Epidemic Content Distribution in Mobile Networks

A study of epidemic content distribution characteristic with social relationship evaluation

LAN SUN

KTH Information and
Communication Technology

Degree project in
Communication Systems
Second level, 30.0 HEC
Stockholm, Sweden

Epidemic Content Distribution in Mobile Networks

A study of epidemic content distribution characteristic
with social relationship evaluation

Lan Sun

Master of Science Thesis

Communication Systems
School of Information and Communication Technology
KTH Royal Institute of Technology
Stockholm, Sweden

Feb 18 2013

Supervisor:
Professor Lieliang Yang
Professor Gerald Q.Maguire Jr

Examiner: Professor Gerald Q.Maguire Jr

Abstract

With the growing popularity of integrating mobile networks and social networks, people now enjoy a freer and more efficient means of communication. Smarter mobile devices facilitate modern human life. In the information age, various new types of information have begun to appear. How to disseminate content to people in a swift and fair way has long been a question. Choosing the right strategy for content distribution is especially crucial for mobile social networks. In this thesis project we use epidemic models for content distribution in mobile social networks.

Stochastic mobility models and an SIR epidemic model are set up in the evaluation. We analyze the impact of various parameters of mobility models and epidemic model on content distribution's success rate and delivery delay. Also, we exploit the social relationships to facilitate content distribution and show the impact of social relationships on content distribution.

Simulations have shown that increasing speed and node number in the mobility models will have positive impact on content distribution success rate as well as decreasing the delay. The infect time limit and infect count limit of the epidemic model are also important for swiftly distributing content while considering energy consumption and fairness for nodes.

In the social relationship simulation, nodes' meeting times during a period of time are calculated and a threshold based on a certain level of meeting times is used for categorizing the friendship relationships between nodes. The results show that it will be easier for a successful distribution to be achieved as the social relationship between nodes gets stronger. Also, the delay shows a decreasing trend until reaching the ideal distribution delay time.

Keywords: content distribution, epidemic, social networks

Sammanfattning

Med den växande populariteten för att integrera mobila nätverk och sociala nätverk, människor njuta nu en friare och effektivare sätt att kommunicera. Smartare mobila enheter underlättar moderna människans liv. I den information som ålder, har olika nya typer av information börjat visas. Hur sprida innehåll till människor påett snabbt och rättvist sätt har länge varit en fråga. Att välja rätt strategi för distribution av innehåll är särskilt viktigt för mobila sociala nätverk. I den här avhandlingen projekt använder vi epidemiska modeller för distribution av innehåll i mobila sociala nätverk.

Stokastiska rörlighet modeller och en SIR-epidemi modell sätts upp i utvärderingen. Vi analyserar effekterna av olika parametrar rörlighet modeller och epidemisk modell påinnehållsdistribution s framgång och leveransförsening. Dessutom utnyttjar vi de sociala relationerna för att underlätta distribution av innehåll och visa hur sociala relationer pådistribution av innehåll.

Simuleringar har visat att ökad hastighet och nodnummer i rörlighet modellerna kommer att ha en positiv inverkan pådistribution av innehåll framgång samt att minska fördröjningen. Den infektera tid och infektera räkna gräns epidemin modellen är ocksåviktiga för att snabbt distribuera innehåll och samtidigt överväga energiförbrukning och rättvisa för noder.

I den sociala relationen simulering är noder möte tid under en tidsperiod beräknas och en tröskel baserad påen viss nivåav mötestiden används för att kategorisera vänskap relationer mellan noder. Resultaten visade att det blir lättare för en lyckad spridning uppnås som den sociala relationen mellan noder blir starkare. Dessutom visar fördröjningen en nedåtgående trend tills den når fördröjningen av en ideal fördelning.

Sökord: innehåll distribution, epidemi, sociala nätverk

Acknowledgements

First of all, I should express my deepest gratefulness to my adviser Professor LieLiang Yang. Thanks so much for you help for providing the topic of this thesis as well as all the kindly suggestions and help during this work.

Also, I would like to express my honor and gratefulness to my supervisor Professor Gerald Q.Maguire Jr. for his kindly help through out the thesis period, especially on all his valuable comments on the report.

I would like to thank Junzhe Tian, for his kindly support including instructing me on using Latex. Thanks to all people that provided help during my simulations, I will always be grateful for your kindness.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

This document requires readers to be familiar with some terms and concepts. Below we give a short description of these acronyms.

A/G	autonomous gossiping
DSR	Dynamic Source Routing
DTN	Delay Tolerant Network
GM	Gauss-Markov
GPS	global positioning service
Mt	meeting time
MSNs	Mobile Social Networks
ODE	Ordinary differential equation
PCS	personal communication service
PDA	personal digital assistant
PDE	Partial differential equation
RDM	Random Direction Mobility
RM	RandomWaypoint Mobility
SEIR	Susceptible-Exposed-Infectious-Recovered
SI	Susceptible-Infectious
SIR	Susceptible-Infectious-Recovered
SIS	Susceptible-Infective- Susceptible
SPM	social pressure metric

Chapter 1

Introduction

Wireless technologies embedded in portable mobile devices have promoted the direct exchange of information (here after referred to as content) among people. Within a short range, content that you might be interested could be directly transmitted to you from a device belonging to another person who happens to be standing, sitting, or passing near enough to you for your device's wireless communication link to successfully transfer this content from this other person's device to your device. In this situation, transferring content between the two portable mobile devices via a short range wireless communication link can be seen as "a close analogy to the oldest and best-studied disease epidemics"[2]. The analogy of information dissemination to a biological epidemic helps us to understand the potential for transferring information over an intermittent network, where the route from a source to a destination is dynamic and unpredictable due to the mobile nodes, movements and due to the limited (and environmentally dependent) communication range of a given radio link. Using an epidemic model to understand content distribution is the main aim of this thesis project. We are specifically going to look at an epidemic content dissemination model in the context of a mobile network environment while considering to social relationships between people.

1.1 Problem description

In this master's thesis, we want to examine content dissemination in the context of a mobile social network. Similar to the spread of a disease during an epidemic, content dissemination will take place in a (physically) mobile network. An example of such a network is a wireless *ad hoc* network. However, the goal for content distribution is different from simply exploiting epidemic routing (see [3] for details of epidemic routing). The main goal of the content dissemination mechanism is to swiftly distribute content to people in a society. In this thesis project, we try to examine an epidemic content diffusion method that can achieve both fairness and efficiently distribute content. In our evaluation of the proposed solution we will analyze a number of different factors of both mobility models and diffusion algorithms that may have impact on the success rate of this method of content distribution. One problem is to find a suitable mobility model for use during our evaluation, and to examine how this mobility model affects the results of the content distribution. Another problem is to define a diffusion algorithm for efficient content delivery which will quickly delivery content to a community of people with the lowest latency, while minimizing the energy consumption of all the devices. The last problem is to study what impact social relationships have on success rate and latency of content distribution.

1.2 Structure of this thesis

An overview of this thesis was given in the previous paragraphs. The rest of the thesis is divided into four chapters.

In Chapter 2, the background work is reviewed to provide a basis for the rest of this thesis and to facilitate the reader's ability to understand the rest of this thesis and its context. Chapter 3 describes the goals, metrics, and solutions proposed for simulations, including the mobility models and infection algorithms that have been used. Social related implementation is also introduced in this chapter. The results of simulations are analyzed and evaluated in Chapter 4. Finally, Chapter 5 concludes with a summary of previous work and give suggestions for future implementation and modifications to the models and algorithms utilized in this thesis. Reflections on this work in terms of economic, social, and ethical issues relevant to this work are also discussed in this final chapter.

1.3 Limitations

There are some limitations for this work. This work use Matlab to simulate node's movement and nodes who meet the conditions can delivery content to each other. The mobility models set up are stochastic models, not real traces. The content delivery time is ignored and we do not simulate large size content delivery which might take up to minutes or hours. Also, there is no mechanisms for notifying nodes when a certain type of content is successful delivered to all. Last but not least, the simulation time is limited due to the simulation environment restrictions.

Chapter 2

Background

As wireless communication has become widely deployed and integrated into everyday human life, a new means for information dissemination among people is available. People can conveniently gather information or communicate with others through their wireless devices. The very high penetration rate of wireless devices, many of which are always on, has lead to a new opportunity to distribute content via wireless networks. Research using, bio-inspired networking has attract a lot of attention. In this chapter, we will introduce related work.

2.1 Epidemic routing

Various protocols have been proposed for opportunistic routing in intermittent networks. In April 2000, Vahdat and Becker's protocol [3] was introduced. This is one of the pioneer works concerning epidemic routing in a partitioned mobile *ad hoc* network. Their flooding based mechanism is based on the theory that if we allow random exchanges of messages among entitles holding different messages, then all messages will be received by all nodes within "a bounded amount of time". The goal of this mechanism is to maximize message delivery rate and to minimize message delivery latency, as well as to minimize the system resources consumed in message delivery. Similar to this protocol, other protocols such as those proposed by Grossglauser and Tse[4] and Groenevelt, Nain, and Koole [5] use variations of a two-hop relay protocol, which can be regarded as an epidemic protocol with hop count of two.

Since flooding mechanisms are commonly used in epidemic routing and their overhead is relative high, a autonomous gossiping (A/G) self-organizing epidemic algorithm for selective information dissemination method is proposed by Datta, Quarteroni, and Aberer in [6]. Instead of flooding the whole network, the A/G algorithm chooses target nodes based on the data items' own profile and the target node's profile in order to selectively spread content. Policies such as migration and replication are used together with a utility value. This utility value is increased or decreased as a reward or punishment for efficient delivery. The completeness of this A/G content-based dissemination was evaluated and it was observed that this algorithm does not guarantee the completeness of data delivery. As a result the A/G self-organizing epidemic algorithm is only suitable for situations where completeness is not critical or when minimizing the overhead is more important than completeness. Another method introduced by Hayashi, Hara, and Nishio in [7] provides an epidemic method that reduces the communication traffic for data dissemination. When mobile hosts communicate with each other, each of the nodes needs to update its data items. Invalidation reports are disseminated to cause nodes to discard old data replicas. It is shown in the paper that compared to the other conventional methods introduced in [7], the proposed method of exploiting an epidemic model generates the lowest traffic for update data dissemination.

As epidemic spreading of context has become an interest of the research community, how information diffuses during the propagation process has been widely discussed. In [8], a definitive metric called "reproductive rate" is determined which can well characterizes the rate at which information diffuses as a function of the node density, mobility, radio range, etc. In this work a mathematical formula is deduced using a Susceptible-Infectious-Recovered (SIR) epidemic routing algorithm under the Random Direction Mobility (RDM) mobility model. Simulation has been used to evaluate the accuracy. Similarly in [9], a simple Susceptible-Infectious (SI) epidemic model is analyzed with a mathematical formulation describing its propagation process, leading to the result that this model can be well modeled by a "deterministic compartmental epidemic model". This model is characterized by the infection rate, where the infection rate has a relationship with node density.

Ordinary differential equations (ODEs) can be used for investigating basic epidemic routing. We can calculate the average delivery delay by using the SIR model, ODEs, and a Markov process [10]. Such mathematical epidemiology has also been used in [11] for modeling the ideal information propagation in a sparsely connected network and a wave-like solution for modeling message propagation is analyzed using a reaction-diffusion partial differential equation (PDE). The PDE models in [11] are the evolution of the density of informed nodes, so this is a replacement for the "well-mixed ODE" in [10] under a uniform density assumption.

If it is still not clear that different mobility models will greatly affect the performance of epidemic routing, then the study in [12] directly analyzes the importance of choosing appropriate mobility models in network performance investigation. Using the Dynamic Source Routing Protocol (DSR), four different mobility models were simulated and the performance metrics such as data packet delivery ratio, end-to-end delay, average hop count, and protocol overhead were compared. Among the four mobility models, there are three entity mobility models: Random Walk Mobility Model, Random Direction Mobility Model, and Random Waypoint Mobility Model. According to the simulation results, Random Waypoint model has the highest data packet delivery ratio, the lowest end-to-end delay, and lowest average hop count. However, this only suggests that Random Waypoint Mobility Model stressed the DSR less than the other two entity mobility models. The paper suggests that the performance of a network protocol should be evaluated with the most suitable mobility model that “matches the expected real-world scenario”.

2.2 Mobile Social Network related work

As social networks grow more significant in human life, recent researches focuses on social relationships more than ever. A mobile social network is considered to be another type of mobile *ad hoc* network with human social relationships considered in the topology. Social interactions between people will strongly affect their movement behavior, as well as the content distribution efficiency. Several researchers have examined human mobility models, such as [13, 14]. Mobility traces are important for analyzing human movement patterns and using data collected from a global positioning service (GPS) or academic experiments “human moves are highly predictable” [13]. Two classes of models for mobility have been utilized. One is a trace-based model that is defined by statistics from traces, however this can be very complex to implement and also it’s narrowed to a specified scenario so it can not express a “general model” for user movement. The other is a synthetic model that takes location preference, time schedules, and social connections into consideration. However, the synthetic model might need validation using existing traces. A mobility model of “Heterogeneous Human Walk” based on an overlapping community structure was introduced in [14]. The tradeoff between reality and complexity made the overlapping communities generated through the synthetic method different than social graphs or real life traces.

Mobile social network friendship relations are explored in [15]. A new metric is introduced which is called "social pressure metric (SPM)". This metric can reflect the node relations with better accuracy by considering the three aspects: meeting frequency, meeting longevity, and regularity. The same paper proposes a routing algorithm and the forwarding is based on the nodes' relationship: "a node will forward data message to nodes that contain the destination nodes in their friendship communities". The routing algorithm was evaluated by two real DTN traces and a synthetic mobility trace, in comparison with several previously used routing algorithms . The results shown that this new routing algorithm out performed than the previous "benchmark algorithms" of "Prophet" in[16],"SimBet" in [17], and "Fresh" in [18] in terms of routing efficiency using the three traces.

Social connections related studies have shown that human networks can be quite complex to represent and network graphs are usually used to express the connections between human beings. Three of the most commonly used graphs to represent different connection structures between individuals are: random graph, scale-free graph, and small-world graph.

A random graph can be used for a "first approximation for networks" whose structure is "unknown except for the number of vertices and edges" [19]. Different probability distributions can be used to create graphs by different random graph models [20]. For a social network in real world, the graphs are not totally randomly structured. "Clustering" exists in the real world social networks. However, a random graph does not possess this clustering. In a scale-free graph, a power law distribution of degrees exists for the vertices [21]. This means that highly connected people will have a higher probability to connect to people newly added to the network. Many networks are scale free networks, such as a social network and biological networks. However, a scale-free network is still not the "ultimate model" for all social networks due to the complex scaling problem of human network [19]. A small world graph is a kind of graph in which most nodes can be reached by a small number of hops or steps. Such a graph is also highly clustered, but the growth of the diameter ("maximum distance separating two vertices in the graph") [19] follows a logarithmic distribution. We can describe this mathematically as L stands for the typical distance of two randomly chosen nodes and the number of nodes is N[22], then L follows a logarithms distribution as follows:

$$L \propto \log N$$

2.3 Epidemic Algorithms

In recent years, researchers shown a great interest in biological systems and believed that nature could provide a good paradigm for modem systems and guide the design of algorithms. A survey by Falko Dressler and Ozgur Akanb [23] illustrates the great potentials for bio-inspired networking. Epidemic algorithms are only one small aspect of this paradigm, which researchers adopted from biological systems. In this section, several main epidemic algorithms are introduced. One of these algorithms will be further considered in our content distribution simulation. The original use of epidemic theory was to predict the spread of disease in nature. A close analogy can be made from a disease spreading between human beings to information dissemination between nodes of a network. The epidemic theory used in information dissemination can be regarded as a "store-carry-forward" paradigm, where a node stores and carries certain information which it forwards to its neighbors within a certain communication range [8]. In each epidemic round, a node might have one of the following three states: Susceptible, Infective, and Recovered. An explanation of these three states follows:

Susceptible The node does not have the specific information, but has a possibility to get it.

Infective The node has the specific information and can infect others, i.e., delivery this information to other nodes.

Recovered The node has a copy of the information, but it will no longer delivery this information to others nor will it receive the same information again.

Different epidemic models characterize different processes of infection, here the process for spreading information. These epidemic models can be divided into different classes based on immunity. The three most commonly used models are: Susceptible-Infective (SI) Model, Susceptible-Infective-Recovered (SIR) Model, and Susceptible-Infective- Susceptible (SIS) Model.

2.3.1 Susceptible-Infective (SI) Model

In this model, all nodes are initially considered to be susceptible. When the infection period starts, the node becomes infected when it receives the information and tries to transmit this information to its neighbors. All infected nodes remains infective until all nodes are infected. Since there is no mechanism for a node to stop infecting others, some method(s) should be employed to the stop of the infection when there are no susceptible nodes in the whole network.

2.3.2 Susceptible-Infective-Susceptible (SIS) Model

The SIS model has the same starting period as the SI model. However, the difference is that in the SIS model, there exists a condition when a node stops spreading the information to its neighbors due to its own temporarily recovered state. In this state, the node will no longer infect others, but must revert to being a susceptible nodes before it is infected again. The SIS model was first introduced by Herbert W. Hethcote in 1973 [24].

2.3.3 Susceptible-Infective-Recovered (SIR) Model

In the SIR diffusion model, nodes are classified into susceptible nodes S(t), infected nodes I(t), and recovered nodes R(t). If N is the total population, then at a given point in time t:

$$N = S(t) + I(t) + R(t)$$

The SIR model is similar to the SIS model except that each node will finally enter to a recovered state, which means this node will not spread information to others. The SIR model has been widely used and we will also implement this model in our simulation as the basic information diffusion/infection model.

In addition to the above three classes, there are also other epidemic models used, such as the Susceptible-Exposed-Infectious-Recovered (SEIR) model and other gossip algorithms. We will not further explain these model in this thesis, more information related to epidemic models can be found in [25].

2.4 Mobility Models

Mobility models play an important role in the simulation and performance evaluation of a protocol or algorithm in a mobile *ad hoc* network environment. There are generally two types of mobility models for simulation: synthetic models and real life traces [26]. Models build up based on mathematical models are regarded as synthetic models. Models such as Random Waypoint Model, Random Direction Model, and Gauss-Markov Model are all widely used synthetic models. On the other hand, real life traces obtained from long term experiments measuring nodes' location or movement can well represent the mobility of nodes under certain circumstances. One of the most popular traces is the Reality Mining project carried out at the MIT Media Laboratory in 2005 [27]. In this project Bluetooth devices are used to collect data about movements of participants. There are also similar projects, such as the "Wireless Topology Discovery project at UCSD" [28], the user mobility experiment using Bluetooth enabled PDAs at the Univeristy of Toronto [29], and the user mobility model measured from traces of Wi-Fi traffic at Dartmouth College[30]. However, the number of publicly available traces open to public is still limited and the scenarios where these traces can be used are quite narrow. Additionally, the data generated through the traces are not flexible since the distribution of speed or density of nodes, etc., are normally fixed to a specified value. For this reason synthetic models are often preferred over experimental traces. Below we will introduce two different synthetic models which are the most used models for simulations: the Random Waypoint Mobility model and the Gauss-Markov Mobility model. These two models were used in this thesis project's simulations.

2.4.1 Random Waypoint Mobility Model

This mobility model has been widely used for simulation [31][32]. In the random waypoint mobility model, the mobile node chooses a destination randomly within the boundary. Then the mobile node will head toward this destination with a speed, which can be uniformly selected from within a predefined interval. The node will pause for a period of time upon arrival at the destination, after this pause the mobile node repeats the procedure with another destination. This model can also be simplified by setting the pause time to zero, as for example in [33], the result is similar to the Random Walk Mobility Model. A long pause time will lead to a more stable network even with high speed movements. When the node heads to anther destination, a new speed is selected. An important problem in the random waypoint model is initialization, which needs careful consideration, in order to avoid warm up problems. In the performance evaluation in [12], the average neighbor percentage in the initial stage is relatively high compared to other times during the simulation. This higher neighbor percentage will lead to overestimation of the performance. There are several solutions for this problem, for example, dropping the initial 1000 seconds of simulation time to ensure the initialization period is removed or saving the previous nodes' locations from a long running simulation and using this as the staring point of future simulations. In this thesis project we have adopted the second solution. Note that we have used the ending positions of one simulation as the starting positions for the subsequent simulation run. As a result it is as if a very long simulation were run with changes in the simulation parameters at different points in the single simulation.

Figure 2.1 illustrates the movement pattern generated by the Random Waypoint Mobility Model used in this thesis. The start position and end position are shown as a green circle and a red cross respectively.

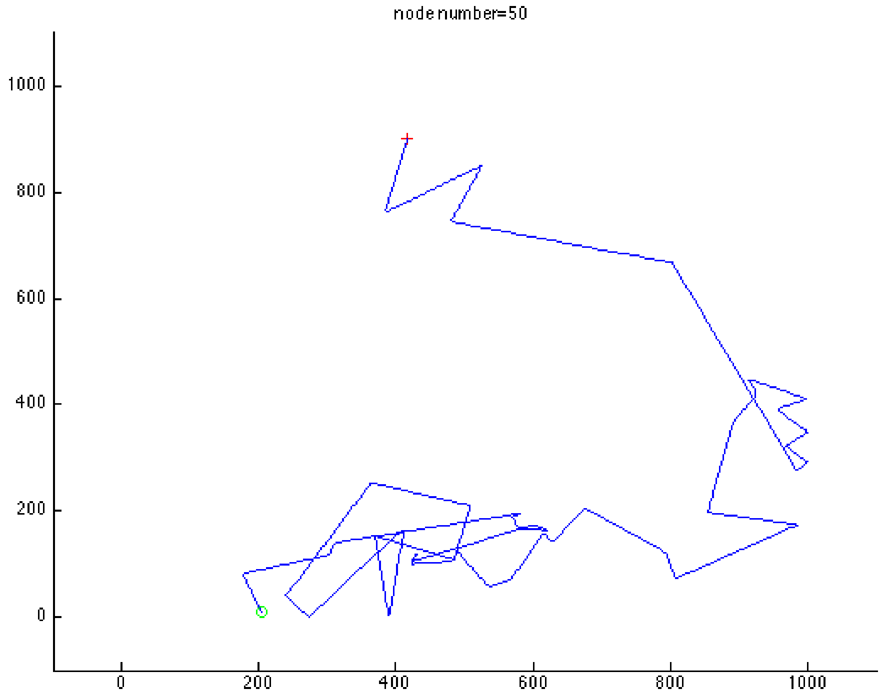


Figure 2.1: The mobility graph of a single node generated from a Random Waypoint Mobility Model with pause time $= 0s$ and speed range $= [0, 5.5]m/s$

2.4.2 Gauss-Markov Mobility Model

The original use of the Gauss-Markov Mobility Model was to predict mobile nodes' future locations. This model was initially introduced by B.Liang and Z. J. Haas in their simulation of wireless personal communication service (PCS)[34]. Initially, the mobile node has a initial speed and direction. After a fixed time interval, the speed and direction will be updated to a new speed and a new direction. The correlation between the new velocity(V_n) or direction(D_n) of n^{th} time interval and the previous velocity(V_{n-1}) or direction(D_{n-1}) of the $(n-1)^{st}$ time interval can be expressed in the following equations:

$$V_n = \alpha V_{n-1} + (1-\alpha)\bar{v} + \sqrt{(1-\alpha^2)}V_{x_n}$$

$$D_n = \alpha D_{n-1} + (1-\alpha)\bar{d} + \sqrt{(1-\alpha^2)}D_{x_n}$$

In the two equations above, α is a parameter between $[0, 1]$ which controls the randomness between the previous and current movement, while $\bar{v}$ and $\bar{d}$ stand for the mean velocity and mean direction respectively and the V_{x_n} and D_{x_n} are random variables from a Gaussian distribution. This method was used in [35] to implement the Gauss-Markov mobility model and it has also been used in this thesis project.

This model can be adapted for different randomness levels using the parameter α. The coupling to the previous location and velocity overcomes the Random Walk model's problem of "sudden stops" and "sharp turns".

Figure 2.2 illustrates the movement pattern generated from the Gauss-Markov Mobility Model used in this thesis project. The start position and end position are shown in green circle and red cross respectively.

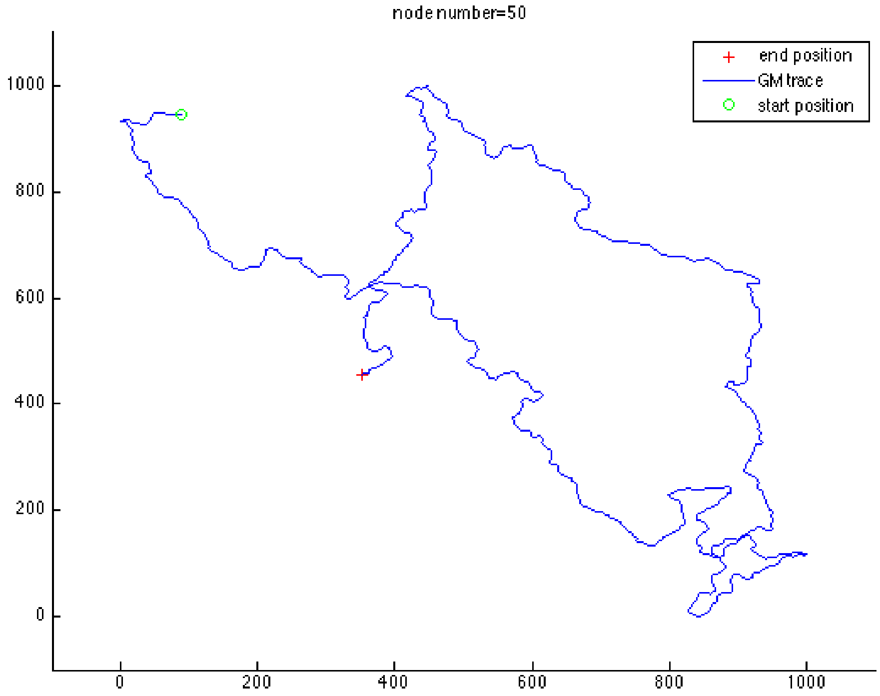


Figure 2.2: The mobility graph of a single node generated from a Gauss-Markov Mobility Model with $\alpha = 0.85$ and speed range $= [0, 5.5]m/s$

In addition to the two models described above, there are several other synthetic mobility models, including: Random Walk Mobility Model, Random Direction

Mobility Model, and some group mobility models. Further details about these other mobility models can be found in [12].

Although widely used, synthetic models have some problems. The major problem is that the results of the synthetic models can differ significantly from the node mobility observed in reality. The movements of nodes in random mobility models are not in accordance with real life human movements. However, in this thesis and our simulations, we will not take these problems into further consideration.

Chapter 3

Method

In order to carry out our study of content distribution using epidemic methods, simulation models need to be set up. The choice of mobility model is fundamental for content distribution in mobile networks since it controls the mobility of the mobility nodes within a certain area. Given suitable parameters, the mobility model should be consistent with the inner connectivities and relationships between nodes. Matlab was chosen for generating mobility models, since it is provides matrix calculations. Another important element in the simulation is the SIR epidemic model for distributing contents to others. This model was also programed in the Matlab environment. In this chapter we will describe these models in more detail.

3.1 Mobility Models

In order to study content distribution in a mobile network with relation to social connections between people, we have adopted the widely used Random Waypoint Mobility Model and Gauss-Markov Mobility Model as the basic mobility models for our mobile nodes. As introduced in section 2.4, these two mobility models are synthetic mobility models. We discuss the implementation details of these two models in the following subsections.

3.1.1 Implementing the Random Waypoint Mobility Model

The first and most important thing when setting up a mobility model is to choose suitable parameters. The parameters in our model were chosen based on several other papers' synthetic model models, specifically those used in [15, 1, 36]. In these papers, a square area of size $1,000 \times 1,000\ m^2$ is used for all the mobility simulations. Several sets of basic parameters are listed in Table 3.1. These parameters were used in the Random Waypoint Mobility Model described in [1].

Table 3.1: Parameters for Mobility Model in [1]

Parameter	Values	Unit
Number of Nodes	50, 100	-
MaxSpeed	5, 50	m/s
Pause Time	0, 1, 100,1500	s
Duration Time	1500	s
Area Size	1000×1000	m^2

In the random waypoint mobility model of [36], the speed range is [0.1, 20] m/s, which is similar to the speed range of [20, 120] m/minute used in [15]. Note that the speed is chosen uniformly randomly over the range. Based on the values above, we implemented our own mobility model using the set of parameters shown in Table 3.2.

Table 3.2: Random Waypoint Mobility Model Parameters

Parameter	Values	Unit
Number of Nodes	50,100,200	-
MaxSpeed range	[0, 20]	m/s
Pause Time	0, 10 ,60, 120	s
Walking Time range	[0, 60]	s
Area Size	1000×1000	m^2
Duration Time	10000	s

One important issue in the implementation of the random waypoint mobility model is the initial startup problem. In order to reduce the over estimation of neighbors during a start up period, a long enough simulation is needed to bring the system into a “warm” state. In this thesis a long simulation was done once and the ending position of all nodes were saved. These saved node positions are used in subsequent simulation as the initial start position for all of the nodes. Then after each simulation run, the last position of all nodes are recorded and saved for as the nodes’ location in the next simulation run. If the next destination is outside the boundary during the simulation, then the direction is changed into a reflection when the node reaches the boundary. The node will keep heading in the reflection direction until the end of this walking period.

3.1.2 Implementing the Gauss-Markov Mobility Model

There are no so many existed Gauss-Markov mobility models implemented due to its complexity in route calculation, since the future speed and direction in the

Gauss-Markov mobility model is calculated based on the past speed and direction. However, this model eliminates the sudden stop and sharp turns problems of the Random Waypoint mobility model, hence we implemented this model for comparison with the Random Waypoint mobility model. The parameters chosen for the Gauss-Markov mobility model are similar to the Random Waypoint mobility model. However, there are several extra parameters for the Gauss-Markov mobility model: the “Degree of Randomness”, “Boundary Range” and the“Time Unit”. Based on the parameters used in earlier papers (specifically [1, 36]), we implement our own Gauss-Markov mobility model using the set of parameters shown in Table 3.3.

Table 3.3: Gauss-Markov Mobility Model Parameters

Parameter	Values	Unit
Number of Nodes	50,100,200	-
MaxSpeed range	[0, 20]	m/s
Degree of Randomness (α)	0, 0.25, 0.5, 0.75, 1	-
Boundary Range	100	m
Time Unit	1	s
Area Size	1000×1000	m^2
Duration Time	10000	s

As introduced in section 2.4.2, the turing parameter α for “Degree of Randomness ” in Gauss-Markov model is used to control the randomness between the previous and current movement. Boundary constraints are implemented in the Gauss-Markov model to eliminate the boundary problem causing mobile nodes to tend to gather at the edge of the boundary while moving This constraint is implemented based on the mechanism used in [35], i.e., when a mobile node moves to within a certain distance of the boundary, denoted as B_r(boundary range), the mean value of direction $\bar{d_n}$ is changed. The value of $\bar{d_n}$ is changed based on the position of the node following the rules shown in Figure 3.1. The main purpose of these rules is so that in different positions we will select different mean direction values to keep the nodes away from the edges. If this boundary problem is not solved, most nodes will simply move around the four edges, which will cause unbalanced node density at the center of the simulation area.

3.2 Epidemic Model

The SIR model was chosen to describe when users distribute content to each other. The reason why we use SIR model compared to the other models is that this SIR model can be good represent the human behavior to new messages. People tend

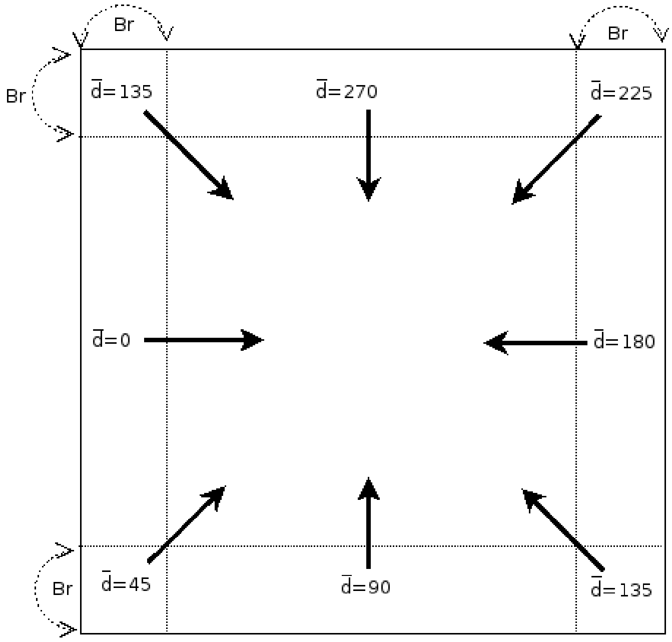


Figure 3.1: The first constraint of boundary problem for Gauss-Markov Mobility Model

to sent new messages to others when they firstly receive them and they will lose interest to delivery it others after a period of time. Also, the SIR model has a removed period which will be good from the energy consumption aspect. As introduced in Chapter 2, the nodes in the SIR model have three stages: susceptible, infected, and recovered. Each node must be in one of these three stages. We use a infect flag I_n to stands for the infection stage of node n. The parameters we used with the SIR Epidemic Model are listed in Table 3.4.

Table 3.4: The SIR Epidemic Model Parameters

Parameter	Values	Unit
Infection Range	30	meter
Infection Status	0, 1, 2	-
Infection Count	[1, 10]	-
Infection Time	[1, 200]	s

The infection range is used to limit when content can be transferred to other nodes. In [15] and [36], the communication range between mobility nodes is set to the same value of 30 meters. If a node is within this infection range, then an infected node can infect others, (i.e., send the content to another node). Here we do not consider the probability of successful transmission and we ignore the size of the content (i.e., we assume instantaneous transmission). The content is consider to be transmitted within a very short time as compared to the time before the next movement. The “Infection Status” can be one of the three different values (0, 1, 2), where 0 stands for the susceptible stage, 1 stands for the infected stage, and 2 stands for the recovered stage. The “Infection Count” is the the maximum number of times each infected node will be triggered to send content during an infection period. The “Infection Interval” is the interval of time during which an infected node will be triggered to send content.

In this SIR epidemic model, we use a simple implementation of infection. In order to achieve fairness for different nodes, the “InfectionTime” and "InfectCount" are set so that each node will not have too heavy a burden for infecting other nodes, i.e, it will not have to transfer content too many times, hence avoiding excessive energy consumption. We implemented unicast transmission for this infection. In unicast of infection, we assume that an infected node will infect others one by one, but within a very short time so that we can ignore the movement of the nodes.

In this model, we first introduce a content item into a selected node. Then this node will infect others during a period, until its infection comes to an end. We consider the system’s status after a fixed period of time, then we try to figure out what the content distribution will be within this period of time. The success rate is calculated at the end of this fixed period of time. There are two situations when infection will end. One situation is that all nodes in the simulation area are infected, which means that the content has been successfully received by all nodes. The other situation is that not all nodes are infected, but there are no infected nodes within the area, which means all nodes that were once infected have transitioned into the recovered stage and there are no longer any nodes in the infected stage which can infect susceptible nodes. In the second situation, we must calculate the success rate of the content distribution during this period.
There are two metrics for evaluating a content distribution period: success rate, delay. These metrics are:

Success rate After an infection period, the successful rate is the percentage of nodes that successfully received the content compared to the total number of nodes.

Delay The time recorded from the start of the infection to the end of the infection period.

3.3 Social network aspect of the implementation

Social behaviors are adopted in mobile network for epidemic means of content distribution in a Mobile Social Network environment [15]. The relationships between nodes are important for sharing information between friends/acquaintances. Unlike the previous model's simple probability of infection, mobile nodes with a social relationship will selectively share their content with other nodes. A simple category is used for categorize two nodes' relationship. There are only two types of relationships considered in our simulation: friends and non-friends. More categories could be applied based on setting different thresholds, but in this thesis project we only use the two types described above. In order to find the relationships among mobile nodes, we use the nodes' meeting time (denoted as M_t) to define the relationship between nodes.

We will firstly analyze the nodes' meeting situations during a fix time period. If two nodes' meeting time is above a certain threshold, we regard the two nodes as having a friendship relationship. If two nodes' meeting time is below the threshold, we regard the two nodes as not having a friendship relationship with each other. Content distribution is only succeed between two nodes with friendship relationship, other than previous fully successful situation which is in regardless of two nodes' social relationship.

In the simulation of a social network, the previous Random Waypoint Mobility Model is used as the mobility model. The simulation of a Mobile Social Network is carried out in two steps. In the first step, the simulation of mobility models were run for a period of time T. We use a period of time to determine the friendship relationship based on the calculation of previous meeting times. Different lengths of test sequences for building up social relationship were implemented during the simulation. In the second step, the infection model is applied with a social threshold for content distribution based on the relationships generated by the different test sequences. The results are collected based upon four parameters: threshold, infect time, infect count, and testing sequence length. These results are shown in section 4.3.

Chapter 4

Analysis

After setting up the various models, we use these models to simulate content distribution. In this chapter we will first analyze the influence of several important parameters in the two mobility models for content distribution. Then we analyze the infection model parameters for content distribution, e.g., success rate and delay. Finally we will simulate content distribution with social relationships and compare the results with the earlier non-social relationship models.

4.1 Evaluation with different parameters for the mobility models

In this section, different parameters of Random Waypoint Mobility (abbreviated as RM) Model and Gauss-Markov Mobility (abbreviated as GM) Model are analyzed to show how these basic mobility model parameters will influence content distribution in terms of the following aspects: successful rate and delay. There are two parameters common to both in the Random Waypoint Mobility Model and the Gauss-Markov Mobility Model: speed and number of nodes.

For the Random Waypoint Mobility Model, the pause time is important and we will examine if this pause time influences the content distribution. For Gauss-Markov Mobility Model, the randomness degree (α) is important and we will examine if it affects the content distribution.

4.1.1 Effect of speed on success rate and delay

In order to analyze the influence of speed, we fixed the other parameters of both RM and GM mobility models. Here we used 100 nodes for both models and the pause time for RM was set to zero, the α value of GM is set to 0.75. We use moderate values of parameter (node number =100 and α=0.75) in the simulation because we don't want to get extreme results. A duration of 10000s is used here for the constraint of the infection time.

We will firstly analyze the random differences for different rounds of simulations. We use four different rounds of 1, 10, 50 and 100 for success rate at the speed of 5m/s. The results of success rate distribution is showed in Figure 4.1.

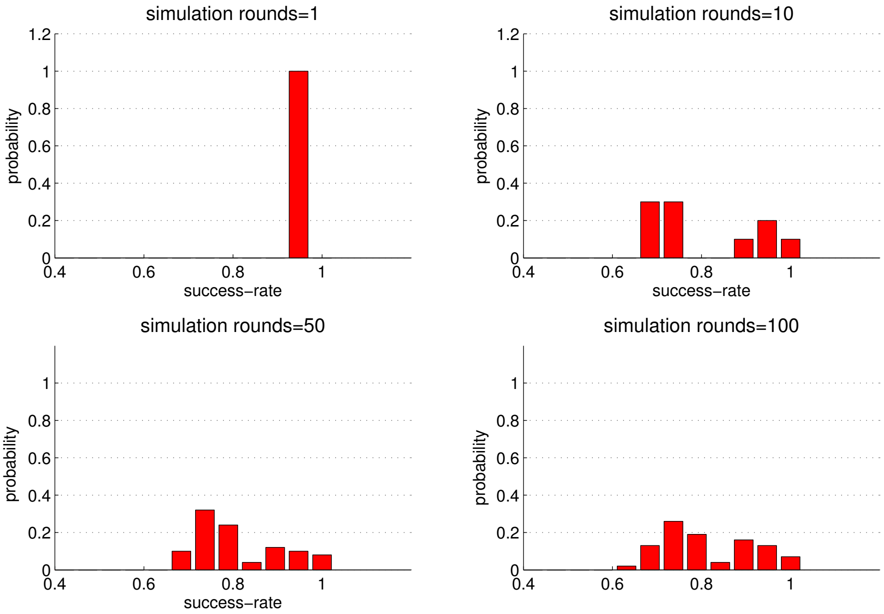


Figure 4.1: Different simulation rounds at speed of 5m/s

From the result we can see that for different rounds, different results of success rates' regions are generated. As the simulation rounds increases, the success rate region covers more situations of results which means the more rounds will enhance better coverage for results. However, as simulation rounds increases, the simulation time increases so we use maximum 100 rounds for the following simulations in terms of both good coverage and efficiency.
]

In the following simulations, each model was simulated for 100 rounds with an infection time of 10000s and the results of these 100 times are shown in the probability density bar charts. The success rates for both models are shown in Figure 4.2 and Figure 4.3. The delays for both models are shown in Figure 4.4 and Figure 4.5.

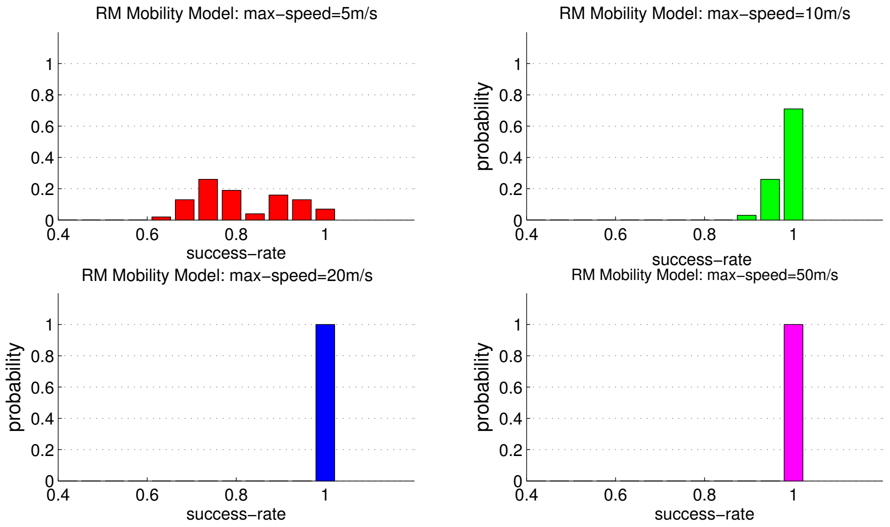


Figure 4.2: RM speed versus success rates for maximum speeds of 5, 10, 20, and 50 m/s

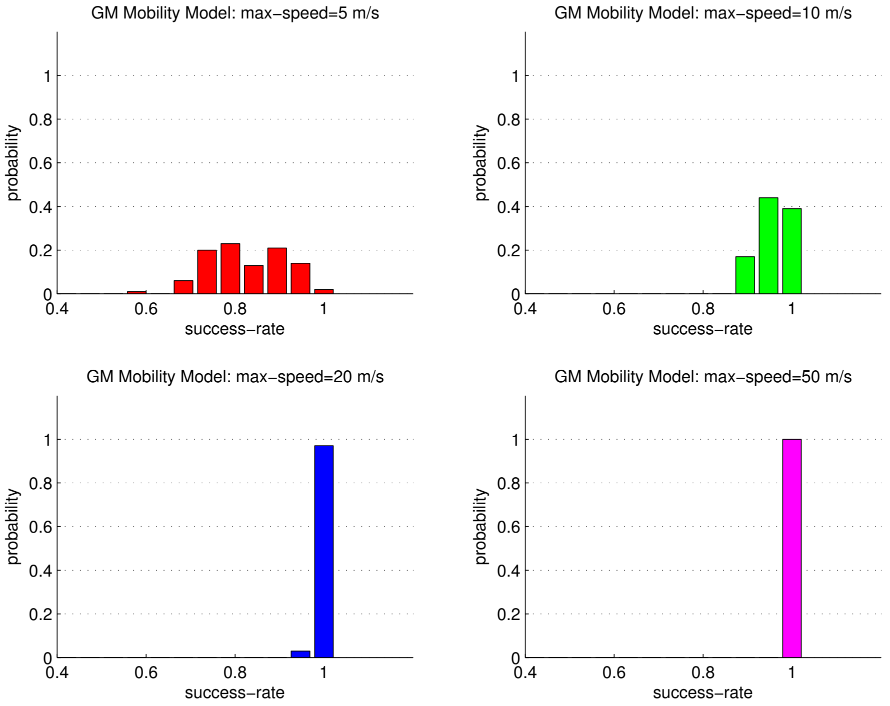


Figure 4.3: GM speed versus success rates for maximum speeds of 5, 10, 20, and 50 m/s

In the bar charts of success rate for RM and GM, as the speed increases from 5m/s to 50m/s, both models shown a trend suggesting that the success rate increases as the node speed increases. For speeds of 20m/s and 50m/s, the probability of success rate was 1 which means that there was 100% of delivery within the 10000s.

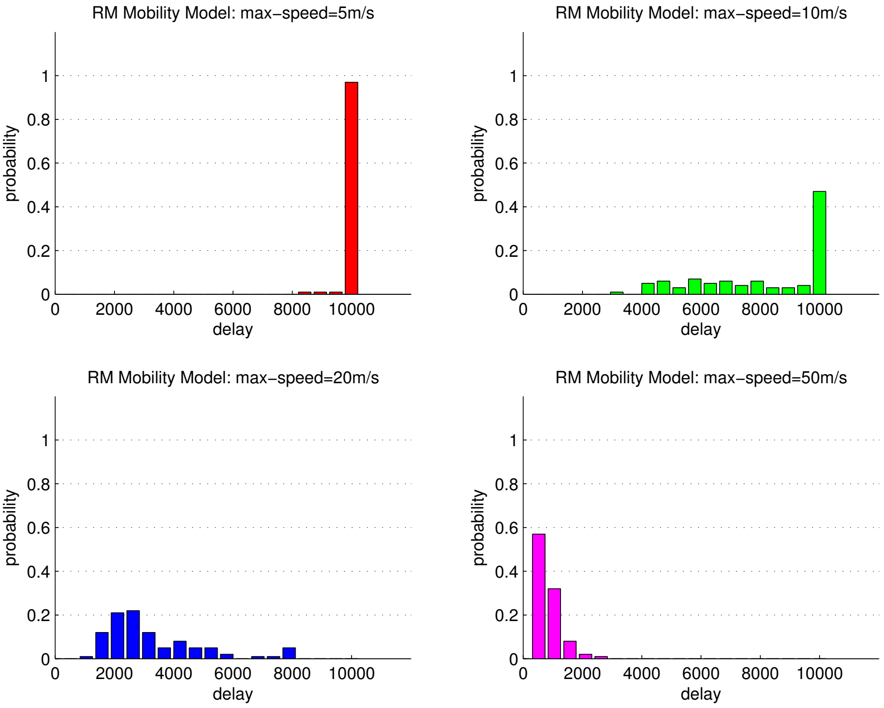


Figure 4.4: RM speed versus delays for maximum speeds of 5, 10, 20, and 50 m/s

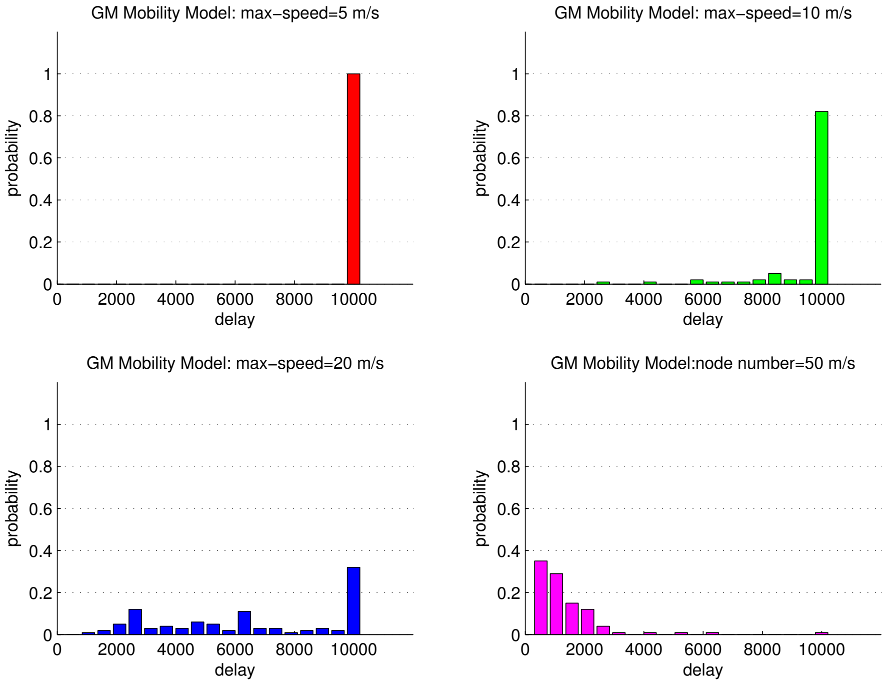


Figure 4.5: GM speed versus delays for maximum speeds of 5, 10, 20, and 50 m/s

In the bar charts of delay for both RM and GM, as the speed increases from 5m/s to 50m/s, both model shown a trend that the delay decreases as the speed increases.

Comparing the same speed for the two different mobility models GM and RM, when the speed is 10m/s, the average delay of GM is slightly higher than for RM. This same relationship is true when the speed is 5m/s and 50m/s. We draw the conclusion that with same speed, the RM mobility model is slightly more sensitive to epidemic content distribution than the GM mobility model. This is be to expected since in the RM model the nodes will shift their positions abruptly, hence there is a high probability of being near enough to infect another node in a short period of time.

4.1.2 Effect of number of nodes on success rate and delay

Within a certain area, the number of nodes will also affect content distribution. Here we fix the other parameters for both RM and GM mobility models and we test the influence of four different numbers of nodes. The same duration of 10000s is used and each model is simulated for 100 simulation runs. The success rate results for both models are shown in Figure 4.6 and Figure 4.7. The delay results for both models are shown in Figure 4.8 and Figure 4.9.

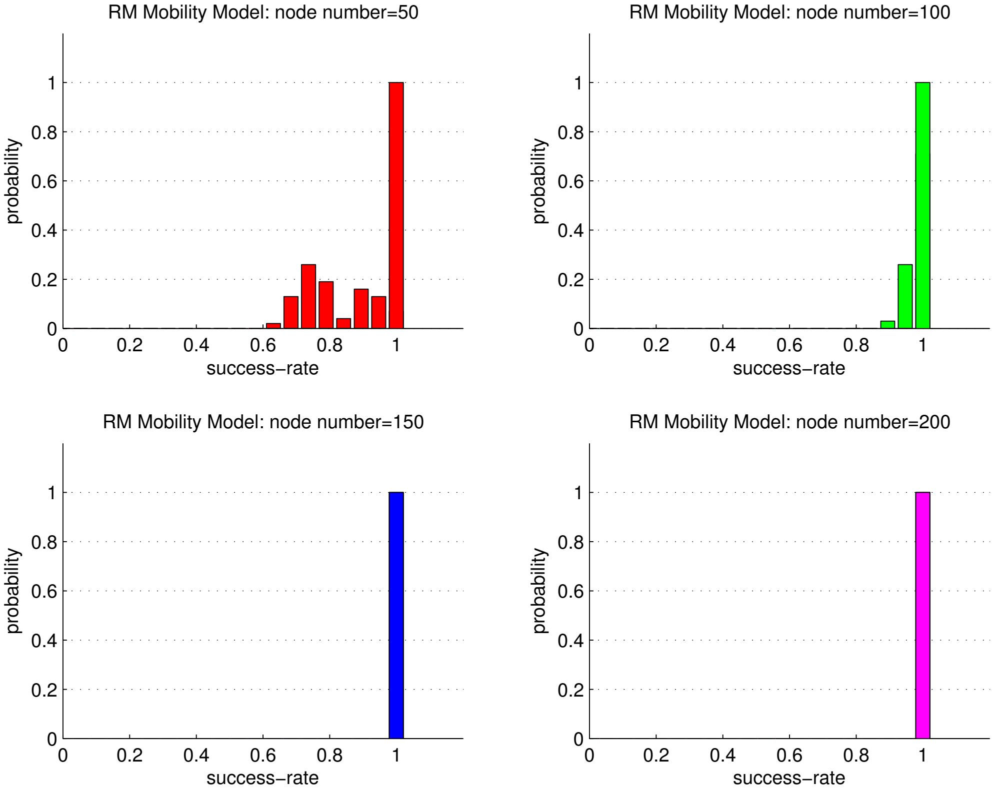


Figure 4.6: Number of nodes node versus success rate for the RM mobility model

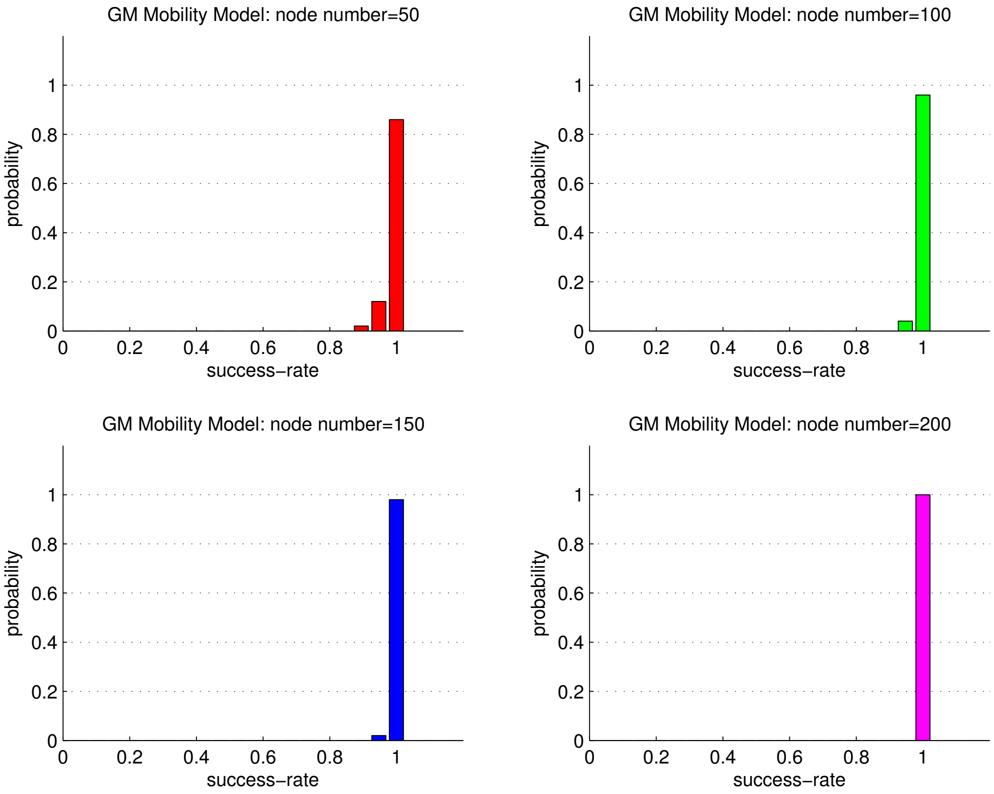


Figure 4.7: Number of nodes versus success rate for the GM mobility model

In the bar charts of success rate for RM and GM, as the number of nodes increases from 50 to 200, both sets of results shown a trend that the success rate increases as the number of nodes increases. This is to be expected because the number of nodes per units area increases, hence the probability of an infection increases.

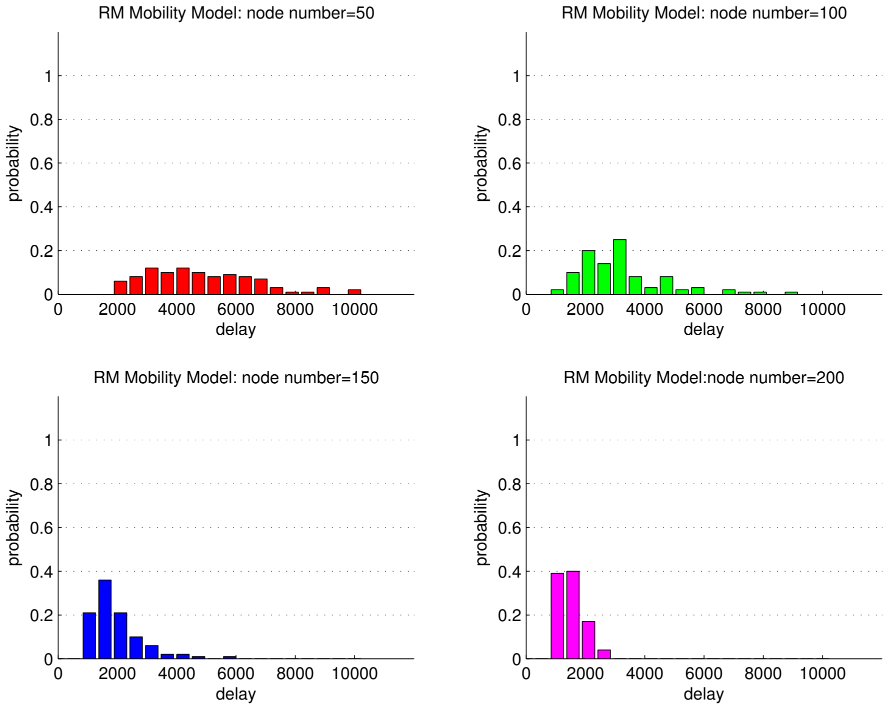


Figure 4.8: Number of nodes versus delay for the RM mobility model

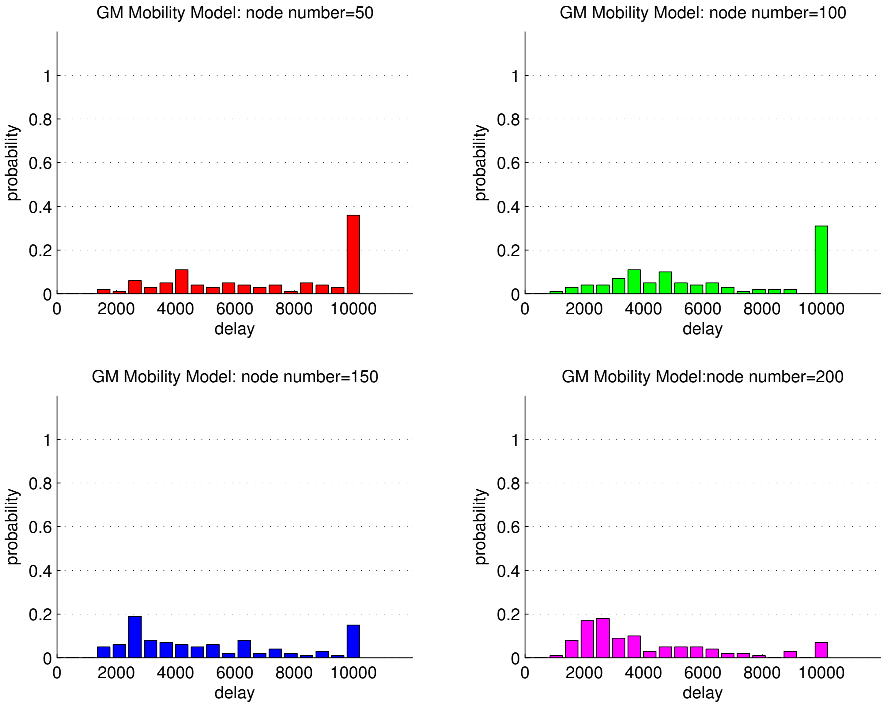


Figure 4.9: Number of nodes versus delay for the GM mobility model

The RM delay results shown an obvious trend that the width of the distribution and the mean of the distribution of the delay time decreased as the number of nodes increased. However, the GM delay results are not as obvious when compared to the RM delay results. The distribution changes slightly as the node number increases. This illustrates that the content distribution using RM mobility model is more sensitive to the number of nodes than GM mobility model.

The GM results also show that there is a very large (but decreasing) number of nodes that experience a very long delay (the peak at 10000s). The mean of the rest of the distribution does decrease as expected with increasing numbers of nodes, since the number of nodes per unit area increases with the increase in numbers of nodes. (Going from 50 to 250 nodes would reduce the average distance between nodes by a factor or 2 - hence the delay should reduce by roughly 1/2.)

4.1.3 RM pause time

The pause time in the RM mobility model will influence the content distribution. As shown in Figure 4.10 and Figure 4.11, it is clear that as expected as the pause time increases, the success rate decreases and the delay increases. Although it was shown in [12] that a longer pause creates a more stable network, here we can see that this will have a negative influence on the content distribution success rate and delay. The reason is that when increasing pause time, the average speed of the node time will be reduced, hence for the same simulation time, the node's movement will be slower with a higher pause time.

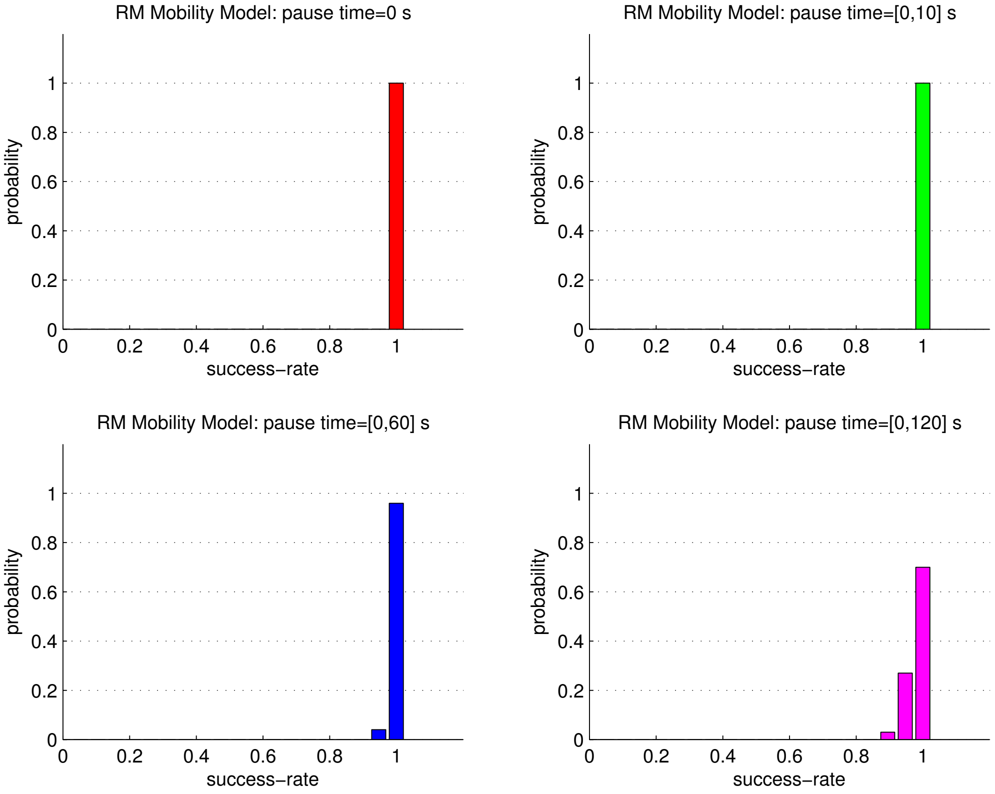
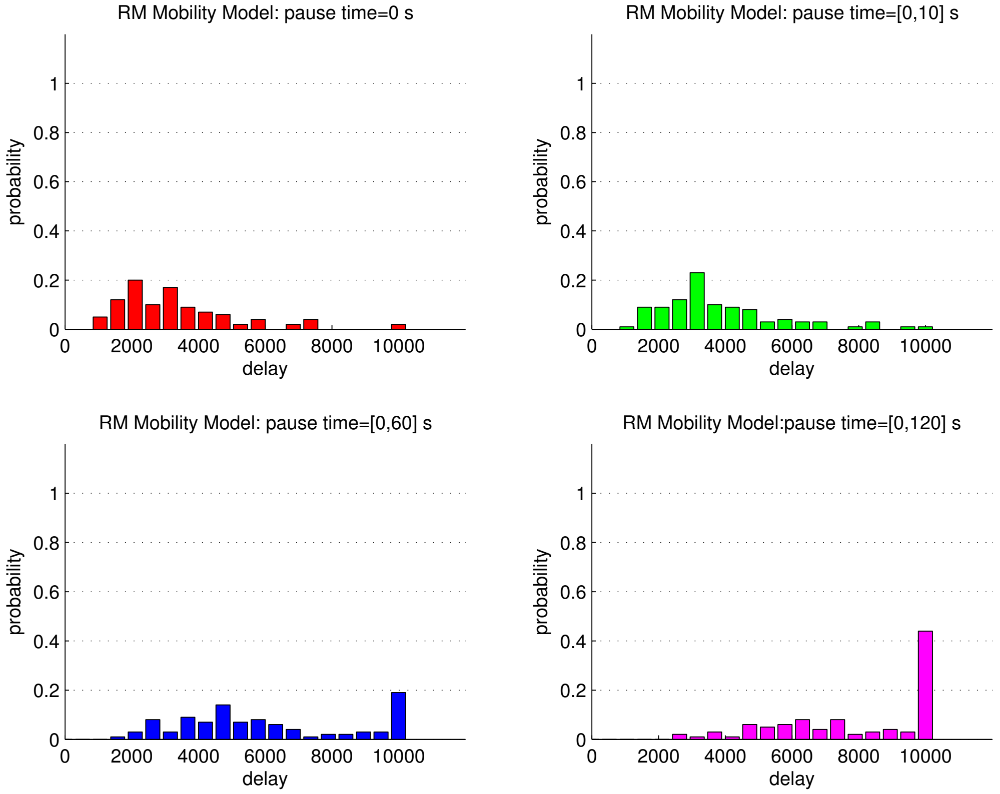


Figure 4.10: RM pause time versus success rate


RM Mobility Model: pause time=0 s

RM Mobility Model: pause time=[0,10] s

RM Mobility Model: pause time=[0,60] s

RM Mobility Model:pause time=[0,120] s

Figure 4.11: RM pause time versus delay

4.1.4 GM alpha

The value of α in the GM mobility model will influence the content distribution. Five different values were used ranging from 0 to 1. The α a tuning parameter of the randomness of the mobility which we introduced in section 3.1.2. The results are shown in Figure 4.12 and Figure 4.13.

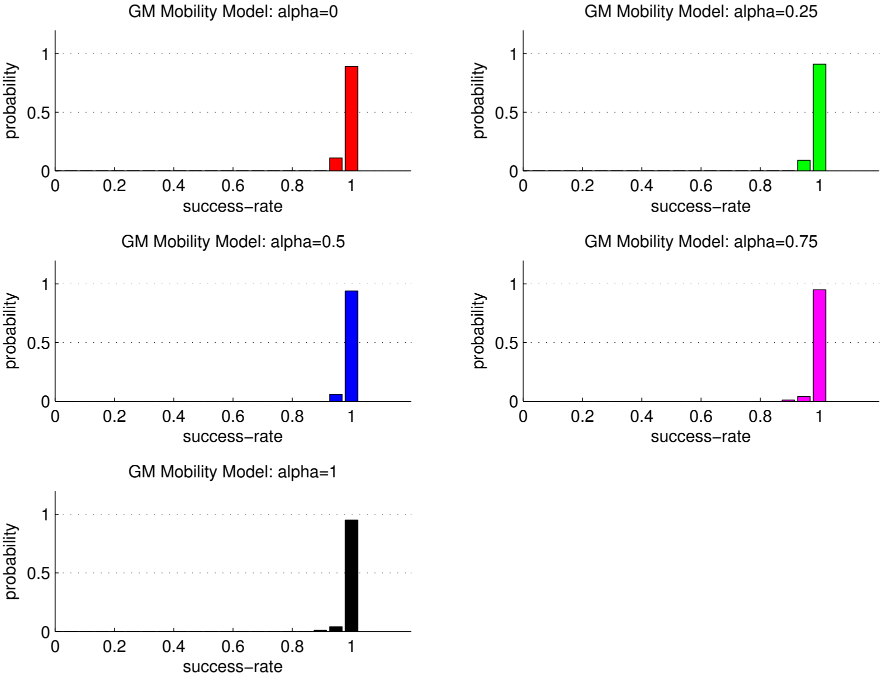


Figure 4.12: GM success rate as a function of α

The distribution for success rate for the five different values of α is almost the same, except for slightly increases in the width of the distribution around 1.

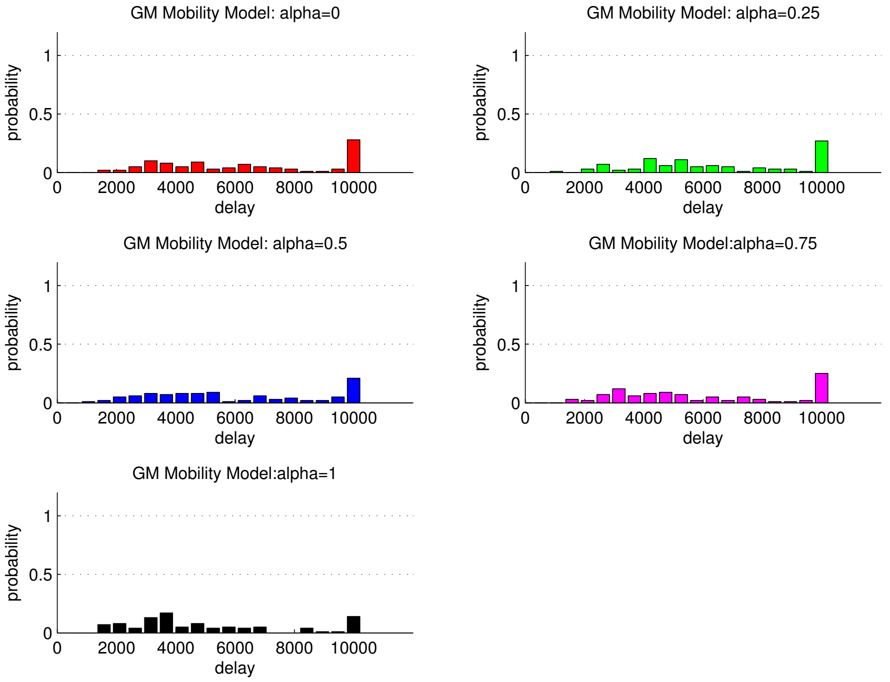


Figure 4.13: GM delay as a function of α

We can see from Figure 4.12 that there is no obvious affect as the α value increases from 0 to 1. As a result the delay time distributions are similar with no clearly trend as the α value increases from 0 to 1. From these results, we can draw the conclusion that for GM mobility, increasing the value of α value will not strongly infect the content distribution success rate or delay.

4.2 Evaluation of Epidemic model

In the simulation, we use the SIR epidemic model for content distribution and we have implemented the SIR epidemic models together with the Random Waypoint Mobility Model. In our analysis we will will consider fairness during content distribution. In this section we will analyze how different parameters in the Epidemic model affect content distribution. In this section, the results are shown when using the basic SIR epidemic model for content distribution together with a Random Waypoint Mobility Model. In this implementation, two parameters (infect time and the infect count) are introduced. We will examine how these two parameters of two epidemic models will affect the successful rate and delay for content distribution.

The simulation was categorized into four steps in section 4.2.1, section 4.2.2, section 4.2.3, and section 4.2.4. We use the RM mobility models with 100 nodes and a maximum speed of 20m/s in order to quickly distribute content. The infection range is set to 30m. The pause time was set to zero in order to quickly distribute content.

4.2.1 Content distribution without infect time and infect count limit

In the first step, we simulated the epidemic model with no limitation of infection time or infection count under the predefined Random Waypoint Mobility Model. In this situation, the infection period can be regarded as the SI epidemic model and there no nodes are removed during this infection period. The results are shown in the table below.

Table 4.1: SIR Epidemic Model distribution without infect time and infect count limit

Parameter	Values	Unit
success rate	1	-
delay	157	s

The results above are ideal results and the delay is rather small as each infected node will infect other susceptible nodes once they meet. But in reality, considerations about constraints are important from both the cost and efficient aspects. It would be rather unfair for all nodes to infect others without limit. For this reason, two parameters (infect time and infect count) are introduced. We will explore whether these two parameters contribute to fairness and efficiency.

4.2.2 Content distribution with different infect count limits

In the second step, we limit the “infect count” of our epidemic model and use the same mobility model traces generated during the previous step. We consider different limits where each node may infect others *N* times (N=1,2,3,4,5,6,7,8,9,10) before becoming the removed nodes from the set of infectors. The total simulation time is 10000s. The first result shown the infect count versus delay over the whole simulation time. The second result illustrates each node’s infection period (which we called the infect time for each node) as a distribution in ascend order. The results are shown in Figure 4.14, Figure 4.15, and Figure 4.16.

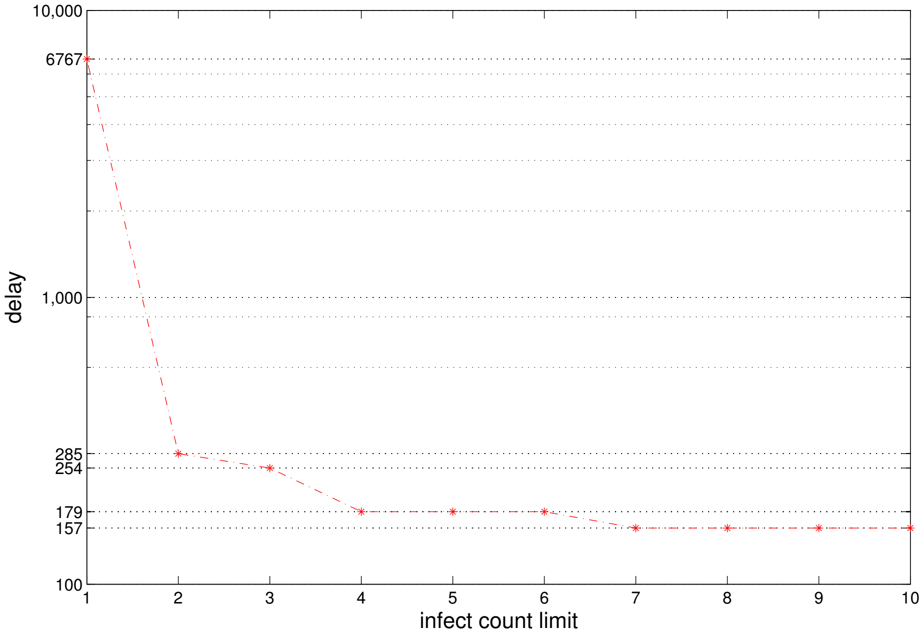


Figure 4.14: RM SIR: different infect count limits versus delay

From the result we see that each infect count limits’ delay is less than the total simulation time (10000s), which means that all content was successfully distributed within the simulation time. We do not show the success rate since the success rate is 1 for all infect count limits. The delay shows a decreasing trend as the infect count limits increases until with a limit of 6 infections the delay reaches the previous ideal delay time (of 157s) without any constraint on the infection count.

Here we analyze how the infect count limit will affect each node's infection period. Here we show a comparison of node's in an ascending order of infection periods. The results are shown in Figure 4.15 and Figure 4.16. The figure of every node's infection period distribution (a) is an overall of the results for all different count limit while the figure of every node's infection period distribution (b) is a closer look for infect count limit above 2.

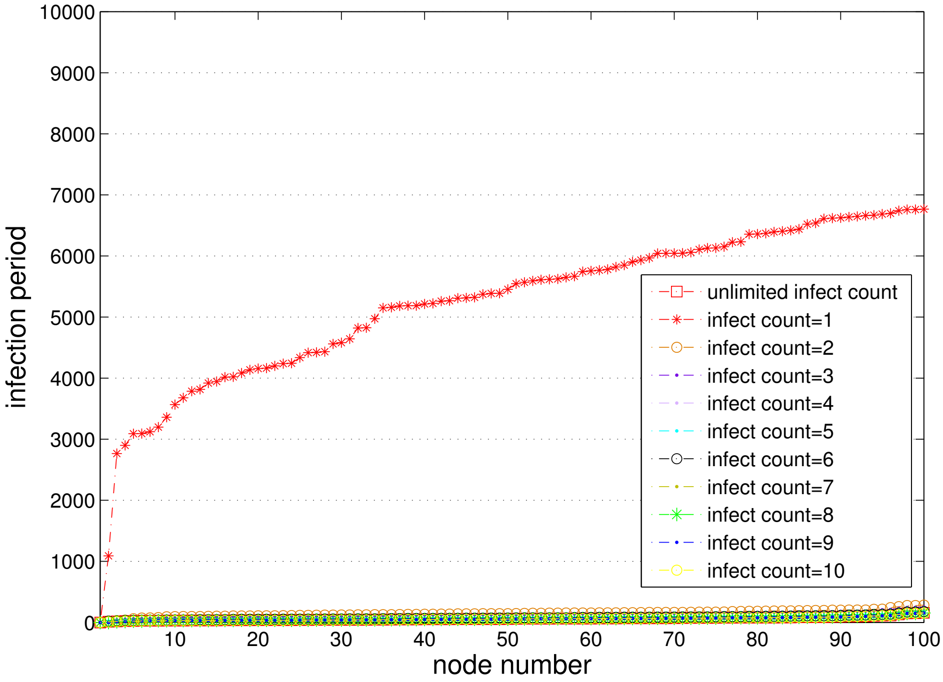


Figure 4.15: RM SIR: every node's infection period distribution (a)

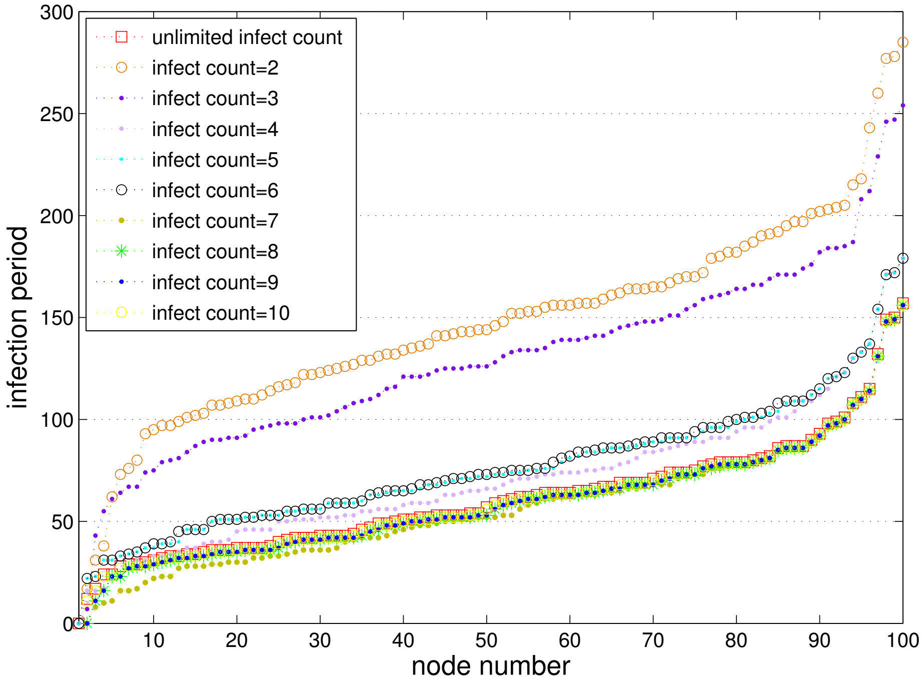


Figure 4.16: RM SIR: every node's infection period distribution (b)

It is clear from these results that nodes' infection period has a decrease trend as the infect count limit increases. When the infect count limit equals to 1, the infection period for each node is relative high, which is the situation shown in Figure 4.15. A closer look is displayed in Figure 4.16, where we can see that when the infect count limit increases from 2 to 5, the infection period decreases. However, as the infect count continually increases, the trend is not so obvious. The unlimited infect count limit behavior is almost the same as the infect count limits of 8, 9, and 10. There is a interesting discovery that when infect count equals to 7, the infection period is slightly lower than the other infect count limits as well as the unlimited infect count situation. From this result, we can see that increasing the infect count limit does not always lead to the lowest infection period for the nodes. There exists an infect count limit (in our case is infect count=7), when we can get a slightly shorter infection period while preserving the shortest delay (delay=157 for infect count 7). This situation occurs is simply because of the fact that with an infection count of 7 that the distribution has already reached the minimum delay, hence increasing the infection count limit does not have any further effect on delay but increases the time that a node is infected.

4.2.3 Content distribution with different infect time limits

In the third step, we apply the “infect time” to the same epidemic model, and each node might infect others for a period “infect time” before being removed. In the meanwhile, the infect count is set to unlimited. Content can only be delivered during the infect time. We use several different infect time values (infect time= 10, 20, 30, 40, 50, 60, 70, 80, 100, 120, 150, 160,180, 200s) for our simulations. The results are shown in Figure 4.17, Figure 4.18, and Figure 4.19.

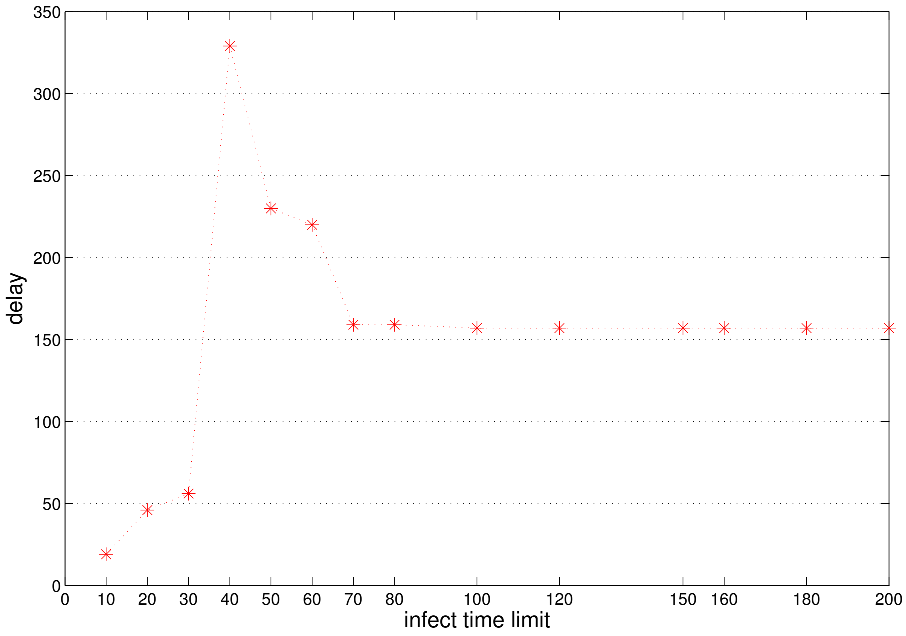


Figure 4.17: RM SIR: different infect time limits versus delay

From Figure 4.16 we can see that as the infect time increases, the delay first increases to a peak, then decreases until reaching the limit time (this limit time is the same as previous mentioned when there was unlimited infection time). The reason the delay first increases is that the infect time is so limited at the beginning that the infection is quite hard to distribute. As infect time limit increase, more nodes will be infected and nodes will have a longer period of infect time to distribute their content to others before being removed, thus the time increases as the infection period becomes longer. As more nodes are being infected, the chance of infecting new nodes decreases so there will be no added delay (since the content is already reaching every node). However, as the infect time increases, each node will have greater probability of meeting other nodes during their infection period, so the delay time decrease before reaching the limited of 157s. When the infect time is above 100s, the delay remains at 157s. The success rate associated with each infect time is shown in Figure 4.18.

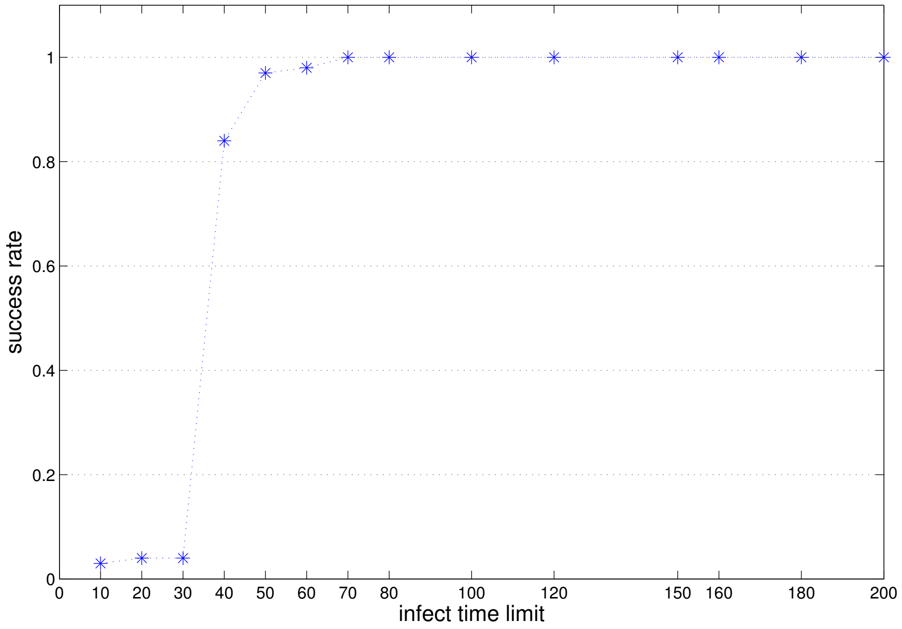


Figure 4.18: RM SIR: different infect time versus success rate

The success rate shows an increasing trend as the infect time increases. When the infect time is above 70s, the success rate is 1, which means the content is successfully distributed to all nodes once the infect time is above 70s.

Figure 4.19 shows the infect time from 10s to 120s. We simulation the infect times of 150, 180, and 200, but they shown the same results as an infect time 120s, so we use theinfect time 120s to represent the infect time above it. The results of an infect time 120s is the same as the unlimited infect time result. It is clear from the figure that the distribution of infection count for each node is quite uneven. Roughly 45% nodes do not infect others. Less than 10% of the nodes will infect more than 4 others.

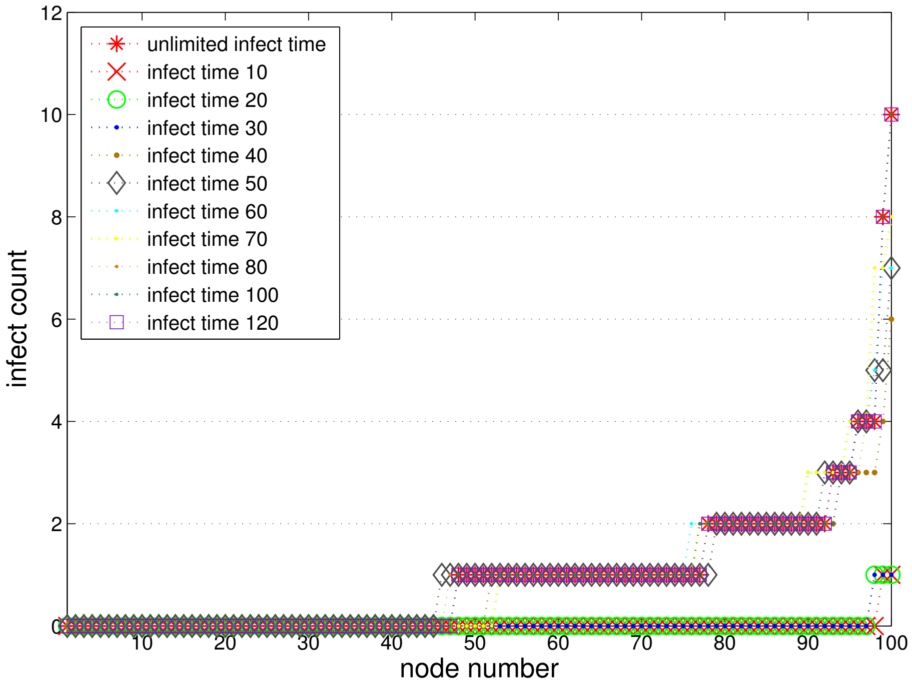


Figure 4.19: RM SIR: every node's infection count distribution

4.2.4 Discussion about content distribution with both different infect count and infect time limits

In this subsection we explore the fairness of transmissions between nodes. We implemented both infect time and infect count limits as infection constraints on content distribution in our fourth step. From the earlier results, we set the infect count limit denoted by N ($N = 1,2,3,4,5,6,7,8,9,10$) and the infect time limit denoted by T_i ($T_i = 5,10,20,30,40,50,75,100,150,200$). The success rate and delay results are shown respectively in Figure 4.20, Figure 4.21, and Figure 4.22.

From the Figure 4.19, we can see that when the infect count limit is not 1, then for every infect time limit above 50s, the success rate is above 0.9. The affect of the infect count limit is not very obvious compared to the affect of the infect time limit. When the infect time limit is 70s and infect count limit is 7, then the success rate reaches 1.

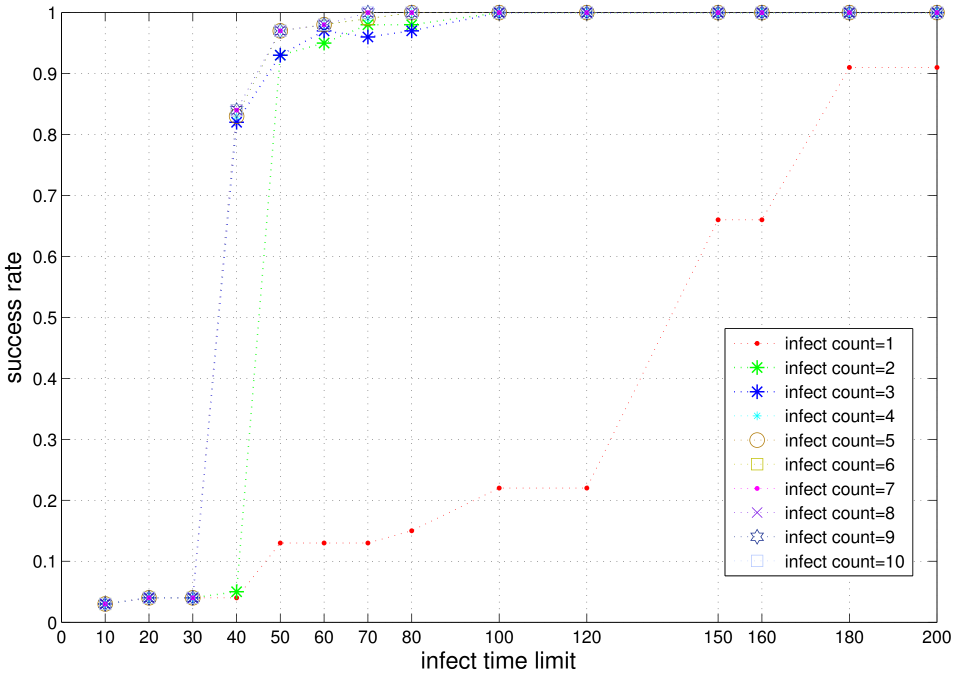


Figure 4.20: RM SIR: different infect count and infect time's success rate

In Figure 4.21 and Figure 4.22, the delay is displayed with different infect time limits and infect count limits.

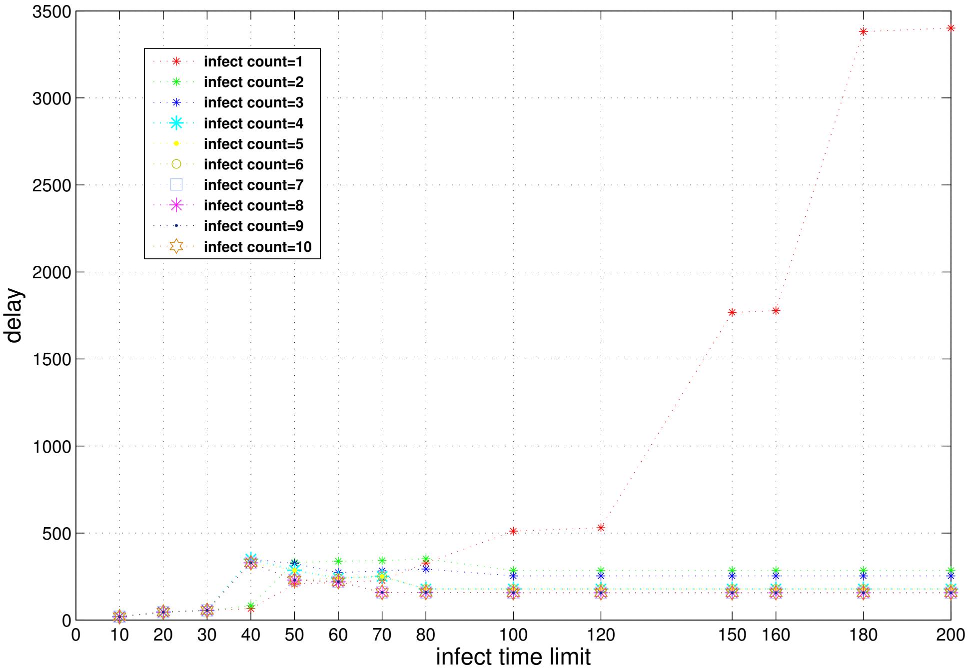


Figure 4.21: RM SIR: lowest cost with limited infect count and infect time limits on overall delay

From Figure 4.20, we can see that except for the infect count 1, all delays are below 500s. The delay shows a trend of decreasing as the infect count limit increases, which is the same as earlier results. Next we take a close look at infect count limits from 2 to 10 in Figure 4.22.

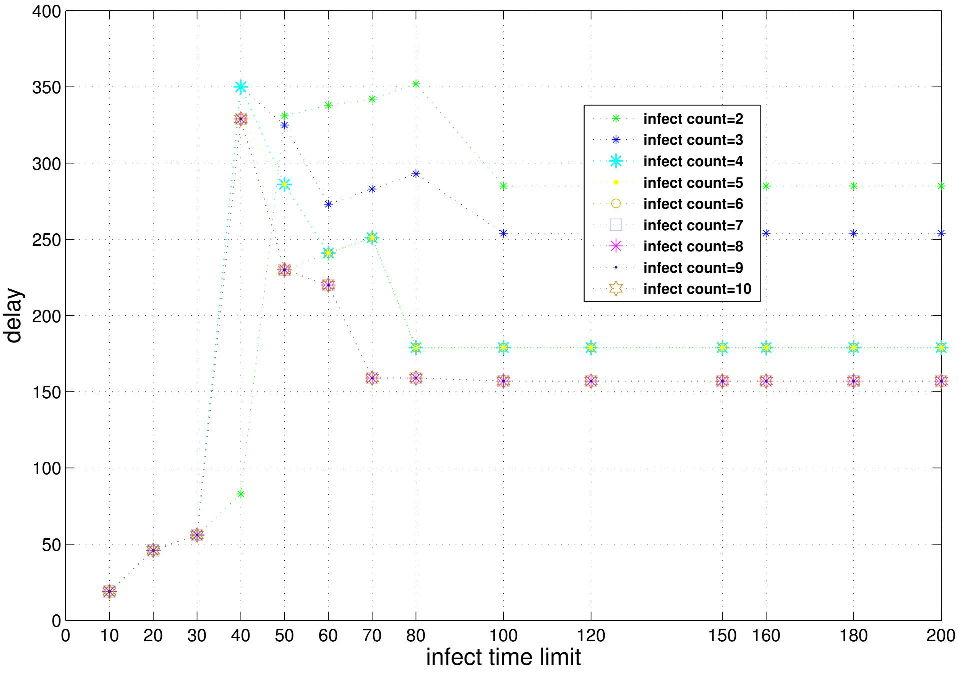


Figure 4.22: RM SIR: lowest cost with limited infect count and infect time's delay

From a close look at Figure 4.21, we can see that for a fixed infect time limit, the delay is the same as for an infect count limit over 7. The lowest delay is firstly achieved with an infect count limit of 7 and an infect time limit of 100s.

4.3 Evaluation with Social Network Implementation

In this section, a simple "Friendship evaluation method" is implemented. The simulation time for calculating people's friendship relationship is 10000s. And the simulation time for content distribution is 1000s. Using 10000s for social relationship building up provides an acceptable time bound for whole the simulation. The content distribution simulation time is 1000s since we want to examine the content distribution within a limited time range which is shorter compared to the time for building up social relationships.

4.3.1 Results of average node meeting time

We firstly analyzed the median meeting times using the Random Waypoint Mobility Model. Different simulation times' median meeting times are shown in Figure 4.23.

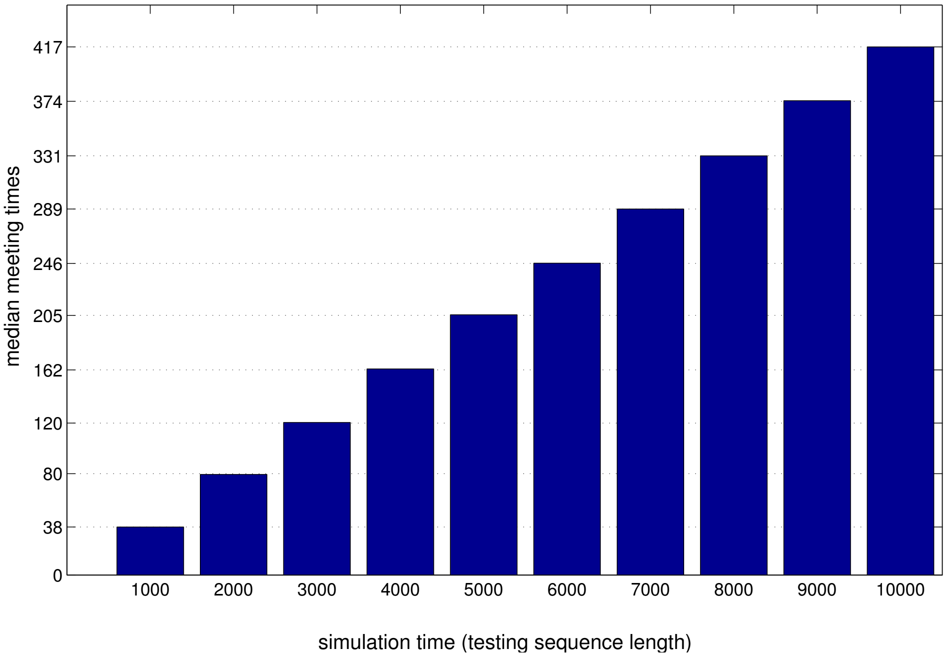


Figure 4.23: RM median meeting times vs simulation time

From Figure 4.22, it is clear that there is almost a linear increase in nodes' median meeting times as the total simulation time increases. This figure shows an almost linear trend in the relationship between simulation time and nodes' median meeting times for the Random Waypoint Mobility Model.

4.3.2 Results of different threshold situation

In the following simulations, a predefined simulation time was selected as the testing sequence time to calculate a node's social relationship. We use different duration of testing sequence times for different levels of social relationships. The length varied between 1000s to 10000s. From the results in Figure 4.23, it is clear that the median meeting times increase as the length of the testing sequence time increases. The length of the testing sequence time can also be regarded as people's familiarity level in a social network. In a social network, human interactions are important for the formation of social relationships. Relationships between people will be gradually enhanced if they are becoming familiar between each other within a society. The mobility model generated by the Random Waypoint Mobility Model can be a simplified model of the process of people becoming more familiar with each other. However, here we ignore other factors which might affect people's relationships during the relationship formation period since the mobility model used here is simply a stochastic model without specific implementations of the social aspects of the nodes.

As introduced in section 3.3, a threshold value is introduced. This threshold is a fixed meeting time value, which acts as a boundary value to define two nodes' relationship. If two nodes' meeting time is above this threshold, then content can be delivered from one node to the other, otherwise the content will not be delivered.

The affect of different threshold values on content distribution are discussed in the following. The results of different thresholds versus the content distribution success rate and delay are shown in Figure 4.24 and Figure 4.25. The content distribution time is limited to 1000 s and four different thresholds 100, 200, 300 and 400 are compared in the figure. The infection time limit of the epidemic model was fixed to 800s and infect count limit was fixed to 8 to facilitate content distribution.

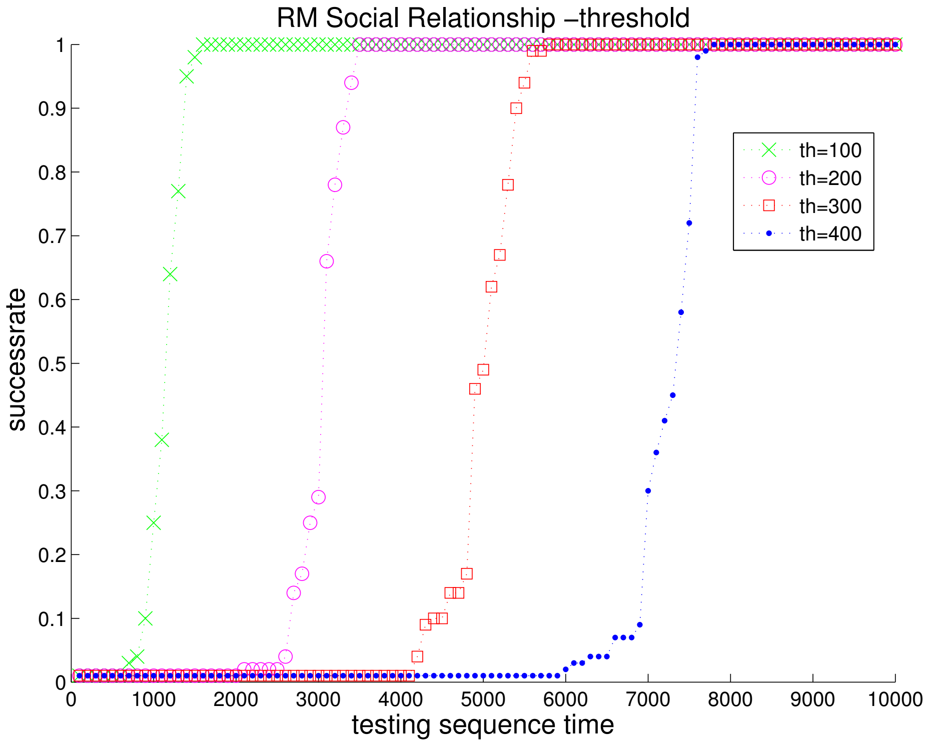


Figure 4.24: RM social threshold versus duration of relationship testing sequence

Figure 4.25 shows that as the threshold increases from 100 to 400, a longer testing sequence time is needed if the content is to be successfully distributed to all nodes (i.e, to achieve a success rate of 100 percent). Also, for each different threshold value, if we want the content distribution to be successful, then we should first find a suitable testing sequence time, above which there is the possibility of success. Since the testing sequence time represents people's familiarity level, this model also shows the social meaning of people's familiarity level increase higher, with a higher probability of successful content distribution.

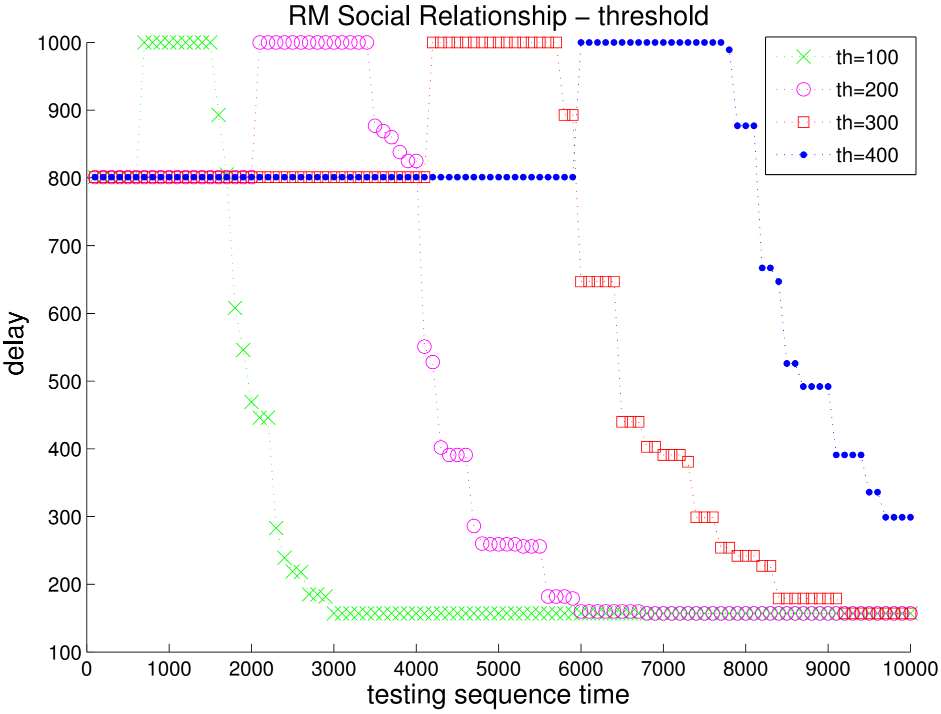


Figure 4.25: RM social threshold versus duration of relationship testing sequence

It is more clear in Figure 4.25 the delay is 800s (a preset infection time) from the beginning, and it remains 800s until the testing sequence time reaches a certain limit and then the delay increases to 1000s (which is the preset time bound for content distribution). The content can be delivered during the 1000s even though the distribution may not be 100% successful. As the testing sequence time increases, the delay drops dramatically from 1000 s to around 150s. The four different threshold shown the same decreasing trend in figure 4.25 until they reach the lowest delay. This result suggests that as people become more familiar with each other, the content distribution delay can be reduced using our model. Before reaching the lowest delay, we compare the two figures (Figure 4.24 and Figure 4.25) when the threshold is 100s. After 1500s of testing sequence time, the success rate reaches 1 and remains 1, but the delay drops as the testing sequence duration increases. A similar trend occurs for thresholds of 200, 300, and 400. We can see that for a fixed threshold, when the success rate reaches 1, there is also a certain period during which the delay decreases before reaching the lowest value.

4.3.3 Results of different infect time situation

In previous our analysis of the threshold, we fixed the infect time to be 800s. Here we compare different infect time values to show the importance of the infect time on content distribution while considering different social relationship level (i.e., the length of the testing sequence period). We fix the threshold to be 300 and the infect count is unlimited. Seven different infect time limits are compared and the results are shown in Figure 4.26 and Figure 4.27.

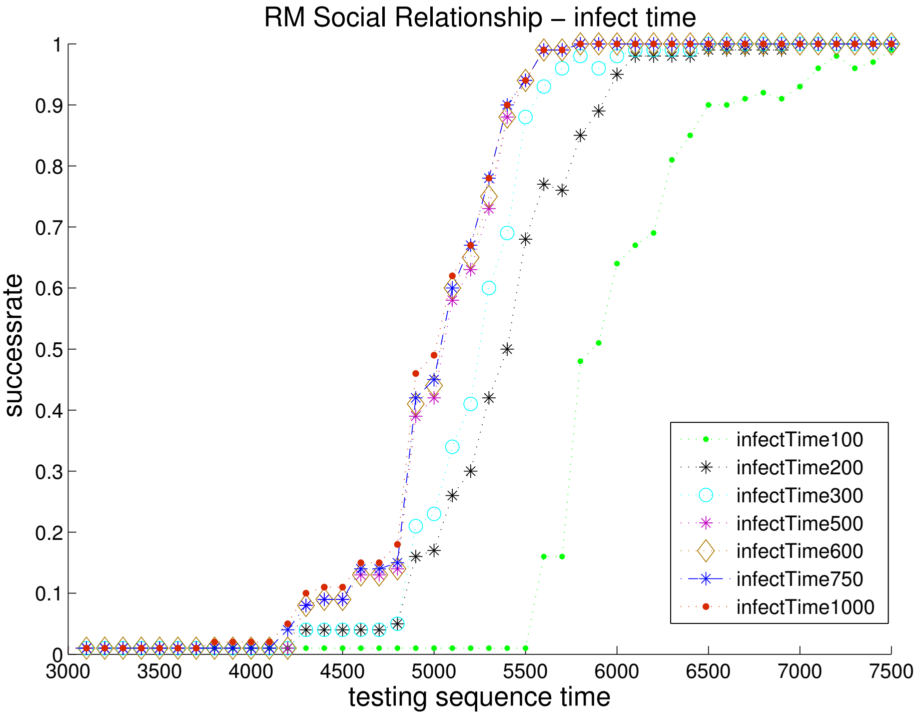


Figure 4.26: RM infect Time versus relationship testing sequence duration (in seconds)

For each infect time curve, the success rate in Figure 4.26 shows an increasing trend as the testing sequence time period increase until the success rate reach 1, which means that the content is fully distributed to all nodes in the society. Comparing different infect time limit values, with a testing sequence time of 5000s for example, the success rate shows a trend of increasing as the infect time increases from 100s to 1000s, even the content is not fully distributed. From the figure we can see that infect time limits of 500s, 600s, 750s, and 1000s share the same success rate (as the success rate equals 1) when the testing sequence length is over 5800s. From a cost efficiency point of view, if we need to fully distribute content based on a certain level of social relationship, we can determine the lowest infect time limit for which the social relationship is sufficient enough for successfully distribution.

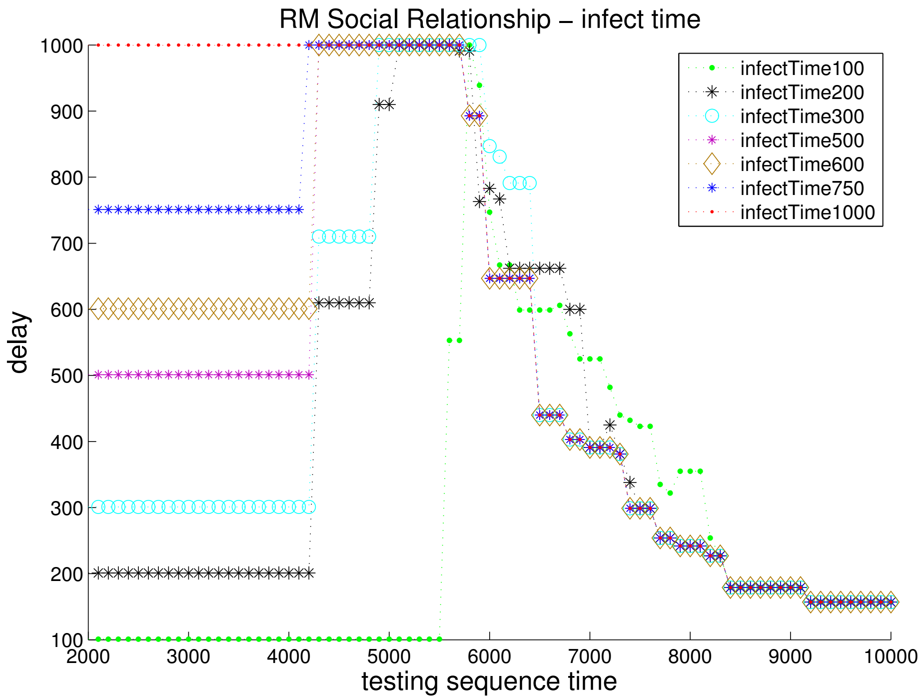


Figure 4.27: RM infect Time limit versus relationship testing sequence duration in seconds

In Figure 4.26, we can see the delay time for the beginning period equals the infect time. That is because the social relationship is too weak (i.e., the testing sequence is too short) to reach the threshold (threshold=300s) required for distributing content. At this period, the infected nodes will be removed when their infect time limits are reached, hence they will not able to infect new nodes so the infect time is the same as the delay time when no more suspected nodes exist, which means that the distribution process is finished. As the testing sequence time increases, nodes will be able to transmit content to others but only until the end of the simulation (1000s), If the content can not be successfully distributed, the delay equals 1000s (the simulations finish time). As social relationships increase, more nodes reach the threshold needed to transmit so the delay starts to drop, which means the success rate reaches to 1. The delay will drop as the social relationships increase until to a limiting value when all nodes have really high relationships so that the threshold value no longer has any effect on any nodes, so the ideal content transmission procedure will be applied to this situation.

4.3.4 Results of different infect count situation

Similar to previous simulation, here the threshold is set to 300s and same mobility trace is used. The infect time limit is set to 1000s, the same as the simulation time to make sure that this infect time limit will not affect the infect count limit's results. Ten different infect count limits (1 to 10) are tested. The results are shown in Figure 4.28 and Figure 4.29.

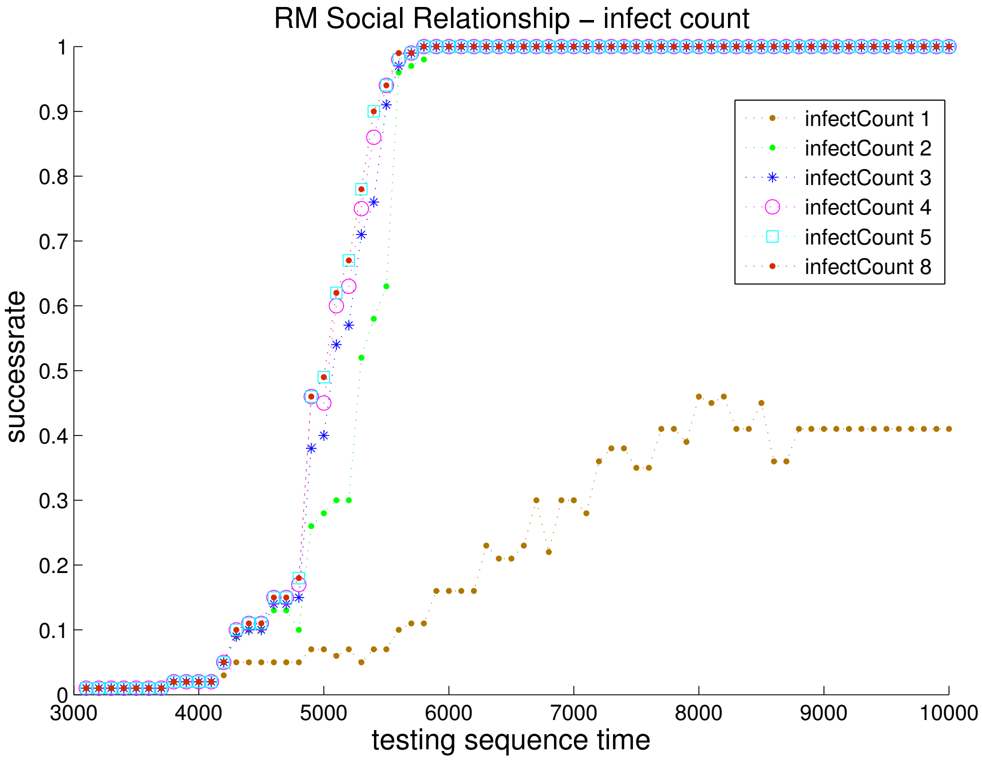


Figure 4.28: Success rate: RM infect Count versus relationship testing sequence duration in seconds

In these results we shown only six different infect count limits, because when the infect count limit is above 6, the same results are obtained so we only show the results of an infection count limit of 8 (as it represents all of the higher limits). We can see from Figure 4.28 that when the infect count limit is equal to 1, content can not be successful distribution during the simulation time (1000s). As the infect count limit increases, then above a certain social relationship with a testing sequence time 5700s in the figure, all infect count limits will lead to successful content distribution. This result shows that when infect count limit is above 6, then the success rate will be the same even as the testing sequence duration increases. It is clear that changing the infect time limit will have a slight affect on the success rate, which is not as we expected except for the situation of an infect count limit of 1.

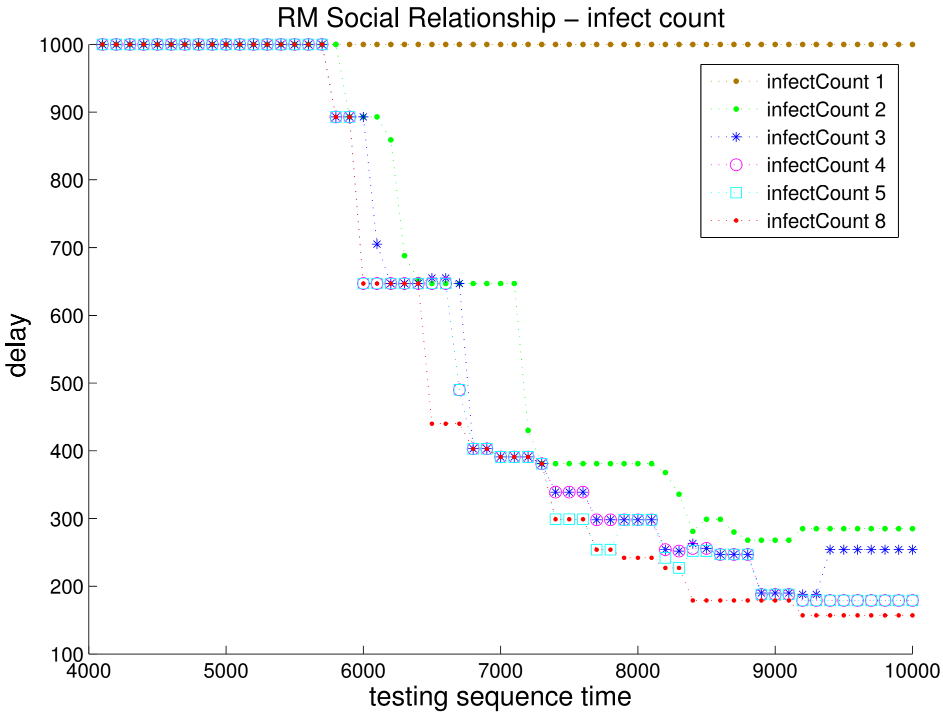


Figure 4.29: Delay: RM infect Count versus relationship testing sequence duration in seconds

Concerning the delay time in Figure 4.29, the infect Count 1 is always 1000s (simulation time) since the content can not be fully distributed no matter how the testing sequence changes. The delay is 1000s at the beginning for all infect count limits, then when the success rate reaches 1, the delay starts to decrease. For the same testing sequence time duration, the infect count limit will slightly affect the delay if it does not reach the delay limit. But as the infect count limit increases, this affect will be weaker and will be the same for all infect count limits is above 6. This result shows delay decreases as the infect count increases until the infect count limit reaching a limit. Also for all infect count limits, the testing sequence time duration is important for the delay as the social relationship gets stronger, as the delay decreases. However, there is an interesting abnormal situation on the lower right corner for infect count limit 2 and 3. As testing sequence time increases, the delay time fluctuates a bit. We can see that although social relationships level increases as testing sequence time increases, if the infect count limit is not large enough, which node is infected within the infect count limit will cause slight variance to the delay. But the overall delay decreasing trend is same as the testing sequence time increases.

Chapter 5

Conclusions and Future work

In this chapter, we will firstly state the conclusions of this project. Then we will discuss the possible future work. The chapter will end with some reflections on economic, social, and ethical issues associated with this thesis project.

5.1 Conclusion

In general, this thesis aims to study content distribution characteristics when using an epidemic distribution method and it studies the impact of several different parameters. Two stochastic mobility models were set up in a Matlab environment to provide a basic mobility environment for content distribution and an SIR epidemic model was set up for content dissemination.The impact of various parameters were evaluated for both mobility models and epidemic models. The results obtained in this thesis project are based upon statistical analysis. Unlike previous studies of epidemic routing, content distribution focuses on swiftly distributing information content to all people in a society, so the success rate and the delay time for content distribution were evaluated by simulation. The observed results as a function of the various parameters impact on both Random Waypoint Mobility Model and Gauss-Markov Mobility model showed that increasing of the speed of nodes and the number of nodes in a fixed area will have positive impact on content distribution. Longer pause time for the Random Waypoint mobility model will lead to a longer delay for content distribution. In the study of Gauss Markov mobility model, the results shown that increasing of the value of α will not have an obvious impact on the content distribution success rate or delay time. The epidemic model was further explored under constraints of infect time limits and infect count limits. In these studies each node may infect another within a limited period of time or a limited infection period. The results showed that the infect time limit for each node will be important to the success rate and latency of content distribution. If the infect time limit is too short, then distribution will be unsuccessful. However, if the infect time limit is too long, then each node will suffer from a long time while searching for other nodes, which is not be a good option from an energy consumption point of view. On the other hand, the infect count of each nodes can be limited to achieve fairness for each node. From the results concerning every node's infection count distribution, we can see that most of the nodes will not be able to contribute to delivering content to others; some other nodes will contribute highly to delivering content to others. As the infect count limit increases, the delivery delay will be shorter, thus there should be a trade off between delay time and fairness in content distribution. When each node's infection period is fixed and we need the distribution to be finished as soon as possible, it is better to increase the infect count limit so that some node will expend some more energy to distribute content to other nodes.

In addition to the above results, we also tried to simulate how social relationships will influence content distribution. Based on the mobility model, we implemented a social relationship evaluation method based on nodes' meeting time during a period of time. The Random Waypoint mobility model was used in the simulation. A threshold value is used to define a node's relationship. The testing sequence were generated based on a node's movement within a period of time. Different testing sequence durations were used to represent different levels of node's relationships. The obtained results showed that as the testing sequence increases (which means that the social relationship between nodes increases), then success rate increases until reaching 1, which means the content is fully distributed to all nodes. The delay time showed a trend of decreasing until reaching to a minimum delay. As the relationship threshold increases, a higher barrier is created for content distribution, so that a longer testing sequence time is needed if the content is still to be fully distributed. The infect time limit and infect count limit are also tested in the social relationship model. If the infect time is long enough, then the social relationship will have less impact on the success rate. But the delay time will obviously be influenced by the testing sequence length (social relationship). Higher social relationships will shorten the delay time for successful content distribution until the delay reaches a certain limit. As for infect count limit, the results do not show an obvious impact on the infect count for the success rate as the testing sequence increases. The delay time, however, will be strongly influenced as the infect count limit increases. From the results obtained from the social relationships, we showed the social relationship's important impact on content distribution. As the social relationship between nodes gets stronger, it is easier for a successful distribution to be achieved. For a successful distribution, as the social relationship between nodes gets stronger, the delay time decreases until reaching the delay of an ideal distribution.

Above all, this project has answered the questions brought up at the beginning of this thesis. The models set up under the Matlab environment can be reused by others who are interested in this topic or these models can be used to solve other problems.

5.2 Future work

In this thesis, the mobility models are both stochastic models which may not represent a specific situation or environment in real life. For these reasons the mobility model should be further improved to better suit a real scenario. Some real life traces could be taken into considerations. Also, this thesis does not explore or discuss situations with multiple infection nodes and multiple items of content. So these would be interesting topics to do research on. In this thesis project, we took a step forward in exploring social relationship's impact on content distribution. However, the distribution criteria was based on meeting times, which could be improved by considering other more relevant criteria. Last but not least, similar simulations results of other works can be compared with our model in future. In our project we use similar parameters in another work about epidemic routing, but our purpose is to swiftly distribute content in a society other than epidemic routing. So in future study, we might find some more proper simulations results from other works for comparison.

5.3 Required reflections

This study reflected the content distribution characteristics using epidemic models in a mobile network and explored the impact of social relationship a content distribution. Various parameters are examined for content distribution success rate and speed. The trade off of energy consumption and fairness issues are also addressed in the work which will have potential benefits for users (in terms of low cost and fairness) in future implementation. This study illustrate several basic constraints for content distribution between people. The social relationship evaluation in this work provides a new insight in content distribution among mobile social networks. The mobility models and epidemic models generated under Matlab environment in this project can be reused for other researchers in future study of this project or other projects, which will provide a good foundation for further researches. Considering the limitations in the work, the problems listed in future work can also provide good motivations for the continuous study.

Bibliography

[1] Jinthana Ariyakhajorn, Pattana Wannawilai, and Chanboon Sathitwiriyawong. A comparative study of random waypoint and gauss-markov mobility models in the performance evaluation of manet. In *Communications and Information Technologies, 2006. ISCIT '06. International Symposium on*, pages 894 –899, 18 2006-sept. 20 2006.

[2] Jon Kleinberg. Computing: The wireless epidemic. *Nature*, 449(7160), September 2007.

[3] Amin Vahdat and David Becker. Epidemic routing for partially connected ad hoc networks. Technical Report CS-200006, Duke University, Department of Computer Science, Durham, NC, USA, April 2000.

[4] M. Grossglauser and D. Tse. Mobility increases the capacity of ad-hoc wireless networks. In *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1360 –1369 vol.3, 2001.

[5] Robin Groenevelt, Philippe Nain, and Ger Koole. The message delay in mobile ad hoc networks. *Performance Evaluation*, 62:210 – 228, 2005. In Performance 2005, 24th International Symposium on Computer Performance, Modeling, Measurements and Evaluation.

[6] Anwitaman Datta, Silvia Quarteroni, and Karl Aberer. Autonomous gossiping: A self-organizing epidemic algorithm for selective information dissemination in wireless mobile ad-hoc networks. In Mokrane Bouzeghoub, Carole Goble, Vipul Kashyap, and Stefano Spaccapietra, editors, *Semantics of a Networked World. Semantics for Grid Databases*, volume 3226 of *Lecture Notes in Computer Science*, pages 126–143. Springer Berlin Heidelberg, 2004.

[7] Hideki Hayashi, Takahiro Hara, and Shojiro Nishio. On updated data dissemination exploiting an epidemic model in ad hoc networks. In AukeJan

Ijspeert, Toshimitsu Masuzawa, and Shinji Kusumoto, editors, *Biologically Inspired Approaches to Advanced Information Technology*, volume 3853 of *Lecture Notes in Computer Science*, pages 306–321. Springer Berlin Heidelberg, 2006.

[8] Zijie Zhang, Guoqiang Mao, and B.D.O. Anderson. On the information propagation in mobile ad-hoc networks using epidemic routing. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1 –6, dec. 2011.

[9] Abdelmajid Khelil, Christian Becker, Jing Tian, and Kurt Rothermel. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, MSWiM '02, pages 54–60, New York, NY, USA, 2002. ACM.

[10] Xiaolan Zhang, Giovanni Neglia, Jim Kurose, and Don Towsley. Performance modeling of epidemic routing. *Computer Networks*, 51(10):2867 – 2891, 2007.

[11] D.J. Klein, J. Hespanha, and U. Madhow. A reaction-diffusion model for epidemic routing in sparsely connected manets. In *INFOCOM, 2010 Proceedings IEEE*, pages 1 –9, march 2010.

[12] Tracy Camp, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. *Wireless Communications & Mobile Computing (WCMC): Special Issue On Mobile Ad Hoc Networking: Research, Trends And Applications*, 2:483–502, 2002.

[13] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. Human mobility models for opportunistic networks. *Communications Magazine, IEEE*, 49(12):157 –165, december 2011.

[14] Shusen Yang, Xinyu Yang, Chao Zhang, and E. Spyrou. Using social network theory for modeling human mobility. *Network, IEEE*, 24(5):6 –13, september-october 2010.

[15] Eyuphan Bulut and Boleslaw K. Szymanski. Exploiting friendship relations for efficient routing in mobile social networks. *Parallel and Distributed Systems, IEEE Transactions on*, 23(12):2254 –2265, dec. 2012.

[16] Anders Lindgren, Avri Doria, and Olov Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 7(3):19–20, July 2003.

[17] Elizabeth M. Daly and Mads Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '07, pages 32–40, New York, NY, USA, 2007. ACM.

[18] Henri Dubois-Ferriere, Matthias Grossglauser, and Martin Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, MobiHoc '03, pages 257–266, New York, NY, USA, 2003. ACM.

[19] Vincent Jaquet and Marek Pechal. *Epidemic spreading in a social network.* ETH Zurich, Zurich, Switzerland, December 2009. Available as `http://www.soms.ethz.ch/teaching/MatlabSpring10/projects/hs2009_1440_jaquet_pechal_epidemic_spreading.pdf`, Last Access Dec,2012.

[20] *Wikipedia Random graph.* Available as `http://en.wikipedia.org/wiki/Scale-free_network`, Last Access Dec,2012.

[21] *Wikipedia Scale-free network.* Available as `http://en.wikipedia.org/wiki/Scale-free_network`, Last Access Dec,2012.

[22] *Wikipedia Small-world network.* Available as `http://en.wikipedia.org/wiki/Small_world_network`, Last Access Dec,2012.

[23] Falko Dressler and Ozgur B. Akan. A survey on bio-inspired networking. *Computer Networks*, 54(6):881–900, April 2010.

[24] Herbert W. Hethcote. Asymptotic behavior in a deterministic epidemic model. *Bulletin of Mathematical Biology*, 35(0):607–614, February 1973.

[25] Zhifeng Sun. *Epidemic spreading survey.* Northeastern University, Boston, Massachusetts, USA, February 2009. Available as `http://www.ccs.northeastern.edu/home/austin/writeups/epispread.pdf`, Last Access Dec,2012.

[26] Mirco Musolesi and Cecilia Mascolo. Mobility models for systems evaluation. In BenoÃ®t Garbinato, Hugo Miranda, and LuÃs Rodrigues, editors, *Middleware for Network Eccentric and Mobile Applications*, pages 43–62. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[27] Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, September 2009.

[28] Marvin McNett and Geoffrey M. Voelker. Access and mobility of wireless PDA users. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9(2):45–55, April 2005.

[29] Jing Su, Alvin Chin, Anna Popivanova, Ashvin Goel, and Eyal de Lara. User mobility for opportunistic ad-hoc networking. In *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*, WMCSA '04, pages 41–50, Washington, DC, USA, 2004. IEEE Computer Society.

[30] Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. In *Proceedings of the 10th annual international conference on Mobile computing and networking*, MobiCom '04, pages 187–201, New York, NY, USA, 2004. ACM.

[31] J.J. Garcia-Luna-Aceves and M. Spohn. Source-tree routing in wireless networks. In *Network Protocols, 1999. (ICNP '99) Proceedings. Seventh International Conference on*, pages 273 – 282, oct.-3 nov. 1999.

[32] DavidB. Johnson and DavidA. Maltz. Dynamic source routing in ad hoc wireless networks. In Tomasz Imielinski and HenryF. Korth, editors, *Mobile Computing*, volume 353 of *The Kluwer International Series in Engineering and Computer Science*, pages 153–181. Springer US, 1996.

[33] Young-Bae Ko and Nitin H. Vaidya. Location-aided routing (lar) in mobile ad hoc networks. *Wirel. Netw.*, 6(4):307–321, July 2000.

[34] B. Liang and Z.J. Haas. Predictive distance-based mobility management for multidimensional pcs networks. *Networking, IEEE/ACM Transactions on*, 11(5):718 – 732, oct. 2003.

[35] Viswanath Tolety and Viswanath Tolety. *Load Reduction in Ad Hoc Networks Using Mobile Servers*. Master's thesis, Colorado School of Mines, 1999. Available as `http://libra.msra.cn/Publication/420695/load-reduction-in-ad-hoc-networks-using-mobile-servers`, Last Access Dec,2012.

[36] Ibrahim A. Hemadeh and Lieliang Yang. *A Popularity-Aided Routing Protocol for Mobile Social Networks and Its Performance*. Master's thesis, University of Southampton, 2012. unpublished.

TRITA-ICT-EX-2013:31

www.kth.se