Efficient Data Dissemination in Mobile P2P Ad-Hoc Networks for Ubiquitous Computing

Do Hoon Kim, Myoung Rak Lee, Longzhe Han, Hoh Peter In
Department of Computer Science and Engineering
Korea University, Seoul, 136-701, Korea
{karmy01, nail2727, longzhehan, ydchung}@korea.ac.kr

Abstract

In this paper we propose Rank-Based Broadcast (RBB) algorithm using High Order Markov Chain (HOMC) with weight values in mobile p2p ad-hoc networks (MOPNET) for ubiquitous. RBB using HOMC focuses on the first step to improve the performance of Ubiquitous broadcast computing. The proposed model is characterized by its ability to optimize the receiving query and sending report for broadcasting with probabilistic priority by Markov assumption. Moreover, our model enhanced the reliability of broadcasting by weight value. In this paper we also set up a mobile p2p ad-hoc networks by Qualnet simulator which created the value of the specific configuration of mobile units and estimated the performance our model with other algorithm or ideal case.

1. Introduction

As the evolution of broadband wireless networks and mobile devices, such as WiMAX, Wi-Fi, Personal Digital Assistant (PDA) and cell phone, more and more applications which are typically used for wired networks and desktop computers are also highly required for mobile environment in Ubiquitous broadcast computing. For example, mobile database systems[2] are designed by Database researchers, practitioners, and commercial organizations to enable users to access and manage databases by mobile devices at anywhere and anytime. The computing power, memory capacity, and storage of PDA and cell phones are significantly increasing and these devices are capable to do database tasks quite efficiently. The advances in wireless networks have made it possible to break the spatio-temporal limitations and connect to networks. Although introduction mobility into database systems can provide more promising functionalities and services to users, it imposes new issues for database management systems in Ubiquitous environment which are not existed in conventional systems.

There are two basic types of conventional database system architectures: centralized and distributed. The centralized database systems consist of only one server (we can think the clustered server farms logically as one virtual server) and a number of clients. The server deals with all tasks of the database, such as storage allocation, transaction management, query processing, backup, and load balancing. Clients pose queries to the server by predefined interfaces, for example SQL. In distributed database systems, several independent centralized servers connect each other by broadband links.

Table 1. Comparison of data distribution schemes

<table>
<thead>
<tr>
<th>Activity</th>
<th>Full Partition</th>
<th>Full Replication</th>
<th>Partial Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data locality</td>
<td>Low</td>
<td>Medium</td>
<td>Very high</td>
</tr>
<tr>
<td>Serialization</td>
<td>Easy</td>
<td>Moderate</td>
<td>Difficult</td>
</tr>
<tr>
<td>Global consistency</td>
<td>Easy</td>
<td>Moderate</td>
<td>Difficult</td>
</tr>
<tr>
<td>Reliability</td>
<td>Low</td>
<td>Moderate</td>
<td>Very high</td>
</tr>
<tr>
<td>Recovery cost</td>
<td>Low</td>
<td>Moderate</td>
<td>Very high</td>
</tr>
</tbody>
</table>

According to data distribution schemes, the architecture of distributed database systems can be further divided into: full Partition, partial replication and full replication. In full partition scheme, each server holds a part of the whole database. However, in full replication scheme, every server has a copy of the whole database. And the database is partitioned and a subset of partitions is replicated at more than one
server in partial replications. Table 1 shows different characteristics of these three schemes.

Because mobile users could access database system at different locations, mobility introduces another dimension to conventional data distribution: Location-Dependent Data (LDD) and Location-Independent Data (LID). LID is the type of data which is managed by conventional database systems; the value of data does not depend on the specific location, such as personal bank account information, Social Insurance Number (SIN) and personal profile information should be same while information accessing locations maybe different. With increasing widespread Wireless LANs, mobile devices can construct an ad-hoc network (MANET) to share and distribute the data (LDD) which is highly related that location where the users are. The dynamic and infrastructureless architecture tightly fits the LDD requirements. However the mobile devices usually have relatively low computing power, battery and network bandwidth, an efficient data management scheme is needed. To gain better insight to performance of mobile device for constructing of ad-hoc and communication with each node, we propose Rank-Based Broadcast using Higher Order Markov Chain in Mobile P2P Network for Ubiquitous broadcast computing and compare with exist RBB algorithm and ideal case. Our proposed model expresses priority for data dissemination as a deterministic state maximized in its likelihood by observing the strongly related interested query of user and its response such as report.

The rest of the paper is organized as follows. Related work is described in Section 2. Our proposed model is presented in Section 3. A case study is described and compared with exist algorithm in Section 4, and Section 5 concludes the paper.

2. Related Work

2.1. Rank-Based Broadcast Algorithm (RBB)

As mentioned in the introduction, Mobile Ad-hoc Network (MANET) consists of a number of wireless mobile nodes, which freely corporate with each other and forms a network without any pre-existing infrastructure. Paper [3] discussed the data communication issues in MANET database systems. Also a novel architecture, mobile peer-to-peer database, was proposed in [1], [5]. P2P and MANET systems share the key concepts of self-organization and decentralization, and need to provide communication functionalities in a completely decentralized environment. In both systems, no central infrastructure is required, which manage and coordinate the network, and also mobile nodes could freely join in/leave to network. The mobility might affect the topology of the network continually changing. The data flooding or broadcasting mechanism is required in these systems, in order to maintain and update the network structure, and to distribute data across the network. Finally the mobile nodes in both networks not only act as hosts but also routers to forward data and discover routes.

Despite these similarities, there are also several differences between P2P and MANET systems. The most important one is the different layers they work on. If comparing with TCP/IP protocol suite, MANET works at network layer, and P2P operates at application layer. Because database systems usually run on application layer, so we focus on data management issues on mobile P2P database systems. Generally pull or push methods are used for data distribution through flooding or broadcasting. In [1][6], a Rank Based Broadcast (RBB) method was proposed for data distribution in mobile P2P databases. In RBB method, mobile nodes could receive several pull messages (called query message) from other nodes and generate push messages (called report message). When mobile nodes received enough queries or they have adequate amount of reports, a broadcast is triggered. The broadcasting messages are decided by the RBB algorithm, which is defined as following:

Let \( Q_1, Q_2, ..., Q_n \) be the queries in mobile node \( O \)'s queries relation. The rank of a report \( a(R) \) at \( O \) at time \( t \) is:

\[
rank(a(R)) = \sum_{i=1}^{n} \text{match}(a(R), Q_i)
\]  

(1)

The satisfaction of A report \( a(R) \) to query \( Q \) can be decided if \( O(a(R)) = \text{true} \) and \( \text{match}(a(R), Q) = 1 \). Otherwise \( O(a(R)) = \text{false} \) and \( \text{match}(a(R), Q) = 0 \). By using RBB, the mobile nodes can broadcast the most “wanted” data by other nodes and improve the performances of data distribution.

2.2. Higher Order Markov Chain (HOMC)

In the following, we assume that each data point \( X^m \) in a categorical data sequence takes values in the set

\[
\mathcal{M} = \{1, 2, ..., m\}
\]

and \( m \) is finite, i.e., a sequence has \( m \) possible categories or states. The conventional model for a \( k \)-th order Markov chain has \( (m - 1)m^k \) model parameters. The major problem in using such kind of model is that the number of parameters (the transition probabilities) increases exponentially with respect to the order of the
model. This large number of parameters discourages people from using a higher-order Markov chain directly. In Raftery\cite{4} proposed a higher-order Markov chain model which involves only one additional parameter for each extra lag. The model can be written as follows:

\[ R = P(X^n = j_0 \mid X^{n-k} = j_1, \ldots, X^{n-1} = j_k) = \sum_{i=1}^{k} \lambda_i q_{j_0 j_i} \]

where \[ \sum_{i=1}^{k} \lambda_i = 1 \] (2)

and \( Q = [q_{ij}] \) is a transition matrix with column sums equal to one, such that

\[ 0 \leq \sum_{i=1}^{k} \lambda_i q_{j_0 i} \leq 1, j_0 j_i \in M \] (3)

The constraint in (3) is to guarantee that the right-hand-side of (2) is a probability distribution. The total number of independent parameters in this model is of size \( (k+m^2) \). Important point is that we have to consider the weight value for user’s interested query. So, we proposed the extended Raftery’s model to a more general higher-order Markov chain model by allowing \( Q \) to vary with different lags. Here we assume that the weight \( \lambda_i \) is non-negative such that

\[ \sum_{i=1}^{k} \lambda_i = 1 \] (4)

It should be noted that equation (2) can be re-written as

\[ X^{(n+k+1)} = \sum_{i=1}^{k} \lambda_i q X^{(n+k+1-i)} \] (5)

where \( X^{(n+k+1)} \) is the probability distribution of the states at time \( (n+k+1-i) \). Using (4) and the fact \( Q \) is a transition probability matrix, we note that each entry of \( X^{(n+k+1)} \) is in between 0 and 1, and the sum of all entries is also equal to 1. In Raftery’s model, it does not assume \( \lambda \) to be non-negative and therefore the additional constraints (3) should be added to guarantee that \( X^{(n+k+1)} \) is the probability distribution of the states. Raftery’s model in (5) can be generalized as follows:

\[ X(n+k+1) = \sum_{i=1}^{k} \lambda_i q X^{(n+k+1-i)} \] (6)

The total number of independent parameters in the new model is \( (k + km^2) \). We note that if

\[ Q_1 = Q_2 = \ldots = Q_k \] (7)

then (7) is just the Raftery’s model in (5).

In the model we assume that \( X^{(n+k+1)} \) is depends on \( X^{(n+i)} \) \( (i = 1, 2, \ldots, k) \) via the matrix \( Q_i \) and weight \( \lambda_i \). One may relate \( Q_i \) to the \( i \)-step transition matrix of the process and we will use this idea to estimate \( Q_i \). Here we assume that each \( Q_i \) is a non-negative stochastic matrix with column sums equal to one. In this paper this assumption used for priority of broadcasting in MOPNET.

3. Rank-Based Broadcast using Higher Order Markov Chain in Mobile P2P Network

In this chapter, Rank-Based Broadcast using Higher Order Markov Chain in Mobile P2P Network (RBB using HOMC in MOPNET) is proposed. The conceptual overview of the model is like Figure 1.

![Figure 1. Overview of RBB using HOMC in MOPNET](image-url)
hybrid algorithm. To establish components of Rank-Based Broadcast using Higher Order Markov Chain one by one, the following sub tasks are required.

- **Sender.** All related states such as the transition matrix of interested query from each node trained in this part. So, we calculate the higher order Markov model follow equation.

The first order Markov model can expend to \( n \)-th order Markov model with change of past continuous user’s interested pattern. In Figure 2, when it expends the continuous user’s interested propensity with time until \( n \)-th, we can call \( n \)-th order Markov model.

**Figure 2. Higher Order Markov Chain Process**

That is a higher order Markov chain. So, if we consider the higher order, we represent the generated higher order Markov chain as following.

\[
P(X(t + 1, d) = x_j | X = x_i, X(t - 1, d)) = \frac{n_{x_j}}{\sum_{i=1}^{n} n_{x_i}} \quad (8)
\]

where \( t = time, d = distance, n = order. \)

Also, we compute the broadcast size \( k \) in this step. That is, it is possible to include the total number of reports and queries in the broadcast if the broadcast is triggered. Also, it can be adapted to the length of the time period from the last broadcast of moving object \( O \). According to this concept, we consider the expected value \( E(x) \) of random variable for broadcast size \( k \) as following equation [1].

\[
2\pi k \int_{0}^{r} \delta (1 - p) \lambda r^2 \left( 2q \left( \frac{d}{r} \right) + (n - 2q \left( \frac{d}{r} \right)) \times (2r + 1) \right) \, dr \leq E(x) \leq \sum_{i=1}^{n} n_{x_i} \quad (9)
\]

So, sender is that their report broadcast with native query and broadcast size \( k \) to other node. If we take this value as the broadcast size, the observation fashion of broadcast size is dynamically and adaptively adjusted to the length of time between subsequent broadcasts with the length of the time. Furthermore, the broadcast size \( k \) can allocate the parameter of a bandwidth such as only 10% of the available short-range bandwidth to mobile P2P resource discovery. Then the broadcast size is taken to be only 10% of the optimal \( k \). Rank the reports in the reports database using Equation (8). Denote by \( S \) the set of top \( k \)-1 reports.

- **Receiver.** Rank the reports in the reports database using Equation (8). Denote by \( S \) the set of top \( k \)-1 reports. Also, if they communicate each other, they try to re-broadcast to others. So, this step estimate the broadcast trigger with the distance traveled and the change to the database since last broadcast. Moreover, this step try to optimize for reducing of duplicated transmissions by disseminating only new data to old neighbors and old data only to new neighbors. Observe that RBB performs a broadcast without a handshake, thus it has no knowledge of who are its neighbors. Thus it is assumed that when the moving object has traveled a distance of \( 2r \) or more, its set of neighbors is new. So, we can calculate the expected value (i.e. rank of report) by combined two equations like a following process.

At first, we consider the rank function and priority by Markov higher order model like a follow:

\[
rank(a(R)) = \sum_{i=1}^{n} \text{match}(a(R), Q_i)
\]

where

\[
R = P(X(t + 1, d) = x_j | X(t, d) = x_i, X(t - 1, d)) = \frac{n_{x_j}}{\sum_{i=1}^{n} n_{x_i}} \quad (10)
\]

A proposed model with a modified equation of RBB and HOMC; the model’s effectiveness is investigated like a follow:

\[
\text{Rank} = \sum_{i=1}^{n} \text{match} \left( a \left( \frac{n_{x_j}}{\sum_{i=1}^{n} n_{x_i}}, Q_i \right), Q_i \right) \quad (11)
\]

4. Case Study

4.1. Construction of Qualnet Simulation
We configured Mobile p2p ad-hoc network using Qualnet4.0 simulator which created the value of the specific configuration of mobile units for estimating of performance in Figure 3. In simulator we used 6 mobile units to make peer-to-peer environment. As a specific procedure, firstly, we located six mobile units which are connected with mobile p2p ad-hoc network. In particular, at the start time of the simulation n moving objects are randomly placed in a 0.5mile×0.5mile square area.

Each object has a destination which is another random point in the space. Secondly, we choose 4 pairs of source and destination nodes which can communicate each pair of node with constant bit rate. And then, finally set up specific properties like Table 2.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Source (node num')</th>
<th>Destination (node num')</th>
<th>Items to Send</th>
<th>Items Size (Byte)</th>
<th>Interval (s)</th>
<th>Start &amp; End (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBR</td>
<td>7</td>
<td>8</td>
<td>100</td>
<td>512</td>
<td>0.1</td>
<td>1–25</td>
</tr>
<tr>
<td>CBR</td>
<td>9</td>
<td>1</td>
<td>100</td>
<td>512</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>CBR</td>
<td>9</td>
<td>7</td>
<td>100</td>
<td>512</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>CBR</td>
<td>11</td>
<td>10</td>
<td>100</td>
<td>512</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Specific Configuration of Mobile Unit

Each source node sends 512bytes packets to destination node for the simulation time. In the simulator, total simulation duration is 25 second, and 4pairs of mobile unit communicate with p2p communication environment. Each source node send 100 packets of 512 bytes size with 0.1 second intervals to destination node.

4.2. Ranking based on HOMC for broadcasting

For the case study, we selected only user's interested query and its frequency of response report frequency where r1, r2, ..., r8 as Table 3.

Table 3. Preferred calculation using 1-order Markov Chain from data

<table>
<thead>
<tr>
<th>r</th>
<th>r1</th>
<th>r2</th>
<th>r3</th>
<th>r4</th>
<th>r5</th>
<th>r6</th>
<th>r7</th>
<th>r8</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r2</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r4</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>r7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>r8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In Table 3, the 8 x 8 matrix (r1, r2, ..., r8) express the event frequency of response each user. Also the equation (12) is that transition of inferred stochastic characteristic calculated by time line.

\[
\hat{P}(X(t + 1, d) = x_j | X(t, d) = c_i) = \frac{\sum_{n_{ij}} n_{ij}}{\sum_{n_{ij}} n_{ij}} \tag{12}
\]

Also, it calculates rank with time t and t+1 like a Table 4.

Table 4. Used report at time t and t+1 from test data

<table>
<thead>
<tr>
<th>r1</th>
<th>r2</th>
<th>r3</th>
<th>r4</th>
<th>r5</th>
<th>r6</th>
<th>r7</th>
<th>r8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Also, we can get the priority of report for broadcast with weight value like a following Table 5.

As a result, we can choose the r1, and r5. However, r3 is selected. Because the weight value of report (r3) is higher than r1 (1.611 > 1.429) in Table. According to, another rank of reports is decided by above calculation continuously.
In Figure 4 is a comparison among RBB using HOMC, exist RBB and ideal value.

<table>
<thead>
<tr>
<th></th>
<th>r1</th>
<th>r2</th>
<th>r3</th>
<th>r4</th>
<th>r5</th>
<th>r6</th>
<th>r7</th>
<th>r8</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r3</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r4</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>r8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5. Preferred calculation using Weighted Markov Model from trained data.

In Figure 4 is a comparison among RBB using HOMC, exist RBB and ideal value.

5. Conclusion

In this paper, we proposed novel Rank-Based Broadcast(RBB) model which considered Markov chain to disseminate query report efficiently in mobile p2p ad-hoc networks for Ubiquitous Computing. The application of this research resulted as follows; firstly, we could determine the priority of user’s interested query report in mobile p2p ad-hoc networks. Secondly, we implemented dynamic experimental environment by Qualnet simulation. Thirdly, in our algorithm, we proposed broadcasting methodology by priority of data to overcome the limited resources.

As a future work, we need to consider various wireless environments like as IEEE 802.11, 802.15 and 802.16 and others. Moreover, we have to consider more optimized algorithm in our model, because of computational overhead by increasing nodes and query.

6. References