

A Data Replication Scheme based on Primary Copy for Ensuring Data Consistency in Mobile Ad hoc Networks

Aekyung Moon, Han Namgoong, Hyoungsun Kim, Hyun Kim
Software Robot Research Team, ETRI
161 Gajung-dong, Yusong-gu, Taejeon, KOREA
Tel: +82 42 860 6735, Fax: +82 42 860 6790
{akmoon, nghan, kimhs, hkim}@etri.re.kr

Abstract - A mobile ad hoc network (MANET) is a multi-hop wireless network capable of autonomous operations. Since MANET causes frequent network partitions, a data replication is considered as a way to increase the data accessibility and the availability of transaction processing. If the database is replicated among mobile hosts, additional energy is required to keep data consistency. We propose a data replication scheme for MANET, which provides data consistency and increased number of successful committed transactions reducing energy consumption. The proposed scheme updates the data in the primary copy, which guarantees the latest value of the data item. If an application wants the latest version of data, it should refer to the primary copy; otherwise it accesses the local copy of the data. When the application does not require the latest version, our scheme approves the access of the data item to the local copy. It can improve the energy efficiency by reducing the additional number of messages to maintain data consistency and to query data. Furthermore we can expect the performance improvement by alleviating the message traffic or by reducing the number of aborted transactions due to network partitions because the proposed scheme validates any conflicts only on the host of the primary copy.

Keywords: Ubiquitous, Sensor Network, Data Replication, DBMS, Data Consistency

1 Introduction

A MANET is a multi-hop wireless network capable of autonomous operations [11, 12]. The mobility of hosts in MANET can lead to frequent network partitions, which brings in lower data availability than that in conventional networks. Data replication has been used as a basic mechanism for data availability, but existing replication schemes may not be applicable to MANET [9]. They do not consider underlying topology changes, which also incur different access routes to the primary copy, and need to consider the limited energy [1]. Karumanchi [9] proposed a quorum-based approach, which executes updates and query transactions in a number of quorum servers. However, it leads to lower ratios of successful committed transactions and consumes much energy due to frequent network partitions.

We focus how to increase the ratio of successful committed transactions and reduce energy consumption. The proposed scheme updates the data in the host of primary copy, which guarantees the latest value of the data item. If an application wants the latest version of data, it should refer to the primary copy; otherwise it accesses the local copy of the data. When the application does not require the latest version, our scheme approves the access of the data item to the local copy. It can improve the energy efficiency by reducing the additional number of messages to maintain data consistency and to query data. Furthermore we can expect the performance improvement by alleviating the message traffic or by reducing the number of aborted transactions due to network partitions because the proposed scheme validates any conflicts only on the host of the primary copy.

This paper is organized as follows. Section 2 presents related work; in section 3, the proposed scheme is discussed. Section 4 describes the simulation model and performance analysis; finally, section 5 summarizes the conclusion.

2 Related Works

Data replication schemes in MANET emphasize the increase of data accessibility [3, 5], where [5] does not allow updates of data. If updates by different hosts are permitted without constraints, the values of replicated data may diverge. To solve this problem, there are several primary replication schemes: Tamori approach [14], STE (Select then Eliminate) and ETS (Eliminate then Select) by Karumanchi [9], Hara approach [6], and so on. These schemes ensure data consistency in network partitions under several assumptions.

Tamori indicates that it is difficult to have the latest data in all replicated hosts; hence, a timestamp list is used to distinguish old data from new data. Each mobile host has a timestamp list and data list. After updating data, hosts exchange a timestamp list and check for the latest version of stored data. A timestamp list usually returns the latest version of data, but not guarantees it due to different hosts in different network partitions.

STE and ETS are based on quorum approach; updates or queries are considered to be successful if one or more

host returns an acknowledgement of the receipt of updated message. Since there are frequent network disconnections in MANET, the probability of obtaining old data becomes bigger as time passes. Suppose that the information is the number of the hosts which are not reachable from the specific host. ETS guarantees data consistency because updates or queries succeed if all hosts of one quorum are within the partition. However, the rate of successful transactions goes down as the probability of network partitioning increases. Network partitions require high communication overheads even though transactions are only queries.

Hara separates data items as original or replica. Updates are done in the host of original data. In case of a network disconnection, hosts read old values and when hosts connect to the host of original data, version numbers are compared to check whether earlier reads were dirty or not. If consistency is required, it is left to user [2]; if an application is affected by old reads, it must be aborted or rolled back. In ROWA [10], data can not be updated when network partitions occur. Holliday [8] sets up a proxy to cover planned disconnections and allow updates. When a host disconnects, the host appoints another host as voting member, but it is difficult to apply because hosts in MANET do not gracefully incur disconnections.

3 Data Replication Scheme

In this section, we explain our replication scheme, which is based on primary copy (PCRA). In PCRA, update transactions are executed only in the host of the primary copy. Any host can access the latest version of data if it is connected to the host of the primary copy.

3.1 Data Structures and Assumptions

Hosts in a MANET can be classified by their capabilities [4]; LMH (Large Mobile Host) has unlimited resources while SMH (Small Mobile Host) has limited resources. SMH typically caches a portion of the database and LMH contains the entire database. LMH usually takes the primary responsibility for updates of transactions. Assume that n be the total number of data items and m is the total number of LMH. We also choose D and LMH_S as the set of all data items and the set of all LMH respectively, i.e. $D = \{d_1, d_2 \dots d_n\}$ and $LMH_S = \{LMH_1, LMH_2, \dots, LMH_m\}$. Furthermore we assume LMH_j be the host of the primary copy of the data item d_i ; d_i must be updated by LMH_j and if d_i is not accessed due to network partitions, update transactions to d_i are aborted. A transaction is represented by T and a specific transaction k can be written as T_k . Transactions have predefined maximum error values; M_k represents maximum error value for T_k . When the maximum error value is greater than the limit α , query-only transactions can be executed by other LMH, i.e. dirty values are allowed in applications.

Each SMH manages the information of neighboring LMH in N-TBL. Update or query transactions are

forwarded to LMH registered in N-TBL. LMH has primary information table (P-TBL) and update information table (U-TBL); P-TBL has the information of the host of the primary copy and U-TBL registers updated data items of $[data_id, ts(d_k)]$. The $ts(d_k)$ is the logical timestamp to update data item d_k . $P-TBL(d_r)$ denotes the host of the primary copy for data item d_r .

3.2 Primary Copy based Replication Scheme

Figure 1 shows the procedure of query-only transaction. The proposed scheme first checks whether the desired transaction is read-only or update transaction. If the request is a read-only transaction, the host sends it to LMH_i which is a member of N-TBL, i.e. $LMK_i \in N-TBL$. When the incoming request is the update, the host transfers the request to the host of the primary copy. The detailed is as follows:

```

Procedure PCRA-Query Only Transactions:
1.  If ( $M_k > \alpha$ ) then
2.    return each  $d_r$  when  $d_r \in T_k$ ;          /* step 1 */
3.  else {
4.    for (each  $d_r \in T_k$ ) {                    /* step 2 */
5.      if ( $is\_not\_conntecd((P-TBL(d_r))$ )
6.        return abort;
7.    }
8.    for (each  $d_r \in T_k$ ) {                    /* step 3 */
9.      transfer message for  $d_r$  to  $P-TBL(d_r)$ ;
10.     receive the new value  $d_r$  from  $P-TBL(d_r)$ ;
11.     return each  $d_r \in T_k$  when  $d_r \in T_k$ ;
12.   }
13. }

```

Figure 1. The procedure of query-only transaction in LMH.

If the request is read-only transaction, the procedure is as follows. In step 1 (line 2), M_k is the allowed maximum error value of T_k . If M_k is greater than the predefined limit α , LMH_k simply returns to SMH. The d_r may be a dirty value. In step 2 (line 4 ~ 7), the test of $is_not_conntecd(P-TBL(d_r))$ is first checked. If the host of the primary copy of data item d_r is not connected the LMK_i , T_k is aborted. In step 3 (line 8 ~12), if connected, the host of the request will receive the new value of data item d_r from the host of the primary copy of d_r .

In the case of update transaction T_k , SMH requests T_k to LMK_i . The request, i.e. a transaction T_k , consists of two kinds of data items: read data set (RS_k) and write data set (WS_k). The data item in the transaction RS_k is noted as d_r^k , hence $P-TBL(d_r^k)$ represents the host of the primary copy of d_r^k . On the other hand, The data item in the transaction WS_k is noted as d_w^k , hence $P-TBL(d_w^k)$ represents the host of the primary copy of d_w^k . The update procedure is as follows:

- [1] SMH sends an update transaction T , to LMK_i . T_k consists of RS_k and WS_k .
- [2] Upon receiving of T_k , LMK_i checks that each $d_r^k \in RS_k$ and each $d_w^k \in WS_k$ to see whether LMH_k is connected to $P-TBL(d_r^k)$ or $P-TBL(d_w^k)$.

- [3] Upon receiving of T_k , LMH_i checks that each $d_r^k \in RS_k$ and each $d_w^k \in WS_k$ to see whether LMH_k is connected to $P-TBL(d_r^k)$ or $P-TBL(d_w^k)$.
- (1) $\exists d_i$, LMH_i is not connected to $(P-TBL(d_i^k))$, where $d_i^k \in (RS_k \cup WS_k)$
 LMH_i returns an abort message to T_k .
 - (2) $\forall d_i$, LMH_i is connected $(P-TBL(d_i^k))$, where $d_i^k \in (RS_k \cup WS_k)$
 LMH_k and the host of the primary copy are within the same partition. LMH_k accesses the host to process T_k . Process step 3.
- [4] LMH_i requests a read lock for each d_r^k , where $d_r^k \in RS_k$, to $P-TBL(d_r^k)$.
Upon receiving the request, $P-TBL(d_r^k)$ acquires a read lock on d_r^k .
- (1) If there is no lock on d_r^k , grant the lock and send LOCK-OK message to LMH_k
 - (2) If there is a read lock on d_r^k , grant the lock and send LOCK-OK message to LMH_k
 - (3) If there is a write lock on d_r^k , then enqueue the lock request. After this lock is released, send LOCK-OK message to LMH_k
- [5] LMH_i requests a read lock for each d_w^k , where $d_w^k \in WS_k$, to $P-TBL(d_w^k)$.
Upon receiving the request, $P-TBL(d_w^k)$ acquires a write lock on d_w^k .
- (1) If there is no lock on d_w^k , grant the lock and send LOCK-OK message to LMH_i
 - (2) If there is a read lock on d_w^k or a write lock on d_w^k , then enqueue the lock request.
After this lock is released, send LOCK-OK message to LMH_i
- [6] After receiving the message from each LMH_r in $LMH_r \in P-TBL(d_r^k)$ or $P-TBL(d_w^k)$.
 LMH_i sends a commit message to SMH and registers to all d_w^k in WS_k into U-TBL. LMH_i piggybacks this information of U-TBL with a lock request message of next update transactions to LMH_r . On receiving of information U-TBL, LMH_r release all locks for T_k and then update old d_w^k to newer d_w^k . If LMH_i doesn't receive from any LMH_r after timeout, it sends an abort message to SMH.

3.3 Analysis

We analyze the overhead and the qualitative difference of ETS and PCRA only. STE is eliminated from comparison study because it cannot guarantee data consistency unlike ETS and PCRA. We use the first order radio model [7, 15]. The energy dissipation to run the transmitter or receiver circuitry is set by the real experimental values, E_{elec} is = 50 nJ/bit, and the transmit amplifier as $E_{amp} = 100$ pJ/bit/m² [7]. Assume that $E_{Tx}(k, d)$ and $E_{Rx}(k)$ be the energy to transmit a k -bit packet to a distance d and to receive that packet respectively. The energy consumption equations are as follows:

$$E_{Tx}(k, d) = E_{elec} \times k + E_{amp} \times k \times d^2$$

$$E_{Rx}(k) = E_{elec} \times k$$

The cost to transmit a message of k -bit from the host i to j is $C_{ij}(k) = 2 \times E_{elec} \times k + E_{amp} \times k \times d_{ij}^2$, where d_{ij} denotes the distance from host i to host j .

Assume that $\#N$ and $\#Q$ are the number of hops and the number of quorum hosts per quorum set. If $\#N$ is the same in all hosts, the consumed energy by PCRA is equal to $\{\#N \times C_{ij}(k)\}$ because PCRA sends the request to host of the primary copy. ETS dissipates the amount of $\{\#Q \times \#N \times C_{ij}(k)\}$; it sends a request to all quorum hosts in a quorum set. ETS also consumes more communication overheads than that of PCRA.

4 Experiments

We compare the performances by STE, ETS and PCRA. Figure 2 shows the simulation model by CSIM discrete-event simulation package [13]. The simulation parameters are the same as those of [9]. We assume a system of 100 mobile hosts, consisting of 75 SMHs and 25 LMHs. Initially, the 100 hosts are randomly distributed in a square region with 1000 length units.

To add the effect of movement by SMH, we run repeatedly the following two steps: (i) the duration for a host to stay at a location is exponentially distributed with mean 0.5 time units, and (ii) at the end of this duration, the host randomly selects another location within the square region that is at most 75 length units away from its current location, and instantaneously moves to a new location.

The *replica control manager* implements the replication schemes to evaluate performances. We adopt the same quorum size as [9] to implement STE and ETS. 25 LMHs are deployed in a 5×5 square grid and 25 quorums are formed, where each quorum is composed of the union of a row and a column of LMHs. Each quorum has 9 LMHs, i.e. (5 from a row, 5 by a column and one by double count), and is assigned by distinct sequence number from 1 to 25. Each LMH has DQL (Disqualified List) or PCA directory. In PCRA LMHs have PCA directory for the hosts of the primary copy. The transaction generator generates transactions, a sequence of database operation (read or write). The number of operations in a transaction is set to one for simplification. The performance metric is the ratio of successful update transactions to communication overhead, which is the number of messages for update transactions.

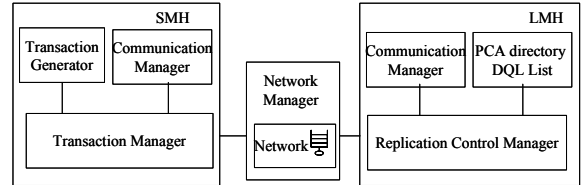


Figure 2. Simulation Model

The performance evaluation is examined in various wireless ranges. Each time a new link is established, the simulator executes Floyd-Warshall algorithm to detect the reachability among hosts and the presence partitions in the network. Figure 3 shows the impact by different wireless

ranges on the number of successful committed update transactions. When the wireless range is from 400 to 500 length units, almost all the update transactions succeed in three schemes due to few partitions in the wireless range.

When partitions occur, the partitions get merged quickly. When the wireless range decreases, the success ratio drops rapidly. Communication overheads by large wireless ranges are represented in figure 4. ETS incurs a very high communication overhead in range of 200 length units, but PCRA and STE make reasonable overheads. PCRA has the least overhead because it sends a single message to the host of the primary copy. In the range of 100 length units, PCRA incurs more communication overhead than that of ETS because ETS sends an update message only if all hosts of at least one quorum are within the partition.

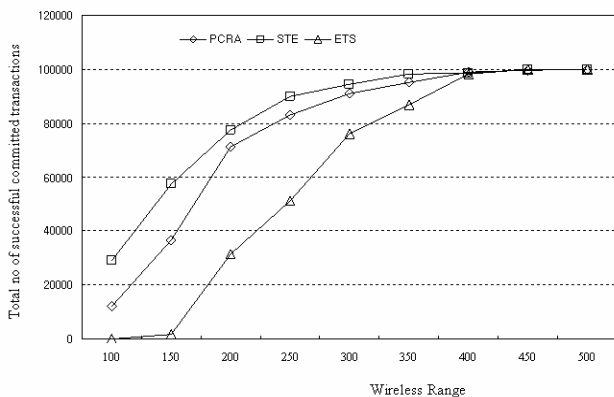


Figure 3. The number of successful committed update transactions with varying wireless range

5 Conclusions

We propose a data replication scheme for MANET, which provides data consistency and increased number of successful committed transactions reducing energy consumption. The proposed scheme updates the data in the host of primary copy, which guarantees the latest value of the data item. If an application wants the latest version of data, it should refer to the primary copy; otherwise it accesses the local copy of the data. When the application does not require the latest version, our scheme approves the access of the data item to the local copy. It can improve the energy efficiency by reducing the additional number of messages to maintain data consistency and to query data. Furthermore we can expect the performance improvement by alleviating the message traffic or by reducing the number of aborted transactions due to network partitions because the proposed scheme validates any conflicts only on the host of the primary copy.

We evaluate the performance of our approach (PCRA), ETS and STE. The results are as follows: for the wireless range from 400 to 500 length units, almost all the update transactions succeed. In 200 length units of the range, ETS incurs the highest communication overhead due to the high probability of network partition.

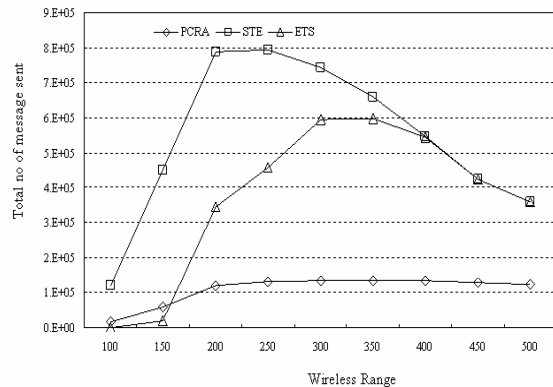


Figure 4. The number of message sent with varying wireless range

6 References

- [1] L. Fife and L. Gruenwald, "Research Issues for Data Communication in Mobile Ad-Hoc Network Database Systems," *SIGMOD Record*, 32(2), pp. 42-47, 2003
- [2] J. Grey, P. Helland, P.O'Neil and D. Shasha, "The Dangers of Replication and a Solution," *Proc. ACM SIGMOD*, pp. 173-182, 1997
- [3] V. Gianuzzi, "Data Replication Effectiveness in Mobile Ad-Hoc Networks," *PE-WASUN'04*, pp. 17-22, 2004
- [4] L. Gruenwald, M. Javed and M. Gu, "Energy-Efficient Data Broadcasting in Mobile Ad-Hoc Networks," *Proc. International Database Engineering and Applications Symp.*, pp. 64-73, 2002
- [5] T. Hara, "Effective Replica Allocation in Ad Hoc Networks for Improving Data Accessibility," *IEEE INFOCOM*, pp. 1568-1576, 2001
- [6] T. Hara and S. Madria, "Dynamic Data Replication Using Aperiodic Updates in Mobile Adhoc Networks," *Proc. Intl. Conf. DASFAA 2004, LNCS 2973*, pp. 869-881, 2004
- [7] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient Communication protocol for Wireless Microsensor Networks," *Proc. System Sciences*, pp. 3005-3014, 2000
- [8] J. Holloday, D. Agrawal and A. Abbdai, "Planned Disconnections for Mobile Databases," *DEXA Workshop*, pp. 165-172, 2000
- [9] G. Karumanchi, S. Muralidharan and R. Prakash, "Information Dissemination in Partitionable Mobile Ad Hoc Networks," *Symposium on Reliable Distributed Systems*, pp. 4-13, 1999
- [10] B. Kemme and G. Alonso, "A New Approach to Developing and Implementing Eager Database Replication Protocols," *ACM Trans. Database Syst.*, 25(3), pp. 333-379, 2000
- [11] S. Nesargi and R. Prakash, "MANETConf: Configuration of Hosts in a Mobile Ad Hoc Network," *IEEE INFOCOM*, pp. 1059-1068, 2002
- [12] C. Perkins, *Ad Hoc Networking*. ADDISON-WESLEY, 2001
- [13] H. Schwrtman, *CSIM User Guide for use with CSIM Revision 16*, MCC, 1992
- [14] M. Tamori, S. Ishihara, T. Watanabe and T. Mizuno, "A Replica Distribution Method with Consideration of the Positions of Mobile Hosts on Wireless Ad-hoc Networks," *Proc. Intl. Conf. Distributed Computing Systems Workshops (ICDCSW '02)*, 2002
- [15] H. Tan and I. Körpeoğlu, "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks," *SIGMOD Record*, 32(4), pp. 66-71, 2003