Research Journal of Applied Sciences, Engineering and Technology 6(9): 1639-1648, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: December 31, 2012 Accepted: February 08, 2013 Published: July 15, 2013

Review of Data Replication Techniques for Mobile Computing Environment

Haroon Shahzad, Xiang Li and Muhammad Irfan
School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Abstract: With the tremendous growth of the portable computing devices and wireless communication, the mobile computing has become extremely popular. The wireless enabled portable computing devices with massive storage capacity and high-end processing capabilities have started to make the extensive use of mobile databases already. The rising popularity in these computing paradigms demands that the mobile computing be reliable enough to ensure the continuous data availability and integrity. However mobility comes at the cost of bandwidth, limited power, security and interference. In modern mobile computing systems data replication has been adopted as an efficient means to ensure the data availability and integrity as well as an effective means to achieve the fault tolerance. Data replication not only ensure the availability of the data but also reduce the communication cost, increase data sharing and increase the safety of the critical data. Furthermore, it also determine when and where (location) to place the replica of data, controlling the number of data replicas over a network for efficient utilization of the network resources. In this study we survey the research work in data replication for mobile computing. We reviewed some of the existing data replication techniques proposed by the research community for mobile computing environment for efficient management of data replicas.

Keywords: Computing algorithm, data replication, data availability, fault tolerance, mobile computing, reliability, reliability engineering

INTRODUCTION

The field of wireless and mobile computing is a conjunction of the personal computing, distributed computing, wireless networks and Internet applications. This integration is supported by a large number of devices and wireless networks, which is based on a continuously and increasing interaction between communication and computing. Mobile computing system is a type of distributed system (Biswas and Neogy, 2010). The success of mobile data communication lies in the expectation to provide different services to users anytime and anywhere. In simple terms mobile computing can be defined as the computing on the go. Mobile computing is a new software paradigm that is of tremendous interest in the Information Technology research community. Today, mobile computing technology is used to link portable computing equipment to corporate distributed computing and other sources of information. Many researchers and scientists from both academia and industry are undertaking efforts to explore new technology for mobile computing and wireless communication, in a heterogeneous wireless and wired network environment, with mobile computing applications (Boukerche, 2006). We begin by considering the requirement for mobility and its cost.

Mobility and portability are important aspects in mobile computing. With its popularity, it is very important that these systems be dependable and fail safe. To address the question of data availability various data replication protocols and techniques have been proposed and developed in the mobile computing systems. Data replication increases data availability and reduces data access latency may be at the cost of data storage. The main goal of mobile computing was to support the anytime, any-form and anywhere computing with the tremendous growth in the mobile technologies. To provide the data to the users with portable computers and mobile phones, the many a more techniques have been proposed for the improvement of QoS. All these requirements made mobile data management, transaction processing and query processing and data dissemination hot topics for research. Pitoura and Chrysanthis (2007) listed three challenges for the research in mobile computing:

Mobility: Hampers the capability of processing at the network layer. The nodes being highly moveable pose greater challenge and a number of complexities.

Limited resources: Limited battery, limited processing capabilities and memory of mobile devices also contributes toward the challenges for mobile computing to be a dependable computing environment.

Intermittent connectivity: The absence of a permanent communication link caused frequent disconnection due to signal strength.

Corresponding Author: Haroon Shahzad, School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

The networking infrastructure in a wireless computing environment can be categorized into two main types i.e., single-hop and multi-hop. In the former infrastructure each of the mobile devices are connected with a stationary host, which corresponds to its point of attachment in the network. The stationary host (MSS) is responsible for all the routing and processing in the network. In the later type an ad-hoc wireless network is formed in which different host participate in routing messages among each other. In the former type a data dissemination tree is formed by the stationary devices (MSSs) while in the later the same is performed by corresponding wireless hosts. The hosts in the tree may store the data and take part in processing. Therefore caching or replicating data at the mobile host or at the stationary nodes of the dissemination tree are important for improving system performance and availability (Pitoura and Chrysanthis, 2007). According to Walke (2002), setting up three neighbouring cells with each of them serving 120° sector (120°×3 = 360°) could help reduce the number of base stations and hence reduced the overall cost. Walke (2002) also defines the spectral efficiency as the traffic capacity unit divided by the product of bandwidth and surface area element and is dependent on the number of radio channels per cell and the cluster size (number of cells in a group of cells):

$$Efficiency = \frac{N_c}{BW \times A_c} \quad (1)$$

where,
N_c = Number of channels per cell
BW = System bandwidth
A_c = Area of cell

The regular algorithms designed to manage the data are not suitable for the mobile environments because the number of clients, bandwidth, connection model, base structure, capacity of the clients and base station may differ from one region to another (Sorkhabi and Shahamfar, 2010). We have observed in the literature that the replication cost and the placement of data replicas have been widely studied for the performance issues (Korupolu *et al*., 1999; Li *et al*., 1999; Qiu *et al*., 2001; Venkataramani *et al*., 2001) by the research community. The data availability issues have also been explored for the traditional database (Barbara and Molina, 1986; Coan *et al*., 1986) concept as well. The active replication techniques have also been discussed at length where the clients communicate by multicast with all replicas (Coulouris *et al*., 2001; Wiesmann *et al*., 2000; Guerraoui and Schiper, 1997; Schneider, 1990). The aim of this study is to provide a comprehensive study that helps the designer to choose the most appropriate data replication strategies for mobile computing environment or design their own replication schemes based on the characteristics of previously done research work. This study surveys the existing data replication protocols that have been proposed in the literature and summarizes some of them. This detail study would further our understanding and knowledge base for the design of more suitable data replication scheme for our environment and would serve as the base line for our research work.

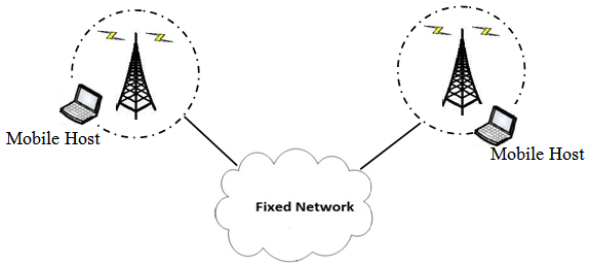

Fig. 1: A typical mobile computing scenario

LITERATURE REVIEW

Mobile computing: Mobile computing can be termed as, "taking a computer or computing device (Mobile Host) i.e., laptop, palmtop, PDA and all necessary files and software out into the field". It uses cell phones to connect to the internet to access data in addition to making voice calls. In general the term mobile computing refers to a variety of device mentioned above that could allow the user to access data from any where anytime. The 'anywhere anytime' access is possible only with a wireless connection and it brings certain limitation to the mobile computing paradigm. Furthermore, this expectation from the mobile computing paradigm brings challenges for the ongoing research to make it dependable. Figure 1 shows a typical mobile computing scenario where Mobile Support Stations (MSS) are connected through a high speed wired network. A Mobile Host (MH) i.e., cell phone, PDA, Laptop could possibly be in a cell of one station at a time. It could move from one cell to another at any time while keeping its session (Data-call) intact.

The communication between the clients and server ought to be minimized to reduce the contentions keeping in view the narrow bandwidth of wireless channels (Su *et al*., 2005). In addition to this the MH mobility also causes the degradation of data access. The mobility of mobile computing devices also changed the way mobile applications access and manages data. Instead of storing data in a central database, data is being moved closer to applications to increase efficiency and autonomy. This trend leads to many interesting problems in mobile database research (Earl, 2007). There has been a continuous research on the subject of Dependability of Mobile Computing to ensure the continuous system availability. To cater for these issues a large number of data replication algorithms have been proposed by the research community for various types of mobile computing scenarios. The data replication has becomes a very important technique for increasing data availability,

scalability and performance of the system hence effecting the QoS. The placement of replica nearby the commuters will reduce the access time and increase availability (Earl, 2007; Grenoble, 2004).

Mobile Data communication: It is the communication of data over the wireless network. Wide area Network (WAN) components may interconnect to transfer data. Generally, wireless data connections used in mobile computing take three general forms. Cellular data service uses technologies such as GSM, CDMA or GPRS and more recently 3G networks such as W-CDMA, EDGE or CDMA2000 are gaining popularity throughout the world. These networks are usually available within range of commercial cellular towers as shown in Fig. 1. Wi-Fi connections offer higher performance, may be either on a private business network or accessed through public hotspots.

Data replication technology: Data Replication in the computing terms means the mutual sharing of information so as to ensure consistency between redundant resources, such as software or hardware components, to improve reliability, fault-tolerance, or accessibility. Data replication in databases has been a hot topic for research and development. The prime objective of the replication is to ensure the easy availability of data in case of any failure or disastrous event. The data replicated at different sites will also increase the speed of access and hence reduce the communication cost of the network. However, replication also brings some extra processing and communication into the system in the form of updates and synchronization of various copies of data at various sites. For large databases many reliable replication tools are available in the market. The data replication techniques (Wolfson *et al*., 1997; Ratner, 1998; Acharya and Zdonik, 1993; Huang *et al*., 1994) for mobile computing take into consideration an environment where Mobile Hosts (MH) access the data at sites in the fixed network like in our strategy and create the replica of data on mobile hosts. These strategies also assume the one-hop communication. If the data is replicated onto a mobile host in this type of communication and if the Mobile Host becomes unavailable for any reason then it can cause the data unavailability hence degrading the performance of overall system. The mobile host failure with replicated data can cause the consistency, availability and accessibility problem. The data replication is one very important aspect in the distributed systems. There exist a number of well cited models for data replication each with its own distinct feature and properties (Marton and Attila, 2009).

Transactional replication model is used for replicating transactional data, for example a database or some other form of transactional data structure. The single-copy serialize model is employed in this case, which defines legal outcomes of a transaction on replicated data in accordance with the overall properties. Another important scheme is State machine replication model. This model assumes that replicated process is a deterministic finite automaton and that atomic broadcast of every event is possible. It is based on a distributed computing problem called distributed consensus and has a great deal in common with the transactional replication model. This is sometimes mistakenly used as synonym of active replication. State machine replication is usually implemented by a replicated log consisting of multiple subsequent rounds of the Paxos algorithm.

Virtual synchrony is another computational model which is used when a group of processes cooperate to replicate in-memory data or to coordinate actions. This is a group based scheme where different members join a group with current data scheme. The process then send multicast to the group members. The data transactions in mobile systems might have to be split into sets of operations due to disconnections and mobility properties and may share their current states and partial results with other Transactions. Therefore, the mobile computing transactions require computations and communications to be supported by stationary hosts. A mobile computations may be divided into a set of actions some of which execute on mobile host while other; on stationary host. When the MH moves from one cell to another, the state of transaction, states of accessed data objects and the location information also move, this process is known as the hand off procedure. The mobile transactions are little different in nature from the normal database system transaction due to the mobility of both the data and users and due to the frequent disconnections. Therefore in order to support mobile transactions, the transaction processing models should accommodate the limitations of mobile computing, such as unreliable communication (due to wireless link), limited battery life, low band-width communication and reduced storage capacity. Mobile computations should minimize the stoppages faced due to the frequent disconnections. Operations on shared data must ensure correctness of transactions executed on both MSS and MH. The blocking of transactions execution on either the MH or MSS must also be minimized to reduce communication cost. A Proper mechanism may also be required to support local autonomy to allow transactions to be processed and committed on the MH regardless of disconnections (Madria and Bhargava, 1997).

DATA REPLICATION TECHNIQUES

It has been established that the data replication technique increase the availability, scalability and performance in the databases. The replication techniques used for the traditional or distributed databases don't suffice the mobile environment because of the volatile nature of the network. As on today, several data replication strategies have been proposed for mobile computing.

In this section we will have deep analysis of some famous approaches devised by different authors, as it's impossible to analyze all approaches here in this study.

Access frequency based methods: Various authors have proposed replication techniques based on data access frequency. Hara (2001) has presented three replica allocation methods for enhancing data accessibility. The replica allocation method is calculated based on a certain criteria which gives the highest possible data accessibility in the whole network. If m1 denote the total number of MH which are connected to each other with either a single hop or multi-hop link and n1(n1 ≤ n) denote the number of types of data items held by a certain MH, m1 then the number of possible combinations of replica allocation is determined by the expression:

$$\left\{\frac{n_1!}{(n_1-C)!}\right\}^{m_1} \quad (2)$$

SAF (Static Access Frequency) which takes into account the access frequency of the corresponding data item. 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. The Mobile Host (MH) does not need to exchange the information with other MHs for replica allocation. Furthermore, the allocation process does not take place after allocation of necessary replica allocation by respective hosts. Therefore, it is said that the method incur the low overhead and low traffic. The problem however with this technique is it creates more duplicate replicas.

DAFN (Dynamic Access Frequency and Neighborhood) which in addition to access frequency also consider the neighborhoods of the Mobile Host (MH), It does determine the replica allocation in the same way as the SAF method then it check if there is a duplication of the data between two neighboring Mobile Hosts (MH) then a host with the lower access frequency to the data changes the replica to another. Hence the duplication is eliminated. DCG (Dynamic Connectivity based Grouping) which consider the network topology in addition to the individual data item access frequency. In addition to the DCG method shares the replicas in larger group of mobile hosts than that of DAFN which considered only the neighbors. Hence it creates MHs that are bi-connected component in the network (Aho *et al*., 1974).

The extension to these techniques i.e., E-SAF, E-DAFN and E-DCG has also been proposed (Hara, 2003) to cater for the updates based on PT value i.e. an average number of access requests for a given data resource. The PT value is given by the formula:

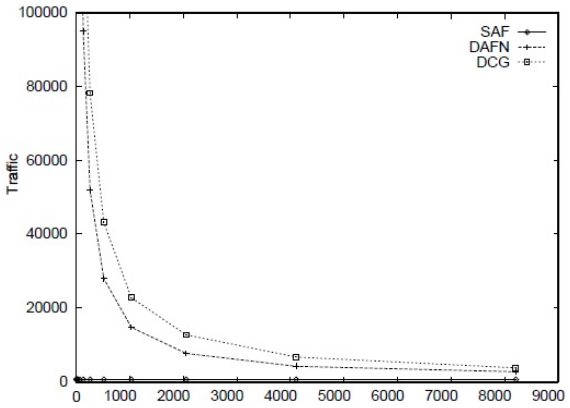

Fig. 2: Relocation period and traffic (Hara, 2003)

$$p_{ij}.\tau_j = p_{ij}.(T_j - t_j) \quad (3)$$

where, p_{ij} denotes the probability that an access request for a certain data item D_j from mobile host M_i is issued at a unit of time, i.e., the access frequency; τ_j denotes the time remaining until D_j is updated next; T_j denotes the update period of D_j; t_j denotes the time that has passed since D_j has been updated at the most recent update period.

The relocation period is different for all the methods which are represented by the Fig. 2.

Pre-Write operation based method: The Data Transactions are mainly in the form of a Read/Write operation in a system. A pre-Write operation before an actual write operation is introduced in order to enhance the availability of the data. Since, it would not change the state of the data but show the value of the data will have after the transaction has been committed. Once all the read and write values have been set by a transaction, it can then commit at a Mobile Host (MH) and the remaining execution is transferred to the fixed station i.e., MSS (Madria and Bhargava, 1997). Pre-committed model doesn't require a roll back and it's not aborted. To save the space and bandwidth, time and energy a Mobile Host (MH) can cache only pre-write values of the data items. According to this model each pre-write operation makes the values of write transaction visible. Pre-writes have different semantics in different environments. The pre-write and write values may match exactly for a very simple operation. The authors' further extends their argument by stating that a transaction is not allowed to terminate after pre-commit and the pre-write provides non-strict execution without aborting the operation. They considered a scenario of two sub-transactions where pw(x), w(x), pr(x) and r(x) are the pre-write value, write value, pre-read value and read-value respectively, for the data object x. The transaction T2 commits before T1. In case T1 aborts after T2 commits, there will not be a cascading abort:

$$T1 \longrightarrow r(x), pw(x) \longrightarrow pc \longrightarrow w(x) \longrightarrow c \quad (4)$$

$$T2 \longrightarrow pr(x) \longrightarrow c \quad (5)$$

The transaction is formulized as:
A transaction T_i is a partial order with ordering $<_I$ where,

$T_i \subseteq \{pr_i(x), r_i(x), pw_i(x), w_i(x) \mid x$ is an object$\}$
$\cup \{pc_i, c_i, a_i\}$

If $\alpha_i \in T_i$ if and only if $pc_i \notin T_i$ and $c_i \notin T_i$

If t is c_i or α_i then for any other operation $p \in T_i$, $p <_i$ tand if t is pc_i then $pc_i <_i c_i$

If $pr_i(x)$, $r_i(x)$, $pw_i(x)$, $w_i(x) \in T_i$ then either $pr_i(x) <_i w_i(x)$ or $w_i(x) <_i r_i(x)$ or $r_i(x) <_i (x)$,or $pr_i(x), <_i w_i(x)$ $w_i(x)$, $pr_i(x) <_i r_i(x)$, $pw_i(x) <_i pr_i(x)$ $pr_i(x)$

In the transaction T_i on the data object x is denoted as $r_i(x)$, while the write as $w_i(x)$ pre-read operation as $pr_i(x)$ and the pre-write operation as $pw_i(x)$ on data object x. The transaction history will include the locking and unlocking operations. Each operation $p_i(x)$ is followed by a locking operation $pl_i(x)$ and unlocking $pul_i(x)$ respectively. The update from a pre-write lock obtained by a certain transaction T_i on object x to the write is shown by:

$$pwl_i(x) \longrightarrow wl_i(x) \quad (6)$$

According to the authors this model helps to deal with the scarce resources such as battery power, limited memory, low bandwidth and frequent disconnection or weak connection. The model does not need to handle aborts by using before-image or compensation. A pre-committed transaction does not abort. In case a pre-committed transaction is forced to abort due to system dependent reasons such as crash or any malfunction, the transaction will be restarted on system restart. The pre-write method also helps the caching of the data during the weak connections. In case the mobile host is disconnected from the network or moved to another cell, the processing at MH can still continue using the pre-write operations. It doesn't have to wait for the commit of previous transaction at the fixed hosts because the fixed hosts also have the values of pre-write so the transaction can still continue execution in case of the disconnection from the Mobile Host (MH). The MH doesn't require the cache as the pre-writes are already available. It doesn't require the undo in case of abort and also simplifies the hand-off procedure.

Madria *et al*. (2000) have addressed the issues pertaining to the queries that are location dependent in mobile computing. They have developed and discussed the hierarchies based on the location specific data which define the mapping among different levels of locations. They have then used these hierarchies as scattered directories in order to speed up the process of finding the relation containing the value of location dependent attribute at a particular location. They have further extended the hierarchies to include the spatial indexes of such attributes. They have also discussed at length the process of location based partitioning and replication to process the queries efficiently.

One-copy two-copy method: In the very early days of mobile computing when it was merely an idea of technological future and no concrete development was made (Huang *et al*., 1994) analyzed the replication allocation methods. They have visited several data allocation algorithms for mobile computing environment including the one-copy and two-copy methods. A family of dynamic allocation methods has also been proposed. These methods need to select the allocation scheme according to he read/write ratio. It can be written as:
If (f(Read)) ↑ then

$two - copy()$
else
$one - copy()$
End

where, f the frequency of the read operation and symbol ↑ indicates the high frequency

The static and dynamic algorithms both are analyzed on the basis of worst-case and the expected case for read and write those are poison distributed. Mainly two cost models have been employed i.e., connection based and the message based. The average expected cost per request AVG_A is computed as:

$$AVG_A = \int_0^1 EXP_A(\theta) d\theta \quad (7)$$

Here θ is the probability that the next request is a write operation between 0 and 1. $EXP_A(\theta)$ is the expected cost of the relevant request. The probability that the next request would be read is mathematically written as:

$$1 - \theta = \frac{\lambda_r}{\lambda_\omega + \lambda_r} \quad (8)$$

In this expression λ_r and λ_ω are read and write distribution parameters and θ is the probability that the next operation is a write.

Access-cost based methods: The access cost is mainly due to low bandwidth, poor connectivity or excessive traffic in the network. A dynamic replication strategy has been proposed by Sun *et al*. (2009) to minimize the access cost called MAC Replication. According to their

proposed scheme first of all a popular file in context of access request is looked for and the average response time is calculated on each file in order to determine which logical resource is to be replicated. MAC scheme take the access cost as a measure. The service ration of the requesting peer is given by SR:

$$SR = \frac{R_p}{\sum_{1}^{n} R_m} \tag{9}$$

where, R_p is the requests processes and the R_m shows all the request messages from 1 - n.

The Reliable value of each peer RV is given by:

$$RV = \frac{online\ \ time}{per\ \ unit\ \ of\ \ time} \tag{10}$$

And the cost for a peer accessing the physical Replica R_i is give by the relation:

$$Cost_i = \frac{filesize \times JobLength_i}{BW_{pi} \times SR_i \times RV_i} \tag{11}$$

The optimized replica is selected through the prediction of the cost in the future. In Eq. 11 BW_{pi} represents the bandwidth.

MAC Replication scheme finds the Replica Scarce Domain in which peers also query the Resource LR_j with high ART which is then followed by appropriate site for MAC replication. The Average Response Time ART of LR_j is mathematically represented as:

$$ART_j = \frac{\sum_{i=0}^{n} \sum_{k=0}^{Time_i} RT_{Ri}[k]}{\sum_{i=0}^{n} Time_i} \tag{12}$$

As a last step MAC replication strategy selects optimized replica through predicting the access cost and replicate it to suitable sites. On contrary to other dynamic data replication strategies, they have proposed a dynamic approach for monitoring the load balancing and query latency to increase the replicas. Furthermore, they have also considered parameters like nodes request frequency, computing capacity etc. to optimize the process of replication. Yu and Vahdat (2002) have presented a minimal cost model for data replication. The replication cost model used by them is given as:

$$Cost = \sum_{i=1}^{n} (\alpha \times usage_i + \beta \times create_i + \gamma \times teardown_i) \tag{13}$$

In Eq. 13 $usage_i$ is the total time that $loction_i$ has a replica and α is the cost/unit time. $create_i$ Denotes the number of replica creations at location$_i$ and β is the creation cost. $teardown_i$ and γ represent the teardown cost of the respective replica.

Shivakumar and Widom (1995) have proposed a minimum cost maximum flow based algorithm for user profile replication based on user movement patterns in a mobile environment. They also shared the detailed analysis and results through (Guerraoui and Schiper, 1997) with research community. They assumed a classical personal communication service system to compute the profile replication strategy based on user movement patterns. The user profile assumed to contain phone number and location information. The minimum-cost maximum-flow algorithm is assumed to run on a central server which after a periodic computation of the replication strategy communicates to the different zones and area code level databases. The replication plan received from the central site is then carried out at area code level database. The said algorithm achieves the faster location look-ups to make a call and increase the overall efficiency. Shivakumar and Widom (1995) and Shivakumar *et al*. (1997) take into consideration the capacity constraints in the databases and network and fixed calling and mobility patterns of users.

User profile based methods: Hara and Madaria (2004) have proposed a dynamic replication scheme based on a-periodic data updates. Their scheme involves user profile, read-write patterns in order to actively adjust the replicas to adjust to the changes in user behavior and network status. They have given an extended set of three replication schemes that actively provides replication services for mobile users. The random updates in each Mobile Host (MH) can be handled by employing the Read-Write Ratios (RWR). It improves access cost, improve response time and achieve high local availability. The implementation of their scheme though requires careful consideration of several issues including the maintenance of access histories and user profiles in addition to the selection of copy, number of replicas and clock synchronization for periodic time check events.

Wu and Chang (1999) have employed user profiles in order to record user read-write patterns to satisfy the individual user information requirement. They calculated the future access pattern by getting the daily schedule of the user mobility pattern and data requirement. The user movement almost always has some pattern and also the accessed data. If the user is not following its schedule (Wu and Chang, 1999) uses the past statistical data for making the replication decision. For the better estimation of the cost they have also proposed the concept of open objects in using read-write histories. The former technique uses the mobile ad-hoc network which is a multi-hop communication network.

User centered active method: A user-centered data replication scheme for managing the data dynamically

is proposed by Wu and Chang (2006). It defines the notion of activity-data dependency for inferring the information requirement of mobile users based on their scheduled activities. They maintain the detailed access statistics of each individual user in his profile. The diverse information maintained by the system include the user movement and daily schedule, user data access pattern based on read/write and future data predictions. They predict the user's current and future data requirements based on the concept of open objects. User profiles and access statistics are used to predict the future data requirements. This scheme maintain the access statistics of every individual user in his/her profile in order to make an appropriate replication decision to cater for the movement and data needs of every user. It also has a unique feature to provide for the daily schedule of the user. This schedule is used to determine the mobility pattern of the user and enable the system to predict replication allocation. Activity data dependency is used to represent the relationship between a type of activity and the corresponding data. The idea of open object is employed in order to represent a user's current and near future data requirement. Their model also allows the emergency events and objects declaration which has to be replicated unconditionally. The purpose of emergency events and objects is mainly for the time critical applications.

Active Data dependency is letting α be an activity and D be a set of data objects, we can say that the activity data dependency $\alpha \rightarrow D$ holds if and only if α uniquely determines the set of data objects D that are required for the successful completion of α (Wu and Chang, 2006).

If $\alpha \rightarrow D$, we say that α determines D or D is determined by α. For the set of activity-data dependencies A, if $(\alpha \rightarrow D) \in A$, we say that α is applicable in A.

It is a very complex and dynamic technique which is employed to predict the user next moves and data requirements which is not easy. According to the authors the use of this scheme can also increase the local availability of the data in addition to reducing the cost and improve upon the response time. Figure 3 (Wu and Chang, 2006) shows the results for access locality and Fig. 4 (Wu and Chang, 2006) shows the results for movement locality.

Wu and Fan (2010) introduced a more comprehensive set of user activity called complex activity to cater for the ever changing user behavior patterns. This scheme is also based on the proactive data management for predictive data allocation. A complex activity according to them is a sequence of location movement, service requests, the co-occurrence of location and service, or the interleaving of all. They have described the emergence of moving patterns of a user as a naturally modeled sequence of locations visited by the user. Mobile users also invoke services

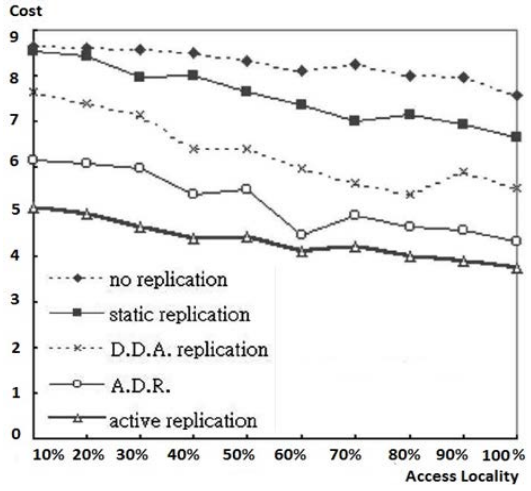

Fig. 3: Cost vs. access locality

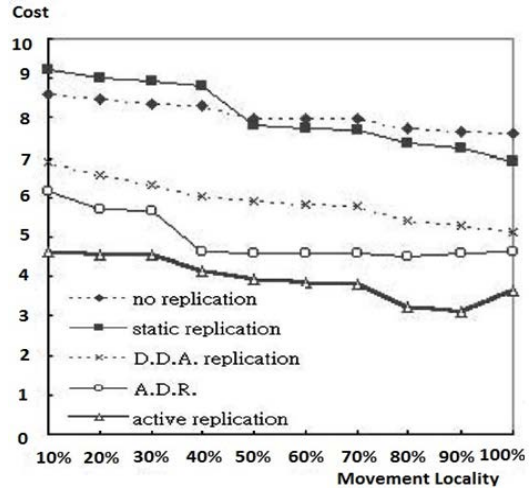

Fig. 4: Cost vs. movement locality

one after another when traveling. Service patterns emerge if a sequence of services is repeatedly invoked by the same or different user. They have characterized the user behavior patterns into three broad categories i.e. Location-only patterns (L-type) which is a sequence of locations that are repeatedly visited by mobile users, Service-only patterns (S-type) which is a sequence of services that are repeatedly invoked by mobile users and the Location-service patterns (LS-type) which is a sequence of location-service pairs that are repeatedly visited and invoked by mobile users. They have argued that the complex user activity can be used to model the user behaviors in order to determine his location and service requirements. They have proposed three data management schemes i.e., proactive pushing, Predictive handoff and precision pre-fetching.

Intra-group method: Zhang *et al.* (2011) have comprehensively studied the data replication problem in mobile tactical networks. They have proposed a new

intra-group data replication scheme and quantified the effects of mobility on different inter-group data replication schemes. They also established certain metrics, which include the average access delay and data availability and the temporal and spatial analysis of these values.

Grid based method: Nukarapu *et al*. (2011) have proposed a data grid model for the formulation of the data replication problem. According to authors data grid is an enabling technology and data replication in a data grid for scientific data intensive application is not only an NP-Hard problem but also cannot be approximated.

COMPARISON OF DIFFERENT ALGORITHMS

In this section, we summarize the algorithms studied and drawn a comparison based on the general performance of different replication algorithms. According to the study we made in the previous sections, from every replication scheme, we take the parameters such as access cost, data availability, traffic and accessibility. The availability and accessibility are interrelated as the increase in the data availability will automatically ease the accessibility. The Access frequency based methods SAF has the lowest traffic while DCG has high accessibility however in a real world scenario an appropriate methods is to be selected based on the system characteristics. The E-SAF, E-DAFN and E-DCG are the special cases of extension and give the replica allocation in an environment where data is updated at a periodical interval. The E-DCG method gives the high accessibility while E-SAF has the lowest traffic. The periodic algorithms have certain drawback like if they use a long time period, data availability may decrease because network partitioning occurs and the replication at that time becomes ineffective. If the time period is short, the traffic may increase because the partition prediction algorithm is unnecessarily executed despite the topology remains unchanged.

The Pre-Write based method gives a disconnected operation and shifts the computing load from MH to MSS. It does give a good utilization of the resources. In One-Copy and Two-Copy Methods if the number of read at MH is higher than the write operations at the fixed, then use a two copy method else use one copy method. In general the static allocation methods are not competitive. As compared to the static schemes the average expected cost of the dynamic allocation methods is low. The Access cost based method is one classic method which decides dynamically for placement of replica. It selects what to replicate in addition to the location and make use of the resources efficiently. The user profile based methods are highly dynamic in nature as the data replicas are dynamically adjusted according to the user behavior and network status. This scheme is more suitable for mobile ad-hoc network i.e. multi-Hop communication system than to a single hop. This scheme however, reduces the access cost and improves the response time. We have discussed the user centered active method which manage the data dynamically. It increases the accessibility. It puts user behavioral data as the main deciding factor for decision making for replica placement. The algorithm is based on the future requirements of the user considering his history. It may become computationally intensive and may require more storage space if the number of users increase significantly. The intra group methods, grid based methods also address the replication problem for a certain scenario and application. We have observed that for a data access for N hopes the complexity is O(n) but if there involves multiple hops, the complexity increases due to the route discovery procedures. So the data placed around the user would decrease the accessibility of the data, increase availability and decrease unnecessary communication between the nodes.

DISCUSSION AND CONCLUSION

Mobile Computing differs from a typical distributed system due to the absence of a physical link and high mobility rate of nodes. In fact the mobility attribute of this type of system define or limit its capabilities by so many ways.

DATA replication is often employed to improve the availability and effectiveness of information services in distributed systems (Helal *et al*., 1996). A replication scheme determines the number (replication level) and location (replication placement) of replicas in a distributed system. Traditional replication schemes (Helal *et al*., 1996; Ciciani *et al*., 1990; Hwang *et al*., 1996) are static in the sense that the number and placement of replicas are predetermined and fixed.

Therefore, typical data replication schemes used for normal distributed system doesn't work for mobile computing systems. Every system has limited resources and these resources need effective utilization. Therefore, it is of utmost importance to manage the replicas of the data effectively and avoid unnecessary duplication of replicas otherwise it is wastage of disk space and creates unnecessary communication between the nodes. Unnecessary replicas could create a problem of data consistency among different replicas at various sites which is to be rectified. A multi-criteria algorithm similar to Sorkhabi *et al*. (2009) may have to be employed to address the problem of consistency in replicated data bases.

To manage the replicas efficiently and reduce the communication cost and overhead, various data replication techniques have been proposed. We have reviewed some of the schemes as it is not possible to cover all of them. These techniques range from static

replica placement to the highly dynamic user mobility based replica placement techniques. From the literature we have learnt that if the process of replica allocation is kept static i.e. the replica allocation is decided beforehand, it may create communication overhead and wastage of disk space by unavoidable and unnecessary duplications of the data across the network. However if we use the highly dynamic schemes that record the user data access pattern based on previous read/write history and employ the user mobility pattern for decision making of replica placement, than it may also create a burden for the system resources which of course are limited. It would create a large amount of user profile data in addition to the actual user data which itself require the huge of amount of storage as well as the computing power. Therefore, we argue that a tradeoff is required between static and dynamic methods. The replication scheme should be devised so as to minimize the load on system and enhance the system performance. We also argue that the replica placement algorithm may not be same everywhere, rather it may be designed according to the user mobility in some specific region for example in University areas, offices, Airports, railway stations etc. Also the user requirement may vary depending on his distance from native (home) cell and in different foreign cells.

The data replication has becomes a very important technique for increasing data availability, scalability and performance of the system hence effecting the Quality of Service (QoS). It is a common method these days used in order to minimize the bandwidth usage and access latency. Presently, the majority of replications methods deal with the read-only data, thus making the replica management mechanism relatively easy (Grenoble, 2004). The volatile nature of the wireless network and mobile computing structure, certain features are introduces into the mobile computing systems. Dynamic replication schemes are more effective as compared to the static as they provide an efficient mechanism for replica placement decision, time and location and number of replicas. These techniques ensure the availability of data as well integrity in a mobile computing environment and reduce the communication cost. However there is a need for a certain tradeoff for the user profile information as the more and more parameters are stored, the process of decision making become more computation intensive and may take the system resources. Furthermore, placing the replicas of critical data near the home location of a mobile device ensure the easy recovery and efficient utilization of system resources.

ACKNOWLEDGMENT

This study was supported in part by the National Natural Science Foundation of China (Grant No. 61100004).

REFERENCES

Acharya, S. and S.B. Zdonik, 1993. An efficient scheme for dynamic data replication. Technical Report, Brown University, USA.

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

Barbara, D. and H.G. Molina, 1986. The vulnerability of vote assignments. ACM T. Comput. Syst., 4(3): 187-213.

Biswas, S. and S. Neogy, 2010. A Mobility-based checkpointing protocol for mobile computing system. IJCSIT, 2(1).

Boukerche, A., 2006. Handbook of Algorithms for Wireless Networking and Mobile Computing. Taylor and Francis Group, LLC.

Ciciani, B., D.M. Dias and P.S. Yu, 1990. Analysis of replication in distributed database systems. IEEE T. Knowl. Data Eng., 2(2): 247-261.

Coan, B., B. Oki and E. Kolodner, 1986. Limitations on database availability when networks partition. Proceedings of the 5th ACM Symposium on Principle of Distributed Computing, pp: 187-194.

Coulouris, G., J. Dollimore and T. Kindberg, 2001. Distributed Systems: Concepts and Design. 3rd Edn., Addison-Wesley, New York.

Earl, O., 2007. A Survey of Mobile Database Caching Strategies. David R. Cheriton School of Computer Science, University of Waterloo, ACM Press.

Grenoble, U., 2004. Mobile Databases: A Selection of Open Issues and Research directions Action members. SIGMOD Record, 33(2).

Guerraoui, R. and A. Schiper, 1997. Software-based replication for fault tolerance. IEEE Comput., 30(4): 68-74.

Hara, T., 2001. A effective replica allocation in ad hoc networks for improving data accessibility. Proceeding of the IEEE Infocom 2001, 3: 1568-1576.

Hara, T., 2003. Replica allocation methods in ad hoc networks with data update. Mobile Netw. Appl., 8: 343-354.

Hara, T. and S.K. Madria, 2004. Dynamic Data Replication using Aperiodic Updates in Mobile Adhoc Networks. Database Systems for Advanced Applications. Springer Berlin, Heidelberg, 2973: P111-136.

Helal, A.A., A.A. Heddaya and B.B. Bhargava, 1996. Replication Techniques in Distributed Systems. Kluwer Academic, ISBN: 0-7923-9800-9.

Huang, Y., A.P. Sistla and O. Wolfson, 1994. Data replication for mobile computers. Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data. Minneapolis, Minnesota, pp: 13-24.

Hwang, S.Y., K.K.S. Lee and Y.H. Chin, 1996. Data replication in a distributed system: a performance study. Proceeding of the Conference Database and Expert Systems Applications, pp: 708-717.

Korupolu, M., G. Plaxton and R. Rajaraman, 1999. Placement algorithms for hierarchical cooperative caching. Proceedings 10th Annual Symposium on Discrete Algorithms.

Li, B., M. Golin, G. Italiano and A.K. Sohraby, 1999. On the optimal placement of web proxies in the internet. Proceeding of the IEEE INFOCOM'99. New York, USA, 3: 1282-1290..

Madria, S.K. and B. Bhargava, 1997. Improving Data Availability In Mobile Computing Using Prewrite Operations. Computer Science Technical Reports. Paper 1369. Retrieved from: http:// docs. lib. purdue. edu/ cstech/ 1369/.

Madria, S.K., B. Bhargava, E. Pitoura and V. Kumar, 2000. Data organization issues for location-dependent queries in mobile computing. Proceedings of the East-European Conference on Advances in Databases and Information Systems (ADBIS-DASFAA '00).

Marton, T. and G. Attila, 2009. Keyspace: A Consistently Replicated, Highly-Available Key-Value Store. Retrieved from: arxiv.org/ pdf/ 1209.3913.

Nukarapu, D.T., B. Tang, L. Wang and S. Lu, 2011. Data replication in data intensive scientific applications with performance guarantee. IEEE T. Parall. Distrib. Syst., 22(8).

Pitoura, E. and P.K. Chrysanthis, 2007. Caching and Replication in Mobile Data Management. IEEE Data Eng. Bull., 30(3): 13-20.

Qiu, L., V. Padmanabahn and G. Voelker, 2001. On the placement of web server replicas. Proceeding of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), 2001: 1587-1596.

Ratner, D.H., 1998. Roam: A scalable replication system for mobile and distributed computing. Technical Report, Computer Science Department, UCLA, USA.

Schneider, F.B., 1990. Implementing fault-tolerant services using the state machine approach: A tutorial. ACM Comput. Surv., 22(4): 299-319.

Shivakumar, N. and J. Widom, 1995. User profile replication for faster location lookup in mobile environments. Proceeding of the ACM MOBICOM'95. Berkeley, CA USA, pp: 161-169.

Shivakumar, N., J. Jannink and J. Widom, 1997. Per-user profile replication in mobile environments: Algorithms, analysis and simulation results. Mobile Networks and Applications, Baltzer Science Publishers, 2: 129-140.

Sorkhabi, V.B. and H. Shahamfar, 2010. Detecting appropriate Replication approach considering Client and Server parameters in mobile environment. Proceeding of the 2nd International Conference on Education Technology and Computer (ICETC).

Sorkhabi, V.B., M.R.F. Derakhshi and H. Shahamfar, 2009. An Algorithm for detecting similar data in replicated databases using Multi Criteria decision making. Proceeding of the IEEE 2nd International Conference on Environmental and Computer Science.

Su, M., C.F. Wang and W.C. Houa, 2005. An approach of composing near optimal invalidation reports. Proceedings of the 6th International Conference on Mobile Data Management (MDM'05), ACM, New York, USA, pp: 116-124.

Sun, X., J. Zheng, Q. Liu and Y. Liu, 2009. Dynamic data replication based on access cost in distributed systems. Proceeding of the 4th IEEE International Conference on Computer Sciences and Convergence Information Technology.

Venkataramani, A., P. Weidmann and M. Dahlin, 2001. Bandwidth Constrained Placement in a WAN. Proceedings of the 20th annual ACM symposium on Principles of Distributed Computing. New York, USA, pp: 134-143.

Walke, B.H., 2002. Mobile Radio Networks: Networking, Protocols and Traffic Performance. John Wiley, West Sussex England.

Wiesmann, M., F. Pedone, A. Schiper, B. Kemme and G.Alonso, 2000. Understanding replication in databases and distributed systems. Proceedings of 20th IEEE ICDCS'2000, IEEE Computer Society, pp: 264-274.

Wolfson, O., S. Jajodia and Y. Huang, 1997. An adaptive data replication algorithm. ACM T. Database Syst., 22(4): 255-314.

Wu, S.Y. and Y.T. Chang, 1999. An active replication scheme for mobile data management. Proceedings of the IEEE 6th International Conference on Database Systems for Advanced Applications (DASFAA'99), pp: 143-150.

Wu, S.Y. and Y.T. Chang, 2006. A user-centered approach to active replica management in mobile environments. IEEE T. Mobile Comput., 5(11): 1606-1619.

Wu, S.Y. and H.H. Fan, 2010. Activity-based proactive data management in mobile environments. IEEE T. Mobile Comput., 9(3): 390-404.

Yu, H. and A. Vahdat, 2002. Minimal replication cost for availability. Proceeding of the 21st ACM Symposium on Principles of Distributed Computing (PODC) USA, pp: 98-107.

Zhang, Y., S. Ray, G. Cao, T.L. Porta and P. Basu, 2011. Data replication in mobile tactical networks. Proceeding of the Military Communication Conference-Track2-Network Protocols and Performance.