Research Article

Improved Message Diffusion Model for Node Coverage Problem of Ad Hoc Network Based on Node Visit Times

Zhe Yang, 1 Lingzhi Li, 1 Shukui Zhang, 1,2 Yong Sun, 1 and Yanqin Zhu 1

1 School of Computer Science and Technology, Soochow University, Suzhou 215006, China
2 State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

Correspondence should be addressed to Zhe Yang; yangzhe@suda.edu.cn

Received 12 July 2013; Revised 24 October 2013; Accepted 7 November 2013

Academic Editor: Hongli Xu

Copyright © 2013 Zhe Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is of great significance to conduct researches on the message diffusion process for the node coverage problem, which can be generally abstracted as a random sampling model in the cooperative communication systems including the ad hoc and unstructured P2P networks. However, the message diffusion in the ad hoc network is not a completely independent random process. When forwarding the messages, the nodes will be influenced by such factors as degrees, visit times, and network connectivity. But the random sampling model does not take these factors into consideration, resulting in the overestimated node coverage degree. Discussing the message diffusion process of the cooperative communication systems like ad hoc network, this paper analyzes the causes of the inaccuracy problems of random sampling model and solves the problems by specially introducing the factors such as node degree and visit times. As for the E-R random network topology, it validates the effectiveness of the model proposed herein in contrast with the simulation experiment results. Compared with the random sampling model, the model proposed herein coincides better with the simulation results of the ad hoc network message diffusion process in the condition of network connectivity and its accuracy can meet the requirements for 3 and 5 visits.

1. Introduction

The nodes of ad hoc and unstructured P2P networks which are the autonomous, peer-to-peer, multihop systems need to forward messages from other nodes as well as their own messages to realize the resource sharing and cooperative communication. The searching and positioning of nodes and resources in the ad hoc network are completed by transmitting the search message within the network. Therefore, the research of message diffusion model plays an important role in the cooperative communication systems like ad hoc network [1]. Also in the social network, the message diffusion model can be used to analyze the transmission process of gossip messages by means of the social relationships of the crowd.

The main objective of analyzing the message diffusion model in the cooperative communication systems is generally to solve the problem of how many messages should be forwarded to reach the specific number of all the nodes or the specific number of messages which are given to confirm the ratio of nodes receiving the messages. In general, it is desperately expected to cover as many nodes as possible with few messages, which is the typical node coverage problem [2].

There are similar researches on node coverage conducting the keyword research through message diffusion in the unstructured P2P network [2, 3]. Although the distances of nodes in the physical networks are quite long, the specific overlay network topology can be built to achieve the link in logics to directly transmit the messages to the remote nodes to improve the efficiency of message diffusion. Take the small world network for instance; its topological structure is just between the regular network and the random network and is characterized by the small network diameter and high node aggregation, resulting in the fast speed of message transmission. Because the message diffusion process of unstructured P2P network is not limited by the physical distance, it is significant to analyze the effects of topological features of the coverage network on the efficiency and costs of message transmission to study the node coverage problems.

But in ad hoc network, the nodes can dynamically join and leave, and sometimes they even can be movable, just like
in the vehicular ad hoc network and mobile sensor network; therefore, the topological structure of the whole network is dynamic and random. And the computing and storage capacity of ad hoc nodes are too weak to receive the whole topological structure and the distribution of other nodes and resources [4], let alone building the overlay network. In addition, the communication capability of ad hoc nodes is limited and they are incapable of remote communication. In this condition, the message diffusion process is limited by the physical distance, so the message can just be transmitted to as many nodes as possible through the neighbor nodes. During this process, response time and message overhead are taken into consideration from the perspective of system performance. Methods including flooding [5, 6], random walk [7–13], and multiple random walks [14, 15] can be adopted. The flooding model is fast but its message overhead is huge. The major methods, random walk and multiple random walks, show great performance in node coverage and message overhead, but their response speeds are low.

Generally speaking, when random walk is used in transmitted messages in the ad hoc network, the ultimate state of node coverage, if the number of nodes in the network is large, can be analyzed by the random sampling model including random pick [16, 17] and coupon collector’s problem [18, 19]. But forwarding messages in the method of random walk or multiple random walks is not completely equal to “independent random sampling.” In fact, the forwarding of message is usually connected with the current node state: the probability of visiting a new node in the way of message forwarding is related to the unvisited nodes in the current network as well as the current state of the forwarding nodes. But the independent random sampling process will not take the latter into consideration; thus, when the node coverage rate at a certain time point is analyzed, the estimated value may be high.

This paper makes use of node degree and visit times to solve the problems existing in the random sampling model. The simulation experiment results show that the message diffusion model proposed herein of ad hoc network is in accordance with the reality in the condition of high node coverage. Section 2 discusses the current problems of the random sample model. Focusing on the node coverage in the message diffusion process, Section 3, based on the normal model, explores the effects of node degree and visit times to obtain the model proposed herein. Validating the effectiveness of that model through simulation experiments, Section 4 studies the identical degrees of the theoretical values and the simulation experiment results in the networks of different scales and node degrees. Section 5 makes the conclusion.

2. Random Sampling and Its Shortcomings

At present, the random walk and multiple random walks are two major methods in the research of the cooperative communication networks like ad hoc network. Both methods assume that nodes take the paths different from the message receiving paths to forward messages and messages randomly picked from all the paths (but for the coming paths) to leave the nodes. The only difference between them is that the multiple random walks have multiple paths for messages to leave. In addition to this, there is no essential difference. When the number of nodes is large enough in the network, the limit of node coverage can be analyzed by the random pick [16, 17] and coupon collector’s problem [18, 19]. Node coverage herein refers to the rate of nodes reached by the messages in the network.

2.1. Random Pick Model. Random pick model can be generally described as follows: there are $n$ balls in the box and one ball will be taken out and then returned to the box at a time. After $x$ times, will $u$ balls be taken out without repetition? The general result of this question [16, 17]: in the initial condition of $x = 0$ and $u = 0$, $u$ can be described as the function equation of $x$:

$$u(x) = n \left(1 - e^{-x/n}\right).$$

(1)

One typical application of analyzing the node coverage with the random pick model is to estimate the success rates of the search of nodes and resources: if $r$ target resources are randomly distributed in the ad hoc network, the probability of finding a target resource after sending $x$ messages, that is, the success rate of resource search $p$, can be described as the function equation of $x$:

$$p(x) = 1 - \left(1 - \frac{u}{n}\right)^r = 1 - (e^{-x/n})^r.$$  

(2)

2.2. Coupon Collector’s Problem. Coupon collector’s problem is another typical random sampling process [19], which can be generally described as follows: if there are $n$ kinds of coupon and each kind of coupon has an equal chance to be taken out, the $x$ times will be needed to collect $u$ kinds of coupon on condition that one coupon will be taken out at a time.

As we all know, $x$, number of times to collect $u$ kinds of coupon, is the function equation of $n$:

$$x(n) = n \sum_{i=1}^n \frac{1}{k} = n \cdot H_n,$$

(3)

where $H_n$ is the harmonic series of $n$, and if $n \rightarrow \infty$, then

$$H_n = \ln n + \gamma + O\left(\frac{1}{n}\right).$$

(4)

Here $\gamma$ is Euler-Mascheroni constant and $\gamma \approx 0.5772156649$. Experiment number to collect $u$ kinds of coupon, $x$, is the function equation of $u$:

$$x(u) = n \left(\sum_{i=1}^n \frac{1}{k} - \sum_{i=1}^{n-u} \frac{1}{k}\right)$$

$$= n \left(H_n - H_{n-u}\right)$$

$$= n \left(\ln n - \ln (n-u) + O\left(\frac{1}{n}\right) - O\left(\frac{1}{n-u}\right)\right).$$

(5)

When $n \rightarrow \infty$ and $u \rightarrow \infty$, we can neglect the latter 2 items; then

$$x(u) = n \ln \frac{n}{n-u}.$$  

(6)
2.3. Problems of Random Sampling Model. This shows that the limit of node coverage can be quantitatively analyzed by random sampling models including random pick model and coupon collector’s problem if there are numerous nodes in the ad hoc network \((n \to \infty)\) and the messages transmitted cover most of the nodes \((u \to \infty)\). But neither of these 2 models can exactly describe the node coverage at a time point.

Random sampling model is a limit model instead of a process description model, which is the first problem we may encounter in the analysis process of ad hoc network node coverage. Moreover, the node coverage of ad hoc network will change in the message diffusion process. This message diffusion process is not a completely independent random process, so nodes, when forwarding the messages, will be influenced by node degree and network connectivity. The probability of forwarding the messages to the new nodes will depend on the undiscoversd nodes in the network as well as the current state of the forwarding node. However, random sampling model just assumes that these two samplings are independently random, without taking into consideration the probability that the current nodes may have been visited for many times. As a result, the estimated value may be high if the random sampling model is used to estimate the ad hoc network node coverage at a certain time point.

3. Improved Ad Hoc Message Diffusion Model Based on Visit Times

When analyzing the node coverage process of cooperative communication systems like ad hoc network, the random sampling model does not take the current node state into consideration; in order to solve this problem, this paper introduces the factors such as node degree and visit times. First, presenting a general algebraic model to conduct the quantitative analysis on the node coverage of ad hoc network message transmission, this part analyzes the effects of node degree and visit times on the node coverage to propose the modified model herein.

3.1. General Algebraic Model. Suppose that ad hoc network topology is in compliance with E-R random model [20], as shown in \(G(n, p)\); there are \(n\) nodes in the graph and the link probability between 2 random nodes is \(p\) \((0 < p < 1)\), so the average degree of the node is \(p \cdot (n - 1)\). To make it simple, we assume that the node degree, \(d\), is \(p \cdot (n - 1)\). Imagine that the graphs are connected and nodes take the paths different from the message receiving paths to forward messages, which means the messages randomly picked from all the paths (but for the coming paths) to leave the nodes. At a time point of the message diffusion process, \(x\) refers to the number of messages forwarded and \(u\) refers to the number of nodes that have received the message. As time goes on, \(u\) will enlarge with the increase of \(x\) and gradually tend to approach \(n\). So \(u\) can be described as the function \(u(x)\). At this time, if the node forwards the message to the neighboring nodes, then the probability that the message reaches a new node is \((n - u) / (n - 2)\). After the message is forwarded successfully, the number of nodes which can be covered by the message is \(u + (n - u) / (n - 2)\). When \(n\) is large enough, \(n - 2\) can be replaced by \(n\):

\[
\begin{align*}
\frac{u(x + 1) - u(x)}{n} &= 1 - \frac{u(x)}{n} \\
\Rightarrow u(x + 1) &= u(x) + \frac{n - u(x)}{n}.
\end{align*}
\]

(8)

If \(n \to \infty\), then

\[
\begin{align*}
\frac{u(x + 1) - u(x)}{n} &= 1 - \frac{u(x)}{n} \\
\Rightarrow u(x + 1) - u(x) &= 1 - \frac{u(x)}{n} \\
\Rightarrow u'(x) &= 1 - \frac{u(x)}{n}.
\end{align*}
\]

The solution of the equation is

\[
u(x) = C \cdot e^{-x/n} + n.
\]

(10)

\(C\) is a constant dependent on the initial conditions which also determined its value. Generally, when \(x = 0\), \(u(x) = 0\); then we can get

\[
u(x) = n \cdot \left(1 - e^{-x/n}\right).
\]

(11)

3.2. Parameter Modification. General algebraic model can conduct quantitative analysis on the node coverage at a certain time point in the message forwarding process of cooperative communication systems like ad hoc network. But it does not consider the node degree, resulting in the overestimated value of node coverage. The general algebraic model assumes that, with \(x\) messages having been transmitted to \(u\) nodes, the probability of message transmitted to new nodes, according to (8), equals the ratio of the number of unvisited nodes and the total number of nodes, that is, \((n - u(x))/n\). The assumption is available when the current node is the first node that the message passes by and the node randomly picks a link to forward the message. In this case, the rest of the links are new. So the probability of the message reaching a new node through any link is dependent on the proportion of new nodes in the network. However, the probability that the message will forwarded by the current node to a new node decreases if the message has visited the node. The reason is that the message may take the former visited link to reach the visited nodes with repetition.

So the probability of the message forwarded to new nodes is related to the current node degree. When the message visits
the node again, the higher the node degree, the higher the probability that the message forwarded through unvisited links, and finally the probability of reaching new nodes will be close to the probability of the unvisited nodes in the network, that is, \(^{(n - u(x))}/n\); conversely, the probability of reaching new nodes is lower. In some extreme cases, the parameters \(p_i^1\) and \(q_i^1\) are undetermined, and they, according to definitions, are all connected with node visit times. The calculation processes of these two parameters are as follows.

This paper changes (8): if \(x\) messages are forwarded to \(u\) nodes, the probability that the messages forwarded to new nodes through unvisited links will be \(p(x)\), and (8) will be changed as follows:

\[
u(x + 1) = u(x) + \frac{n - u(x)}{n} p(x).
\]

In the network of \(n\) nodes, if \(x\) messages are forwarded to \(u\) nodes, then the number of visit of every node is different from each other. Just as mentioned before, if the node degree is just \(p(n - 1)\), the probability of unvisited links of every node is different. The more the visit times, the smaller the probability of unvisited links of the node. \(p(x)\), the probability of unvisited links in the network, is related to 2 factors: the probability that any node \(i\) is visited for \(j\) times and the probability of unvisited links of every node in this condition. So \(p(x)\) is the weighted average of visit times of all the nodes and the probability of the unvisited links of nodes in the network.

Before presenting the computing method of \(p(x)\), we firstly give the symbol definitions:

- \(d\): node degree;
- \(q_i^j\): probability that the node \(i\) is visited for the \(j\) time;
- \(p_i^j\): probability that there is at least one unvisited link when the node \(i\) is visited for the \(j\) time;
- \(k_i^j\): the average number of visited links when the node \(i\) is visited for the \(j\) time;
- \(p_i^j(m)\): probability that there are \(m\) \((0 \leq m \leq d)\) visited links when the node \(i\) is visited for the \(j\) time;
- \(p_i^j(n \mid m)\): probability that the number of visited links changes from \(m\) to \(n\) \((0 \leq m \leq n \leq d\) and \(0 \leq n - m \leq 2)\) visited links when the node \(i\) is visited for the \(j\) time.

As a result, the computing equation of \(p(x)\) is

\[
p(x) = \sum_{i=1}^{n} \sum_{j=1}^{x} p_i^j \cdot q_i^j.
\]

### 3.3. Effects of Visit Times

From (13), we may find it is necessary to define 2 parameters to calculate the value of \(p(x)\): \(p_i^j\) and \(q_i^j\). According to the definitions, both of them are related to the visit times of nodes. The calculation processes of these two parameters are as follows.

1. **Probability of Nodes Visited for Multiple Times.** If \(u\) nodes are covered after \(x\) messages are forwarded, then the probability of any node \(i\) visited for \(j\) times is

\[
q_i^j = C_x^{j-1} \cdot \frac{1}{n}^{j-1} \cdot \left(1 - \frac{1}{n}\right)^{x-j+1}
\]

2. **Probability That There Are Unvisited Links after the Nodes That Have Been Visited for Multiple Times.** It is necessary to consider multiple conditions of visit times \(j\) to calculate the probability of unvisited links after the nodes that have been visited for multiple times \(p_i^j\); the following examples give the specific analysis processes, as shown in Figure 1 \((d = 8)\).

   (i) \(j = 1\). The message first visits one node, which means that the visit time of that node \(j\) is 1; at that time, the initial state of the node is stated as shown in Figure 1(a), which satisfies

\[
k_i^1 = 0,
p_i^1(0) = 1.
\]

   Because all the links are new \((k_i^1 = 0)\), the message can pick any one link to reach another new neighboring node after the message reaches the current node; then

\[
p_i^1(2 \mid 0) = 1.
\]

   The node state changes to state \(b\), shown in Figure 1(b), and the number of visited links is 2, which means that the message takes 2 different links to reach and leave the node. Because of the only state to transfer the path, after the first visit of the node, the probability that the number of visited links changes from 0 to 2 is

\[
p_i^1(2 \mid 0) = 1.
\]

(ii) \(j = 2\). If the message secondly visits the node \((j = 2)\), shown in Figure 1(b), there may only be 2 visited links \((k_i^2 = 2)\) for the current node:

\[
k_i^2 = 2,
p_i^2(2) = p_i^1(0) \cdot p_i^1(2 \mid 0) = 1.
\]

At this time, the node randomly picks one link to forward the message which can only take the left \(d - k_i^2\) unvisited links to reach the new node:

\[
p_i^2 = \frac{d - k_i^2}{d} = \frac{d - 2}{d}.
\]

After the message leaves the node again, its link state will change to the third condition in Figure 1(c, \((c_i - c_3))\); its transition condition and probability are as follows.

(a) \(b \rightarrow c_i\) (old in old out). The message visits the node through the visited link twice and leaves the node through another visited link; thus, the number of visited link is 2. Then the conditional probability that the number of visited link is still 2 after the node has been visited twice is

\[
p_i^2(2 \mid 2) = \frac{2}{d} \cdot \frac{1}{d - 1}.
\]
(b) $b \rightarrow c_2$ (old in new out and new in old out). The message visits the node through the visited link for twice and leaves the node through an unvisited link, or the message visits the node through the unvisited link for twice and leaves the node through a visited link. Then the conditional probability that the number of visited link changes from 2 to 3 is

$$p_i^3(3 \mid 2) = \frac{2d-2}{d} \cdot \frac{d-2}{d} \cdot \frac{2}{d-1} = \frac{(2d-2)(d-2)}{d(d-1)} = \frac{2d-2}{d^2-1},$$

(21)

(c) $b \rightarrow c_3$ (new in new out). The message visits the node through an unvisited link for twice and leaves the node through an unvisited link; thus, the number of visited link is 4. Then the conditional probability that the number of visited link changes from 2 to 4 is

$$p_i^4(4 \mid 2) = \frac{d-2}{d} \cdot \frac{d-3}{d-1},$$

(22)

(iii) $j = 3$. If the message visits the node thirdly ($j = 3$), the initial state of the node may have 5 conditions, shown in Figure 1(c, ($c_2-c_3$)), and the numbers of visited links are, respectively, 2, 3, and 4. Because all these states are transferred from state $b$, the probabilities that there are 2, 3 of 4 visited links are, respectively,

$$p_i^2(2) = p_i^2(2) \cdot p_i^2(2 \mid 2),$$

$$p_i^3(3) = p_i^2(2) \cdot p_i^2(3 \mid 2),$$

$$p_i^4(4) = p_i^2(2) \cdot p_i^2(4 \mid 2).$$

(23)

And the mathematical expectation of the visited link number of the node is

$$k_i^3 = \sum_{m=2}^{4} m \cdot p_i^3(m).$$

(24)

The probability that the node takes an unvisited link to forward the message is

$$p_i^3 = \frac{d-k_i^3}{d}.$$  

(25)

After the message leaves the node thirdly, the node state will change from state $c_1-c_3$ to state $(d_1-d_5)$, shown in Figure 1(d); the state transition condition and probability are, respectively,

(a) $c_1 \rightarrow d_1: p_i^3(2 \mid 2) = \frac{2d-2}{d}, \frac{1}{d^2-1},$

(b) $c_1 \rightarrow d_2: p_i^3(3 \mid 2) = \frac{2d-2}{d} \cdot \frac{d-2}{d-1},$

(c) $c_1 \rightarrow d_3: p_i^3(4 \mid 2) = \frac{d-2}{d} \cdot \frac{d-3}{d-1},$

(d) $c_2 \rightarrow d_2: p_i^3(3 \mid 3) = \frac{3}{d} \cdot \frac{2}{d-1},$

(e) $c_2 \rightarrow d_3: p_i^3(4 \mid 3) = \frac{3}{d} \cdot \frac{d-3}{d-1},$

(f) $c_2 \rightarrow d_4: p_i^3(5 \mid 3) = \frac{d-3}{d} \cdot \frac{d-4}{d-1},$

(g) $c_3 \rightarrow d_3: p_i^3(4 \mid 4) = \frac{4}{d} \cdot \frac{3}{d-1},$

(h) $c_3 \rightarrow d_4: p_i^3(5 \mid 4) = 2 \cdot \frac{4}{d} \cdot \frac{d-4}{d-1},$

(i) $c_3 \rightarrow d_5: p_i^3(6 \mid 4) = \frac{d-4}{d} \cdot \frac{d-5}{d-1}.$

(26)

(iv) $j = 4$. If the message visits the node fourthly ($j = 4$), the initial state of the node may have 5 conditions, shown in Figure 1(d, ($d_1-d_6$)) and the numbers of visited links are, respectively, 2, 3, 4, 5, and 6. Because all these states are transferred from state $c_1-c_3$, the probabilities that there are $m$ ($m = 2, 3, 4, 5, 6$) visited links are, respectively,

$$p_i^4(2) = p_i^3(2) \cdot p_i^3(2 \mid 2),$$

$$p_i^4(3) = p_i^3(2) \cdot p_i^3(3 \mid 2) + p_i^3(3) \cdot p_i^3(3 \mid 3),$$

$$p_i^4(4) = p_i^3(2) \cdot p_i^3(4 \mid 2) + p_i^3(3) \cdot p_i^3(4 \mid 3) + p_i^3(4) \cdot p_i^3(4 \mid 4),$$

$$p_i^4(5) = p_i^3(3) \cdot p_i^3(5 \mid 3) + p_i^3(4) \cdot p_i^3(5 \mid 4),$$

$$p_i^4(6) = p_i^3(4) \cdot p_i^3(6 \mid 4).$$

(27)

The mathematical expectation of visited link number of the node is

$$k_i^4 = \sum_{m=2}^{6} m \cdot p_i^4(m).$$

(28)

The probability that the node forwards the message through an unvisited link is

$$p_i^4 = \frac{d-k_i^4}{d}.$$  

(29)

After the message leaves the node fourthly, the conditional probability of the message state transition ($m = 2, 3, 4, 5, 6$) is

$$p_i^4(m \mid m) = \frac{m}{d} \cdot \frac{m-1}{d-1},$$

$$p_i^4(m + 1 \mid m) = 2 \cdot \frac{m}{d} \cdot \frac{d-m}{d-1},$$

$$p_i^4(m + 2 \mid m) = \frac{d-m}{d} \cdot \frac{d-m-1}{d-1},$$

(30)
\( p_i^j (m) \)

\[
\begin{aligned}
&= \begin{cases} 
    p_i^{j-1} (m) \cdot p_i^{j-1} (m \mid m), & m = 2, \\
    p_i^{j-1} (m-1) \cdot p_i^{j-1} (m \mid m-1) + p_i^{j-1} (m) \cdot p_i^{j-1} (m \mid m), & m = 3, \\
    p_i^{j-1} (m-2) \cdot p_i^{j-1} (m \mid m-2) + p_i^{j-1} (m-1) \cdot p_i^{j-1} (m \mid m-1), & m = 2j - 3, \\
    p_i^{j-1} (m-2) \cdot p_i^{j-1} (m \mid m-2), & m = 2j - 2.
\end{cases}
\end{aligned}
\]

(31)

(iv) \( j > 4 \). Generally, if the message visits the node for the \( j \) times (\( j > 4 \)), the value range of \( m \), the number of old node link, is \( m \in [2, 2j - 2] \). The probability that there are \( m \) visited links for the node is shown in (31). The mathematical expectation of visited link number of the node is shown in

\[
k_i^j = \sum_{m=2}^{2j-2} m \cdot p_i^j (m).
\]

(32)

The probability that the node takes an unvisited link to forward the message is shown in

\[
p_i^j = \frac{d - k_i^j}{d}.
\]

(33)

After the message leaves the node, the conditional probability of the transition of its old and new states is shown in

\[
p_i^j (m \mid m) = \frac{m}{d} \cdot \frac{m - 1}{d - 1},
\]

\[
p_i^j (m + 1 \mid m) = 2 \cdot \frac{m}{d} \cdot \frac{d - m}{d - 1},
\]

\[
p_i^j (m + 2 \mid m) = \frac{d - m}{d} \cdot \frac{d - m - 1}{d - 1}.
\]

(34)

3.4. Message Diffusion Model Based on Visit Times. According to Sections 3.2 and 3.3, integrate (12), (13), (14), and (33) to obtain the improved model herein:

\[
\ u(x+1) = u(x) + \frac{n-u(x)}{n} \sum_{i=0}^{n} \sum_{j=0}^{x} p_i^j \cdot q_i^j.
\]
$q_i^j = C_x \cdot \left(\frac{1}{n}\right)^i \cdot \left(1 - \frac{1}{n}\right)^{x-i}$,

\[ p_i^j = \frac{d - k_i^j}{d} \tag{35} \]

Detailed computing processes of $k_i^j$, $p_i^j(m)$, and $p_i^j(n \mid m)$ are shown in Section 3.3 which will not be mentioned here.

Different from the general algebraic model shown in (8), this modified model, when forwarding messages, considers the effects of the current state of node, namely, the visit times of the nodes, on the forwarding probability in the next step. In some extreme conditions, all the links of the current node are visited, so no unvisited link will be reached in the next step. However, in the random pick or coupon collector’s problem, every pick of link is independent, not considering the node state. So this modified model will more accurately describe ad hoc network message diffusion process, especially when the node coverage is high.

4. Experiment and Analysis

4.1. Experimental Environment and Parameter Setting. To validate the ability of this model proposed herein to describe the message diffusion process of ad hoc network, this paper compares the theoretical value of quantitative analysis and the results of simulation experiments to test the deviation. Before presenting the specific experimental results and analyses, it defines the determination processes of simulation experiment environment and parameters.

(1) Experimental Environment. Focusing on the problems of random sampling model in the process of describing the message transmission of ad hoc network, this paper proposes the factors like node degree and visit times for improvement. In order to verify the improved results, it calculates the quantitative analysis results of random sampling model and the model proposed herein in different parameters in Matlab 7.12. Coupon collector’s problem of the sufficient sampling space $n$ and $u$ is equal to the random pick model; therefore, this paper takes the theoretical value of coupon collector’s problem for analysis.

Both coupon collector’s problem and random pick model present the theoretical calculation results, so this paper, based on NS-2 V2.29 network simulation platform, simulates the message diffusion process of ad hoc network. To simulate the ad hoc network of different scales, $n$, the node number in the network, is set to 1000, 2000, and 3000 and the network topology is in compliance with E-R random graph model [20]. This paper emphasizes the message diffusion process, so the simulation does not consider the delay and packet loss and each message data packet is 1 Byte. When the experiment starts, one node is randomly selected from $n$ nodes as the message source and transmits the message to other nodes. All the nodes that receive the message will forward the message again at the same time slot and the experiment ends when the message has been forwarded for $3n$ times. Calculate the current node coverage $u/n$ at each time slot and take the average value of 10 experimental results as the final result.

(2) Node Visit Times—$j$. As shown in (35), this modified model proposes two key parameters $p_i^j$ and $q_i^j$, which are related to $j$, the node visit times. So it is necessary to figure out the visit times of node in the process of ad hoc network message diffusion.

According to (14), in the network of $n$ nodes, if $x$ messages are forwarded and cover $u$ nodes, which means the node coverage is $u/n$, then $q_i^j$, the probability that any node $i$ has been visited for $j$ times, will change with the node coverage, just as Figure 2 shows ($n = 2000$).

We can see that most nodes have been visited for only 1-2 times in the condition of low node coverage ($u/n < 0.5$). When the coverage is 0.5, the probability that the nodes are visited for once ($j = 1$) or twice ($j = 2$) is 50.0% or 34.7%. With the node coverage's ($0.5 < u/n < 0.9$) increase, the nodes will be multiply visited. When the node coverage reaches 0.9, the probability that the nodes have been visited for multiple times ($2 \leq j \leq 5$) is, respectively, 23.0% ($j = 2$), 26.5% ($j = 3$), 20.4% ($j = 4$), and 11.7% ($j = 5$). But from the curve change of $j \leq 5$ in Figure 2, the probability that the nodes have been visited for more than 5 times is low in the node coverage conditions ($u/n < 0.9$). Even when the coverage is 0.9, the probability that the nodes have been visited for no more than 5 times still reaches 91.6%. Only when the coverage is higher than 0.9, the probability that the nodes have been visited for more than 5 times increases. But there are still 51.2% nodes of less than 5 visit times even when the node coverage reaches 0.99. So the simulation experiment does not consider the condition of more than 5 visit times and the visit times are, respectively, set as 1, 3, and 5.

(3) Node Degree—$d$. As mentioned in Section 3.1, this paper assumes that the ad hoc network topology is in compliance with the E-R random graph model. So the degree of nodes is definite, so $d$ was equal to $x$, and the calculation results were shown in Figure 2. However, in the random pick or coupon collector’s problem, every pick of link is independent, not considering the node state. So this modified model will more accurately describe ad hoc network message diffusion process, especially when the node coverage is high.

![Figure 2: Relationship between node visit times $j$ and node coverage $u/n$.](image-url)
with E-R random model; as \( G(n, p) \) shows, \( n \) is the number of nodes and \( p (0 < p < 1) \) is the probability that there is a link between any 2 nodes, so the average degree of the node is \( d = p \cdot (n - 1) \). To make it simple, we assume that the average degree of node is \( p \cdot (n - 1) \). To research the message diffusion process of the network, we need to make sure that the network topology shall be connected; otherwise, some nodes will never be reached.

According to the nature of E-R random model [20],

(i) if \( p \cdot n < \ln n \), then \( G(n, p) \) must be unconnected;
(ii) if \( p \cdot n > \ln n \), then \( G(n, p) \) must be connected;
(iii) if \( p \cdot n = \ln n \), then \( G(n, p) \) may be connected or not.

In consequence, the node number \( n \) and node link probability \( p \) of the ad hoc network must satisfy the condition \( p \cdot n > \ln n \) in the simulation experiment. Then

\[
\begin{align*}
p \cdot n &> \ln n \\
\implies p \cdot n - p &> \ln n - p \\
\implies p \cdot (n - 1) &> \ln n - p \\
\implies d &> \ln n - p.
\end{align*}
\]

As (36) shows, in order to make all the value of \( d \) meet the requirements in different combinations of \( n \) and \( p \), we assume \( d > \ln n \) and \( d > \ln n - p \) are applicable in any condition. According to the curve of \( \ln n \) in Figure 3, only when the value of \( d \) falls in the dark side, \( d > \ln n \). In fact, if \( n < 3000 \), then \( \ln n < 8.01 \) and \( d \) equals 8 to satisfy the requirements. Only if \( 3000 < n < 5000 \), then \( 8.01 < \ln n < 8.52 \) and \( d \) is more than 8. Although the number of node is 3000–5000 in the network, \( d \) is set as 8 and 10 in this simulation experiment to illustrate the effects of this proposed model.

4.2. Results and Comparison. Figures 4(a)–4(c) show the message diffusion simulation experiment results in the conditions of \( d = 8 \), \( n = 1000, 2000, \) and \( 3000 \), which are compared with the theoretical analysis results of coupon collector’s problem and the model (\( j = 1, j = 3, \) and \( j = 5 \)) proposed herein.

\( x \)-axis represents \( x \): the total number of messages forwarded as the simulation experiment continues, and \( y \)-axis represents \( u/n \): the node coverage. As mentioned, the experiment ends after the number of messages forwarded in the network reaches \( 3n \). Take the average value of ten simulation results as the final result.

From Figures 4(a)–4(c), we may find that, random sampling model taking coupon collector’s problem as representative will overestimate the node coverage in any conditions. When node visit number equals 1 \( (j = 1) \), this model will underestimate the node coverage. When \( j \) equals 1, we just need to consider the condition that the message first visits the node. In fact, with messages being forwarded, the nodes will be visited by the messages for multiple times. So the parameter setting of \( j = 1 \) does not comply with the actual condition. From the figure, we can see that, as the experiment continues, the deviation between the curve \( j = 1 \) and the simulation result curve tend to be larger and larger. In contrast, the curves of \( j = 3 \) and \( j = 5 \) are closer to the simulation result curve, especially in Figure 4(a) \( (n = 1000) \).

In addition, in Figures 4(a)–4(c), the deviations of the simulation experiment results and the theoretical results of \( j = 3 \) and \( j = 5 \) become increasingly large. The deviation is the smallest and the simulation results and theoretical values are closest in Figure 4(a). In Figure 4(c), the deviation is the largest, which means that the simulation result cannot validate the analytical result of this model. The parameter value of node degree \( d \) causes the problem.

According to the analysis of Section 4.1, to ensure the link of the experimental network topology, \( d \), the average node degree, must satisfy the requirement of \( d > \ln n \). In the experiment of Figure 4, \( d \) is 8. In Figure 4(c), \( n = 3000 \) and \( \ln n = 8.01 \). According to the parameter setting of \( d \), the network topology may be connected or unconnected. The simulation experiment result is the average value of ten experiment results; while the topologies are not connected in some experiments; some nodes cannot be connected and the node coverage is comparatively low; thus, the large deviation arises. So we repeat the abovementioned experiments under the parameter conditions of \( d = 10 \) and \( n = 1000, 2000 \) and 3000, and the results are shown in Figures 5(a)–5(c).

Also from Figures 5(a)–5(c), we may see that the random sampling model overestimates the node coverage in any case. If \( j \) equals 1, the proposed model will underestimate the node coverage. Compared with Figure 4, Figure 5 shows a small deviation with no problems of Figure 4(c). This means that, in the research process of ad hoc network message diffusion, our model, taking the node visit times and node degree into consideration, is more modified than the random sampling model in the condition of network connectivity; compared with the random sampling model, the result is closer to the real result with sufficient accuracy.

5. Conclusion

In the cooperative communication systems like ad hoc network or the unstructured P2P network, the message diffusion process can be abstracted as the random sampling model. But message diffusion process is not a completely independent random process; nodes, when forwarding the messages, will be influenced by node degree and network connectivity. However, the random sampling model does not take into consideration the condition that the current node may have been visited for times. So the estimated value may be high if the random sampling model is used to estimate the ad hoc network node coverage at a certain time point. With the increase of node coverage and message number, the random sampling is no longer an accurate model.

Exploring the message diffusion process of cooperative communication systems like ad hoc, this paper analyzes the causes of the inaccuracy of random sampling model and specifically introduces factors like node degree and visit times to solve the problem. It validates the effectiveness of the model proposed herein by comparing its results with the simulation experiment results. Our results are just from random graph network model, while, in the practical application,
Figure 3: Relationship between $d_i$ and $\ln n$.

Figure 4: Comparison between the theoretical values and simulation results of message transmission of ad hoc network of different scales ($d = 8$).
the general network structure is similar to the “small world” network and power law network. The quantitative analysis of message diffusion in this network topology is still a challenge, which will be the focus of our research in the future.

Acknowledgments

The work is supported in part by the National Science Foundation of China under Grant nos. 61070169 and 61070170, the Natural Science Foundation of Jiangsu under Grant no. BK2011376, the University Science Research Project of Jiangsu Province under Grant no. 11KJB520017, and the Application Foundation Research of Suzhou of China under Grant nos. SYG201118 and SYG201238.

References

[1] Y. Sun, S. K. Zhang, H. L. Xu, and S. Lin, “Cooperative communications for wireless ad hoc and sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 161268, 2 pages, 2013.

[2] J. Risson and T. Moors, “Survey of research towards robust peer-to-peer networks: search methods,” *Computer Networks*, vol. 50, no. 17, pp. 3485–3521, 2006.

[3] E. K. Lua, J. Crowcroft, M. Pias et al., “A survey and comparison of peer-to-peer overlay network schemes,” *IEEE Communications Survey and Tutorial*, vol. 7, no. 4, pp. 72–93, 2005.

[4] S. K. Zhang, J. X. Fan, J. C. Jia, and J. Wang, “An efficient clustering algorithm in wireless sensor networks using cooperative
communication,” *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 274576, 11 pages, 2012.

[5] N. B. Chang and M. Liu, “Optimal controlled flooding search in a large wireless network,” in *Proceedings of the 3rd International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt ’05)*, pp. 229–237, Trentino, Italy, April 2005.

[6] S. Jiang, L. Guo, X. Zhang, and H. Wang, “LightFlood: minimizing redundant messages and maximizing scope of peer-to-peer search,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 5, pp. 601–614, 2008.

[7] S. Dolev, E. Schiller, and J. L. Welch, “Random walk for self-stabilizing group communication in ad hoc networks,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 7, pp. 893–905, 2006.

[8] Z. Bar-Yossef, R. Friedman, and G. Kliot, “RaWMS—random walk based lightweight membership service for wireless ad hoc networks,” in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC ’06)*, pp. 238–249, Florence, Italy, May 2006.

[9] C. Avin and B. Krishnamachari, “The power of choice in random walks: an empirical study,” *Computer Networks*, vol. 52, no. 1, pp. 44–60, 2008.

[10] R. Beraldi, “Biased random walks in uniform wireless networks,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 4, pp. 500–513, 2009.

[11] M. Zuniga, C. Avin, and B. Krishnamachari, “Using heterogeneity to enhance random walk-based queries,” *Journal of Signal Processing Systems*, vol. 57, no. 3, pp. 401–414, 2009.

[12] N. Bisnik and A. A. Abouzeid, “Optimizing random walk search algorithms in P2P networks,” *Computer Networks*, vol. 51, no. 6, pp. 1499–1514, 2007.

[13] L. Rodero-Merino, A. F. Anta, L. López, and V. Cholvi, “Performance of random walks in one-hop replication networks,” *Computer Networks*, vol. 54, no. 5, pp. 781–796, 2010.

[14] R. Elsässer and T. Sauerwald, “Tight bounds for the cover time of multiple random walks,” *Theoretical Computer Science*, vol. 412, no. 24, pp. 2623–2641, 2011.

[15] T. Sauerwald, “Expansion and the cover time of parallel random walks,” in *Proceedings of the 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC ’10)*, pp. 315–324, Zurich, Switzerland, July 2010.

[16] J. Xie and K.-S. Lui, “Modeling random walk search algorithms in unstructured P2P networks with social information,” in *Proceedings of the IEEE International Conference on Communications (ICC ’09)*, pp. 1–5, Dresden, Germany, June 2009.

[17] V. M. López Millán, V. Cholvi, L. López, and A. Fernández Anta, “A model of self-avoiding random walks for searching complex networks,” *Networks*, 2011.

[18] S. Vasudevan, D. Towsley, D. Goeckel, and R. Khalili, “Neighbor discovery in wireless networks and the coupon collector’s problem,” in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking (MobiCom ’09)*, pp. 181–192, Beijing, China, September 2009.

[19] J. E. Kobza, S. H. Jacobson, and D. E. Vaughan, “A survey of the coupon collector’s problem with random sample sizes,” *Methodology and Computing in Applied Probability*, vol. 9, no. 4, pp. 573–584, 2007.

[20] P. Erdos and A. Rényi, *On the Evolution of Random Graphs*, vol. 5, Publications Mathematician, 1959.