Effective Replica Allocation in Ad Hoc Networks for Improving Data Accessibility

Takahiro Hara

Abstract— **Recent advances in computer and wireless communication technologies have led to an increasing interest in ad hoc networks which are temporarily constructed by only mobile hosts. In ad hoc networks, since mobile hosts move freely, disconnections occur frequently, and this causes frequent network division. Consequently, data accessibility in ad hoc networks is lower than that in the conventional fixed networks. In this paper, we propose three replica allocation methods to improve data accessibility by replicating data items on mobile hosts. In these three methods, we take into account the access frequency from mobile hosts to each data item and the status of the network connection. We also show the results of simulation experiments regarding the performance evaluation of our proposed methods.**

Keywords—**ad hoc networks, replica allocation, data accessibility, mobile computing environment**

I. Introduction

Recent advances in radio communication and computer technologies have led to the development of mobile computing environments. In mobile computing environments, by utilizing wireless networks, users equipped with portable computers, called *mobile hosts*, can change their locations while retaining network connections. As one of the research fields in mobile computing environments, there has been an increasing interest in *ad hoc networks* which are constructed by only mobile hosts[2], [4]. In ad hoc networks, every mobile host plays the role of a router, and communicates with each other. Even if the source and the destination mobile hosts are not in the communication range of the two mobile hosts, data packets are forwarded to the destination mobile host by relaying transmission through other mobile hosts which exist between the two mobile hosts. Since no special infrastructure is required, in various fields such as military affair and commerce, many applications are expected to be developed in ad hoc networks.

In ad hoc networks, since mobile hosts move freely, disconnections occur frequently, and this causes frequent network division. Consequently, different fundamental technologies from the conventional fixed networks are needed. For example, if a network is divided into two networks due to the migrations of mobile hosts, mobile hosts in one of the divided two networks cannot access data items held by mobile hosts in the other network. Thus, data accessibility in ad hoc networks is lower than that in the conventional fixed networks. In Figure 1, if the radio link between two mobile hosts at the central part is disconnected, the mobile hosts in the left-hand side and those in the right-hand side cannot access data items D_1 and D_2, respectively. In ad hoc networks, it is a very important issue to prevent deterioration of data accessibility at the point of network division. A possible solution is by replicating data items at mobile hosts which are not the owners of the original data. In Figure 1, if the replicas of data items D_1 and D_2 are allocated at one of the mobile hosts in the opposite networks as shown in Figure 2, every mobile host can access both data items after the network division.

T. Hara is with the Department of Information Systems Engineering, Graduate School of Engineering, Osaka University, Japan. E-mail: hara@ise.eng.osaka-u.ac.jp

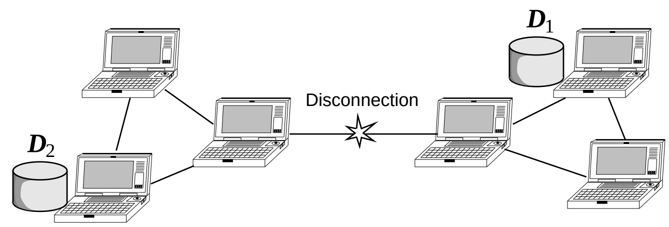

Fig. 1. Network division and data access.

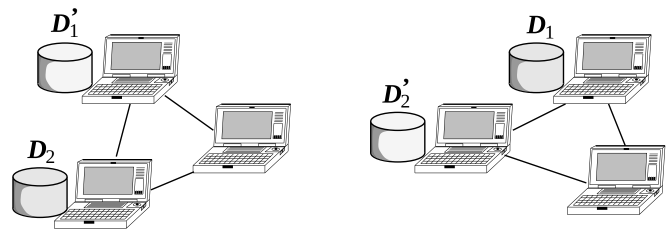

Fig. 2. Effective data replication.

In this way, data replication is very effective for improving the data accessibility. On the other hand, mobile hosts generally have poor resources and thus it is usually impossible for mobile hosts to have replicas of all data items in the network. For example, 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 obtained from the investigation may consist of various types of data such as numerical data, photographs, sounds, and videos. In this case, although it is useful to have the data that other members obtained, it seems difficult for a mobile host to have replicas of all the data.

In this paper, we assume an environment where each mobile host has limited memory space for creating replicas. We propose three replica allocation methods for improving data accessibility. Moreover, we verify the effectiveness of our proposed methods by simulation experiments.

The remainder of the paper is organized as follows. In section II, we introduce some related works and compare them with our proposed methods. In section III, we propose replica allocation methods. In section IV, we show the results of simulation experiments. Finally, in section V, we summarize this paper.

II. Related Works

Several protocols for improving the connectivity in ad hoc networks have been proposed in IETF (Internet Engineering Task Force) so far. DSDV (Destination-Sequenced Distance Vector)[20], [21] is a distance vector routing protocol where each host obtains the route information by using periodical broadcast. When a host communicates with another host, it starts to transmit data packets by source routing based on the route information. AODV (Ad hoc On demand Distance Vector)[22] and DSR (Dynamic Source Routing)[5], [15] are routing protocols where each host obtains the route information by flooding a route request packet when it communicates with another host. ZRP (Zone Routing Protocol)[10] and CBRP (Cluster Based Routing Protocol)[14] are routing protocols where all mobile hosts are divided into several groups and each host periodically obtains route information within a group. When a host communicates with another host in the same group, it starts to transmit data packets by source routing based on the route information. When a host communicates with another host in another group, it floods a route request packet to search for the route. Some other routing protocols have also been proposed by several research groups[7], [9], [13], [17], [18], [19], [24]. All the above protocols have been proposed to improve the connectivity among mobile hosts in the network level. Thus, these protocols are useful for applications where users equipped with mobile hosts directly communicate with each other, e.g., video conference systems. However, as mentioned in section I, in ad hoc networks, there are also other applications where mobile hosts access data items held by other mobile hosts. In these applications, the above protocols can only improve the connectivity among mobile hosts which are connected to each other, but cannot do anything when the network is divided as in the case in Figure 1. On the other hand, our proposed replica allocation methods specialize in such applications and appropriately allocate replicas of data items before the division of the network. Thus, our proposed methods can improve data accessibility. To our best knowledge, studies for improving data accessibility in ad hoc networks have never been done before us.

In the research field of distributed database, many strategies for data replication have been proposed so far[6], [8], [11], [16], [25], [26], [27], [28]. In distributed database systems, data replication offers the benefits of shortening the response time for database write/read operations and improving the data availability. The former is achieved by replicating data and performing database operations locally. In distributed database systems, shortening the response time is considered one of the most important issues, and thus, most replication strategies proposed so far address this issue. This issue includes the propagation of update operations to replicas. The latter is achieved by replicating databases and using the replicas when the site which holds the original database fails. This approach is considered to be similar to our approach, because both approaches address to improve data availability (accessibility). However, since system failures do not frequently occur in distributed database systems in fixed networks, it is usually sufficient to create few replicas of a database, and thus no special strategy is required. On the other hand, since frequent division of the network is a special characteristic in ad hoc networks, our approach takes it into account and is completely different from that in distributed database systems. Shortening the response time and propagating update operations are also significant issues in ad hoc networks. We are planning to address them in our future work.

In the research field of mobile computing, several strategies for data replication have been proposed so far[3], [12], [23]. These strategies assume an environment where mobile hosts access databases at sites in a fixed network, and create replicas of data on the mobile hosts because wireless communication is more expensive than wired communication. They address issues of keeping consistency between original data and its replicas and propagating update operations to the replicas with low communication costs. They are considered to be similar to our approach, because both approaches create replicas on mobile hosts. However, the strategies in [3], [12], [23] assume only one-hop wireless communication, and thus, they are completely different from our approach which assume multihop communication in ad hoc networks.

III. Replica Allocation Methods in Ad Hoc Networks

In this section, the basic system model used to discuss our proposed methods is described. Then, we propose three replica allocation methods for improving data accessibility in ad hoc networks.

A. System model

The system environment is assumed to be an ad hoc network where mobile hosts access data items held by other mobile hosts as the originals. Each mobile host creates replicas of the data items, and maintains the replicas in its memory space. When a mobile host issues an access request to a data item, the request is successful in either case: (i) the mobile host holds the original/replica of the data item or (ii) at least one mobile host which is connected to the request issue host with a one-hop/multihop link holds the original/replica. Thus, first, the request issue host checks whether or not it holds the original/replica of the target data item. If it does, the request succeeds on the spot. If it does not, it broadcasts the request of the target data item. Then, if it receives reply from other host(s) which holds the original/replica of the target data item, the request is also successful. Otherwise, the request fails.

In this system environment, we also make the following assumptions:

- We assign a unique *host identifier* to each mobile host in the system. The set of all mobile hosts in the system is denoted by $\boldsymbol{M} = \{M_1, M_2, \cdots, M_m\}$, 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.
- Data is handled as a data item which is a collection of data. We assign a unique *data identifier* to each data item located in the system. The set of all data items is denoted by $\boldsymbol{D} = \{D_1, D_2, \cdots, D_n\}$, where n is the total number of data items and D_j $(1 \leq j \leq n)$ is a data identifier. All data items are of the same size, and each data item is held by a particular mobile host as the original.
- Each mobile host has memory space of C data items for creating replicas excluding the space for the original data item that the host holds.
- The data items are not updated. This assumption is for the sake of simplicity. In the abovementioned example of digging investigation, new data is inserted but no update occurs. There are also other applications where the newness

of data is not a serious problem, such as weather information.

- The access frequencies to data items from each mobile host are known, and do not change.

B. Replica allocation methods

Based on the assumptions in section III-A, first we consider to determine the replica allocation which gives the highest data accessibility in the whole network. Let m_1 denote the number of mobile hosts which are connected to each other (with one-hop/multihop links) and n_1 ($n_1 \leq n$) denote the number of kinds of data items (replicas) held by the m_1 mobile hosts. Thus, the number of possible combinations of replica allocation is expressed by the following expression:

$$\left\{ \frac{n_1!}{(n_1 - C)!} \right\}^{m_1}. \qquad (1)$$

In order to determine the optimal allocation among all possible combinations, we must analytically find a combination which gives the highest data accessibility considering the following parameters: (i) the access frequency from each mobile host to each data item, (ii) the probability that each mobile host will participate in the network and will disappear from the network, (iii) the probability that each two mobile hosts connected by a (one-hop) link will be disconnected, and (iv) the probability that each two disconnected mobile hosts will be connected by a one-hop link. Even if some lopping is possible, the computational complexity is very high, and this calculation must be done every time when the network topology changes due to the mobile host migration. Moreover, among the above four parameters, the three latter ones cannot be formulated in practical because mobile hosts move freely. For these reasons, we take the following heuristic approach:

- Replicas are relocated in a specific period (*relocation period*).
- At every relocation period, replica allocation is determined based on the access frequency from each mobile host to each data item and the network topology at the moment.

Based on this approach, we propose three replica allocation methods which differ in emphasis put on access frequency and network topology:

1. *SAF (Static Access Frequency)* method: Only the access frequency to each data item is taken into account.
2. *DAFN (Dynamic Access Frequency and Neighborhood)* method: The access frequency to each data item and the neighborhood among mobile hosts are taken into account.
3. *DCG (Dynamic Connectivity based Grouping)* method: The access frequency to each data item and the whole network topology are taken into account.

B.1 The SAF method

In the SAF method, each mobile host allocates replicas of C data items in descending order of the access frequencies. At the time of replica allocation, a mobile host may not connect to another mobile host which has an original or a replica of a data item that the host should allocate. In this case, the memory space for the replica is retained free. The replica is created when a data access to the data item succeeds or when the mobile host connects to another mobile host which has the original or the replica at a relocation period.

TABLE I
ACCESS FREQUENCIES TO DATA ITEMS.

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
D_7	0.38	0.32	0.37	0.33	0.40	0.32
D_8	0.22	0.33	0.21	0.23	0.24	0.17
D_9	0.18	0.16	0.19	0.17	0.24	0.21
D_{10}	0.09	0.08	0.06	0.11	0.12	0.09

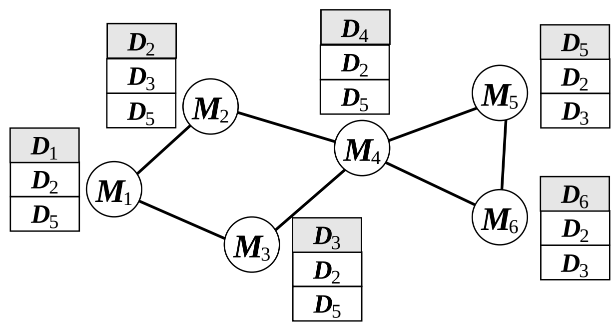

Fig. 3. An example of executing the SAF method.

Now, let us suppose that six mobile hosts ($M_1, \cdots, M_6$) exist and M_i ($i = 1, \cdots, 6$) holds D_i as an original. The access frequency of each mobile host to each data item is shown in Table I. Figure 3 shows the result of executing the SAF method. In this figure, a circle denotes a mobile host, a straight line denotes a wireless link, a gray rectangular denotes an original data, and a white rectangular denotes a replica allocated.

In the SAF method, mobile hosts do not need to exchange information with each other for replica allocation. Moreover, replica relocation does not occur after each mobile host allocates all necessary replicas. As a result, this method allocates replicas with low overhead and low traffic. On the other hand, since each mobile host allocates replicas based on only the access frequencies to data items, mobile hosts with the same access characteristics allocate the same replicas. However, a mobile host can access data items or replicas held by other connected mobile hosts, and thus it is more effective to share many kinds of replicas among them. Therefore, the SAF method gives low data accessibility when many mobile hosts have the same or similar access characteristics.

B.2 The DAFN method

To solve the problem with the SAF method that there are many replica duplication, the DAFN method eliminates the replica duplication among neighboring mobile hosts. First, this method preliminary determines the replica allocation in the same way as the SAF method. Then, if there is replica duplication of a data item between two neighboring mobile hosts, a mobile host with lower access frequency to the data item changes the replica to another replica. Since the neighboring status changes as mobile hosts move, the DAFN method is executed at every relocation period. The algorithm of this method is as follows:

1. At a relocation period, each mobile host broadcasts its host identifier and information on access frequencies to data items. After all mobile hosts complete the broadcasts, from the received host identifiers, every host shall know its connected mobile hosts.
2. Each mobile host preliminary determines the allocation of replicas based on the SAF method.
3. In each set of mobile hosts which are connected to each other, the following procedure is repeated in the order of the breadth first search from the mobile host with the lowest suffix (i) of host identifier (M_i). When there is duplication of a data item (original/replica) between two neighboring mobile hosts, and if one of them is the original, the host which holds the replica changes it to another replica. If both of them are replicas, the host whose access frequency to the data item is lower than the other one changes the replica to another replica. When changing the replica, among data items whose replicas are not allocated at either of the two hosts, a new data item replicated is selected where the access frequency to this item is the highest among the possible items.

At a relocation period, a mobile host may not connect to another mobile host which has an original or a replica of a data item that the host should allocate. In this case, the memory space for the replica is temporary filled with one of replicas that have been allocated since the previous relocation period but are not currently selected for allocation. This temporary allocated replica is chosen among the possible replicas where the access frequency to the replica (data item) is the highest among them. If there is not a possible replica to be temporary allocated, the memory space is retained free. When a data access to the data item whose replica should be allocated succeeds, the memory space is filled with the proper replica.

Figure 4 shows an example of executing the DAFN method in the environment given by Table I and Figure 3. In Figure 4, a dark gray rectangular denotes a replica which is allocated for eliminating replica duplication. In more detail, the following changes of replicas allocated occur between every combination of two neighboring mobile hosts:

$$\begin{array}{ll} M_1\text{-}M_2: & D_2 \to D_7\ (M_1),\ D_5 \to D_8\ (M_2) \\ M_1\text{-}M_3: & D_5 \to D_8\ (M_3) \\ M_2\text{-}M_4: & D_2 \to D_7\ (M_4) \\ M_3\text{-}M_4: & \text{No duplication} \\ M_4\text{-}M_5: & D_5 \to D_8\ (M_4) \\ M_4\text{-}M_6: & \text{No duplication} \\ M_5\text{-}M_6: & D_2 \to D_7\ (M_5),\ D_3 \to D_1\ (M_6) \end{array}$$

Here, M_i-M_j denotes that the replica duplication is eliminated between two mobile hosts M_i and M_j, and $D_k \to D_l$ (M_i) denotes that the replica allocation is changed from D_k to D_l at host M_i. While six kinds of replicas are allocated in the whole network in the SAF method, eight kinds of replicas are allocated in the DAFN method. This is because the DAFN method eliminates replica duplication, and thus many kinds of replicas can be shared among the neighboring mobile hosts. As a result, the data accessibility is expected to be higher than that in the SAF method.

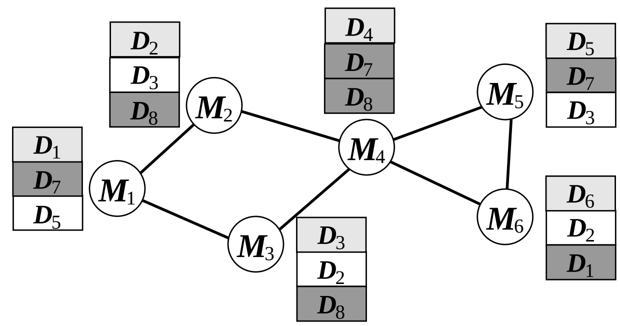

Fig. 4. An example of executing the DAFN method.

However, the DAFN method does not completely eliminate replica duplication among neighboring hosts because it only executes the elimination process by scanning the network once based on the breadth first search. In the example in Figure 4, we can see the replica duplication of D_8 between M_2 and M_4 and between M_3 and M_4, and D_7 between M_4 and M_5. Moreover, if the network topology changes during the execution of this method, the replica relocation cannot be done at mobile hosts over disconnected links. Both the overhead and the traffic are higher than the SAF method because at each relocation period, mobile hosts exchange information and relocate replicas.

B.3 The DCG method

The DCG method shares replicas in larger groups of mobile hosts than the DAFN method that shares replicas among neighboring hosts. In order to share replicas effectively, each group should be stable, i.e., the group is not easily divided due to changes of network topology. From this viewpoint, the DCG method creates groups of mobile hosts that are *biconnected components*[1] in a network. Here, a biconnected component denotes a maximum partial graph which is connected (not divided) if an arbitrary node in the graph is deleted. By grouping mobile hosts as a biconnected component, the group is not divided even if one mobile host disappears from the network or one link is disconnected in the group, and thus it is considered that the group has high stability.

The DCG method is executed at every relocation period. The algorithm is as follows:

1. At a relocation period, each mobile host broadcasts its host identifier and information on access frequencies to data items. After all mobile hosts complete the broadcasts, from the received host identifiers, every host knows the connected mobile hosts.
2. In each set of mobile hosts which are connected to each other, from the mobile host with the lowest suffix (i) of host identifier (M_i), an algorithm to find biconnected components is executed. Then, each biconnected component is put to a group. If a mobile host belongs to more than one biconnected component, i.e., the host is an *articulation*

point, it belongs to only one group in which the corresponding biconnected component is first found in executing the algorithm.

3. In each group, an access frequency of the group to each data item is calculated as a summation of access frequencies of mobile hosts in the group to the item. This calculation is done by the mobile host with the lowest suffix (i) of host identifier (M_i) in the group.
4. In the order of the access frequencies of the group, replicas of data items are allocated until memory space of all mobile hosts in the group becomes full. Here, replicas of data items which are held as originals by mobile hosts in the group are not allocated. Each replica is allocated at a mobile host whose access frequency to the data item is the highest among hosts that have free memory space to create it.
5. After allocating replicas of all kinds of data items, if there is still free memory space at mobile hosts in the group, replicas are allocated in the order of access frequencies until the memory space is full. Each replica is allocated at a mobile host whose access frequency to the data item is the highest among hosts that have free memory space to create it and do not hold the replica or its original. If there is no such mobile host, the replica is not allocated.

At a relocation period, a mobile host may not connect to another mobile host which has an original or a replica of a data item that the host should allocate. In this case, in the same way as the DAFN method, the memory space for the replica is temporary filled with another replica, and it is filled with the proper one when a data access to the corresponding data item succeeds.

Figure 4 shows an example of executing the DCG method in the environment given by Table I and Figure 3. In this example, two groups consisting of M_1, M_2, M_3, M_4 (G_1) and M_5, M_6 (G_2) are created. Table II shows the access frequencies of the two group which are calculated from Table I. In this figure, a dark gray rectangular denotes a replica which is allocated in the second cycle. By executing the DCG method, all ten kinds of replicas are allocated in the whole network.

Compared with the DAFN method that shares replicas among neighboring hosts, the DCG method shares replicas in larger groups of mobile hosts which has high stability. Since many kinds of replicas can be shared, the data accessibility is expected to be higher.

However, since the DCG method consists of three steps; (i) broadcasting host identifiers, (ii) determining the replica allocation, and (iii) notifying it to all hosts in the group, this method takes the largest time among the three methods to relocate replicas. Therefore, the probability is higher that the network topology changes during executing this method, and in this case, the replica relocation cannot be done at mobile hosts over disconnected links. Moreover, both the overhead and the traffic are higher than the other two methods because at each relocation period, mobile hosts exchange information and relocate replicas in a wide range.

IV. Simulation Experiments

In this section, we present simulation results regarding performance evaluation of our proposed methods.

TABLE II
Access frequencies of groups.

Data	Mobile host						Group	
	M_1	M_2	M_3	M_4	M_5	M_6	G_1	G_2
D_1	0.65	0.25	0.17	0.22	0.31	0.24	1.29	0.55
D_2	0.44	0.62	0.41	0.40	0.42	0.46	1.87	0.88
D_3	0.35	0.44	0.50	0.25	0.45	0.37	1.54	0.82
D_4	0.31	0.15	0.10	0.60	0.09	0.10	1.16	0.19
D_5	0.51	0.41	0.43	0.38	0.71	0.20	1.73	0.91
D_6	0.08	0.07	0.05	0.15	0.20	0.62	0.35	0.82
D_7	0.38	0.32	0.37	0.33	0.40	0.32	1.40	0.72
D_8	0.22	0.33	0.21	0.23	0.24	0.17	0.99	0.41
D_9	0.18	0.16	0.19	0.17	0.24	0.21	0.70	0.45
D_{10}	0.09	0.08	0.06	0.11	0.12	0.09	0.34	0.21

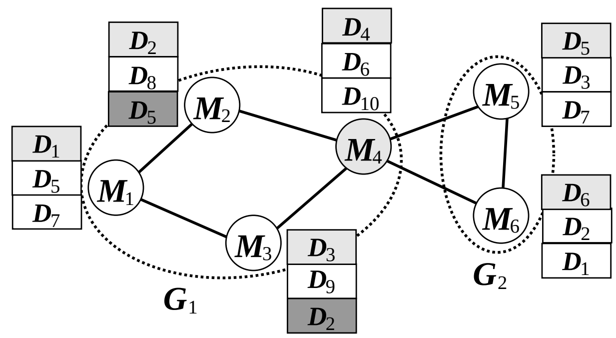

Fig. 5. An example of executing the DCG method.

A. Simulation model

In this subsection, we explain the simulation model.

Mobile hosts exist in a size 50×50 flatland. Each host randomly moves in all directions, and the movement speed is randomly determined from 0 to d. The radio communication range of each mobile host is a circle with the radius of R. Both the number of mobile hosts and kinds of data items in the whole network are 40 ($\boldsymbol{M} = \{M_1, \cdots, M_{40}\}$, $\boldsymbol{D} = \{D_1, \cdots, D_{40}\}$). M_i ($i = 1, \cdots, 40$) holds D_i as the original. Each mobile host creates up to C replicas. An access frequency of each mobile host to D_i is p_i of either of the following three cases:

Case 1: $p_i = 0.5(1 + 0.01i)$.
Case 2: $p_i = 0.025i$.
Case 3: p_i is determined as a positive value based on the normal distribution with mean $0.5(1 + 0.01i)$ and standard deviation σ.

Case 1 represents a situation where every mobile host has the same access characteristics and access frequencies vary in a small range. Case 2 also represents a situation where every mobile host has the same access characteristics, but access frequencies vary in a wide range. Case 3 represents a situation where there exists the scatter of access characteristics of mobile hosts. When $\sigma = 0$, case 3 equals to case 1. As σ gets larger, the difference of access characteristics among mobile hosts gets larger.

In the DAFN and the DCG methods, replicas are periodically relocated based on the relocation period T. Table III shows parameters and their values used in the simulation experiments.

TABLE III

PARAMETER CONFIGURATION.

Parameter	value	
d	1	
R	7	(1∼19)
C	10	(1∼39)
T	256	(1∼8192)

Each parameter is basically fixed to a constant value, but it is changed in a range represented by the parenthetic values in one of the simulation experiments. Because a simulation experiment changing the movement speed, d, relatively shows the same result as that changing the relocation period T, d is not changed in the simulation experiments. For the same reason, the number of mobile hosts and kinds of data items are also not changed, i.e., simulations changing them relatively show the same results as those changing R and C, respectively.

In all simulation experiments, we examine the average data accessibility and the total traffic of each of the three methods, the SAF method, the DAFN method, and the DCG method, during 50,000 units of time. Here, we define the traffic as the total hop count of data transmission for allocating/relocating replicas.

B. Effects of the relocation period

First, we examine the effects of the relocation period on each of the three methods. Here, access frequencies of mobile hosts are determined based on case 1 (p_i = 0.5(1 + 0.01i)). Figure 6 and Figure 7 show the simulation results. In both figures, the horizontal axis indicates the relocation period, T. The vertical axes indicates the data accessibility and the traffic, respectively.

From Figure 6, the DCG method gives the highest data accessibility, and the DAFN method gives the next. Since the degree of replica duplication is high in the SAF method, the DAFN method and the DCG method give a drastic improvement. The relocation period has a little effect on the three methods. Contrary to the intuitive expectation, short relocation period does not give high data accessibility. This shows that sharing various replicas is very important but the determined replica allocation itself is not so important in this environment.

From Figure 7, the DCG method gives the highest traffic, and the DAFN gives the next. The traffic caused by the two methods is inversely proportional to the relocation period.

From these results, since the data accessibility is not sensitive to the relocation period, the relocation period should be set to a certain large value in this environment.

We have also done the same simulation experiment based on case 2. Figure 8 and Figure 9 show the simulation results. These results show almost the same characteristics as those based on case 1. However, in Figure 8, the difference in the data accessibility between the DAFN method and the DCG method is smaller than that in Figure 6. This is because in case 2, access frequencies vary in a wide range, and thus replicas of data items with low access frequencies are allocated in the DCG method.

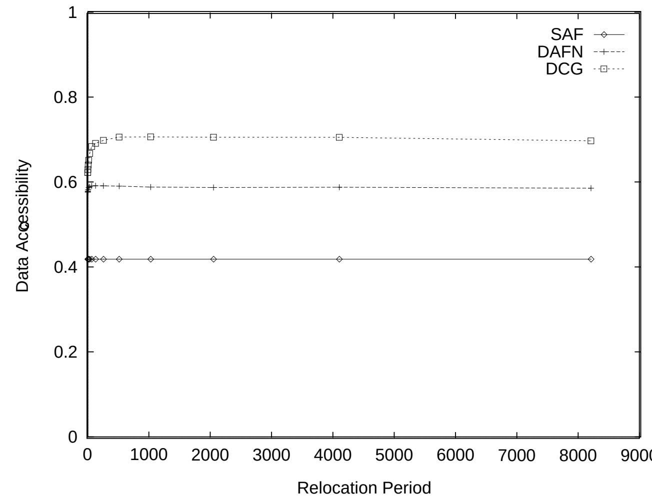

Fig. 6. Relocation period and data accessibility (Case 1).

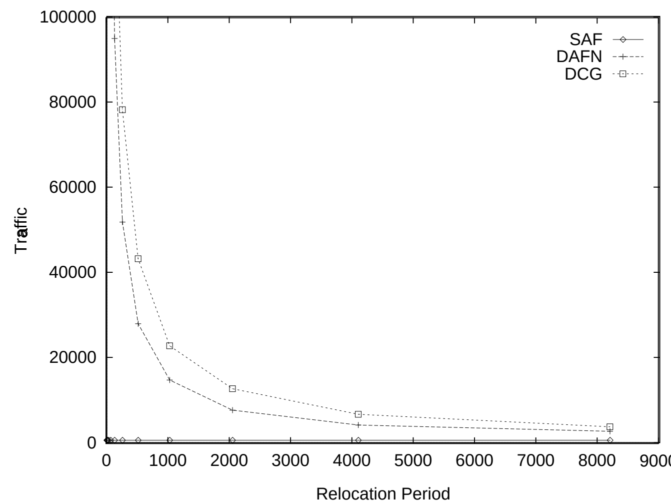

Fig. 7. Relocation period and traffic (Case 1).

C. Effects of the scattering access characteristics

Next, we examine the effects of the scattering access characteristics of mobile hosts on each of the three methods. In this simulation experiment, access frequencies of mobile hosts are determined based on case 3, and the standard deviation, σ, is changed. When σ = 0, access frequencies to data items are determined based on case 1 (p_i = 0.5(1 + 0.01i)) where the order of access frequencies is the same at every mobile host, i.e, every mobile host has the same access characteristics. As σ gets larger, the scatter of the order of access frequencies gets larger, and thus, the difference of access characteristics among mobile hosts gets larger.

Figure 10 and Figure 11 show the simulation results. In both figures, the horizontal axis indicates the value of σ. The vertical axes indicates the data accessibility and the traffic, respectively. These results show that the scatter of access characteristics has little impact on the DCG method.

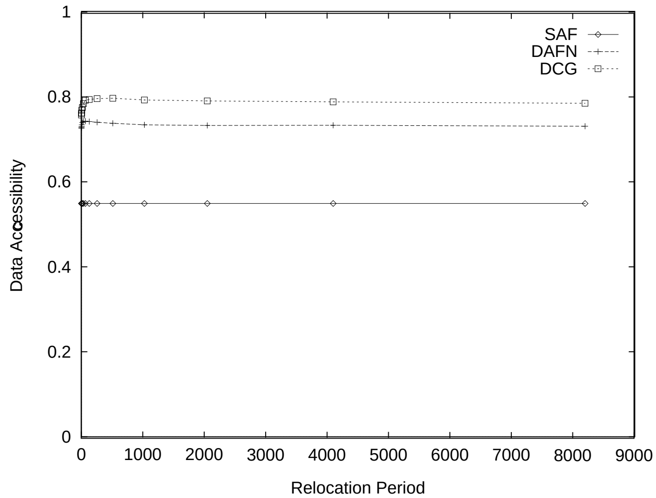

Fig. 8. Relocation period and data accessibility (Case 2).

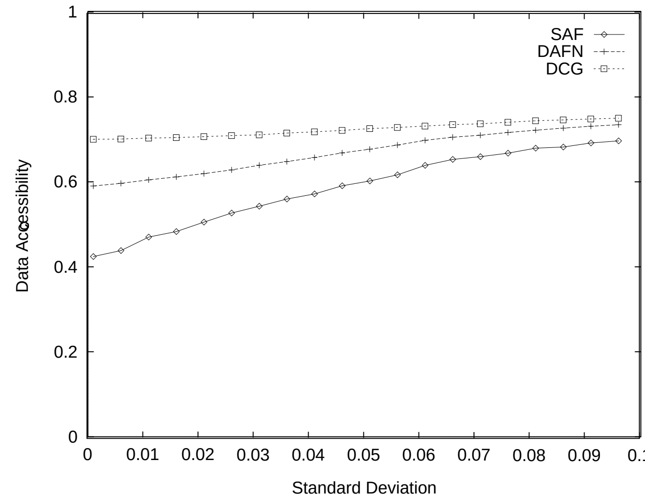

Fig. 10. Scatter of access characteristics and data accessibility.

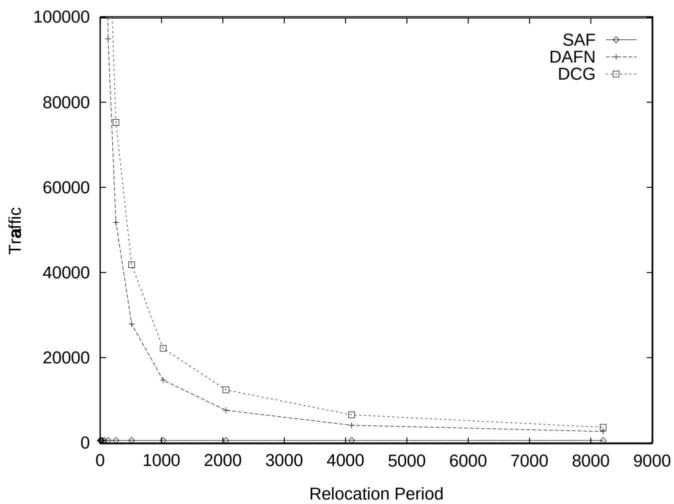

Fig. 9. Relocation period and traffic (Case 2).

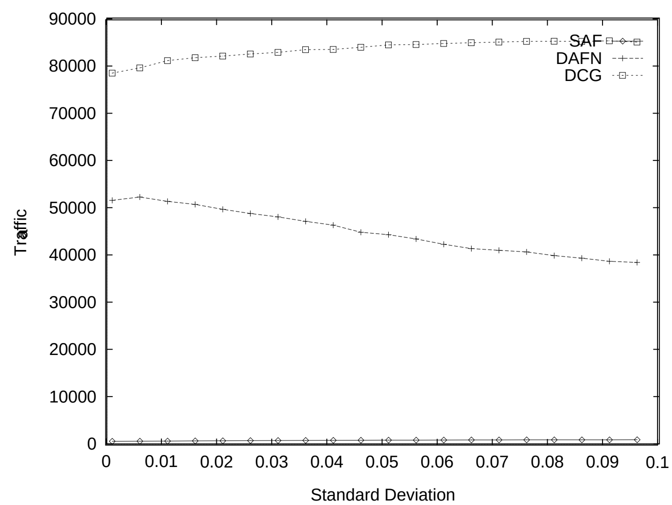

Fig. 11. Scatter of access characteristics and traffic.

From Figure 10, as the difference of access characteristics gets larger, the difference in data accessibility of the three methods gets smaller. This is because when the scatter of access characteristics is large, each mobile host allocates replicas of different data items in the SAF and the DAFN methods, and thus mobile hosts which are connected to each other instinctively share the replicas.

This fact is also seen from Figure 11. As the difference of access characteristics gets larger, the traffic caused by the DAFN method gets smaller. This shows that when the scatter of access characteristics is large, the degree of replica duplication is small in the DAFN method.

D. Effects of the radio communication range

We examine the effects of the radio communication range of mobile hosts on each of the three methods. Figure 12 and Figure 13 show the simulation results based on case 1. In both figures, the horizontal axis indicates the communication range, R. The vertical axes indicates the data accessibility and the traffic, respectively.

From Figure 12, as the radio communication range gets larger, the data accessibility also gets larger in every method. In most cases, the DCG method gives the highest data accessibility. When the communication range is very small, every method gives almost the same data accessibility. This is because the number of mobile hosts connected to each other is small, and thus replica relocation rarely occurs. When the communication range is very large, every method also gives almost the same data accessibility. This is because most mobile hosts are connected to each other, and thus mobile hosts can access original data items in most cases.

From Figure 13, as the radio communication range gets larger, the traffic caused by the DAFN method and the DCG method also gets larger at first, but it gets smaller from a certain point. When the radio communication range is very small, the traffic caused by these two methods is small. This is because the

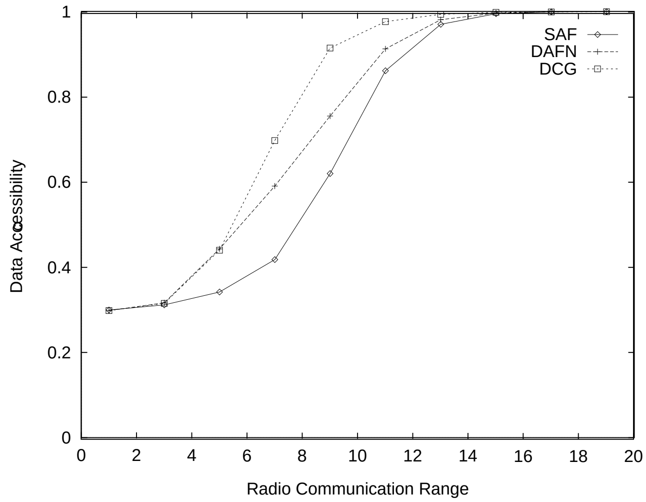

Fig. 12. Radio communication range and data accessibility.

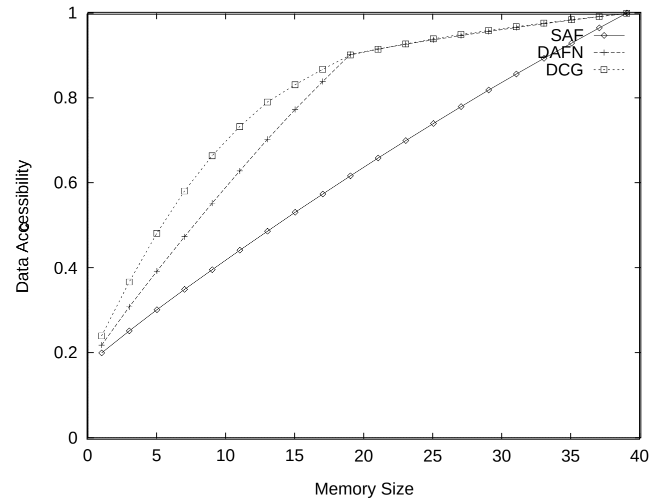

Fig. 14. Memory size and data accessibility.

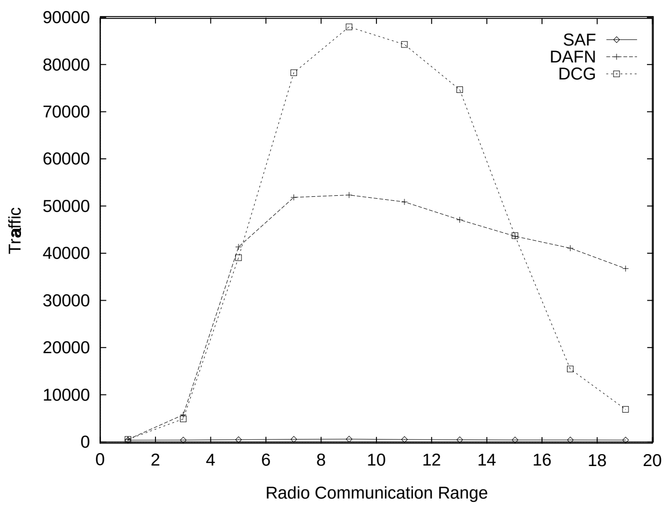

Fig. 13. Radio communication range and traffic.

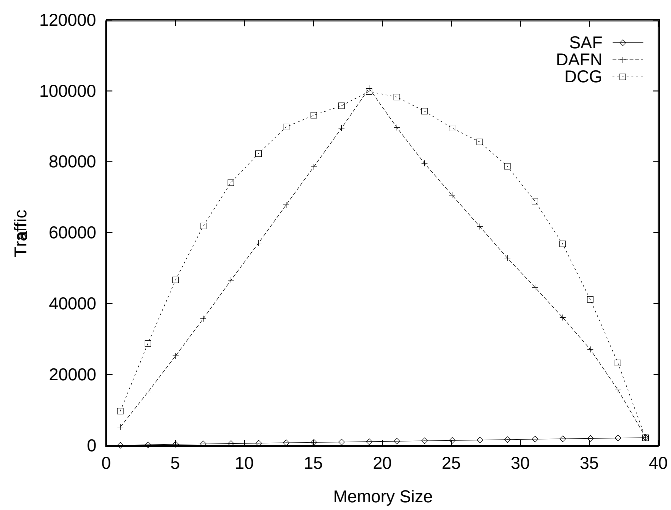

Fig. 15. Memory size and traffic.

number of mobile hosts connected to each other is small, and thus replica relocation does not cause large traffic. When the radio communication range is very large, the DCG method gives smaller traffic than the DAFN method. This is because in the DCG method, the number of mobile hosts in a group is very large (40 in most cases) and thus replica relocation rarely occurs.

E. Effects of the size of memory space

Finally, we examine the effects of the size of memory space on each of the three methods. Figure 14 and Figure 15 show the simulation results based on case 1. In both figures, the horizontal axis indicates the memory size of a mobile host, C. The vertical axes indicates the data accessibility and the traffic, respectively.

From Figure 14, as the memory size gets larger, the data accessibility also gets larger in every method. In most cases, the DCG method gives the highest data accessibility, and the DAFN method gives the next. The accessibility of the SAF method are linearly affected by the memory size. The accessibility of the DAFN method are also linearly affected by the memory size at first, then it shows almost the same value as that of the DCG method.

From Figure 15, as the memory size gets larger, the traffic caused by the DAFN method and the DCG method also gets larger at first, but it gets smaller from a certain point. When the memory is very small, replica relocation does not cause large traffic because the number of replicas created is small. When the memory size is very large, the two methods cause small traffic because each host holds replicas of most data items and thus replica relocation rarely occurs.

V. Conclusion

In this paper, we have discussed replica allocation in ad hoc networks to improve data accessibility. We have proposed three replica allocation methods which take into account the access frequencies to data items and the network topology. In the SAF method, a mobile host allocates replicas with high access frequencies. In the DAFN method, replicas are preliminary allocated based on the SAF method, and then the replica duplication is eliminated among neighboring mobile hosts. In the DCG method, stable groups of mobile hosts are created, and replicas are shared in each group.

The simulation results show that in most cases, the DCG method gives the highest accessibility, and the SAF method gives the lowest traffic. In a real environment, a proper method among the three methods should be chosen based on the characteristics of the system.

As part of our future work, we are planning to address replica allocation in environments where update of data items occurs.

Acknowledgments

The author would like to express his sincere appreciation to Professor S. Nishio and Associate Professor M. Tsukamoto of Osaka University for invaluable comments on this work.

This research was supported in part by Research for the Future Program of Japan Society for the Promotion of Science under the Project "Advanced Multimedia Content Processing (JSPS-RFTF97P00501)".

References

[1] A.V. Aho, J.E. Hopcroft, and J.D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, 1974.

[2] D.J. Baker, J. Wieselthier, and A. Ephremides, "A distributed algorithm for scheduling the activation of links in a self-organizing, mobile, radio network," Proc. IEEE ICC'82, pp.2F6.1-2F6.5, 1982.

[3] D. Barbara and T. Imielinski, "Sleepers and workholics: Caching strategies in mobile environments," Proc. ACM SIGMOD'94, pp.1-12, 1994.

[4] J. Broch, D.A. Maltz, D.B. Johnson, Y.C. Hu, and J. Jetcheva, "A performance comparison of multi-Hop wireless ad hoc network routing protocols," Proc. Mobicom'98, pp.159-164, 1992.

[5] J. Broch, D.B. Johnson, and D.A. Maltz, "Dynamic source routing in ad hoc wireless networks," Internet Draft, draft-ietf-manet-dsr-00.txt, 1998.

[6] M.J. Carey and M. Livny, "Distributed concurrency control performance: A study of algorithm, distribution, and replication," Proc. 14th VLDB Conference, pp.13-25, 1988.

[7] Y.L. Chang and C.C. Hsu, "Routing in wireless/mobile ad-hoc networks via dynamic group construction," ACM-Baltzer Journal of Wireless Networks, Vol.5, No.1, pp.27-37, 2000.

[8] A.W. Fu and D.W.L Cheung, "A transaction replication scheme for a replicated database with node autonomy," Proc. 20th VLDB Conference, pp.214-225, 1994.

[9] M. Gerla and J.T.C. Tsai, "Multicluster, mobile, multimedia radio network," ACM-Baltzer Journal of Wireless Networks, Vol.1, No.3, pp.255-265, 1995.

[10] Z.J. Haas and M.R. Pearlman, "The zone routing protocol (ZRP) for ad hoc networks," Internet Draft, draft-ietf-manet-zone-zrp-01.txt, 1998.

[11] T. Hara, K. Harumoto, M. Tsukamoto, and S. Nishio, "Dynamic replica allocation using database migration in broadband networks," Proc. IEEE ICDCS 2000, pp.376-384, 2000.

[12] Y. Huang, P. Sistla, and O. Wolfson, "Data replication for mobile computer," Proc. ACM SIGMOD'94, pp.13-24, 1994.

[13] Y.C. Hu and D.B. Johnson, "Caching strategies in on-demand routing protocols for wireless ad hoc networks," Proc. Mobicom 2000, pp.231-244, 2000.

[14] M. Jiang, J. Li, and Y.C. Tay, "Cluster based routing protocol (CBRP) functional specification," Internet Draft, draft-ietf-manet-cbrp-spec-00.txt, 1998.

[15] D.B. Johnson, "Routing in ad hoc networks of mobile hosts," Proc. IEEE Workshop on Mobile Computing Systems and Applications, pp.158-163, 1994.

[16] N. Krishnakumar and A.J. Bernstein, "High performance escrow algorithms for replicated databases," Proc. 18th VLDB Conference, pp.175-186, 1992.

[17] C.R. Lin and M. Gerla, "A distributed architecture for multimedia in a multihop dynamic packet radio network," Proc. IEEE Globecom'95, pp.1468-1472, 1995.

[18] M. Nishizawa, H. Hagino, T. Hara, M. Tsukamoto, and S. Nishio: "A routing method using uni-directional link in ad hoc networks," Proc. Int'l Conference on Advanced Computing and Communications (ADCOM'99), pp.78-82, 1999.

[19] V. Park and S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," Proc. IEEE Infocom'97, pp.1405-1413, 1997.

[20] C.E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," Proc. ACM SIGCOMM'94, pp.234-244, 1994.

[21] C.E. Perkins and P. Bhagwat, "Destination-sequenced distance-vector," Internet Draft, draft-ietf-manet-dsdv-00.txt, 1998.

[22] C.E. Perkins and E.M. Royer, "Ad hoc on demand distance vector (AODV) routing," Internet Draft, draft-ietf-manet-aodv-02.txt, 1998.

[23] E. Pitoura and B. Bhargava, "Maintaining consistency of data in mobile distributed environments," Proc. IEEE ICDCS'95, pp.404-413, 1995.

[24] K. Scott and N. Bambos, "The self-organizing wireless network (SWAN) protocol for communication among mobile users," Proc. IEEE Globcom'95, pp.355-359, 1995.

[25] M. Stonebraker, "Concurrency control and consistency in multiple copies of data in distributed INGRES," IEEE Transactions on Software Engineering, Vol.3, No.3, pp.188-194, 1979.

[26] R.H. Thomas, "A majority consensus approach to concurrency control for multiple copy databases," ACM Transaction on Database Systems, Vol.4, No.2, pp.180-209, 1979.

[27] O. Wolfson and A. Milo, "The multicast policy and its relationship to replicated data placement," ACM Transaction on Database Systems, Vol.16, No.1, pp.181-205, 1991.

[28] O. Wolfson and S. Jajodia, "Distributed algorithms for dynamic replication of data," Proc. ACM PODS'92, pp.149-163, 1992.