

Weighted Angular Distance Based Cache Replacement Strategy for Location-Dependent Data in Wireless Environment

Parameswaran Raman, Raghavendra Prasad, Nadarajan R, Mary Magdalene Jane

ABSTRACT

Data caching at mobile clients is an efficient mechanism to enhance data accessibility and reduce access cost. In this paper, we study the issues of cache replacement for location-dependent data under a geometric location model. Considering factors like probability, geometric location, direction of client motion and data distance, we propose a cache replacement policy WIDAAP (Weighted Inverse Distance Angle and Area-wise Probability). We have conducted a comparative performance study with existing popular cache replacement policies, taking into account the randomness of motion and the data distance motivated by user preference. The experimental results show that the proposed replacement scheme is effective when the user preference is considered and the policy significantly outperforms the conventional replacement policies.

Key Words: Cache Replacement, Mobile Computing, Location-Dependent Information Services, Performance Evaluation.

1. Introduction

Seamless mobility and ability to gather information about the users' immediate surrounding, has contributed to the growing popularity of a new kind of information services, called Location-Dependent information services (LDISs) [5,6]. By including location as a part of user's context information, many value-added applications targeted specifically at mobile users especially in geographic, traffic, logistics, tourist, emergency and disaster management systems can be provided.

LDISs, being wireless in nature are plagued by mobility constraints like limited bandwidth, client power and intermittent connectivity [3, 4, 11]. Data caching at mobile clients is an effective antidote for the above cited limitations [1,4]. A data replication mechanism in which copies of data are brought to a mobile unit as a response to a query and retained at the cache for possible use by subsequent queries drastically improves data accessibility and minimizes the access costs.

Limited memory of mobile devices poses a restriction on the cache size, making it impossible to retain all the accessed data

items in the cache. A cache replacement policy thus becomes inevitable in order to utilize the available cache effectively.

In this paper, we propose a new cache replacement policy for Location-Dependent data under geometric location model called WIDAAP. The policy takes into consideration the valid scope area of a data value, direction of client's motion, and distance of data from the user location, user preference and current relevance of the data subject to update-cycles. To evaluate the performance of proposed Location-Dependent data caching strategy, a simulation is carefully designed and a series of simulation experiments are conducted. The experiment results show that the proposed strategy outperforms the conventional replacement policies significantly.

The paper is organized as follows: Section 2 explores terminologies and assumptions, Section 3 introduces the conventional and proposed cache replacement policies, Section 4 deals with the performance evaluation. Section 5 concludes the paper.

2. Assumptions and Important Terminologies

We assume a standard cellular mobile network consisting of mobile clients and fixed hosts, and the client can have seamless mobility across the network. We

term the geographical area covered by the mobile system as the *service area*. The *data values* for *data items* vary depending on the geographical region. For instance, “Nearest Hospital” is a data item and the data values for this item vary based on the location where the query is issued.

A two-dimensional geometric location model is assumed wherein the location of a mobile client can be determined using systems such as Global Positioning System (GPS) [7]. We have adopted the *valid scope* and *scope distribution* approach as in [12]. The *valid scope* of an item is defined as a region within which the item value is valid. The set of valid scopes for all of the item values of a data item is called the *scope distribution* of an item. In this paper the data items are assumed to be of fixed size and read only. However, the relevance of the data item may be determined or altered during update cycles. This *data relevance* has a significant impact on the usability of data. For example, result of the “nearest restaurant” query may become obsolete if the nearest restaurant happens to be closed or unavailable for service.

3. Location-Dependent Cache Replacement Policies

Conventional cache replacement policies, such as LRU and LFU, rely on the temporal locality of client’s access pattern. However, in mobile networks where clients utilize location dependent information services, clients access pattern do not only exhibit temporal locality, but also exhibit dependence on location of data, location of the client and direction of the client’s movement [8,12].

In addition, these traditional cache replacement policies evict only the most unlikely data item from the cache. In our policy, we try to bring in data items which are more likely to be used rather than merely evicting only the one which is least likely. Therefore, we evict not only a single least likely data item, rather remove a set of such data items and replace them

with a corresponding number of data items which are more likely to be used.

Relying solely on temporal locality when making cache replacement decisions will result in poor cache hit ratio in LDIS. Most of the existing cache replacement policies use cost functions to incorporate different factors including access frequency, update rate, size of objects, etc., however very few of these policies account for the location and movement of mobile clients. Cache replacement policies such as LRU, LFU and LRU-k [2,9], only take into account the temporal characteristics of data access, while policies such as FAR [10] only deal with the location dependent aspects of cache management but neglect the temporal properties. Only policy which does consider both spatial and temporal properties of data objects is the PAID policy [12]. Also, direction of the client’s motion plays a decisive role in determining the how likely a data item will be used in the near future. Our policy takes into consideration, all the above mentioned parameters and also another vital factor which is data relevance.

3.1 Proposed WIDAAP Policy

Early cache replacement policies like LRU, LFU, and LFU-K [2, 9] considered *access probability* to be the most important factor wherein the data item with the least access probability was considered the victim for eviction. Subsequent policies such as PAID [12] focused on two other factors namely *data distance* and *valid scope area*.

3.1.1 Access Probability

Access Probability is the probability of a particular data item being queried again within the same valid scope. Calculation of access probability employs the exponential aging method. Two parameters are maintained for each data item i , a running probability (P_i) and the time of the last access to the item ($t_{last, i}$). P_i is initialized to 0. When a new query is issued for data item i , P_i is updated using the following formula:

$$P_i = \frac{\alpha}{(t_{current} - t_{last,i})} + (1 - \alpha)P_i$$

Where $t_{current}$ is the current system time and α is a constant factor to weight the importance of the most recent access in the probability estimate.

3.1.2 Valid Scope Area

Valid scope area refers to the geometric area of the valid scope of a data value. For location-dependent data, valid scope areas are an indicator of the access probabilities for different data values. Larger the valid scope area of the data, higher the probability that the client requests this data as the client has a higher chance of being in large regions than small regions.

Our proposed policy introduces further important dimensions namely angle of bearing, beeline-bearing coefficients and supplement variable which take care of direction of client motion, data distance impact and data relevance respectively.

In the following paragraphs, we discuss the above factors and analyze their impact on cache performance.

3.1.3 Angle of Bearing

Angle of Bearing refers to the angular distance (in degrees), θ between the client's current direction and direction vector of a particular valid scope. In a location-dependent service, the server responds to a query with the suitable value of the data item according to the client's current movement. As such, when the valid scope is in the same direction as the client's current motion, the data has a higher chance of becoming usable again. Thus, the angle between the client's current direction of motion and the direction of the valid scope, θ gains importance in any location-dependent service. We term this angle as the angle of bearing that gives an estimate of how the valid scopes are directionally located relative to the client's current direction of

motion. A valid scope located in the same direction as the client's motion is assigned $\theta = 0$, and henceforth θ increases for any deviation from the client's direction of motion. For example, for a valid scope lying in a direction directly opposite to the client's direction of motion, $\theta = 180$.

3.1.4 Beeline Distance

Beeline distance D is the linear distance of a valid scope from the client's current location. In a location-dependent data service, when the valid scope of a data value is far away from the client's current location, the data is less likely to be used again. Hence, the cache replacement should take into account all necessary beeline distances before victimizing any data in the cache for replacement.

3.1.5 Beeline and Bearing coefficients

In location-dependent data services, the beeline distance may gain precedence over the angle of bearing or vice-versa depending on the need and type of service. Beeline and Bearing coefficients, α_d and α_θ are thus added as weights to D and θ , to indicate this behavior. For example, in an emergency situation, a service available in a direction directly opposite to the client's current direction of motion may still be preferred to the same service available at a longer distance but in the same direction of client's motion.

3.1.6 Supplement Variable

In location-dependent services, the relevance of data also plays a very important role in the cache replacement strategy. Taking into account location-dependent data whose relevance changes dynamically creates a possibility of this data becoming obsolete due to unavailability. Therefore, a dynamic supplement variable v is introduced, which acts as an indicator of how prudent the data is. The supplement variable v is a measure of data relevance. This supplement variable is supplied by the service provider

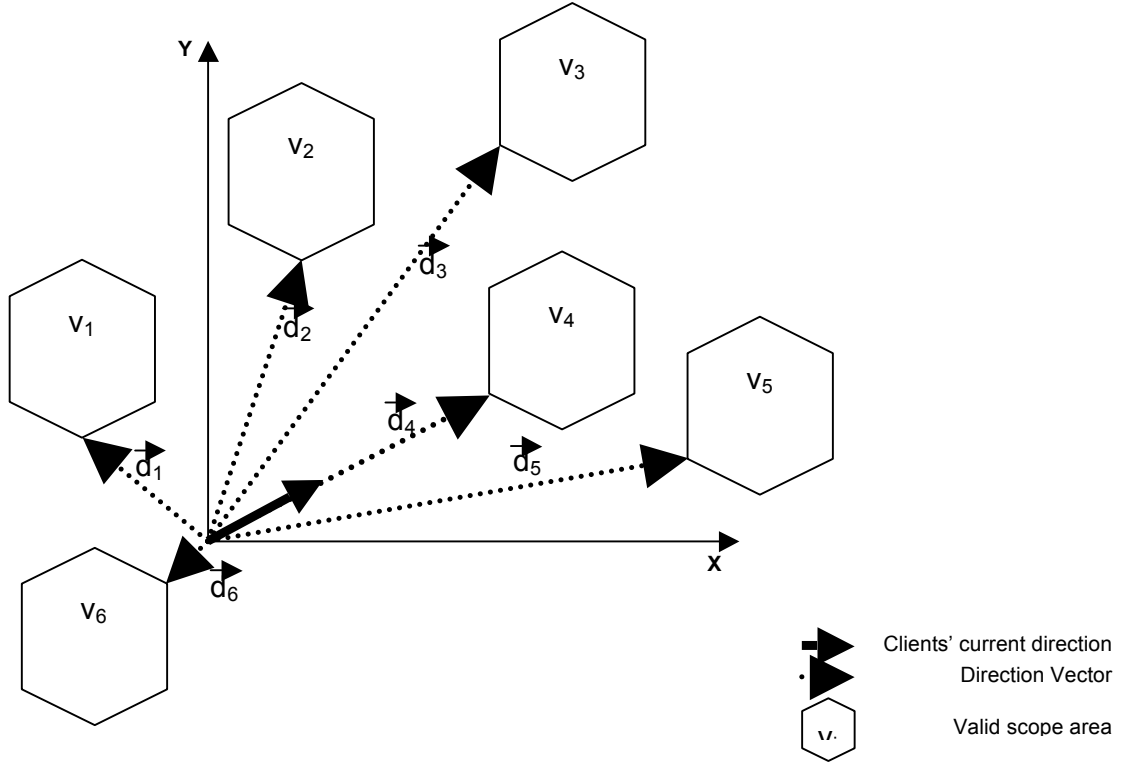


Figure 1: Impact of user preference on cache replacement

and can signify any scenario being just a numerical value. Supplement-variable-management by itself becomes a wide area of research. We plan to devise efficient and optimal supplement-variable-management strategies in near future.

We thus formulate a cost-function which ranks the valid scopes based on the above analyzed factors. The cost-function for data value j of item i is defined as follows:

$$c_{i,j} = \frac{P_i \cdot A_{i,j}}{\alpha_d \cdot D(V_{i,j}) + \alpha_\theta \cdot \theta(V_{i,j})} \times \frac{1}{v}$$

where P is the access probability, A is the valid scope area, D is the beeline distance θ is the angle of bearing, α_d and α_θ are the beeline and bearing coefficients respectively and v denotes the supplement variable.

Figure 1 shows a sample scope distribution consisting of six valid scopes V_1, V_2, V_3, V_4, V_5 and V_6 . The mobile client is initially located at the origin O and his direction of motion is assumed to be towards the valid scope V_4 . We assign distance vectors d_i to

each of the valid scopes V_i which are used to compute the beeline distances and angle of bearing. The access probabilities and area of the valid scopes are already available in the service provider's database. The cost function is now applied to each of these and a rank list is generated.

We now illustrate the impact of user preference on the ranking. Consider the valid scopes V_6 and V_4 . V_6 is closest to the client's current location but is directionally opposite to the client's motion. V_4 is farther from the client when compared to V_6 but is in the same direction as the client's motion. We also assume that the valid scope areas and access probabilities are approximately the same for both.

When the user prefers distance, i.e a situation of emergency where he would prefer to access the data item closest to him irrespective of the direction, the beeline coefficient α_d is given a higher value thus making data item V_6 rank higher over V_4 .

3.2 Process Flow

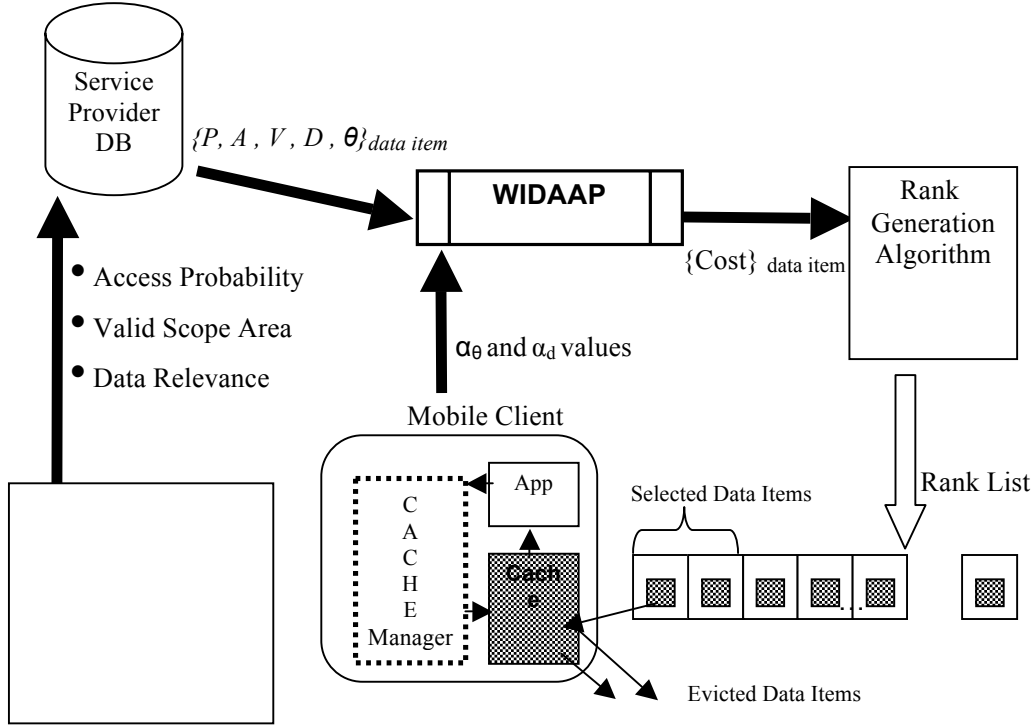


Figure 2: Process Flow of the proposed cache replacement policy

We represent the flow of the WIDAAP process in a block diagram as shown in Figure 2.

The access probabilities, valid scope areas and data relevance are identified from the scope distribution beforehand and maintained in the service provider's database. User preference is provided by the mobile client by adjusting the α_θ and α_d values accordingly. The cost function is thus calculated for all the required data item values and passed onto a Rank Generation Algorithm (RGA) which is a standard sorting algorithm. The cache eviction and replacement is then done using the algorithm summarized below.

The algorithm takes as input the rank list generated by the RGA and performs a step-by-step eviction and replacement.

The aim of this process is to make sure that when one cache miss occurs, the cache is made less vulnerable to another miss by populating it with the most likely data item values. The overheads of heavy cache manipulation are negligible when

compared to the benefits gained by improving the cache hit ratio.

Algorithm

Input : Rank list, list of current items in the cache

Output : List of items in the cache

Procedure :

1: loop $i = 0$ to all items in the cache

1.1 Determine lowest ranked item in the cache, C_{lowest}

1.2 Compare C_{lowest} with the highest ranked item in the rank list, $R_{highest}$.

1.3 If $Rank(C_{lowest}) < Rank(R_{highest})$ then

1.3.1 Replace C_{lowest} with $R_{highest}$

1.3.2 Remove $R_{highest}$

1.4 If $Rank(C_{lowest}) \geq Rank(R_{highest})$ then

1.4.1 Exit

End loop

4. Performance Evaluation

The proposed WIDAAP cache replacement strategy is evaluated based on its performance as compared with the existing policies such as LRU, FAR and PAID.

We take into account the following prominent factors while configuring the simulation model and make the initialization.

PARAMETER	SETTINGS
SIZE OF RECTANGULAR SERVICE AREA	200*200, 400*400
NUMBER OF ITEMS AVAILABLE IN THE DATABASE	100
NUMBER OF VALUES PRESENT FOR EACH ITEM AT DIFFERENT LOCATIONS	50, 95
MINIMUM SPEED OF CLIENT	1, 3 UNITS OF DISTANCE PER SEC
MAXIMUM SPEED OF CLIENT	2, 6 UNITS OF DISTANCE PER SEC
FRACTION OF CACHE SIZE TO THE ACTUAL DATABASE SIZE	10 %
AVERAGE TIME INTERVAL BETWEEN TWO CONSECUTIVE QUERIES	50 SEC
TIME DURATION FOR WHICH CLIENT MOVES AT CONSTANT VELOCITY	100 SEC
UPLINK BANDWIDTH	15 KBPS
DOWNLINK BANDWIDTH	120 KBPS

Table 1: Configuration Parameters

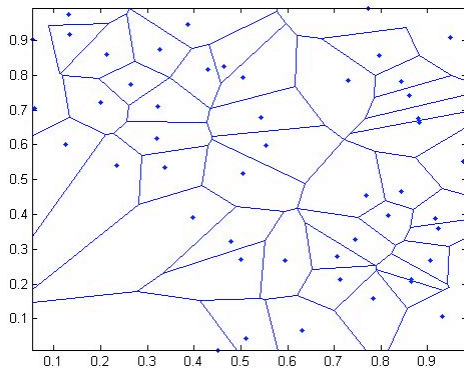


Figure 3a: Scope distribution S_1

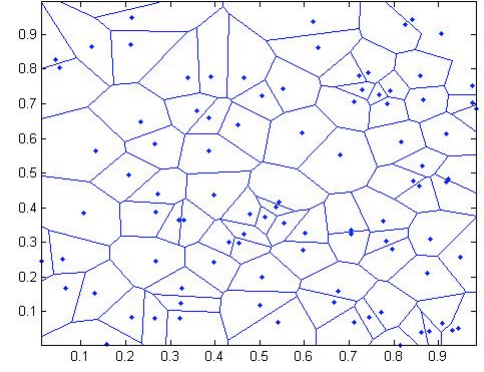


Figure 3b: Scope distribution S_2

Scope distributions S_1 and S_2 with 50 and 95 valid scopes are considered (see Fig 3a, 3b). S_1 depicts a situation where the valid scopes are larger in size while S_2 contains smaller valid scopes. We study the performance of cache replacement policies under the above varied scenarios. Throughout the performance evaluation, *cache hit ratio* has been considered as the evaluation metric for cache replacement as it gives a clear idea of the extent to which cache hits have been managed. Higher this *cache hit ratio*, higher is the local data availability, less is the corresponding uplink-downlink resource costs and better the service. Here, we compare the proposed WIDAAP policy to existing policies LRU, FAR and PAID.

4.1 Effect of changing Query Interval

Query interval is the time interval between two consecutive client queries. We vary the query interval from 20 to 220 seconds. Figures 4a, 4b show the cache performance for Scope Distributions S_1 and S_2 respectively.

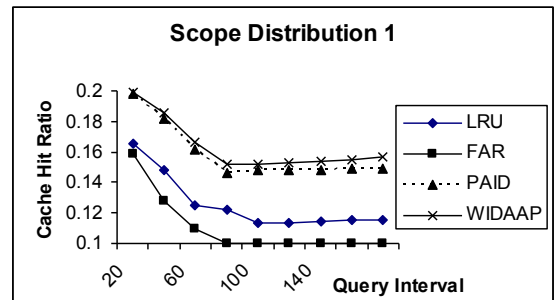


Figure 4a: Cache hit ratio Vs Query Interval, S_1

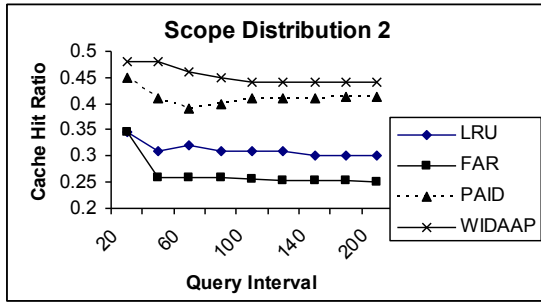


Figure 4b: Cache hit ratio Vs Query Interval, S₂

The graphs clearly show that when the query interval is increased most of the policies perform poorly as the client tends to move more between successive query intervals. PAID performs better than the conventional policies as it takes the distance into consideration, but it still fails when the client tends to make more random and direction oriented movements. WIDAAP takes into consideration both the direction and distance and hence performs better especially in S₂ wherein the valid scopes are greater in number and smaller in size.

4.2 Effect of changing Moving Interval

Moving interval is the time span for which the client maintains constant velocity. We vary the moving interval from 50 to 550 seconds. Figures 5a, 5b show the cache performance for Scope Distributions S₁ and S₂ respectively

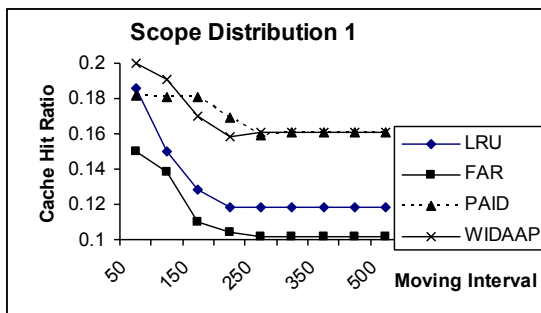


Figure 5a: Cache hit ratio Vs Moving Interval, S₁

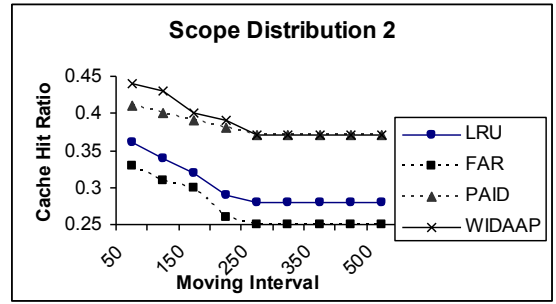


Figure 5b: Cache hit ratio Vs Moving Interval, S₂

The increase in moving interval implies that the client tends to move more distance with constant velocity and hence with less randomness. The graphs clearly show a drastic improvement made by PAID and WIDAAP when compared to conventional policies as both of them take into consideration the distance. When the Scope distribution is like S₂ WIDAAP performs better than all the other policies because we predict the valid scopes that might be used based on direction also which does not change too much because of large moving intervals.

4.3 Effect of Cache Size

In these set of experiments we check the performance of the policies under different cache sizes and analyze the results. Figures 6a, 6b show the cache performance for Scope Distributions S₁ and S₂ respectively

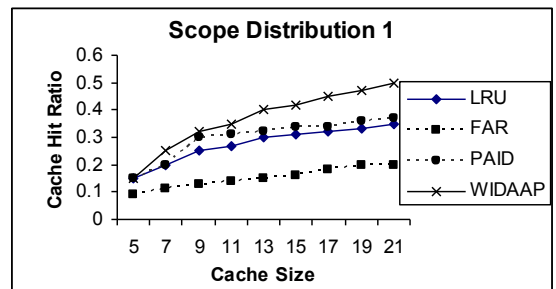


Figure 6a: Cache hit ratio Vs Cache Size, S₁

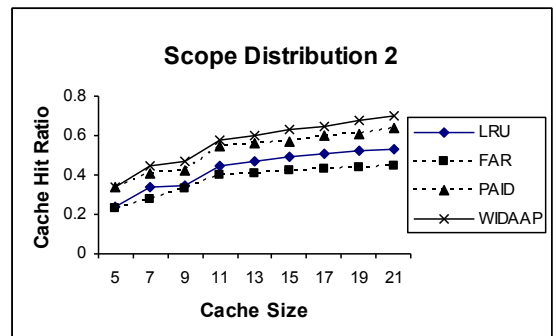


Figure 6b: Cache hit ratio Vs Cache Size , S_2

With increase in cache size, more number of probable data items can be accommodated in the cache, thus improving the cache performance. Since WIDAAP focuses on bringing more likely data into the cache, it makes the best use of the increase in cache size, even when the motion is random, the query intervals large and the valid scopes small in size.

5. Conclusion

In this paper, we have proposed a new cache replacement policy that considers factors like access probability, valid scope area, data distance and direction of client motion. Also a new factor, data relevance was introduced and its impact on the cache performance was analyzed.

A series of simulation experiments have been conducted to evaluate the proposed cache replacement strategy. The results indicate that the proposed policy shows a substantial performance improvement over traditional cache replacement policies like LRU, FAR and PAID.

By changing focus from evicting only the worst data item element from the cache to bringing in the more likely data items into the cache has resulted in overall improvement of caching efficiency. The policy also provides the user an option to select the service based on his need, urgency or convenience.

In this paper we have assumed fixed size and read-only data items, the possibilities of extending the policy to accommodate data of varying size and subject to regular updates are fertile areas for further research. Incorporating a pre-fetch or hoarding mechanism into the proposed policy are also areas which require further study.

References:

- [1] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik, "Broadcast Disks: Data Management for Asymmetric Communications Environments," Proc. ACM SIGMOD Conf. Management of Data, pp. 199-210, 1995.
- [2] A. Balamash and M. Krunz, "An Overview of Web Caching Replacement Algorithms", IEEE Communications Surveys & Tutorials, 6(2), 2004.
- [3] D. Barbara, "Mobile Computing and Databases: A Survey", IEEE Trans. on Knowledge and Data Engg, 11(1), 1999.
- [4] D. Barbara and T. Imielinski, "Sleepers and Workaholics: Caching Strategies in Mobile Environments", In Proc. of SIGMOD, pp. 1-12, 1994.
- [5] K. Cheverst, N. Davies, K. Mitchell, and A. Friday, "Experiences of Developing and Deploying a Context-Aware Tourist Guide: The GUIDE Project," Proc. Sixth Ann. ACM/IEEE Int'l Conf. Mobile Computing and Networking (MobiCom 2000), pp. 20-31, 2000.
- [6] M.H. Dunham and V. Kumar, "Location Dependent Data and Its Management in Mobile Databases," Proc. Ninth Int'l Workshop Database and Expert Systems Applications, pp. 414-419, 1998.
- [7] I. A. Getting, "The Global Positioning System", IEEE Spectrum, 12(30), pp. 36-38, 43-47, 1993.
- [8] D.L. Lee, W.-C. Lee, J. Xu, and B. Zheng, "Data Management in Location-Dependent Information Services", IEEE Pervasive Computing, 1(3), 2002.
- [9] E. O'Neil and P. O'Neil, "The LRU-k page replacement algorithm for database disk buffering", Proceedings of the ACM SIGMOD,

pp.296-306,1993

[10] Q. Ren and M.H. Dhunham,
“Using Semantic Caching to Manage
Location Dependent Data in Mobile
Computing”, ACM/IEEE
MobiCom, 210-221, 2000.

[11] L. Yin, G. Cao and Y. Cai, “A
generalized Target-Driven Cache
Replacement Policy for Mobile
Environments”, In Proceedings of
the IEEE Symposium on Applications
and the Internet, ,pp. 14-21, 2003

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