Adjusted Replica Allocation in Ad Hoc Networks for Improving Data Accessibility

Min Meng[1], Xiaoling Wu[1], Jie Yang[1], Byeong-Soo Jeong[2], Sungyoung Lee[1]
Department of Computer Engineering
Kyung Hee University
Suwon, South Korea
E-mail: [1]{mengmin, xiaoling, yangjie, sylee}@oslab.khu.ac.kr, [2]jeong@khu.ac.kr

***Abstract* — In ad hoc networks, disconnections occur frequently. In this paper, we allocate the data replicas according to time and space. In temporal method, we store the original data, the median data in a specific time period and the replica with the second highest frequency among all the other data on the mobile hosts to improve data accessibility. In spatial method, we store the original data, the median data among all the neighbors and the replica with the second highest frequency. The advantages of storing the median data are for fault tolerance to eliminate the outliers so that we can get more accurate information of a specific time period or a region, to save memory spaces and energy of the mobile hosts.**

***Keywords* — Adjusted replica allocation, ad hoc network, data accessibility.**

1. Introduction

In mobile ad hoc networks, mobile hosts move freely and disconnections occur frequently. In disconnected environment, if one mobile host wants to access the data stored on the other mobile host or get the other regional information, it may fail. We can copy the data on the other mobile hosts to improve data accessibility. In this paper, we categorize the data replica allocation methods according to time and space. In temporal method, we store the original data, the median data in a specific time period and the replica with the second highest frequency among all the other data on the mobile hosts to improve data accessibility. In spatial method, we store the original data, the median data among all the neighbors and the replica with the second highest frequency among all the other data on the mobile hosts to improve data accessibility. The purposes of storing the median data lie in two aspects, one for fault tolerance to eliminate the outliers so that we can get more accurate information of a specific time period or a region. The other purpose is to save memory spaces and energy of the mobile hosts. Instead of using frequencies of the data item to allocate the replicas, we use the median to eliminate the abnormal frequencies. These abnormal frequencies may be caused by terrible communication conditions, so information transmission can't be progressed fluently. Thus the mobile host would keep accessing the data item stored in other mobile host. Actually in normal conditions, this mobile host may not need this data so frequently. To avoid storing this kind of data with extremely high frequencies caused by unexpected communication errors, we can use median filter to get the median value of the data in a specific period or among all the neighbors and store the data as replica on the other mobile hosts.

The remainder of the paper is organized as follows. In section 2, we introduce the related works. In section 3, we propose adjusted replica allocation methods in ad hoc networks for improving data accessibility. In section 4, we describe the experimental model of our methods. In section 5, we summarize the paper and future works.

2. Related works

Three replica allocation methods were proposed based on the access frequency and network topology [1].

a. SAF (Static Access Frequency) The mobile host stores the original data with the highest access frequency and replicas with the descending order of access frequencies.

b. DAFN (Dynamic Access Frequency and Neighborhood) The mobile host allocates replicas based on SAF. And the replica duplications among neighbors are eliminated as much as possible.

c. DCG (Dynamic Connectivity based Grouping) The mobile hosts are organized into groups using biconnected components. By comparing the frequencies of a specific data among the group members, store the data on the mobile host with the highest frequency.

In these methods the replica are allocated according to access frequencies. Then the three methods are extended to adapt to an environment where each data item is periodically updated [2]. PT value is used instead of access frequencies in these extended methods. P denotes the probability that an access request for data item D_j from mobile host M_i is issued at a unit of time, i.e., the access frequency; T denotes the time remaining until D_j is updated next. The PT value represents the average number of access requests which are issued for D_j until D_j is updated next. The three methods can be further extended to environment with aperiodically updated data items [3] and correlated data items [4]. Data priorities among correlated data items are considered in calculation of access frequencies. These replica allocation methods can be used to predict the locations of data items [9]. Several studies use median filter for fault tolerance [6, 7, 8]. Some similarities and differences of sensor networks and mobile ad hoc networks are proposed in [5].

In this paper, we propose temporal and spatial replica allocation methods. The first advantage of our adjusted methods is for fault tolerance. The former methods didn't consider the connection of the ad hoc network and the correctness of the frequency. Some abnormal frequencies may be caused by terrible communication conditions, so information transmission can't be progressed fluently. Thus the mobile host would keep accessing the data item stored in other mobile host. Actually in normal conditions, this mobile host may not need this data so frequently. To avoid storing this kind of data with extremely high frequencies caused by unexpected communication errors, we can use median filter to get the median value of the data and store the data on the mobile host for the other mobile host to replicate. Through this improvement we are not dependent so much on the frequency. Median values can better reflect the real situation of a specific time period or a region.

The second advantage of our adjusted methods is to save memory spaces and energy of the mobile hosts. In the former methods, every time the frequency table updates, a large portion of the replicas should be updated.

3. Adjusted replica allocation methods

3.1 Assumptions

Let us suppose a situation where a research project team which carries out investigation of digging constructs an ad hoc network in a mountain. The results from the investigation may consist of various types of data such as numerical data, photographs, sounds, and videos. In our temporal and spatial methods, we only consider numerical data as our example, such as the changes of the temperature with the depth of the digging.

We assume the system environment is an ad hoc network. When a mobile host accesses a data item, it will be successful if it holds the data item or it is connected to the mobile host which holds the data item as original.

The additional assumptions are as follows:

- Every mobile host has a unique host identifier $M = \{M_1, M_2, \ldots, M_m\}$, where m is the total number of mobile hosts and M_j ($1 \leq j \leq m$) is a host identifier. Each mobile host moves freely.
- Each data item has a unique data identifier $D = \{D_1, D_2, \ldots, D_n\}$, where n is the total number of data items and D_j ($1 \leq j \leq n$) is a data identifier. The original of each data item is held by a particular mobile host.
- Each mobile host has limited memory space.
- The access frequencies to data items from each mobile host are known by empirical data.

3.2 Adjusted replica allocation methods

The data at a specific time should be stored as the original data on the mobile host. We categorize the replica allocation methods according to time and space. In temporal method, on a mobile host M_i, we store its original data, the median value of the data collected in a specific time period, and a replica with the second highest frequency stored on the other mobile hosts. In spatial method, on M_i, we store its original data, the median value among all M_i's neighbors and M_i, and a replica with the second highest frequency stored on the other mobile hosts. The number of data is odd, the median is the middle one; the number of data is even, the median is the average of the middle two.

3.2.1 Temporal replica allocation method

Let us suppose that six mobile hosts ($M_1, M_2, \ldots, M_6$) exist. Each mobile host has limited memory space to store only 3 data. Table 1 shows the access frequencies of each mobile host to the data items according to historical statistics.

Table 1. Access frequencies to data items (unit: /s)

Data	Mobile Host					
	M_1	M_2	M_3	M_4	M_5	M_6
D_1	0.65	0.25	0.17	0.22	0.31	0.24
D_2	0.44	0.62	0.41	0.40	0.42	0.46
D_3	0.35	0.44	0.50	0.25	0.45	0.37
D_4	0.31	0.15	0.10	0.60	0.09	0.10
D_5	0.51	0.41	0.43	0.38	0.71	0.20
D_6	0.08	0.07	0.05	0.15	0.20	0.62

Table 2 shows the data from time 1 to 5 of every mobile host.

Table 2. Temporal data of mobile hosts

Time	Mobile Host					
	M_1	M_2	M_3	M_4	M_5	M_6
1	23	10	29	13	17	11
2	25	9	27	15	19	9
3	24	11	28	12	18	11
4	26	12	30	16	19	11
5	23	9	28	14	17	8
median	24	10	28	14	18	11

The temporal replication allocation results are shown in table 3.

Table 3. Temporal replication allocation method

Data stored	Mobile Host					
	M_1	M_2	M_3	M_4	M_5	M_6
Original	23	9	28	14	17	8
Median	24	10	28	14	18	11
Replica	18	28	18	10	28	10

At time 5, for M_1, we store the data at time 5 namely 23 as the original data. Then we get the median value of the data created during time period 1 to 5. Order these values in increasing order, 23, 23, 24, 25, 26, the median should be 24. We store this median value 24 on M_1 as the replica for the other mobile host to store. It is because the median value of all the data best reflects the accurate information created in this time period. It should be stored on the other mobile hosts as a reference when disconnection happens for them to know the real situation happened on M_1 in this time period.

Using the same rule and at the same time 5, we get the median value of all the mobile hosts. Then we check the access frequency table, M_1 accesses M_5 with the second highest frequency. The data with the highest frequency should be its original data. So M_1 replicates the median value on M_5 as a replica allocated to it.

After allocating replicas on every mobile host according to temporal replica allocation method, the final results are shown in figure 1.

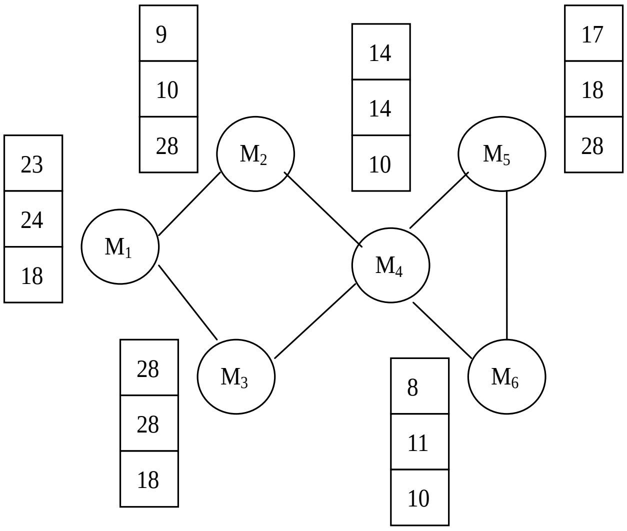

Figure 1. Temporal replica allocation method

In this example, we only store one replica to save memory space of the mobile hosts. However we get more accurate information at other mobile hosts within the specific time period. Because median values reflect more general and accurate information happened in the time period which a single value can't be compared with. We can store more than one replica according to the memory capacity of the mobile hosts.

3.2.2 Spatial replica allocation method

Let us still suppose that six mobile hosts (M_1, M_2, . . . , M_6) exist. Each mobile host has limited memory space to store only 3 data.

In this method, we look at the mobile hosts from the spatial point of view. As in temporal method, we store the data created at a specific time as the original data. We get the median value among the data of a mobile host and all its neighbors. Then we check the access frequency table, and replicate the data with the second highest frequency.

As for M_1, at time 5, 23 is stored as the original data. M_1's neighbors are M_2 and M_3, the median value of these three mobile hosts is 23. Then we check the access frequency table, M_1 accesses M_5 with the second highest frequency. The data with the highest frequency should be its original data. So M_1 replicates the median value on M_5 as a replica allocated to it.

After allocating replicas on every mobile host according to spatial replica allocation method, the final results are shown in figure 2.

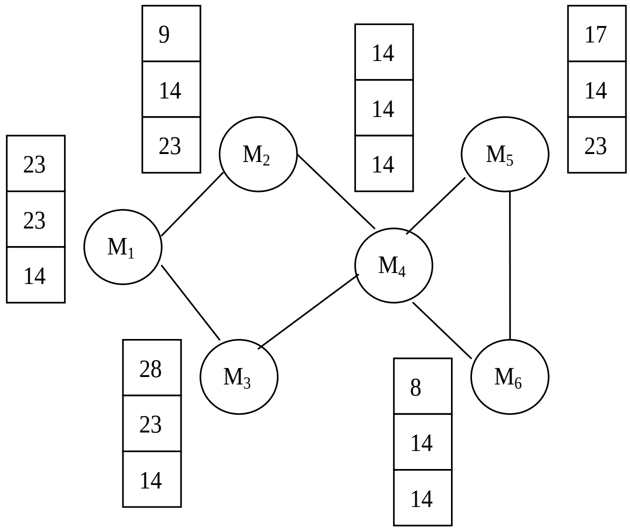

Figure 2. Spatial replica allocation method

In this example, we only store one replica to save memory space of the mobile hosts. However we get more accurate information of the other regions at the specific time. The median values reflect more general and accurate information in a region centered by a specific mobile host which a single value can't be compared with. We can store more than one replica according to the memory capacity of the mobile hosts.

This method is especially appropriate for the largely and densely deployed ad hoc networks. Through the median value of a specific mobile host and all its neighbors, we can get the information of the region centered by the specific mobile host.

4. Experiments

In this section, we describe our experiment model, the pseudo codes of the experiments and the flowcharts of our methods.

4.1 Experiment model

Mobile hosts are randomly deployed in a size of 50×50 flatland. The radio communication range of each mobile host is a circle with the radius of R. One mobile host can not access the data stored on the other mobile host if they are beyond each other's communication range. One mobile host accesses the data on the other mobile hosts according to some specific routing path with access frequency of certain distributions. Data accessibility means whether it can get the data on the other mobile hosts. Traffic means the communication costs to access the data. The data is local, and then the traffic is zero. If the data is not local, the traffic is calculated by the sending and receiving messages. The coordinates of all the mobile hosts are known and it is easy to get the distance of two mobile hosts with the Euclidean Distance formula.

Assume the coordinates of two mobile hosts M_1 and M_2 are (x_1,y_1) and (x_2, y_2) respectively. The distance between them is as follows.

$$d(M_1, M_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

If their distance is less than or equal to the communication radius R, they are neighbors and can directly communicate with each other.

There is another parameter called relocation period, which means after some time, the topology of the mobile ad hoc networks will change. The replicas need to be reallocated.

The following is the pseudo codes of the temporal and spatial replica allocation methods.

```
Algorithm 1 Temporal replica allocation method
Input: access frequency, relocation period
Output: data accessibility, traffic
Begin
  for a period of time
    for every mobile host
      calculate medians
      choose routing path
      according to access frequency
      return data accessibility, traffic
End

Algorithm 2 Spatial replica allocation method
Input: access frequency, relocation period
Output: data accessibility, traffic
Begin
  for a specific time
    for every mobile host
      calculate medians among neighbors
      choose routing path
      according to access frequency
      return data accessibility, traffic
End
```

4.2 Flowcharts of temporal and spatial replica allocation methods

The flowcharts of temporal and spatial replica allocation method are shown in figure 3 and figure 4 respectively.

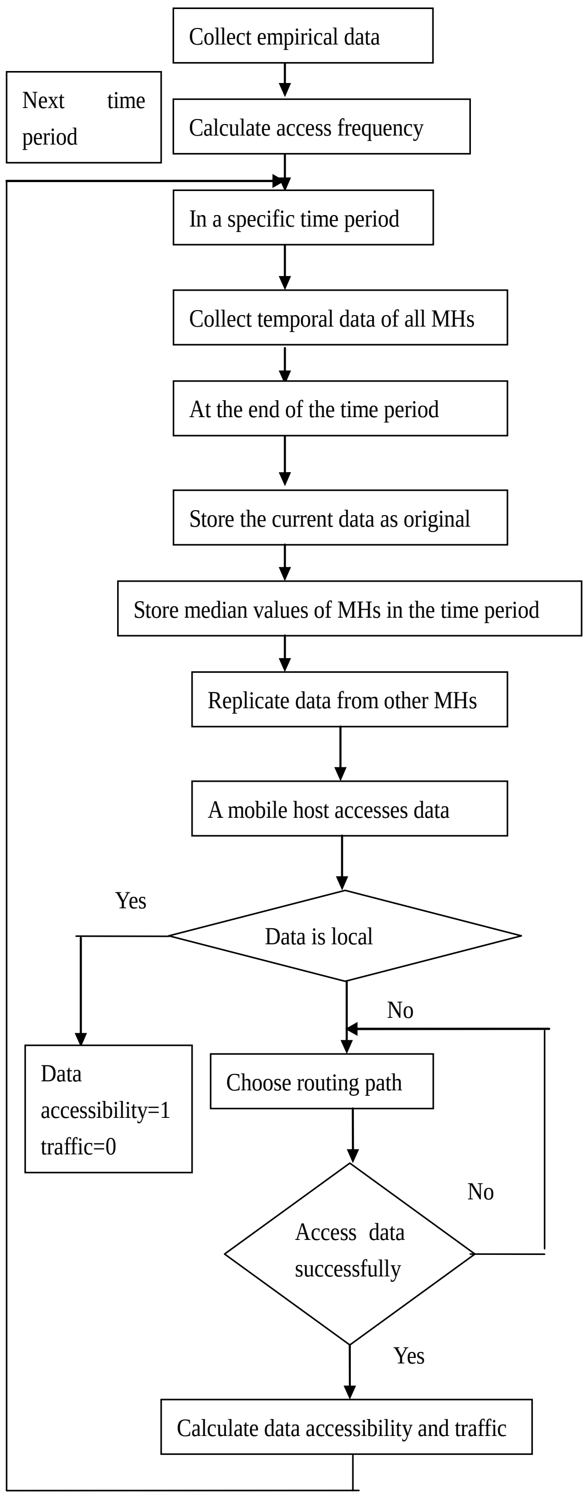

Figure 3. Flowchart of temporal replica allocation method

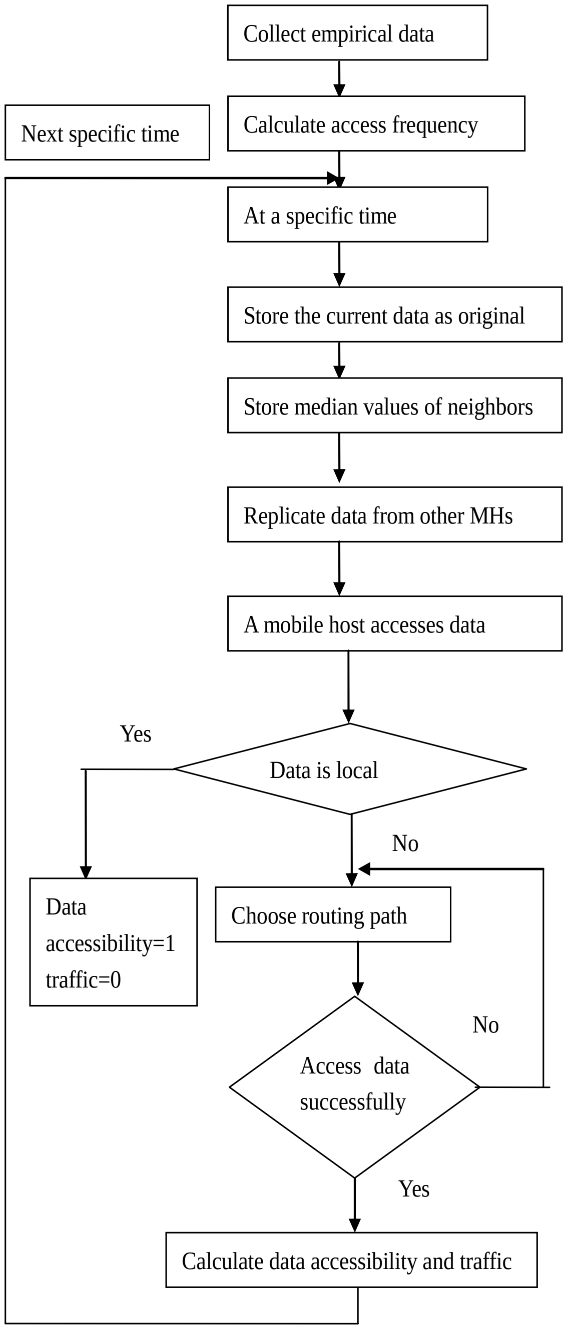

Figure 4. Flowchart of spatial replica allocation method

5. Conclusions and future works

In this paper, we improved replica allocation methods in mobile ad hoc networks. We use median filter to allocate replications so that the replicas stored on the mobile hosts can reflect more accurate information either in temporal or spatial point of view. In temporal method, we store the median values in a specific period of time; in spatial method, we store the median values among the mobile host and all its neighbors. These new methods reduce update frequencies of mobiles hosts to save more energy.

The future works may extend the related methods to wireless sensor networks which have lots of similarities with mobile ad hoc networks, such as limited memory spaces, energy supplies, communication ranges and changing topologies.

REFERENCES

[1] T. Hara, “Effective replica allocation in ad hoc networks for improving data accessibility,” Proc. IEEE INFOCOM, Alaska, 2001, pp.1568-1576.

[2] T. Hara, “Replica allocation in ad hoc networks with periodic data update,” Proc. of International Conference on Mobile Data Management (MDM), Singapore, 2002, pp. 79–86

[3] T. Hara, “Replica allocation methods in ad hoc networks with data update,” ACM-Kluwer Journal on Mobile Networks and Applications, 2003, Vol.8, No.4, pp.343-354.

[4] T. Hara, “Replica Allocation for Correlated Data Items in Ad Hoc Sensor Networks,” SIGMOD Record, Vol. 33, No. 1, 2004, pp38-43.

[5] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” IEEE Communication Magazine, Vol.40, No.8, 2002, pp.102-114.

[6] M. Ding, D. Chen, K. Xing, and X. Cheng, “Localized fault-tolerant event boundary detection in sensor networks,” INFOCOM, Vol. 2, March 13-17, 2005, pp. 902 – 913.

[7] M. Ding, D. Chen, A. Thaeler, and X. Cheng, “Fault-Tolerant Target Detection in Sensor Networks”, IEEE WCNC 2005, Vol. 4, March 13-17, 2005, pp. 2362-2368.

[8] R. Rajagopalan, P. K. Varshney, K. G. Mehrotra and C. K. Mohan, “Fault tolerant mobile agent routing in sensor networks: A multi-objective. optimization approach,” 2nd IEEE Upstate NY workshop on Comm. and Networking , November 2005.

[9] T. Hara, “Strategies for data location management in mobile ad hoc networks,” Proc. IEEE ICPADS, 2005.