# Quantitative Analysis of Partition Statistics and their Impact on Data Replication in MANETs

Ke Shi
*School of Computer Sci. and Tech.*
*Huazhong University of Sci. and Tech.*
*keshi@mail.hust.edu.cn*

Zygmunt J. Haas
*School of Electrical and Computer Eng.*
*Cornell University*
*haas@ece.cornell.edu*

## Abstract

*Data replication is, indeed, a widely used redundancy scheme for improving data accessibility in distributed systems at the cost of increased storage space and communication overhead. In this paper, we study the performance of data replication under mobile ad hoc networking environments in the presence of frequent network partitions by analytical modeling and by statistical analysis of simulation results. In particular, we examine the statistics of network partitions for a number of mobility models, and we propose distribution models to approximate the size of the network partitions. We then establish the relation between the network partitioning pattern and the effectiveness of the data replication scheme, which could be used to dynamically adjust the degree of replication depending on the current network operational conditions, while optimizing the trade-off between storage and data accessibility.*

## 1. Introduction

Nodes in a wireless and mobile ad hoc network (MANET) are connected by wireless links [1] and function not only as end-systems, but also as routers, forwarding transmissions of other nodes. If two nodes are sufficiently close to each other, they can communicate directly; otherwise, the nodes can communicate through other intermediate nodes, provided that there is a path between the two nodes; i.e. there is no network partition which prevents establishment of such a path.

Accessing remote data has been a fundamental application in both fixed and mobile networks. However, distributed data access in MANETs has proved to be a much more difficult problem than in fixed networks, since due to node mobility and due to impairments of wireless transmission, network partitions occur frequently. Such network partitions prevent access to data objects residing in a network partition by nodes which are currently outside this partition. Thus, data accessibility in MANETs can be significantly lower than in fixed networks.

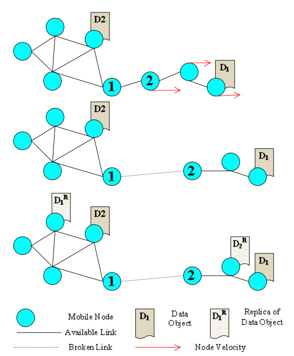



**Fig. 1. Data replication example in MANET**

A possible solution to the low accessibility of data in MANETs is to replicate data objects on several mobile nodes. As depicted in the example in Figure 1, replicas of data objects D1 and D2, which are replicated on nodes in the two network partitions, allows every mobile node in the network to access the data objects, in spite of the network partitioning.

Of course, the extreme solution of replicated every data object on every node in MANETs, is impractical due to limited resources of the nodes in MANETs. However, limited replication is a widely used redundancy scheme for improving the reliability of data access in distributed systems at the cost of increased storage space and communication overhead. Suppose that each data object is replicated into $k$ copies and stored on different network nodes. Of course, to obtain the data object, one needs access only to one out of the $k$ replicas. In fixed networks, where network connectivity is almost 100%, we usually assume that availability of nodes is independent and identically distributed (i.i.d) random variable. The value of $k$ should be set based on the target $\varepsilon$ availability and depending on the average node availability $\alpha$:

$$k = \frac{\log(1-\varepsilon)}{\log(1-\alpha)} \quad (1)$$

Sixth International Symposium on Parallel and Distributed Computing (ISPDC'07)
0-7695-2936-4/07 $20.00 © 2007 IEEE

Based on equation (1), if the availability of network nodes is high, it is usually sufficient to create only a few replicas for achieving high level of data accessibility. Of course, equation (1) does not hold in MANETs, since the availability of nodes is not i.i.d. anymore, and it depends on the location of the nodes. Furthermore, the average node availability cannot be assumed anymore to be high. Consequently, the solution of limited replication needs to be reexamined for the MANET case.

In this paper, we evaluate the performance of data replication in MANETs. In particular, the paper examines the detailed statistics of network partitions, including PMFs, and then attempts to categorize the relationship between network partitioning pattern and data replication effectiveness. The contributions of this study are as follows: 1) Characterizing the statistics of network partitions, 2) Modeling the distributions of the size of network partitions, and 3) Establishing an analytical model based on empirical data (validated by simulation results), to describe the relationship between network partitioning pattern and data replication effectiveness.

The rest of the paper is organized as follows: Section 2 provides an overview of related work. Section 3 introduces the network model used in our work, with data accessibility formally defined. Section 4 presents our simulation setup, while the simulating results are discussed in section 5. Section 6 and 7 derives an analytical model relating the network partitioning patterns to the data replication effectiveness. Conclusions and future work are given in section 8.

## 2. Related Work

Many strategies (e.g., [2] and [3]) for evaluating the effectiveness of replication and tuning the degree of replication to meet special performance requirements have been proposed in the literature. However, these studies relate to wired network, where network partitions seldom occur.

In the field of mobile computing, several strategies (e.g., [4] and [5]) taking the intermittent connectivity of nodes explicitly into account for data replication have been proposed. The goal is to reduce the cost of wireless communication and to compensate for its reduced reliability, by trading off availability, consistency, and communication costs. However, these studies assume only one-hop wireless communication.

Hara ([6] and [7]) proposed a series of replicating schemes in ad hoc networks. In order to guarantee data accessibility upon network partitioning, these works focus on optimizing the location of data replicas within a network and are based on the assumption that access frequencies to data items from each node are known and are fixed. This assumption limits the applicability of the schemes.

Similarly to Hara's work, Wang [8] and Huang [9] considered the problem of replica allocation. Their approach takes into consideration topological information and data replication occurs only when necessary according to certain partition detection schemes. However, it is assumed that the locations and velocities of all nodes are known.

Gianuzzi ([10]) addressed the issue of evaluating data replication effectiveness in MANETs. In his work, a formula is derived, which can be used to evaluate the replica availability, using the probability density functions describing the partitioning of a MANET. However, the random waypoint mobility model used in his work has been proved to fail to provide "steady state". This could lead to unreliable results [16].

Several studies were published that address to global connectivity problem in ad hoc networks ([11], and [12]) through the use of results from percolation theory ([13]). In these studies, nodes are distributed in an infinite area according to a Poisson point process. It has been shown that here exists a critical node density¸ such that when the node density is smaller than this critical node density¸ then the network is composed of infinitely many finite-size connected components [1], while when density is greater than the critical value, there exists a unique infinitely connected component.[2] The characteristics of the finite components that exist below the critical connectivity have not been studied.

Hähner ([14]) investigated partitioning patterns in MANETs through extensive simulation under different mobility models. He concludes that the relatively high number of partitions still needs to be considered even for the very high node density. However, the average number of partitions, without considering the size of the partitions, is insufficient to describe the characteristics of network partitioning.

## 3. System Model

We assume that the nodes in a MANET are homogeneously and randomly deployed within a specific region, termed coverage area according to uniform distribution. The reception range of the node is indicated as a circle with radius *r* around the transmitter.[3] We do not consider in this work small-scale or large-scale fading. Therefore, if the distance

[1] A connected component corresponds to a network partition
[2] Which implies no network partitioning
[3] This is the so-called *protocol reception* model.

Sixth International Symposium on Parallel and Distributed Computing (ISPDC'07)
0-7695-2936-4/07 $20.00 © 2007 IEEE

between two nodes is $r_0$, then if $r_0 <= r$, the two nodes can communicate directly.

An ad hoc network is modeled as a random geometric graph $G(n, r, A)$, where $n$ vertices (nodes) that form the set $V$ of vertices are randomly placed within the coverage area $A$ according to uniform distribution, and any two vertices with distance of $r$ or less between them are connected by an edge (a direct communication link). A network partition is a subset $NP \subseteq V$ where: (1) a path exists between all pairs of vertices $n_i$, $n_j \in NP$, and (2) no path exists between any pair of vertices $n_i \in NP$, $n_k \in V\text{-}NP$.

The degree of a node $v$, denoted as $d(v)$, represents the number of neighbors of $v$, i.e., the number of links. A node of degree $d = 0$ is isolated, i.e., it has no neighbors. The average node degree is defined as:

$$d = \frac{1}{n}\sum_{v \in G} d(v) \qquad (2)$$

In a $G(n, r, A)$, average node degree is:

$$d = (n-1) \times \frac{\pi r^2}{A} \qquad (3)$$

We seek to obtain the distribution of partition size to describe the characteristics of network partitioning in MANETs. Since partition size is a discrete value with finite range from 1 to $n$ ($n=\|N\|$), the probability mass function (PMF) of the partition size, $p(k) = P(\|PN\|=k)$, is utilized to describe this distribution.

Given the above PMF, the average size of network partitions in the network $S_{NP}{}^a$ ($S_{NP}{}^a = \|PN\|$) is calculated as:

$$S_{NP}{}^a = \sum_{k=1}^{n} p(k) \times k \qquad (4)$$

Data accessibility is considered as the ability to obtain data information from other nodes, and to access those data when needed. It is used as the metric to evaluate the effectiveness of data replication in MANETs. As shown in equation (6), we define data accessibility $\alpha_d$ as the ratio of successful access requests to the number of all access requests issued by all the nodes in the MANET.

$$\alpha_d = \frac{\text{\# of successful data access requests}}{\text{\# of all data access requests issued by all the nodes}} \qquad (5)$$

## 4. Simulation Settings

In our simulation, mobile nodes move inside of 1000 [m] by 1000[m] bounded or torus-shaped square area following random waypoint or random walk mobility model [15]. The initial positions of nodes are uniformly selected. The radio communication range of each mobile node is a circle with the radius of $r$, 100 [m].

With random waypoint model, a node's speed $v$ is selected randomly from the range of 1-10[m/s]. The pausing time is 100 [s]. With random walk model, a node's direction and speed are chosen from pre-defined ranges, [0,10m/s] and [0,360°], respectively. Each movement occurs in a constant time interval, 100 [s], at the end of which a new direction and speed are chosen.

The mobility models used in our study do not include group mobility model, as high data availability can be achieved by allocating a replica in every group and re-allocating replicas when certain partitions separate or merge based on prediction. Recent studies ([16]) also have shown that the random waypoint mobility model fails to provide "steady state". Nevertheless, we use this model for comparison purposes with other studies.[4]

By keeping the number of nodes constant and varying the transmission radius, the different values of average node degree are produced.

The simulation period is 100,000 [s]. The samples of partitions are taken every 5 [s] throughout a simulation. For every experiment, the simulation runs 10 times.

## 5. Observation and Analysis

### 5.1. Partition size PMFs

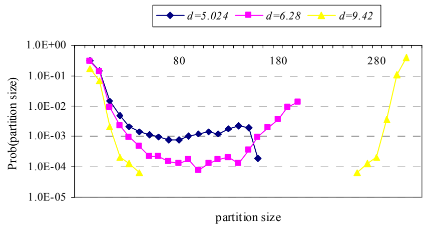
From the results of our simulation, we observed multi-modal distributions of partition size. For different values of average node degree $d$, partition size can be approximated by different distributions.

Based on previous research on random geometric graph and percolation theory ([14], [15] and [16]), there exists a critical value of average node degree, $d_c$, such that if $d \geqslant d_c$, the resulting random geometric graph is connected with high probability.[5] The numerically obtained $d_c$ value in a infinite two dimension area is: $d_c \approx 4.5$.

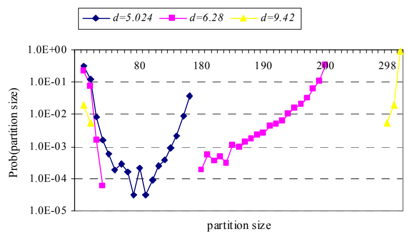
As shown in Figure 2, we observe that if $d$ is greater than $d_c$, the partition size has a bimodal distribution: the probability that the partitions of large size or small size exist is high, and the probability that the partitions of medium size exist is low. This result also demonstrates the network is almost surely connected when $d_c$ is high.

[4] But in torus-shaped area, which eliminates the velocity decay problem.

[5] We say that an event $V_k$, describing a property of a random structure which depends on a parameter $k$, holds with high probability, if $P(V_k) \rightarrow 1$ as $k \rightarrow \infty$.

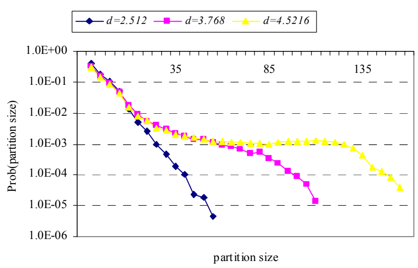



(a) Boundary area

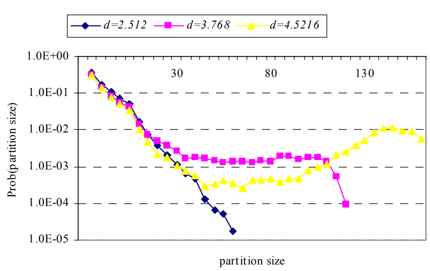



(b) Torus-shaped area

**Fig. 2. The distribution of partition size in stationary ad hoc network with *d* > 4.5**




(a) Boundary area




(b) Torus-shaped area

**Fig. 3. The distribution of partition size in stationary ad hoc network with *d* is around or below $d_c$**

The difference between Figure 2(a) and Figure 2(b) illustrate the effect of borders on the PMF of the partition size. In torus-shaped area, the nodes near the border may connect to the nodes near the opposite border, which means addition of extra links that may connect nodes not already in the same partition. Therefore, the separation of distribution curve emerges earlier in Figure 2(b) than in Figure 2(a). With the same d value, the probability that the partitions of medium size exist is smaller in Figure 2(b) than in Figure 2(a).

If $d$ is around or below $d_c$, we observe the distribution of partition size is scale-free power-law like: partitions of all size can be encountered. As shown in Fig 3, when $d$= 4.526, the small hump at large partition size is always present, which illustrates that the probability that a largest partition containing most nodes (i.e., more than 90% percent of nodes) is almost 1. With the decreasing of d value, the size of the largest partition becomes small. However, there still exist a few large partitions containing most of the nodes. The difference between Figure 3(a) and Figure 3(b) is the result of border effects, as described before.

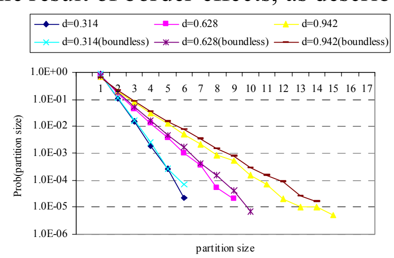



**Fig. 4. The distribution of partition size in stationary ad hoc network with *d* < 1**

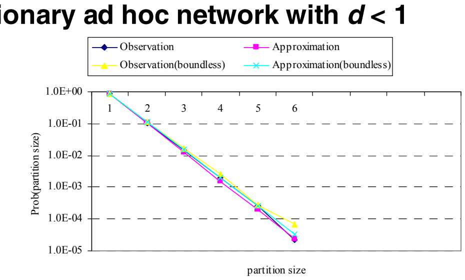



**Fig. 5. The distribution of partition size in stationary ad hoc network with *d*=0.314 and corresponding geometric approximation**

When $d$ is decreased significantly below $d_c$, the probability that the partitions of large size exist becomes very small as shown in Figure 4. There exist a large number of small partitions composed of few nodes. When $d$ is small, the node only has few neighbors. Therefore, the possibility that many nodes form a large partition is small. The difference between the distribution curves in boundary area and torus-shaped area demonstrates that border effects still have impact on the partition size.

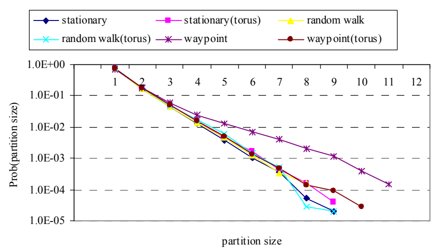
When $d$ is less than 1, we find that the distribution of partition size can be approximated by the geometric distribution:

$$p(k) = P(S_{NP} = k) = p(1 - p)^{k-1} \qquad (6)$$

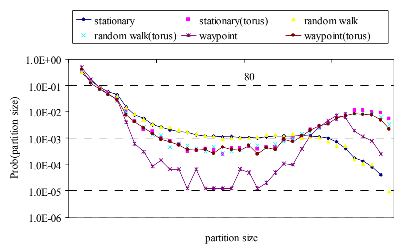
with $\chi^2$ values greater than 0.999. Figure 5 shows the observed distributions and the corresponding geometric approximations.

The size of coverage area is also changed to study its effects on network partitioning. We find the partitioning pattern does not depend on the size of the coverage area. The impact of the border effects on the distribution of the partition size is more obvious with smaller $r^2/A$ value. Due to length limitation, we do not show the simulating results here.

### 5.2. The Impact of Mobility




(a) $d$=0.628




(b) $d$=4.526

**Fig. 6. The distribution of partition size under different mobility models**

Node mobility is considered as one of the important issues affecting the performance of ad hoc networks. Intuitively, node movement can cause one node leave previously affiliated partition and join new partition, which may impact the distribution of the partition size. Figure 6 illustrates the distribution curves of partition size under different mobility models. We observe that except for the random waypoint model, other mobility models do not significantly impact on the distribution of partition size. The distribution curve for random walk model is the same as the distribution curve for stationary ad hoc network. In torus-shaped area, the distribution curves for random walk model, random waypoint model, and stationary ad hoc network are all the same.

This interesting phenomenon can be explained by the fact that the distribution of partition size is determined by the spatial distribution of nodes. Two random geometric graphs with the same $A$, $r$, $n$ and spatial distribution of nodes are considered the same. In our simulation, the initial spatial distribution of nodes is uniform. If the spatial distribution of nodes continues to be unchanged, the distribution of partition size does not change as well. Random walk mobility model does not change the spatial distribution of nodes and random waypoint model in torus-shaped area does not change the spatial distribution of nodes either.

## 6. Effectiveness of Data Replication

In this section, first provide an intuitive explanation of data replication effectiveness in different cases. Then we propose an analytical model for evaluating the effectiveness of data replication based on an empirical formula derived from the PMF of the partition size.

### 6.1. Intuitive explanation of data replication effectiveness

When average node degree is greater than $d_c$, only a few replicas suffice to achieve large data accessibility because most nodes belong to the largest partition. For the same reason, increasing the number of replicas cannot significantly further improve the data accessibility in this situation. It is also impossible to guarantee 100% data accessibility because small isolated partitions always exist unless every node carries all the data replicas.

When average node degree is less than $d_c$, the distribution of partition size is scale free power-like, and there exist a few larger partitions containing most nodes in the network. Large data accessibility can be achieved through allocating enough replicas to these partitions. Replication is then an effective approach to improve data accessibility.

If average node degree is less than 1, the distribution of partition size is approximately geometric. In this situation, the size of most partitions is very small and the relationship between the number of replicas and data accessibility is approximately linear. This implies that replication is not an efficient approach to improve data accessibility.

### 6.2. Analytical model

Consider a mobile wireless ad hoc network modeled by a $G(n, r, A)$ and nodes that attempt to access data objects each represented by nr replicas uniformly distributed in the nodes within the network. Data accessibility $\alpha_d$ defined in section 3 can be considered as the probability of the successful data access, $P_{da}$.

The sufficient condition for a successful data access is that the node initiating data access request has the replicas of requested data object or that the node initiating data access request is located in the same partition with at least one node that has the replicas of the requested data object. Therefore,

$$P_{da\cdot} = P_r + (1- P_r) \times P_{NP} \tag{7}$$

where $P_r$ is the probability that the node initiating data access request has the replicas of requested data object, $P_{NP}$ is the probability that the node initiating data access request is located in the same partition as at least one node that has the replicas of the requested data object.

Since data replicas are uniformly distributed, $Pr = n_r/n$. Given the distribution of partition size, $P_{NP}$ can be computed as follows:

$$P_{NP} = \sum_{i=1}^{N_{NP}} P(a, NP_i) \times P(r, NP_i) \tag{8}$$

where $P(a, NP_i)$ is the probability that the node initiating data access request is in a certain partition $NP_i$ and $P(r, NP_i)$ is the probability that this partition contains at least one node that has the replicas of the requested data object. Since all the nodes are uniformly distributed, the probability that the node initiating data access request is in a certain partition is proportional to the partition size, $P(a, NP_i) = S_{NPi}/n$. Since data replicas is also uniformly distributed in the nodes within the network, then $P(r, NP_i)$ can be computed as follows:

$$P(r,NP_i) = 1 - \frac{\binom{n - n_r - 1}{S_{NPi} - 1}}{\binom{n - 1}{S_{NPi} - 1}} \tag{9}$$

Finally, data accessibility is:

$$\alpha_d = \mathrm{P}_{da} = \frac{n_r}{n} + (1 - \frac{n_r}{n}) \times \sum_{i=1}^{N_{NP}} (\frac{S_{NPi}}{n} \times (1 - \frac{\binom{\mathrm{n - n_r - 1}}{\mathrm{S_{NPi} - 1}}}{\binom{\mathrm{n - 1}}{\mathrm{S_{NPi} - 1}}})), \tag{10}$$

where $N_{NP}$ and $S_{NPi}$ can be obtained from the approximated distribution discussed in the last section.

The resulting data accessibility is shown in Figure 7. When $d$ is greater than $d_c$, only a few replicas suffice to achieve large data accessibility because most nodes belong to the largest partition. For example, when $d$ =6.28, the data accessibility reaches 0.97 when replication ratio is 0.01; i.e., the data is replicated on 1% of nodes. In this situation, only 3% gain in the data accessibility can be achieved, no matter what is the increase in the number of replicas.

When $d$ is around or below $d_c$, improving the number of replicas can greatly improve data accessibility. For example, when $d$=3.768, the data accessibility increases from about 0.244 when data is replicated on 1% of the nodes, to 0.7738 when data is replicated on 10% of the nodes.

However, when $d$ decreases further, the improvement in the data accessibility becomes less significant. For example, for $d$=0.628, the data accessibility increases from about 0.16 when the data is replicated on 10% of the nodes, to 0.64 when data is replicated on 50% of the nodes. The improvement is almost linear with the improvement in replication ratio. This is because the probability of partitions of large size is very small. Using these results, it is possible to relate the average node degree, the number of replicas, and the required data accessibility. For example, for $d$=2.512, the data should be replicated on 20% of the nodes to achieve 70% data accessibility. And for $d$=4.5216, the data should be replicated on 13% of the nodes to achieve 90% data accessibility.

Until now, the issue of managing consistency among the replicas has not considered in our study. However, the data accessibility as calculated by equation (10) can represent the effectiveness of data replication well only for read access only. In a real environment, data items are generally updated, and mobile nodes may access an out-of-date replica, while the original value has been updated.

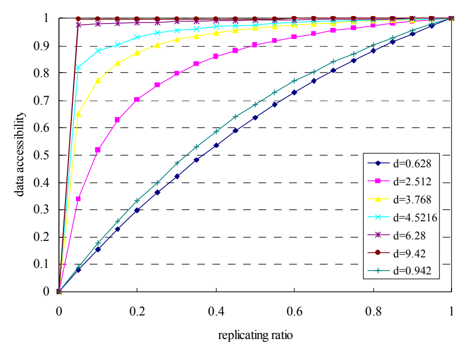



**Fig.7. Data accessibility for different *d* and $n_r/n$**

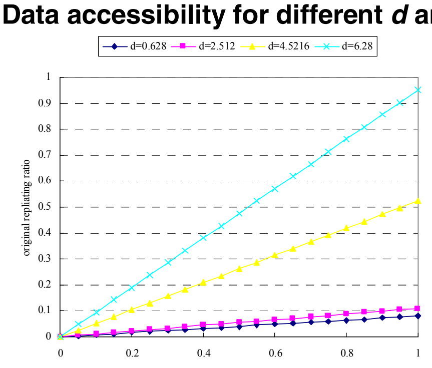



**Fig. 8. $n_{lr}/n$ for different *d***

This stale access can not be considered as successful access. The possibility of stale access is affected by many factors, such as data access pattern, data update pattern, network topology and partitioning pattern, replica ratio, consistency, etc. To simplify the problem, we make the following assumptions: 1) the nodes immediately propagate received data update to other nodes hosting the same replica, and 2) data updates can be fulfilled instantaneously once the nodes issuing these updates connect to the nodes hosting the data copies. Based on these assumptions, the nodes that host the same data copy in the same partition as the updating source will obtain the latest version of data - we use $n_{lr}$ to represent the number of such nodes; nodes in the other partitions cannot obtain this latest version of data. Data accesses to these $n_{lr}$ nodes are valid accesses. $n_{lr}$ can be computed as follows:

$$n_{lr} = \sum_{i=1}^{N_{NP}} (S_{NPi}/n \times (n_r \times S_{NPi}/n)) = \frac{n_r}{n^2} \sum_{i=1}^{N_{NP}} (S_{NPi})^2 \qquad (11)$$

where $S_{NPi}/n$ represents the probability that the updating node belongs to a partition of size $n$, and $n_r \times S_{NPi}/n$ represents the average number of replicas in this partition. Between two subsequent updates, only $n_{lr}$ nodes have the latest data copies.

Figure 8 shows the actual replication ratio, termed the final valid replicating ratio, which is $n_{lr}/n$. When $d$ is greater than $d_c$, the connectivity is large, and most nodes and replicas belong to the largest partition. Therefore, most replicas are valid under our previous assumption; for example, when $d$ is 6.28, the final valid replicating ratio is 0.095, which is only a little less than the original replicating ratio of 0.1. With the decrease in $d$, the number of network partitions increases, the average partition size decreases, and the inconsistency becomes more and more severe. For example, when $d$ is 2.512, the final valid replicating ratio is approximately only 10% of the original replicating ratio.

Using $n_{lr}$ instead of $n_r$ to represent the number of replicas in equation (10), we can now obtain a more accurate estimate of data accessibility. When consistency is considered in MANETs, increasing the number of replicas may not increase data accessibility. The larger the number of replicas is, the larger is the possibility that every network partition has some replicas, and the larger is the possibility that some replicas will be stale when data is updated.

Based on our analysis, data accessibility has an upper limit less than 1 because final valid replicating ratio $n_{lr}/n$ cannot reach 1 except when the network is fully connected. This upper limit decreases with the decrease in the average node degree. We assume that the overhead of maintaining consistency is proportional to the number of replicas. Therefore, in MANETs where network partitioning occurs frequently, a new approach to maintaining data consistency is needed, an approach that considers the trade-off between performance and overhead.

# 7. Further Discussion

To evaluate equation (10), empirical data of partition pattern needs to be collected. We now derive a simplified version of equation (10) based on the results obtained in the section 5 and 6. When average node degree is greater than $d_c$, the probability that data access requests are successful, when initiated by a node not in the largest partition, approaches zero unless the node itself carries replica of requested data object. This is so, because the size of other than the largest partition is very small. Equation (10) then simplifies to:

$$\alpha_a^h \approx \frac{n_r}{n} + (1 - \frac{n_r}{n}) \times \frac{S_{NP\max}}{n} \times (1 - \frac{\binom{n - n_r - 1}{S_{NP\max} - 1}}{\binom{n-1}{S_{NP\max}-1}}) \le \alpha_d \qquad (12)$$

$$S_{NPmax} = \max_{i=1,2,\ldots,N_{NP}} \{S_{NPi}\}$$

Similarly, equation (12) simplifies to:

$$n_{lr} = \frac{n_r}{n^2} \times (S_{NP\max})^2 \qquad (13)$$

Only the size of the largest partition is needed.

When the average node degree is less than 1, the distribution of partition size is approximately geometric. Based on the characteristics of random geometric graph, an approximate solution for $p$, the probability that the partitions contain only one node, can be obtained.

Consider a MANET of n nodes with $N_{NP}$ network partitions, where $p(k)=p\times(1-p)^{k-1}$ is the distribution of the size of a partition, then $N_{NP}\times p(k)$ partitions contain k nodes and

$$\sum_{k=1}^{n} N_{NP} \times p(k) \times k = n. \qquad (14)$$

Then,

$$N_{NP} = \frac{n}{\sum_{k=1}^{n} p(k) \times k} \approx \frac{n}{\sum_{k=1}^{\infty} p(k) \times k} = \frac{n}{\sum_{k=1}^{\infty} p \times (1-p)^{k-1} \times k} = np \qquad (15)$$

Pick one node; the probability to find m other nodes inside its transmission range is equal to:

$$p_{nb}(m) = \binom{n-1}{m} q^m (1-q)^{n-m-1} \approx (\lambda^m / m!) e^{-\lambda} \qquad (16)$$

where $q = \pi r^2/A$ and $\lambda = (n-1)\times q$ for small value of $q$. The probability of one node being isolated, i.e., $m$=0, is $e^{-\lambda}$. The probability that the partitions contain only one node can be computed as:

$$p = \frac{n \times p_{nb}(0)}{N_{NP}} = \frac{n \times e^{-\lambda}}{np} = e^{-\lambda} / p \qquad (17)$$

$$p = \sqrt{e^{-\lambda}} \qquad (18)$$

Based on equation (18), we can get the approximate value of $p$. Table 2 shows that the approximate solution is very close to the value obtained through simulation. Thus, the partition statistics used in formula (10) can be derived using the approximate geometric distribution, as expressed by equations (18).

When the average node degree is between 1 and $d_c$, the distribution of partition size is scale-free power-like. A simple formula for this case is difficult to derive and one still needs to depend on empirical data, as per equations (10).

**Table 1. Approximate solution for p**

| | | $d$=0.314 | $d$=0.628 | $d$=0.942 |
|---|---|---|---|---|
| Simulation | Boundary | 0.87797 | 0.76390 | 0.66967 |
| | Torus-shaped | 0.86778 | 0.74492 | 0.64368 |
| Approximate solution | | 0.86823 | 0.74208 | 0.63423 |

## 8. Conclusion and Future Work

In this paper, we examined the statistics of partition size in MANETs across different mobility models and proposed a simple analytical model that describes the relationship between network partitioning pattern and data replication effectiveness.

In MANETs, data accessibility is also dependent on the size of the requested data object that determines the time needed to transmit the data object between the node initiating the request and the node hosting the data object. In future work, we will study the stability of network partitioning, defined as the time duration during which the nodes stay in the same network partition. We plan on developing a richer analytical model, which considers the size of the requested data object to predict the performance of data replication.

## 9. Acknowledgments

This work is supported by SRF for ROCS, SEM and the Program of Key International Science and Technology Cooperation Projects, Hubei.

## 10. References

[1] Homepage of IETF MANET working group, http://www.ietf.org/html.charters/manet-charter.html

[2] Mondal, A., Kitsuregawa, M., “Effective Dynamic Replication in Wide-Area Network Environments: A Perspective,” DEXA, 2005

[3] Ranganathan, K., “Improving data availability through dynamic model-driven replication in large peer-to-peer communities,” CCGRID, 2002

[4] J. Jing, etc., “An adaptive cache invalidation method in mobile client/server environments,” Mobile Networks and Applications 2(2), 1997

[5] David Ratner, etc., “Replication Requirements in Mobile Environments,” Mobile Networks and Applications, 6(6), 2001

[6] Takahiro Hara, “Replica allocation methods in ad hoc networks with data update,” Mobile Networks and Applications, Vol.8, No.4, 2003

[7] Takahiro Hara, “Dynamic Data Replication Using Aperiodic Updates in Mobile Ad hoc Networks,” DASFAA, 2004

[8] K.H. Wang and B. Li, “Efficient and guaranteed service coverage in partitionable mobile ad-hoc networks,” INFOCOM, 2002

[9] Jiun-L. H., etc., “Exploring group mobility for replica data allocation in a mobile environment,” IKM, 2003

[10] Vittoria G., “Data Replication Effectiveness in Mobile Ad Hoc Networks,” PEWASUN’04, pp: 17-22, 2004,

[11] P. Gupta, “Critical Power for Asymptotic Connectivity in Wireless Networks,” Stochastic Analysis, Control, Optimization and Applications

[12] Dousse, P. Thiran and M. Hasler., “Connectivity in ad hoc and hybrid networks,” INFOCOM 2002

[13] Stauffer, D., etc., “A. Introduction to Percolation Theory,” 2nd ed. London: Taylor & Francis, 1992

[14] Jörg Hähner, etc., “A quantitative analysis of partitioning in mobile ad hoc networks,” ACM Performance Evaluation Review, Vol. 32, No. 1, 2004

[15] Tracy C., etc., “A Survey of Mobility Models for Ad Hoc Network Research,” WCMC: Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, no. 5, 2002

[16] Christian B.,etc., “The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks,” IEEE Trans. On Mobile Computing, 2(3), 2003