\beta)$.

Then $J_{\bar{c},\cdot}(j,\beta) > J_{c,\cdot}(j,\beta)\ \forall\ j < i$.

Proof It follows from the proof of Theorem 1 that $J_{c,\cdot}(i+1,\beta) > J_{\bar{c},\cdot}(i+1,\beta)$ if and only if $f(\beta+\cdot) < \gamma(i+1)(i+2)$. Since $J_{\bar{c},\cdot}(i,\beta) > J_{c,\cdot}(i,\beta)$, we get

$$\Leftrightarrow \gamma i(i+1) \leq f(\beta+\cdot) < \gamma(i+1)(i+2). \tag{5.13}$$

This means that $i+1$ is the forwarding limit corresponding to $\beta+\cdot$. Note that $\gamma i(i+1) \leq f(\beta+\cdot)$ implies that $j(j+1) \leq f(\beta+\cdot)$ for every $j < i$. Hence, $J_{\bar{c},\cdot}(j,\beta) > J_{c,\cdot}(j,\beta)\ \forall\ j < i$.

Chapter 6

Appendix 2

We investigate the effect of the magnitude of the step-size λ on the steps of the algorithm.

Case $1:\ \beta - \lambda \geq \frac{1}{\sqrt{\alpha}}$

Since $f(\beta)$ is a decreasing function for $\beta > \frac{1}{\sqrt{\alpha}}$, for this case

$$f(\beta + \lambda) > f(\beta) > f(\beta - \lambda).$$

Hence, $u = \{\cdot, \lambda_-\}$ will be the best decision.

Case $2:\ \beta + \lambda \leq \frac{1}{\sqrt{\alpha}}$

Since $f(\beta)$ is an increasing function for $\beta < \frac{1}{\sqrt{\alpha}}$, for this case

$$f(\beta - \lambda) > f(\beta) > f(\beta + \lambda).$$

Hence, $u = \{\cdot, \lambda_+\}$ will be the best decision.

Case 3: $\beta = \frac{1}{\sqrt{\alpha}}$

Since $f(\beta)$ is minimum at $\frac{1}{\sqrt[2]{\alpha}}$, for this case

$$f(\beta - \lambda) > f(\beta),\ \text{and}\ f(\beta + \lambda) > f(\beta).$$

Hence, $u = \{\cdot, 0\}$ will be the best decision.

Case $4:\ \beta > \frac{1}{\sqrt{\alpha}}$ **and** $\beta - \lambda < \frac{1}{\sqrt{\alpha}}$

Since $f(\beta)$ is minimum at $\frac{1}{\sqrt{\alpha}}$, $u = \{\cdot, 0\}$ will be the best decision if $f(\beta) < f(\beta - \lambda)$, $u = \{\cdot, \lambda_-\}$ will be the best decision if $f(\beta) > f(\beta - \lambda)$, and both decisions are equivalent if $f(\beta) = f(\beta - \lambda)$. Note that

$$
\begin{aligned}
& f(\beta) < f(\beta - \lambda) \\
\Leftrightarrow \quad & \frac{1 + \alpha\beta^2}{\beta} < \frac{1 + \alpha(\beta - \lambda)^2}{(\beta - \lambda)} \\
\Leftrightarrow \quad & \frac{\lambda(1 + \alpha\beta(\lambda - \beta))}{(\beta - \lambda)\beta} > 0 \\
\Leftrightarrow \quad & \beta - \lambda < \frac{1}{\alpha\beta}.
\end{aligned}
$$

Hence, $u = \{\cdot, 0\}$ will be the best decision if $\beta - \lambda < \frac{1}{\alpha\beta}$. Similarly, $u = \{\cdot, \lambda_-\}$ will be the best decision if $\beta - \lambda > \frac{1}{\alpha\beta}$. Note that $u = \{\cdot, \lambda_-\}$ and $u = \{\cdot, 0\}$ will become equivalent decisions if $\beta - \lambda = \frac{1}{\alpha\beta}$.

Case $5:$ $\beta < \frac{1}{\sqrt{\alpha}}$ **and** $\beta + \lambda > \frac{1}{\sqrt{\alpha}}$

Since $f(\beta)$ is minimum at $\frac{1}{\sqrt{\alpha}}$, $u = \{\cdot, 0\}$ will be the best decision if $f(\beta) < f(\beta + \lambda)$, $u = \{\cdot, \lambda_+\}$ will be the best decision if $f(\beta) > f(\beta + \lambda)$, and both decisions are equivalent if $f(\beta) = f(\beta + \lambda)$. Note that

$$
\begin{aligned}
& f(\beta) < f(\beta + \lambda) \\
\Leftrightarrow \quad & \frac{1 + \alpha\beta^2}{\beta} < \frac{1 + \alpha(\beta + \lambda)^2}{(\beta + \lambda)} \\
\Leftrightarrow \quad & \frac{-\lambda(1 + \alpha\beta(-\lambda - \beta))}{(\beta + \lambda)\beta} > 0 \\
\Leftrightarrow \quad & \beta + \lambda > \frac{1}{\alpha\beta}.
\end{aligned}
$$

Hence, $u = \{\cdot, 0\}$ will be the best decision if $\beta + \lambda > \frac{1}{\alpha\beta}$. Similarly, $u = \{., \lambda_+\}$ will be the best decision if $\beta + \lambda < \frac{1}{\alpha\beta}$. Note that $u = \{\cdot, \lambda_+\}$ and $u = \{\cdot, 0\}$ will become equivalent decisions if $\beta + \lambda = \frac{1}{\alpha\beta}$.