Mobile Database Cache Replacement Policies: LRU and PPRRP

Hariram Chavan[1] and Suneeta Sane[2]

[1] Information Technology, Terna Engineering College, Mumbai University, India
[2] Computer Technology, V. J. T. I., Mumbai, India
chavan.hari@gmail.com
sssane@vjti.org.in

Abstract. The continued escalation of manageable, wireless enabled devices with immense storage capacity and powerful CPUs are making the wide spread use of mobile databases a reality. Mobile devices are increasingly used for database driven applications such as product inventory tracking, customer relationship management (CRM), sales order entry etc. In some of these applications Location Dependences Data (LDD) are required. The applications which use LDD are called Location Dependent Information Services (LDIS). These applications have changed the way mobile applications access and manage data. Instead of storing data in a central database, data is being moved closer to applications to increase effectiveness and independence. This trend leads to many interesting problems in mobile database research and cache replacement policies. In mobile database system caching is an effective way to improve the performance, query delay and save bandwidth since new query can be partially executed locally. However, variable data sizes, data updates and frequent client disconnections make cache management a challenge. The conventional method of cache replacement is Least Recently Used (LRU) which takes into account only temporal characteristics of data items. This paper takes into account the parameters like minimum access probability, maximum distance and scope invalidation for cache replacement. These parameters are used and an enhanced version of Prioritized Predicted Region based Cache Replacement Policy (PPRRP) is simulated for the temporal and spatial characteristics. The implementation of both the policies has been done using java. The simulations of both policies reveal the outperformance of PPRRP over LRU.

Keywords : Mobile Database, Cache replacement policy, LRU, PPRRP, LDD, LDIS.

1 Introduction

The fast development of wireless communications systems and advancement in computer hardware technology has led to the seamlessly converged research area called mobile computing. The mobile computing research area includes the effect of mobility on system, users, data and computing. The seamless mobility has opened up new classes of applications and offers access to information anywhere and anytime.

N. Meghanathan et al. (Eds.): CCSIT 2011, Part I, CCIS 131, pp. 523–531, 2011.
© Springer-Verlag Berlin Heidelberg 2011

In client-server model in a mobile computing environment, the clients are mobile units (MUs) that have limited local resources such as notebooks, personal digital assistants (PDA) etc. This environment provided a number of new applications that have been inspiring researches on query processing in mobile databases systems (MDB). The MU communicates with data servers through a wireless link, accessing information at anytime and anywhere. The wireless network is susceptible to frequent disconnections, low-quality communication and limited bandwidth.

The class of data or the data whose value is functionally dependent on location (Location → Data value) is called Location Dependent Data (LDD) [14]. For example traffic reports or weather information of a city.

In cellular system a MU is free to move around within the entire area of the coverage. Its movement is random and therefore its geographical location is unpredictable. This situation makes it necessary to locate the mobile unit and record its location when client would like to access something form server. Location Dependent Information Services (LDIS) provide users with the ability to access information related to their current location. They include the location as a part of user's context information [15]. For example, a tourist on a trip to a new city can get help from his MU. He can query based on his personal interest the nearest restaurant, nearest railway station, nearest theater, etc, and can get the response on his MU. Using LDIS leads to many challenges inherent to mobile environments. These challenges include limited bandwidth, limited client power and intermittent connectivity.

The most promising solution to deal with these problems is the data caching technique that will store data copies of frequently used data items in the clients MU. Caching is an effective technique to improve data accessibility and to reduce access cost by caching of frequently accessed data item on client side. On the other hand, it is impossible to hold all accessed data items in the cache due to limitations of cache size on MU. Therefore, it is very important to have efficient cache replacement algorithm to locate a proper subset of data items for eviction from cache.

The cache management policy considers two dimensions: consistency and replacement policies. Cache invalidation aims to keep data consistency between the client's cache and the server databse. The cache replacement policy determines which data should be removed from the cache when there is no ample space to hold a new data item.

For the traditional cache replacement policy, the access probability is the most important factor normally considered. Most Recently Used - MRU, Least Recent Used - LRU and Least Recent Used at K Time - LRU-K [7] are examples of the replacement policy. For a mobile computing environment these solutions cannot be convenient since they have only temporal characteristics, generate high network traffic and too much power consumption.

Querying the database directly is a common, robust method for accessing databases on the Internet. As high-speed wireless networks continue to proliferate, the limitations of bandwidth asymmetry and disconnection have a diminishing effect on mobile database design. Thus, the bulk of mobile database research continues to consider limited resources, limited energy, frequent disconnections and bandwidth asymmetry as device characteristics. The common assumption is that accessing central databases for every operation can be prohibitively expensive and can significantly impact that response time and throughput of a mobile application. Caching data and then performing local queries on the cache is used in many cases. Caching increases the performance of local queries, but comes at a cost of complicated cache maintenance strategies.

This paper deals with describing and implementing PPRRP and compares its result with the standard method. The structure of the paper is as follows: section 2 describes Related Work, section 3 describes Mobile Database System framework. Section 4 taxonomy for cache replacement, section 5 gives performance evaluation and section 6 concludes the paper.

2 Related Work

LRU-K [7] method is an approach to database disk buffering. The basic idea of LRU-K is to keep track of the times of the last K references to popular database pages. This historical information is used to statistically estimate and to discriminate pages that should be kept in buffer. The LRU-K takes into account the temporal characteristics of data access.

The Furthest Away Replacement (FAR) [5] depends on the current location and the direction of the client movement. FAR organizes data items in two sets : In-direction and out-direction. The selection of victim for replacement is based on the current location of user. The assumption for replacement is the locations which are not in the moving direction and furthest from the user will not be visited in near future. The limitation of FAR is it won't consider the access patterns of client and is not very useful for random movement of client.

Probability Area Inverse Distance (PAID) [4] considers both spatial and temporal properties of data for replacement. The cost function for replacement considers the area of valid scope, data distance from current client location to the valid scope and the access probability of client data. However, PAID relies only on the current client movement direction and does not take into account updates to data items.

Prioritized Predicted Region based Cache Replacement Policy (PPRRP) [2] predict valid scope area for client's current position, assigns priority to the cached data items, calculates the cost on the basis of access probability, data size in cache and data distance for predicated region.

Data management in mobile computing environments is especially challenging for the need to process information on the move, to cope with resource limitation and to deal with heterogeneity. Among the applications of mobile data management, LDIS have been identified as one of the most promising area in research and development [12]. The problem has also been studied under various terms such as location-aware, context-aware, or adaptive information systems. Many of the previous work on LDIS treated location as an additional attribute of the data tables. LDIS queries can be processed like ordinary queries except with additional constraints on the location attribute. Caching techniques specially tailored for LDIS or mobile computing environments in general have also been a major research area. Semantic caching techniques employed semantic descriptions of cached items to facilitate better cache admission and replacement decisions that are responsive to the user movement.

3 Mobile Database System Model

We assume a cellular mobile network similar to that Mobile Network Architecture Fig. 1. It consists of mobile client (MC) containing data centric applications roam between wireless cells and accesses a centralized database (fixed host). Some of the fixed hosts, called mobile support stations (MSSs), are augmented with wireless interfaces. The wireless channel is separated into two sub channels: an uplink channel and a downlink channel. The uplink channel is used by MCs to submit queries, while the downlink channel is used by MSSs to answers from the server to target mobile client.

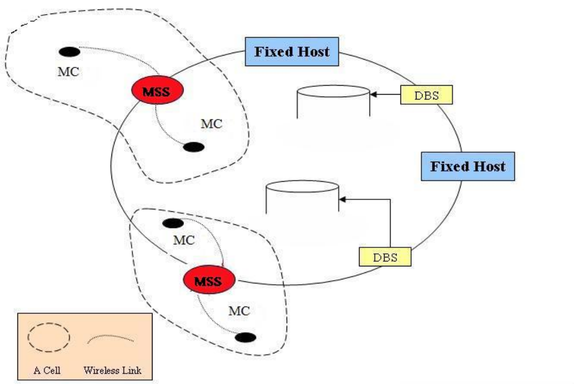

Fig. 1. Mobile Network Architecture

4 Taxonomy for Cache Replacement

In mobile database the taxonomy [13] for cache replacement policy is as shown in Fig. 2.

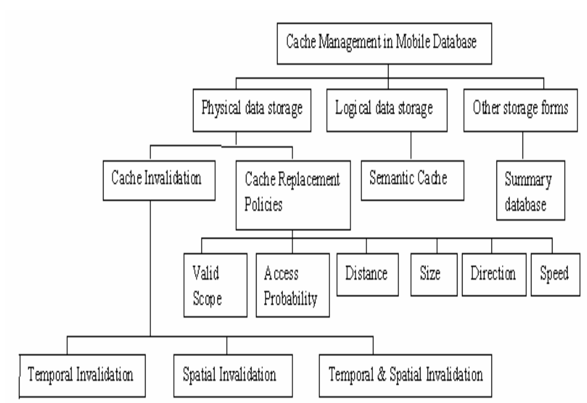

Fig. 2. Taxonomy for cache replacement policies in mobile database

We initiate our classification considering the client cache contents. This will lead to three main categories : physical, logical and other storage forms.

In the physical data storage model the MU cache contents are copies (records or page) of data items from the server. The attempt in this paper is to stick to the physical data storage rather than logical data storage.

The cache invalidation strategies are: Temporal, spatial and temporal & spatial. The Temporal invalidation considers many parameters such as architecture, server type, invalidation method, handoff etc. For example in case of handoff the cache can be kept or completely invalidated.

Spatial invalidation crops up when data values stored in cache become invalid because of the client moves to a new location area (region). The maintenance of a valid cache when the clients move is called location-dependent cache invalidation and a data item can have different value depending on its location [4].

The main feature considered in the spatial invalidation is the valid scope (or valid area). The valid scope of an item value is defined as the set of cells in which the item value is valid. There are different forms to relate the data value with its valid area.

The system set up for this paper deals with following assumptions: Let the space under consideration be divided in physical subspaces as

$$S = \{L_1, L_2, \ldots, L_N\}$$

There are data items with

$$D = \{D_1, D_2, \ldots ,D_M\}$$

Such that each data item is associated with valid scope which is either a set of one or more subspaces from S. Formally shown as

$$D_k = \{L_{k1}, L_{k2}\} \text{ where } L_{k1}, L_{k2} \text{ belong to } S$$

Fig. 3 show the representation of valid scope and Table 1 gives data instance for the same.

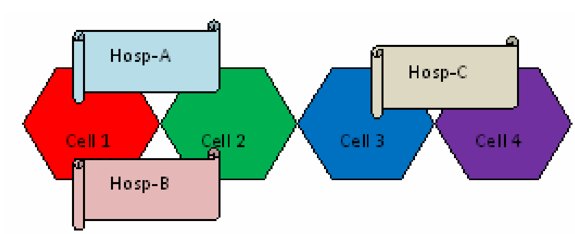

Fig. 3. Valid scope representation for Hospital

Table 1. Valid scope for Hospital data instance

Data instance	Valid Scope
{ nearby-Hosp , {A,B} }	{ 1 , 2 }
{ nearby-Hosp , {C} }	{ 3 , 4 }

The advantage is the complete knowledge of the valid scopes.

The MDS is developed with the following data attributes - valid scope, access probability and distance related to respective location. This centralized database is percolated to MSS in order to make sure that the data relevant in the valid scope is available in the nearest MSS thus making sure that the concept of Prioritized Region is taken into account while considering the cache replacement policy. We have valid scope as an attribute for the location to reduce computation and save the bandwidth. The search for the actual data is done in order as – first select the data as per region (location), get its valid scope and then data with higher access probability will be selected for access so that most probable data item is fetched to the cache. The access probability is updated in database with each access to the cached data item.

When cache does not enough space to store queried data item then, space is created by replacing existing data items from cache based on minimum access probability, maximum distance and scope invalidation. If data items have same valid scope then replacement decision is based on minimum access probability. If valid scope and access probability is same then replacement decision is based on maximum data distance. In some cases, data size plays an important role in replacement. If fetched data item size is large enough and requires replacing more than two data items from cache then replacement is based on maximum equivalent size with minimum access probability, maximum distance and scope invalidation.

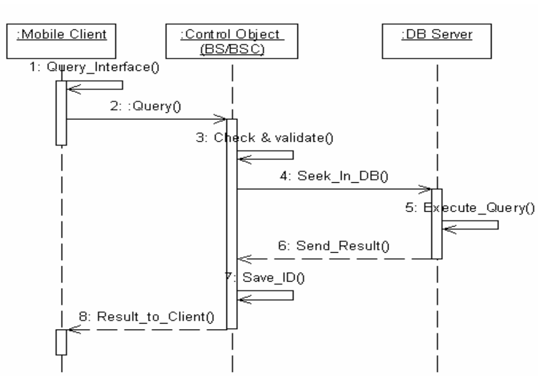

Fig. 4. Sequence diagram for Information Request

The selection of appropriate data item and fetching is done at server side to reduce the computation and save power of client. Fig. 4. shows a scenario of the request for information such as food, ATM, movies, petrol, and hospital etc. The client sends request for information. The server validates the client. For authenticated clients Database Server (DBS) responds with requested information.

5 Performance Evaluation

For implementation, database is created with different regions with locations, location specific resources such as Hospital, Restaurant, ATM, Movies, Blood Bank, Police, Fire Station, Medical 24x7 and resources with different specialty such as Child and Maternity for hospital and as applicable for each resource.

For our evaluation, the results are obtained when the system has reached the stable state, i.e., the client has issued at least 1000 queries, so that the warm-up effect of the client cache is eliminated. We have conducted experiments by varying the query interval and cache size. Query interval is the time interval between two consecutive client queries. In this set of experiments, we vary the query interval from 10 seconds to 200 seconds.

Fig. 5. show the performance of enhanced LRU and PPRRP. To enhance the performance of LRU we have valid scope as added criteria. To enhance the performance of PPRRP we have critical stage-size as additional criteria. The performance is evaluated with respect to change in mean Query Interval. We observe that the performance of PPRRP is far better than LRU. As the query interval increases, cache hit ratio decreases, because the client would make more movements between two successive queries, thus has low probability to remain in the same valid scope queried previously when a new query is issued.

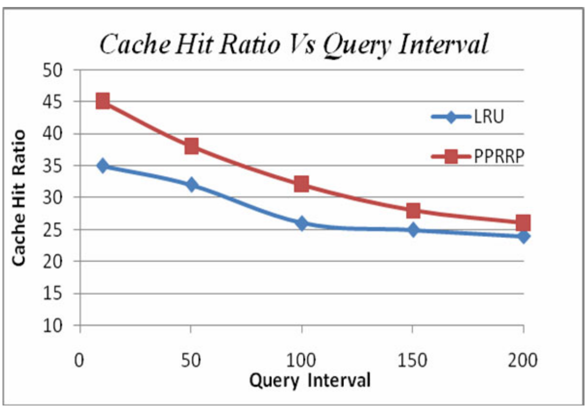

Fig. 5. Cache hit Ratio vs Query Interval

Fig. 6. depicts the effect of cache size on performance of LRU and PPRRP replacement policies. The performance of PPRRP will be increased substantially with increase in cache size so this result could be used to decide the optimal cache size in the mobile client.

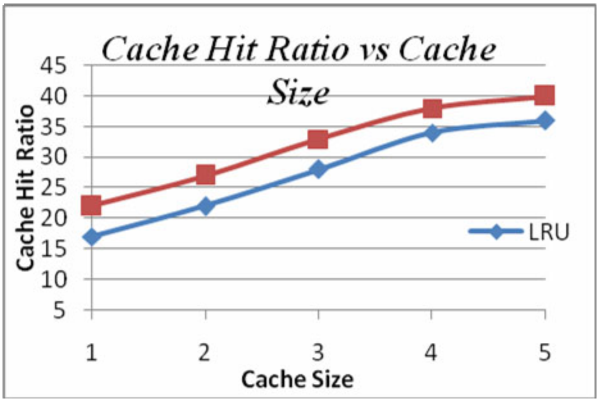

Fig. 6. Cache hit Ratio vs Cache Size

6 Conclusion

In this paper, we have presented a performance evaluation of LRU and PPRRP cache replacement policies. The LRU is temporal while PPRRP takes into account both the spatial and temporal properties of client movement and access patterns to improve caching performance. Simulation results for query interval and cache size show that the PPRRP has significant improvement in the performance than the LRU. In our future work we would like to incorporate Furthest Away Replacement (FAR) as part of comparison with PPRRP and our own Markov Model based cache replacement policy for prediction and replacement.

References

[1] Barbara, D.: Mobile Computing and Databases: A Survey. Proc. of IEEE Trans. on Knowledge and Data Engg. 11(1) (1999)

[2] Kumar, A., Misra, M., Sarje, A.K.: A Predicated Region based Cache Replacement Policy for Location Dependent Data In Mobile Environment. IEEE, Los Alamitos (2006)

[3] Oliver, E.: A Survey of Mobile Database Caching Strategies. David R. Cheriton School of Computer Science University of Waterloo

[4] Zheng, B., Xu, J., Lee, D.L.: Cache Invalidation and Replacement Strategies for Location-Dependent Data in Mobile Environments. Proc. Of IEEE Trans. on Comp. 51(10) (2002)

[5] Ren, Q., Dhunham, M.H.: Using Semantic Caching to Manage Location Dependent Data in Mobile Computing. In: Proc. of ACMIIEEE MobiCom, pp. 210–221 (2000)

[6] Dar, S., Franklin, M.J., Jonsson, B., Srivastava, D., Tan, M.: Semantic Data Caching and Replacement. In: Proceedings of the 22nd VLDB Conference Mumbai, Bombay, India (1996)

[7] O'Neil, E., O'Neil, P.: The LRU-k page replacement algorithm for database disk buffering. In: Proc. of the ACM SIGMOD, pp. 296–306 (1993)

[8] Fife, L.D., Gruenwald, L.: Research Issues for Data Communication in Mobile Ad-Hoc Network Database Systems. Brigham Young University – Hawaii, Computer Science Department, Laie, HI 96762
[9] Xu, J., Hu, Q., Lee, W.-C., Lee, D.L.: Performance Evaluation of an Optimal Cache Replacement Policy for Wireless Data Dissemination. IEEE Transactions On Knowledge And Data Engineering 16(1) (January 2004)
[10] Zhang, J., Gruenwald, L.: Spatial and Temporal Aware, Trajectory Mobility Profile Based Location Management for Mobile Computing, [HREFI], `http://students.ou.edu/Z/Jianting.Zhang-1/Html/STAMobility.htm`
[11] Zebchuk, J., Makineni, S., Newell, D.: Re-Examining Cache Replacement Policies. IEEE, Los Alamitos (2008) 978-1-4244-2658-4/08/$25.00 ©2008
[12] Wu, S.-Y., Wu, K.-T.: Dynamic Data Management for Location Based Services in Mobile Environments. In: Proceedings of the Seventh International Database Engineering and Applications Symposium (IDEAS 2003), pp. 1098–8068. IEEE, Los Alamitos (2003) 1098-8068/03 $17.00 ©2003
[13] Mânica, H., de Camargo, M.S.: Alternatives for Ccache Management in Mobile Computing. In: IADIS International Conference Applied Computing (2004)
[14] Dunham, M.H., Kumar, V.: Dependent Data and its Management in Mobile Databases. In: 9th International Workshop on Database and Expert Systems Applications (DEXA 1998), Vienna, Austria, August 26-28, p. 414 (1998)
[15] Hiary, H., Mishael, Q., Al-Sharaeh, S.: Investigating Cache Technique for Location of Dependent Information Services in Mobile Environments. European Journal of Scientific Research 38(2), 172–179 (2009)