放送・通信融合環境における効率的な問合せ処理方式

蔡 菁[†] 寺田 努[‡] 原 隆浩[†] 西尾章治郎[†]
† 大阪大学大学院情報科学研究科マルチメディア工学専攻
〒565-0871 大阪府吹田市山田丘 1-5
E-mail: {cai, hara, nishio}@ist.osaka-u.ac.jp
‡ 大阪大学サイバーメディアセンターサイバーコミュニティ研究部門
〒567-0047 大阪府茨木市美穂ヶ丘 5-1
E-mail: tsutomu@cmc.osaka-u.ac.jp

本稿では，広帯域のプッシュ型放送通信，狭帯域の基地局プル型無線通信，その中間のプル型放送通信の3つの通信方式を統合的に利用できる環境を想定し，これらのデータ配信方式を選択的に利用する効率的な問合せ処理方式を提案する．本稿では，提案方式を用いることで，従来方式に比べて問合せの平均待ち時間および成功率が向上することをシミュレーション評価により確認した．

A Hybrid Approach of Broadcast Data Delivery and On Demand Wireless Communication

Jing CAI[†], Tsutomu TERADA[‡], Takahiro HARA[†], and Shojiro NISHIO[†]
† Dept. of Multimedia Eng., Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita-shi, Osaka, 565-0871 Japan
‡Cybercommunity Division, Cybermedia Center, Osaka University
5-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

In this paper, we propose a new effective query processing method through our Hybrid Wireless Broadcast (HWB) model, which combines push and pull based broadcast and pull-based point-to-point wireless communication. Our method provides a flexible and complementary information service in different bandwidth and different service range. The results of simulation studies show the HWB approach shortens the average waiting time and enhances the performance of the system.

1 Introduction

With the rapid growth of mobile computers and the advance of mobile communication technology, there is an increasing requirement for high efficiency and high quality mobile information services. On the other hand, frequent disconnection, lower communication bandwidth, and limited energy, storage and computation power are still current challenges facing the mobile computing environment. In addition, the current mobile information services mostly are based on client-server or point-to-point mechanism. However, they will give rise to the overload of server and bandwidth, when numerous mobile users concurrently demand for the services.

As an effective information dissemination method, data broadcast has received a lot of attention in recent years. Mainly because it can scale up to an arbitrary number of mobile users, and thus facilitates efficient bandwidth usage. Accordingly, wireless data broadcast is suitable to disseminate public information, such as stock quotation, news, weather and traffic information, to massive mobile users. Therefore, taking advantage of broadcast for mobile information services is an elegant solution to address the issues of mobile computing.

A large number of studies on data broadcast have been performed, which focus on the push-based broadcast[1], or the balancing of the push and pull based broadcast[2, 3]. Some studies discuss the hybrid networking based on wireless data broadcast[4, 5, 6]. They normally assume that the base station provides both push-based broadcast and pull-based unicast channel, in which the on demand response is limited in point-to-point wireless communication; and the information service of push-based broadcast is also limited in the local scope of base station. There is no study considering combination of the broadcast and wireless communication in a different service range.

The purpose of this paper is to explore a new approach further into hybrid networking. We contribute a new effective query processing method through our Hybrid Wireless Broadcast (HWB) model, which

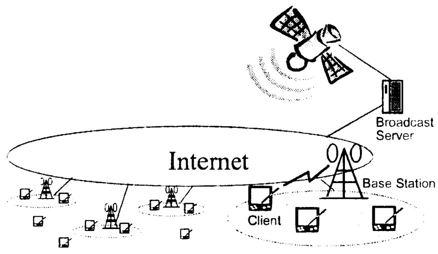

Figure 1: HWB Environment

combines push and pull based broadcast and pull-based point-to-point wireless communication. Mobile users can access the push-based broadcast; also can pull the information from the unicast wireless communication or from the pull-based broadcast. By utilizing the different advantage of the three data delivery ways, our method can provide a flexible and complementary information service in different bandwidths and different service ranges.

The remainder of the paper is organized as follows. The detail of our proposed HWB approach is presented in Section 2. Simulation model developed for the performance evaluation is described in Section 3. Section 4 gives a set of experimental results. Finally, this paper concludes with Section 5.

2 HWB Approach

2.1 Communication Environment

As Figure 1 shows, we assume that our proposed HWB approach is based on a hybrid network environment involving a broadcast server which broadcasts information in a large scope, and lots of the base stations which connect with the fixed network and provide bidirectional point-to-point wireless communication within a limited area. A portable terminal, such as mobile phone, PDA, and palmtop, is supposed to be able to receive the broadcasting information from the broadcast server, as well as the wireless information from the base station. A large number of clients holding with a portable terminal can acquire information from the base station; furthermore, can access information on air from the wireless broadcast.

2.2 Communication Model

Figure 2 indicates the communication model of the HWB approach. Broadly speaking, it is divided into two main kinds of information dissemination: the broadcasting information from the broadcast server is in a large scope, and the wireless information from the base station is in a limited area. Furthermore, the bandwidth for broadcast is classified as the main channel which sequentially broadcasts the whole data, and the sub channel which is used for the pull-based

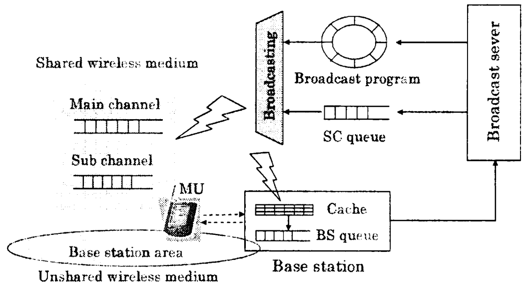

Figure 2: Communication Model

broadcast. Consequently, mobile clients can receive the response from one of the above three data delivery ways: push-based broadcast, pull-based broadcast, and pull-based point-to-point wireless communication. Their different features are illustrated in the following.

The main channel, which provides the push-based broadcast, sequentially and periodically broadcasts the whole data of the broadcast server. Bandwidth does not need to be scaled as the increasing number of mobile users accessing the channel. On the other hand, any access request would not get responded to until the required data arrives. It may be not able to acquire the reply quickly, even though the number of queries is too small.

The sub channel is used for the pull-based broadcast, which sequentially but not periodically broadcasts on-demand data. Any on-demand data broadcast in the sub channel can also be accessed by a large number of mobile users. Therefore, it is efficient for the responding of queries, when many clients request the same data.

The wireless channel of the base station offers pull-based point-to-point wireless communication. Both the sub channel and wireless channel are used to respond on-demand data. However, unlike the pull-based broadcast, the wireless channel is unshared due to the point-to-point communication. On the other hand, each base station has a cache and provides service only for the mobile clients in its responsible area, who normally have some common interest in the local data. As a result, it is good at increasing the cache hits, when caching a lot of local data.

The three data delivery ways have some different features. Besides the communication methods, and the service ranges being different as stated above, the bandwidths are also different. The main channel has a high-bandwidth; the wireless channel has a low-bandwidth; whereas the sub channel has a middle-bandwidth. Taking advantage of these different features, our proposed HWB approach can provide a complementary information service.

2.3 Query Processing

Clients submit data requests to the base station, which is responsible for the query processing. Based on our proposed HWB communication model, we explain how to process the query in this subsection.

When the base station receives a request, the query processing is performed as the following procedure:

Step 1: Calculate the three waiting times respectively. For the wireless channel, the base station needs to check its cache. If there is the requested data in the cache, the waiting time is calculated in accordance with the length of the base station queue. The waiting time of the main channel is calculated according to the relative location of the required data in the broadcast program. As to the sub channel, the waiting time is determined by the location of the data in the sub channel queue, if the requested data has been placed into the queue; or determined by the length of the sub channel queue, if it has not been existed in the queue.

Step 2: Compare the three waiting times, and select the shortest one to reply to the query.

Step 3: According to the compared result, take a corresponding action. If the wireless channel is the shortest, the base station will put the request into its waiting queue, and respond to the query by itself; if the sub channel is the shortest and the data item has not been placed in the sub channel queue, the base station will transfer the query to the broadcast server through the Internet. In other case the base station will not respond or transfer the query, that is to say, the required data will soon be broadcasted from the main channel or sub channel.

3 Simulation Model

To examine our proposed HWB approach thoroughly, the HWB environment stated in Section 2 is modeled in our simulation model, which will be used to evaluate the performance of our approach and other related data delivery approaches. The simulation model specifically is developed into the client model, the base station model, and the server model.

3.1 Client Model

In mobile computing environments, it is important to model how frequently clients issue a query, and in what pattern to issue each query. Taking these into consideration, the client model is described in more detail in the following query pattern and access pattern.

Query Pattern

The exact number of clients is not specified, but instead all the client population is modeled as a single module that generates the independent query interval. The generation of query interval follows the Poisson distribution. The value of the average query interval can be changed to create the effect of a dynamic system workload. The smaller the average query interval, the more frequent the clients' requests, and the heavier the workload of the system.

Access Pattern

Clients independently issue their requests. However, usually they have a biased interest in the public information of their current area, for instance, the local weather or traffic information. Generally, requests follow the 20/80 rule, in other words, a small popular data set is accessed by the majority of queries. For that reason, all the data items of the database are divided into some data groups, in accordance with the number of the base station. It is assumed that the clients in the same base station area have a higher tendency to issue a query from its own data group, which is relevant to its base station; and a lower tendency to request from other data groups. The query tendency can be changed in our experiment, but is normally set as 80%.

Moreover, the query deviation is a key factor of access pattern. In this research, two different access patterns: uniform query and skewed query are provided for the queries inside the data group. The uniform query randomly queries a data item inside the data group. On the other hand, the skewed query employs the Gaussian distribution, with the center of hot spot μ and the deviation σ of hot spot center, namely Gaussian(μ , σ). The value of μ is set as the center of each data group, while the value of σ can be varied to reflect the skewness of the client access patterns. The queries from clients become increasingly skewed as σ decreases.

3.2 Base Station Model

Clients send all queries to the base station. The base station takes responsibility for the query processing and cache management.

Each base station has a cache, which stores local data requested by clients in its area. The LRU rule is used for cache replacement. Moreover, the base station processes the queries. As depicted in Section 2.3, it needs to calculate and compare the three waiting times for the main channel, sub channel, and wireless channel; and then according to the compared results, it takes a different action: respond by itself, transfer the query to the broadcast server, or do nothing.

3.3 Server Model

The broadcast server manages the broadcast of the main channel and the sub channel, which run at the different bandwidths. The broadcast program of the main channel is fixed in random sequence; whereas the on demand data of the sub channel is dynamically changed as the client's request. When the server receives the request for some data item from the base station, it will insert that data item into its sub channel queue.

4 Experiments and Results

In this section, we use the simulation model stated above to demonstrate the characteristics of the proposed HWB approach. The experiments are based on examining some important factors for each part of the simulation model: the client model, the base station model, and the server model. All these important factors, which affect the data access performance of the mobile computing systems, are induced into the four aspects: frequency of data requests, number of data items in the database or in the cache of base station, data access patterns, and the bandwidth of data delivery. By varying each of the above mentioned factors, our experiments evaluate two main performances: the average waiting time and the success rate of the query.

To evaluate our proposed HWB approach more precisely and objectively, some other approaches are introduced into our experiments: random WB approach, push/pull approach, push/w approach, pure pull approach, and pure push approach.

The communication mode of the random WB approach is the same as the proposed HWB approach: both of them hold the base station cache and offer the three data delivery ways, which are main channel, on-demand sub channel and on-demand point-to-point wireless channel, as a response to every request issued from the mobile clients. The only difference is the selection method of query processing. Random WB approach randomly adopts one way from the above three data delivery ways. Instead of random choice, HWB approach selects the optimal way with the shortest waiting time for each query. By comparing with random WB approach, the association between HWB architecture and HWB processing can be clearly clarified.

In addition, other hybrid approaches are used to compare with the HWB approach. Push/pull approach provides push and pull based broadcast, respectively using main channel or on-demand sub channel; while push/w approach uses push-based broadcast with main channel, and point-to-point wireless channel. Both push/pull approach and push/w approach adopt the better way, with much shorter waiting time, from those two possible data delivery ways to respond to each query. Moreover, pure pull approach and pure push approach also help to observe the characteristics of pure pull-based broadcast using only on-demand sub channel, and pure push-based broadcast with main channel.

Table 1: Parameter Settings

Parameters	Values
Database Size [Data Items]	5000
Data Item Size [KB]	100
Number of Base Station	10
Cache Size of BS [Data Items]	100
Main Channel Bandwidth [Mbps]	100
Sub Channel Bandwidth [Mbps]	10
Wireless Bandwidth [Mbps]	1
Time Slot [D/Bm]	20000
Query Interval [ms]	200~2000
Data Group Size [Data Items]	500
Query Tendency [%]	80
Deviation for Gaussain	10
Time Out [s]	10

In the experiments, it is assumed that the parameters normally are fixed to some main values, except for the evaluating parameter factor. The primary parameters and their main values used in our experiments are presented in Table 1. The number of data items in the database is 5000; all data items have an equal size. The number of the base stations is 10; the query tendency inside the data group of each base station is 80%. The skewed query is finished by setting the deviation of the Gaussian distribution as 10 data items. To provide the different bandwidths, the main channel is 100Mbps, the sub channel is 10Mbps, and wireless is 1Mbps. Each experiment runs 20000 time slots with the average query interval varying from 200ms to 2000ms; while a time slot is the time that a data item is broadcast through main channel. Furthermore, for the experiments related to success rate, the time out is set as 10 seconds.

4.1 Impact of Query Frequency

In the first experiment, we evaluate the performance of the average waiting time and the success rate under different workloads, by varying average query intervals from 200ms to 2000ms. The results are shown in Figures 3 and 4. As the query interval increases the workload decreases, therefore the average waiting time declines for almost all the approaches, only the pure push approach has no change. Observing the whole process of the varying query interval, normally the proposed HWB approach outperforms the other approaches, of which the average waiting time is the shortest and the success rate is the high-

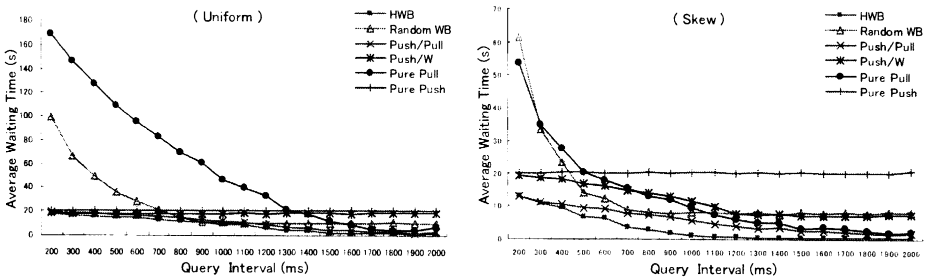

Figure 3: Waiting Time vs. Query Interval

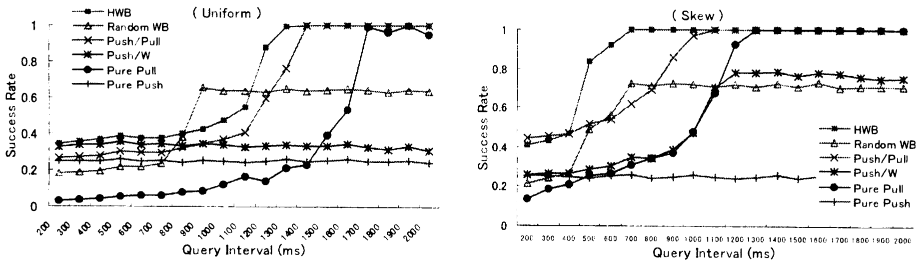

Figure 4: Success Rate vs. Query Interval

est. In the end of the curve, the difference between HWB approach and two other approaches, which are the push/pull approach and the pure pull approach, become very small. The reason is that the amount of queries for the whole system becomes relatively small when the query interval is too long. In this case, dominantly the sub channel is used to answer the requests similarly for the three approaches.

The performance between the HWB approach and the random WB approach is quite different, even though they have the same communication mode. As Figure 4 shows, the highest success rate for the random WB approach is only around 0.7, which is lower than the 1 of HWB approach. Because the method of query processing is rather different; for the HWB approach, all the time it only selects the one with the shortest waiting time from the three data delivery ways stated above; on the other hand, for the random WB approach it not only may select the shortest one, but also maybe the longest one, because it just randomly chooses one. Furthermore, in some cases, its performance may be below the pure pull approach or others. An example is shown in Figure 3: the average waiting time of the random WB approach is longer than the pure pull approach, when the query interval is below 300ms. In that case, the amount of queries for the whole system is so huge that the randomly selected times of the wireless channel also increase, so that the long length of the base station queue gives rise to the long waiting time.

Additionally, there are some other important results. When comparing the two different access patterns under the varying query interval, almost all the approaches perform better in the skewed case rather than in the uniform case, except for the pure push approach. For instance, in Figure 3, under the uniform query at the 200ms query interval, the average waiting time of the pure pull approach is around 170s; whereas under the skewed query, it is only around 50s. The best account for the skewed access pattern greatly improving the performance of pure pull approach, especially when the workload is very heavy, is that the on demand sub channel can be shared by many users, and the length of the queue of the sub channel dramatically decreases when the queries are very skewed. In addition, for the push/w approach in the uniform case the highest success rate is below 0.4; whereas in the skewed case it is above 0.7. The skewed access pattern is helpful for the caching of the base station, which results in the better performance.

As for the proposed HWB approach, its good characteristic performs even better in the skewed case. The most likely explanation is that the HWB approach combines push and pull based broadcast and pull- based wireless communication. As stated above, the skewed access pattern is helpful for the shared on

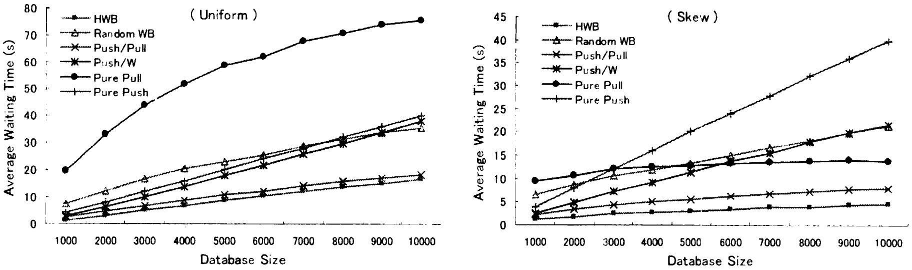

Figure 5: Waiting Time vs. Number of Data Items in Database

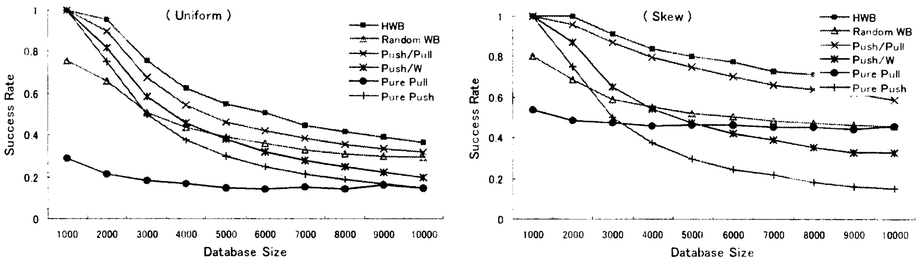

Figure 6: Success Rate vs. Number of Data Items in Database

demand sub channel, as well as the caching of the base station; therefore, the HWB approach can outperform others, particularly in the skewed case, by utilizing these two advantages. An example is shown in the right side of Figure 4, at the 700ms query interval, the success rate of the HWB approach is approaching 1; however, it is only 0.6 for the push/pull approach, and it approaches 1 only when the query interval is above 1100ms; whereas for the pure pull approach, it needs above 1300ms. Additionally, the highest success rates for the other approaches are all below 0.8. Accordingly, the proposed HWB approach generally outperforms the others as the workload increases.

4.2 Impact of Database Size

Secondly, we examine the influence of the number of data items in database from 1000 to 10000. Figure 5 shows that the average waiting time of almost every method, except for the pure pull approach, is proportional to the number of data items, but at different slopes. Ranking from the biggest gradient, it is the pure push approach, push/w approach, push/pull approach and HWB approach. The gradient of each curve clearly reveals its performance, the smaller the gradient the shorter the average waiting time, which relies on its structure. Pure push approach sequentially broadcasts the every data item of the database using the main channel. Based on the broadcast of the main channel, the other approaches add one or two other data delivery ways. The push/w approach, and push/pull approach respectively add the on demand wireless channel of the base station, or the pull-based broadcast of the sub channel. The characteristic of the latter is better than the former, which results in its smaller slope. Moreover, the HWB approach adds both the on demand wireless channel and the on demand sub channel; therefore it outperforms the others with the smallest slope. In the skewed case, the difference of the slope becomes bigger, which can be explained by the fact that each approach acquires the different favor from the skewed query.

The performance for all the approaches declines as the database size increases, but the declining degree for the HWB approach is the smallest. In other words, the HWB approach has the shortest average waiting time and the highest success rate (cf. Figure 5, 6). In the skewed case, it performs better due to its good feature of the combination . To some extent the other approaches also improve their performance. For example, for the pure pull approach (cf. Figure 6), in the uniform case the success rate is around 0.2; while in the skewed case it rises to about 0.5. The reason is that under the uniform query there is a marked increase in the queue of the sub channel for the required data items as the database size increases, so that the average waiting time sharply increases in the left of Figure 5. Whereas in the skewed case, there is only

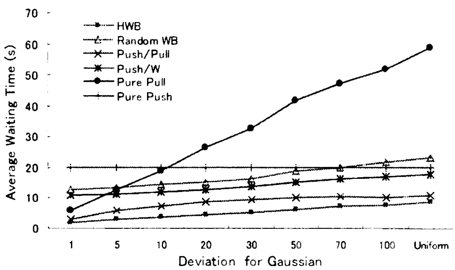

Figure 7: Waiting Time vs. Query Deviation

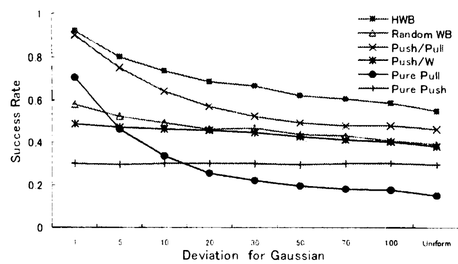

Figure 8: Success Rate vs. Query Deviation

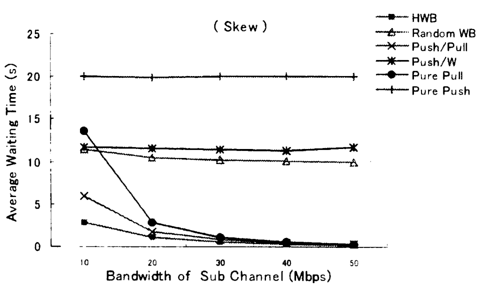

Figure 9: Waiting Time vs. Bandwidth of Sub Channel

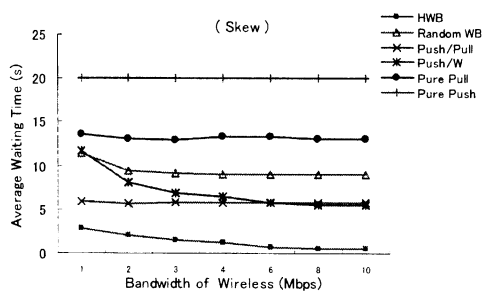

Figure 10: Waiting Time vs. Bandwidth of Wireless

a slight increase for the queue despite the increase of the database size, because the broadcast of the sub channel can be shared by the skewed queries, consequently the average waiting time increases slowly in the right of Figure 5. Only one exception is the pure push approach, which has the same behavior in the uniform case and in the skewed case. It is because the pure push approach should broadcast the whole data items of the database in both cases, and the main channel is the only way to disseminate the data.

4.3 Impact of Access Pattern

This subsection evaluates a key factor of access pattern, namely query deviation stated in Section 3.1. By varying the deviation of the Gaussian distribution, the examinations about average waiting time and success rate are illustrated in Figures 7 and 8. As the deviation is smaller, in other words, the queries become increasingly skewed, the performances of almost all upgrade: especially there is a large improvement in the pure pull approach; however the pure push approach has no change. Across the entire region of the evaluation, the HWB approach still outperforms other approaches.

4.4 Impact of Pull Bandwidth

The evaluation about pull bandwidth is examined by varying the bandwidth of the sub channel form 10 to 50Mbps, or the bandwidth of wireless channel from 1 to 10Mbps (cf. Figure 9, 10), noting only under the skewed query. It is shown that the HWB approach always performs best, however the changing of pull bandwidth has a small influence on it.

The waiting times for the HWB approach, push/pull approach and push/w approach decrease as the bandwidth of the sub channel is increased. However, when the bandwidth is larger than 30Mbps the amount of decrease becomes smaller, and the difference is also very little. This is because in this case the bandwidth of the sub channel is huge enough to respond to almost all the requests.

In addition, the average waiting time of push/w approach improves much as the bandwidth of the wireless channel increases; it even performs a little better than push/pull approach but far from the HWB approach when the bandwidth of wireless channel is larger than 6Mbps. The reason is that each base station holds its own cache and wireless channel to provide service only for the local clients. On the other hand, the sub channel is only one, which is shared

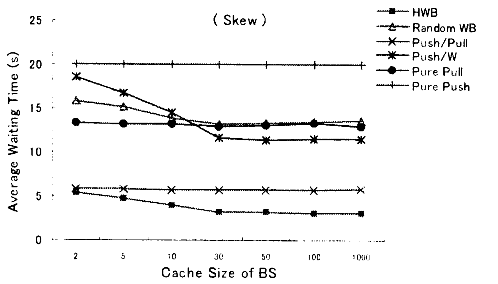

Figure 11: Waiting Time vs. Cache Size

by all the clients in a large scope. Therefore, it can be fairly certain that the on demand wireless channel cannot be replaced by the on demand sub channel.

4.5 Impact of BS Cache Size

To find out the influence of the cache size of the base station, the average waiting time is evaluated only for the skewed queries, by varying the number of data items in the cache from 2 to 1000. Figure 11 shows that as the increase of the cache size, there are some decline for the HWB approach and push/w approach. However, they almost change little when the cache size is larger than 30 data items. Accordingly, there is no necessary for the cache size being above 5 times larger than the size of the deviation of the skewed query.

4.6 Summary of Experiments

We conclude from all of the above evaluation results that our proposed HWB approach shortens the average waiting time and enhances the performance of the system; at the same time, even when the system workload increases, the degrading degree of the HWB approach usually is the smallest. In other words, the HWB approach works effectively and outperforms other approaches all the time. It is clear that the skewed access pattern dramatically improves the performance of the sub channel; and it is also useful for the caching of the base station. Meanwhile, in the skewed case, the HWB approach performs even better by taking advantage of both on demand sub channel and on demand wireless channel. Furthermore, in comparison with the Random WB approach, it is clarified that two core parts of the HWB approach, namely HWB communication model and HWB query processing stated in Section 2, cannot be divided. Finally, it is obvious that only the database size has an influence on the performance of the pure push approach; the other factors, such as query interval, skewed or uniform query, pull bandwidth and the cache size have no impact on it.

5 Conclusion

In this paper, we have put forward a novel information delivery mechanism to contribute hybrid networking through our Hybrid Wireless Broadcast (HWB) model, which combines push and pull based broadcast and pull-based point-to-point wireless communication. Moreover, based on the HWB model we have proposed an effective query processing method, which can provide a flexible and complementary information service in different bandwidths and different service ranges. Furthermore, a simulation model has been developed to evaluate the performance of the data delivery system. The experimental results showed that our proposed HWB approach shortens the average waiting time and enhances the performance of the system; particularly, it performs more effectively for the skewed access pattern; and the degrading degree usually is the smallest, even when the system workload increases.

As future work, the benchmark of the optimal selection from the proposed three data delivery ways, not only simply considering the waiting time of query, we will take the response cost into account to investigate more effective query processing algorithms.

Acknowledgments

This research was supported in part by Grant-in-Aid for Scientific Research (B) (15300033) and (16016260) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and as part of the 21st Century Center of Excellence Program of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] S. Acharya, R. Alonso, M. Franklin, S. Zdonik, Broadcast disks: Data management for asymmetric communication environments, Proc. ACM SIGMOD'95, pp.199-210, May 1995.

[2] S. Acharya, M. Franklin , S. Zdonik, Balancing push and pull for data broadcast, Proc. ACM SIGMOD'97, pp.183-194, May 1997.

[3] K. Stathatos, N. Roussopoulos, J.S. Baras, Adaptive Data Broadcast in Hybrid Networks, Proc. VLDB'97, pp.326-335, Aug. 1997.

[4] Q.L. Hu , D.L. Lee , W.-C. Lee, Performance evaluation of a wireless hierarchical data dissemination system, Proc. Mobicom'99, pp.163-173, Aug. 1999

[5] S. Buchholz, A.Schill, T.Ziegert, A simulation study of update techniques for cyclic data broadcast, Proc. MSWiM, pp.115-122, July 2001.

[6] C.W. Lin, H.Hu, D.L. Lee, Adaptive realtime bandwidth allocation for wireless data delivery, Wireless Networks 10, pp.103-120, 2004.