PINE-guided Cache Replacement Policy for Location-Dependent Data in Mobile Environment

Mary Magdalene Jane.F
Research Scholar
PSG College of Technology
India
janejayakumar@gmail.com

R Parameswaran
Dept of Comp.Applns
PSG College of Technology
India
parameshr@gmail.com

R Nadarajan
Professor
PSG College of Technology
India
nadarajan_psg@yahoo.co.in

Maytham Safar
Asst. Professor
Kuwait University
Kuwait
maitham@mac.com

Caching frequently accessed items on the mobile client is an important technique to enhance data availability and to improve data access time. Cache replacement policies are used to find a suitable subset of items for eviction from the cache due to limited cache size. The existing policies rely on Euclidean space and consider Euclidean distance as an important parameter for eviction. However, in practice the position and movement of objects are constrained to spatial networks where the important distance measure is the network distance. In this paper we propose a cache replacement policy, which considers the network density, network distance and probability of access as important factors for eviction. We make use of an already proven technique called Progressive incremental network expansion to compute the network distance more efficiently. A series of simulation experiments have been conducted to evaluate the performance of the policy. Results indicate that the proposed cache replacement scheme performs significantly better than the existing policies FAR and PAID and WPRRP.

Categories and Subject Descriptors

H.3 [**Information Storage And Retrieval**]: Information Search and Retrieval - *Selection process*

General Terms

Management, Performance, Experimentation.

Keywords

Cache replacement, location dependent information services, mobile computing.

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 based services (LBS) [5, 6, 10]. 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.

PETRA'08, July 15-19, 2008, Athens, Greece.

finding the nearest hospitals or clinics based on their location and health needs, and even provide them with driving directions and real-time traffic information.

LBS users face challenges that are inherent to mobile environments like limited bandwidth, client power and intermittent connectivity [3, 4, 15]. Data caching at mobile clients is an effective antidote for the above cited limitations [1, 4]. It is 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 thereby improving data accessibility and minimizing 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.

The existing Cache replacement policies such as LRU, LFU and LRU-k [2, 12] take into account the temporal characteristics of data access. FAR [13] takes into account the spatial properties and removes the data item that is farthest from the client's location. PAID [16] (Probability Area Inverse Distance) considers both spatial and temporal properties of data objects. The cost function for replacement considers the area of valid scope, data distance from current client location to the valid scope and the probability of data access. The Weighted Predicted Region based Cache Replacement Policy (WPRRP) considers data size and the weighted data distance of an item along with the probability of data access.

These policies rely on Euclidean spaces, where the distance between two objects is determined by their relative position in space. However, in practice, clients move on pre-defined paths (e.g., roads) that are specified by an underlying network. This means that the shortest network distance between a query object and a point of interest(POI) (e.g., a mobile client and a hospital) depends on the connectivity of the network rather than the objects' coordinates. [8] In such cases conventional policies which use Euclidean distance as a parameter for eviction could give only rough estimates. This paper proposes spatial network based cache replacement policy (PINE-CRP) which takes into consideration the density of the valid scope, the network distance between the client and a point of interest based on PINE and the access probability. Cache hit ratio, which is defined as the ratio of number of queries answered by the client cache to the total number of queries generated by the client is employed as the primary performance metric. A high cache hit ratio indicates high local data availability which reduces data access cost.

2. PROBLEM FORMULATION

Location dependent data refers to data whose value is dependent on some reference location, which in most cases is the location of the mobile user who generates the query. A data item, in the context of this study, refers to one type of location dependent data (e.g. hospitals) and usually has different instances. Each data instance is valid only in some specific region which is termed as valid scope in [16]. It represents a bounded area within which a data instance is valid.

The valid scopes are generated using the Voronoi diagram in the plane. The Voronoi diagram divides the plane into areas which are closest to each point of interest (hospital). In this problem, the distance is approximated by the direct distance in the Euclidean plane. However, the Voronoi diagram in the plane is inadequate for solving the location problem. Consider the problem to locate the nearest hospital in a city when an accident occurs. In this case the travel time from the current location to the nearest hospital is very crucial Here the nearest neighbours should be located based on the underlying spatial network. The distances should be measured on the road network in such sensitive problems. The road network matches actual situations more than the Euclidean plane.

3. ASSUMPTIONS AND TERMINOLOGIES

3.1 Valid Scope

Spatial networks (e.g., road networks) can be modeled as weighted graphs where the nodes of the graph represent road intersections or points of interest and the edges connecting the nodes represent the roads. The weights can be the distances between nodes or they can be the time it takes to travel between the nodes. For simplicity we assume the graph to be undirected. We define valid scope as the network region in which a data item value is valid. When spatial networks are modeled as graphs, the valid scope can be represented as a sub graph. The valid scopes are generated using network voronoi diagrams.
In general, a Voronoi diagram of a set of points is a collection of regions that divide up the plane. Each region corresponds to one of the generator point (point of interest), and all the points in one region are closer to the corresponding generator point than to any other generator point [16].

"A Network Voronoi diagram is defined for graphs and is a specialization of Voronoi diagrams where the location of objects is restricted to the links that connect the nodes of the graph and the distance between objects is defined as their shortest path in the network rather than their Euclidean distance."[7] The Network Voronoi diagram considers distances only in networks, not in the plane. It divides the network, not the space, into Voronoi cells. A Voronoi cell in a network is the set of nodes and edges that are closer to one Voronoi generator (Point of Interest) than to any other [11].

3.2 Network Distance

Network distance from the current location of the mobile client to a point of interest refers to the least cost distance (shortest path) from the client to the point of interest. When the mobile client requests for a location dependent service, we assume that he would choose the shortest/least cost path from his current position to the point of interest (e.g. restaurant). The points of interests which are far away from the user either in his direction of motion or in inverse direction will have a lower chance of become usable in near future. This implies that points of interest with higher network distance would be more probable for eviction. This is highly suitable for health care services because the user would prefer to get to the closest hospital irrespective his current direction of motion.

3.3 Network Density

In road networks, the client has a higher chance of remaining in a region with densely connected roads and larger area for a longer time than a region with smaller area and sparsely connected roads. We argue that the density of a valid scope has an impact on the access probabilities for different data values and it should take into consideration the distances between roads and also their degree of connectivity.

The weighted network density (W_d) is computed for every valid scope as

$$W_d = \frac{2.\sum w_i}{n(n-1)}$$

Where w_i represents the weight of each edge in the graph and n is the number of vertices in the graph.

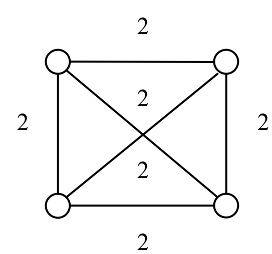

Figure 1. Valid Scope 1.

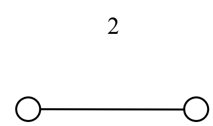

Figure 2. Valid Scope 2.

Figures 1 represent s a street network where the weighted density is 2.

A typical highway is represented with two nodes and a single edge where n (n-1)/2 takes the value 1. So WD becomes equal to the distance between two nodes. WD is 2 for Fig 2. This implies that the user would stay in a region with high WD for a longer time and therefore the higher the probability that the user requests this data.

3.4 Progressive Incremental Network Expansion (PINE)

PINE [14] is a technique based on Dijkstra's algorithm used to locate k-nearest neighbors of a query point using network Voronoi diagram and network expansion algorithm. It performs network expansion starting from the query point q and examines the interest points in the order they are encountered. This approach is also based on the properties of the network Voronoi diagrams and precomputation of the network distances for a very small percentage of the nodes in the network. PINE reduces the problem of distance computation in a very large network, into the problem of distance computation in a number of much smaller networks plus some online "local" network expansion. The main idea behind this approach is to first partition a large network in to smaller/more manageable regions (using network Voronoi diagram). Next, the inter distances for each cell are pre-computed. This will reduce the pre-computation time and space by localizing the computation to cells and handful of neighbor-cell node-pairs. This expansion method utilizes the inter-cell pre-computed distances to find the actual network distance from the query point to the objects in the surrounding area, hence saves on computation time. PINE has to thus access only the border-to-border (inter-cell) and border-to-POI (intracellular) distances which are stored with the polygons. Only the query-to-border distance is computed online.

4. PROPOSED CACHE REPLACEMENT POLICY

A cache replacement policy should choose an item with low access probability, less network density and longer network distance for eviction. The replacement of data items from the cache is done based on a cost function. The cost function in the policy (PINE-CRP) of a data value j of item i is defined by:

$$C_{i,j} = \frac{P_i \times W_d(V_{i,j})}{N_d(V_{i,j})}$$

Where W_d is the weighted network density of the valid scope V, N_d is the network distance between the client's current position to the *POI (Point Of Interest)* considered and Pi is the access probability. This policy will evict the data with least cost during each replacement. The policy would choose the data with less access probability, less network density and greater network distance for eviction.

5. IMPLEMENTATION ISSUES

5.1 Calculation of access probability

Calculation of access probability employs the exponential aging method as in [16]. 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.

5.2 Computing Network Distance using PINE

Computing network distance using Dijkstra's algorithm is expensive both in terms of space as well as time. The new idea is to compute N_d using *PINE*. PINE algorithm computes the corresponding network distance for each nearest-neighbor that it generates. The N_d for each POI from the query point Q_p can be broken down into three segments and summed up as:

$$N_d = d(Q_p, G_q) + [d(G_q, b_1) + d(b_1, b_2) + \ldots + d(b_{n-1}, b_n) + d(b_n, POI)]$$

where, G_q is the generator point of the network voronoi of Q_p, b_1, b_2, …, b_n are the border points of the network voronoi explored by PINE. The bracketed terms represent the intercell border-border distances (assuming *n* border points here) and border-generator distance which are all pre-computed. These can be readily used by PINE, the only distance to be calculated on the fly is $d(Q_p, G_q)$ which is the distance from the query point to its voronoi generator. At this juncture, we make use of the euclidean distance between Q_p and G_p, $d_{euclidean}(Q_p, G_p)$. This assumption of ours that the euclidean distance between Qp and G_p is comparable to their network distance holds good since we do the approximation only within the current network voronoi which is considerably small in area. Employing Dijkstra's algorithm to calculate $d(Q_p, G_p)$ would mean additional space and time overheads.

The new process is both computationally and spatially less expensive as the pre-computation is performed only for the border points of each network voronoi separately and in real world scenarios the ratio of total number of border points to the total number of nodes in the network is small and thus the pre-computation is performed for each network voronoi separately and not across all network voronoi. This process is shown in figure 3.

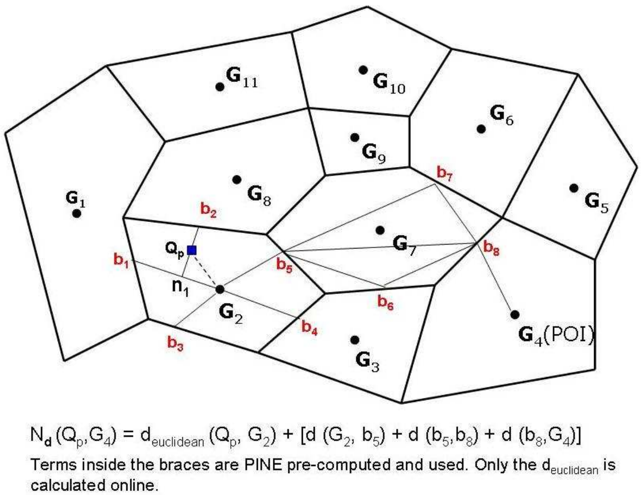

Figure 3. Calculation of the network distance between query point and Point of Interest

6. SIMULATION MODEL

The System execution model, client execution model and server execution model is similar to the one adopted by [16]. Our Simulator is implemented in C++. Scope distributions of the data items are generated based Network Voronoi Diagrams and contains 200 generator points randomly distributed in a street map of Kuwait. The mobile client with fixed cache size is modeled with two independent processes: query process and move process and its access pattern over different items follow a Zipf distribution [2]. The server is modeled by a single process that services the requests from clients on FCFS service principle. The bandwidth of the uplink channel is assumed to be 19.2 kbps and that of downlink channel is 144kbps.

7. PERFORMANCE EVALUATION

In this section, the proposed cache replacement policy is evaluated against FAR, PAID and WPRRP using the simulation model described in the previous section. Cache hit ratio is employed as the primary performance metric.

We have conducted experiments by varying the query interval, moving interval and cache size. Query interval is the time interval between two consecutive client queries. In this set of experiments, we vary the mean query interval from 20 seconds to 200 seconds.

As illustrated in Fig. 4, when the query interval is increased, the performance of all the schemes degrades. We assume that the user is moving at a uniform speed, and if the query interval is longer the probability of reissuing a query in the same valid scope is less. The new policy shows a better hit ratio than other policies when the query interval is small. By dividing the network into network voronoi cells and applying PINE we could accurately generate and retain the nearest neighbors in cache.

We analyze the performance of the replacement policies when the moving interval is varied. The longer the moving interval, the less frequently the client changes velocity and, hence, the less random the client's movement. When the moving interval takes a small value, the user might be cruising and his movement is rather random.

Our policy performs better than other policies for small moving intervals where the user moves more randomly in a dense network. PINE ensures that the points of interests in the vicinity of the user are not evicted irrespective of the direction of motion.

When the moving interval takes a large value, the movement of the client behaves more like a pre-defined trip on highway. Network density ensures that the points of interest on highways are retained there by showing an increase in hit ratio when compared to FAR, PAID and WPRRP.

Effect of cache size on performance of replacement policies is shown in Fig.6. The performance of all policies shows substantial improvement with increased cache size. The results of Fig 6 could be used to decide the optimal cache size in the client.

Figure 4. Cache Hit Ratio Vs Query Interval

Figure 5. Cache Hit Ratio Vs Moving Interval

Figure 6. Cache Hit Ratio Vs Cache Size

8. CONCLUSION

In this paper, we have proposed a new cache replacement policy that considers the density of the valid scope and network distance irrespective of the direction of motion. We have applied PINE to compute the network distance more efficiently. Simulation results show that the new policy shows a substantial improvement in performance than the existing policies FAR, PAID and WPRRP. Consideration of an already proven algorithm like PINE improves the performance significantly. In our future work, we would like to incorporate a pre-fetch or hoarding mechanism into the proposed policy.

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 on Database and Expert Systems Applications, pp. 414-419, 1998.

[7] M. Kolahdouzan and Cyrus Shahabi " Voronoi-Based K Nearest Neighbor Search for Spatial Network Databases ", Proceedings of the 30th VLDB 2004. Pages : 840–851

[8] M. Kolahdouzan and Cyrus Shahabi "Alternative Solutions for Continuous K Nearest Neighbor Queries in Spatial Network Databases" *GeoInfo(9)*, No . 4, December 2005, pp. 321-341

[9] A. Kumar, M.Misra and A.K.Sarje "A weighted cache replacement policy for location dependent data in mobile environments", Proc. of ACM symposium on Applied computing, 2007,pp 920 - 924

[10] D.L. Lee, W.-C. Lee, J. Xu, and B. Zheng, "Data Management in Location-Dependent Information Services", IEEE Pervasive Computing, 1(3), 2002.

[11] A. Okabe, B. Boots, K. Sugihara, S.N. Chiu, Spatial Tessellations, Concepts and Applications of Voronoi Diagrams, John Wiley and Sons Ltd., 2nd edition, 2000, ISBN 0-471- 98635-6.

[12] 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

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

[14] M. Safar, K Nearest Neighbor Search In Navigation Systems, Journal of Mobile Information Systems (MIS), 2005, IOS Press

[15] 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.

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