Vol. 47 No. SIG 19(TOD 32) IPSJ Transactions on Databases Dec. 2006

Regular Paper

A Hybrid Data Delivery Method of Data Broadcasting and On-demand Wireless Communication

Jing Cai,† Tsutomu Terada,† Takahiro Hara† and Shojiro Nishio†

As the recent advances in the wireless technologies and mobile terminals, mobile users equipped with mobile devices are able to access wireless services through 3G cellular network, WiFi hotspot, or WiMAX link, as well as through satellite digital broadcast or terrestrial digital broadcast. By effectively taking advantage of these complementary communication modes, we explore a new hybrid data delivery model, i.e., Hybrid Wireless Broadcast (HWB) model to benefit from the optimal combination of the push-based and pull-based broadcast and on-demand point-to-point wireless communication. The HWB model can provide a flexible and complementary information service in different bandwidths and service ranges, and greatly improve the responsibility, scalability, and efficiency of the system. The results of simulation study show the proposed HWB approach achieves a significant improvement in system performance.

1. Introduction

In recent years, many wireless technologies such as IEEE 802.11, IEEE 802.16, 3G mobile communication, and data broadcast become available. On one hand, mobile users equipped with wireless terminal, such as cellular phone, PDA, and laptop PC, can acquire information services through 3G cellular network, WiFi hotspot, or WiMAX link. Although these technologies vary in many aspects, we look at them as a base station infrastructure network (abbr. as BS network). In such a network, mobile clients access wireless services through the base station, i.e., wireless access point, which connects with the wired network and serves the clients within the responsible area.

On the other hand, wireless information service through digital broadcast, such as satellite digital broadcast or terrestrial digital broadcast, becomes available in more and more widespread areas. Most recently, mobile digital broadcast has been available over the air. The world's first satellite digital broadcasting reception by mobile devices, namely mobile broadcasting MobaHO, starts service from October, 2004 in Japan; meanwhile, Japanese terrestrial digital broadcasting service for cellular phone starts for the first time in the world in April, 2006.

To coincide with the launch of mobile digital broadcast service, many corresponding equipments successively appear in the market, e.g., DoCoMo P901iTV, au W33SA W41H, Vodafone 905SH cellular phone, VAIO type T laptop, MBCO MBT0102A PC card tuner for installing in notebook PC [19)~22)], etc. Therefore, mobile users holding these emerging mobile devices are able to access wireless services through the BS networks, as well as through data broadcast, which significantly extends the available information services for a single mobile terminal. However, as of today, these two communication modes are utilized separately so that it is impossible to enjoy services that benefit from the optimal integration of the BS network and data broadcast.

In fact, the BS network and data broadcast have some complementary features, and it is beneficial to put these features to good use. For the BS networks, geographical decentralization is particularly suited to provide local services to the clients within the responsible region. Location based service is a good example. However, in such a network, the wireless service is unicast communication, and based on the client-server model. The concurrent information requirements of the numerous clients will give rise to the overload of the server and narrowband bandwidth, because all the queries have to be responded by point-to-point communication, even though they request for the same information. In contrast, data broadcast is suitable to disseminate popular information to public, such as stock quotation and news, since broadcast

† Graduate School of Information Science and Technology, Osaka University

can scale up to an arbitrary number of clients, and facilitate efficient bandwidth usage. However, broadcast is generally insensitive to the special conditions of mobile clients. For instance, the answer to a query on stock price is the same, no matter where the data is related and where the client is located. Therefore, it can provide more effective information services, if these two communication modes are integrated efficiently.

So far, some studies discussed the hybrid networking based on wireless data broadcast [6),8),11),15)]. They normally assumed that the base station provides push-based broadcast and pull-based unicast communication, in which the on-demand response is limited in point-to-point wireless communication, and the push-based broadcast is limited in the local scope of the base station. None of research considered to provide a complementary information service in different service ranges.

The purpose of this paper is to explore an effective information dissemination in a new hybrid networking by taking advantage of the emerging mobile/wireless communication. Based on the BS networks, we regard the base station as a local server, which is responsible for serving the mobile clients in its liable region. Moreover, we expand the BS networks by integrating with the broadcast of the server, which provides service to the clients in a global region. In such a decentralized system, the broadcast server manages the information for broadcasting, whereas the base stations take charge of query processing. This is because mobile clients' queries being processed at the base stations can alleviate the workload of the server; otherwise, if a client communicates with the broadcast server directly, the workload of a single server will become quite heavy when the clients' queries greatly increase.

In this paper, we contribute a Hybrid Wireless Broadcast (HWB) model, which combines push-based and pull-based broadcast with pull-based point-to-point wireless communication to provide a flexible and complementary information service in different bandwidths and different service ranges. This combination makes the system more efficient, responsible and scalable. Based on the proposed HWB model, we further put forward an effective query processing method. Furthermore, we develop a simulation model and conduct a set of experiments to evaluate the performance of the proposed HWB approach.

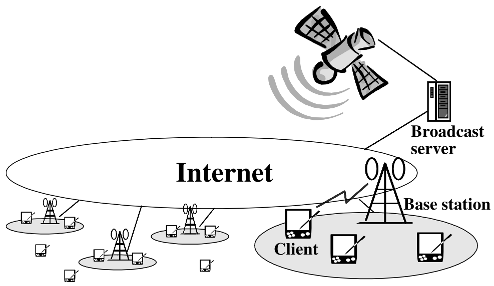

Fig. 1 HWB environment.

The remainder of the paper is organized as follows. The detail of the HWB approach is presented in Section 2. Simulation model is described in Section 3. Section 4 gives a set of experimental results. Section 5 introduces the related work. Finally, this paper concludes in Section 6.

2. HWB Approach

2.1 HWB Environment

We assume the proposed HWB approach is based on a hybrid wireless broadcast environment, in terms of the emerging mobile devices being able to access the BS networks, as well as the satellite broadcast or the terrestrial broadcast. As **Fig. 1** shows, the HWB environment comprises a broadcast server, lots of base stations, and a large number of mobile clients. The broadcast server broadcasts information in a large scope; while the base stations connect with the broadcast server through the internet and provide wireless communication to the clients in their local areas. Mobile clients are supposed to be able to access information from the base station and the broadcast server.

In such a hybrid network, mobile users holding a portable terminal such as PDA and cellular phone can access wireless services through the base station, like in a BS network, meanwhile can acquire information via the broadcast of the server, like in a satellite or terrestrial broadcast.

2.2 HWB Model

Figure 2 further indicates the communication model of the HWB approach. Broadly speaking, the HWB communication is divided into two communication modes: the broadcast from the server to mobile clients, and the unicast communication between the base station and mobile clients. Specifically, the bandwidth for broadcast is classified as the main channel

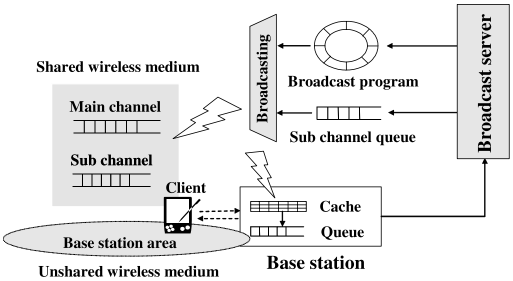

Fig. 2 Communication model.

and the on-demand sub channel. Each base station has a cache and provides a bidirectional point-to-point communication via the wireless channel. Consequently, mobile clients can send requests to the base station through the uplink of the wireless channel, and receive the responses from one of the three downlinks: the push-based main channel, the pull-based sub channel, or the pull-based wireless channel. In a HWB system, the base stations take charge of selecting an optimal way to respond to the queries of clients, while the broadcast server manages the information for broadcasting. The proposed HWB approach attempts to provide a complementary information service, by taking advantage of the different features of the three data delivery ways.

The main channel, which provides the push-based broadcast, sequentially and periodically broadcasts the data of the broadcast server. The response of the main channel is not affected by the system load, namely the number of the queries. However, clients need to wait until the desirable broadcast arrives. The average response time depends on the volume of broadcast.

The sub channel is used for the pull-based broadcast to meet the individual requirements of global clients, which sequentially but not periodically broadcasts the on-demand data. Any on-demand broadcast can be shared by all clients. Therefore, it is efficient to respond to the queries which require for the same data. However, response latency depends on the number of different requests.

The wireless channel of the base station offers pull-based point-to-point wireless communication to meet the individual requirements of local clients. However, unlike the on-demand sub channel, the on-demand response via the wireless channel cannot be shared due to the point-to-point communication. Hence, response latency depends on the number of queries.

Besides the communication method and the service range being different, the bandwidths among the three channels 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.

2.3 Query Processing

Clients submit data requests to the base station. Each base station is responsible for processing the queries of the local clients.

The bandwidths for the main channel, the sub channel, and the wireless channel are denoted by B_m, B_s, and B_w, and the waiting times by T_m, T_s, and T_w, respectively. For the on-demand sub channel and wireless channel, we employ a queue to manage the responses, since the service for the clients' requests mostly complies with the "first come first served." In addition, we assume the broadcast server containing N data items with equal size D. In the case of the different sizes of the broadcast data, the algorithm for calculation is the same, but needs to accumulate the total size for all the corresponding broadcast items.

The procedure for query processing is as follows.

Step 1: Calculate the three waiting times with a suitable formula.

Assume that N_t is the position of the data item in the broadcast program, which is currently being broadcast through the main channel; N_i is that of the requested data item.
If the requested data item has not been broadcast in the current period, i.e., $N_i \geq N_t$, then

$$T_m = (N_i - N_t) \times D/B_m,$$

which is the period of time that the requested data item will be broadcast through the main channel. If the requested data item has been broadcast in the current period, i.e., $N_i < N_t$, then

$$T_m = (N - N_t + N_i) \times D/B_m,$$

which is the time adding the current remained broadcasting period to the time of the data items in the next period to be broadcast until the requested data item.

As to the sub channel, if the requested data item has been placed into the queue, the waiting time is calculated by

$$T_s = N_s \times D/B_s,$$

which is the time of the data items in the sub channel queue from the first data item to the requested one to be broadcast through the sub channel, where N_s is the position of the requested data item in the sub channel queue. Otherwise, if the requested data item has not existed in the queue, it is calculated by

$$T_s = L_s \times D/B_s,$$

which is the time of the whole data items in the sub channel queue to be broadcast, where L_s is the length of the sub channel queue, i.e., the number of the pending data items in the sub channel queue.

For the wireless channel, the base station needs to check its cache. If there is the requested data in the cache, the waiting time of the wireless channel is calculated by

$$T_w = L_w \times D/B_w,$$

which is the time of the whole data items in the base station queue to be transmitted through the wireless channel, where L_w is the length of the base station queue. Otherwise, the wireless channel cannot be used for the query, because the waiting time is infinite, i.e., $T_w \to \infty$.

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

Step 3: According to the selection 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. Otherwise, the base station will not respond or transfer the query, that is to say, the requested data will soon be broadcast through the main channel or the sub channel.

We ignore the time for query processing, because it does not need to perform a very complicated process or calculation, and the time for query processing is very short in comparison with the time for data transmission.

3. Simulation Model

To examine the proposed HWB approach thoroughly, the HWB system is modeled in our simulation model: the client model, the base station model, and the server model.

3.1 Client Model

In the client model, we model how frequently clients issue a query, and in what pattern to issue each query. Here, we assume clients have no capability of caching, and have an interest to extend it in a future study.

Query interval is used to simulate the load of queries of clients. The shorter the query interval is, the more frequent the clients' requests are and the heavier the workload of the system gets. In our simulation, the generation of query interval follows the Exponential distribution, and the varying average query interval for each base station models different system workloads.

On the other hand, information service is better to be grouped according to the different areas, since the clients in different regions normally have heterogeneous accesses, whereas the clients in a same geographical area often have similar interests in the local information, e.g., location based information. Thus, we assume that all the items of the database are divided into several data groups in accordance with the base stations. The clients in the same base station area have a higher tendency to issue a query from the data group relevant to its belonging base station, and a lower tendency to request from other data groups. In addition, the client access patterns are assumed to follow the Zipf distribution, where θ is a parameter named access skewness coefficient and can be varied from zero to one. When $\theta = 0$, it is a uniform distribution, but Zipf distribution is frequently used to model skewed access, which becomes increasingly skewed as θ increases.

3.2 Server Model

The broadcast server manages the broadcast of the main channel and the sub channel. The broadcast schedule of the main channel adopts flat broadcast, i.e., broadcasts the data of the database only once in a cycle. All data items in the database are self-identifying, which can be done by assigning a unique identifier to each data item. For the sake of simplicity, we ignore the varying broadcast of the main channel and the update of data. However, the on demand broadcast of the sub channel is dynamically changed as the clients' requests. When the server receives the request for a certain data item from the base station, it will insert that data item into the sub channel queue, if the requested data item has not been placed in the sub channel queue.

Index technique [8),9)] is employed in our model. The server interleaves index information with data on the broadcast channels to inform the broadcasting information. The index overhead can be ignored, since the index size is much smaller in comparison with the size of the broadcast data.

3.3 Base Station Model

The base stations take responsibility for cache management and query processing for the local clients. Each base station adopts the LRU rule for cache replacement. Through the broadcasting of the main channel or the sub channel, the base stations discard the least recently used items and keep the number of cache size of the most recently queried items in cache. In addition, as depicted in Section 2.3, the base stations need to select an optimal way among the three data delivery ways for responding to the client request, and according to the selected result take a corresponding action: respond by itself, transfer the query to the broadcast server, or do nothing special.

We assume the base stations have the knowledge of the broadcasting. By first retrieving the index information, the base stations are able to obtain the arrival time of the requested data items on the broadcast channel.

In addition, our simulation does not consider the influence of transmission delay. Although a delay may happen in some cases, such as a transmission error of the wireless link and weather influence on the satellite link, the study of this problem is beyond the scope of our research.

4. Experiments and Results

In this section, we use the simulation model stated above to demonstrate the characteristics of the HWB approach. Our experiments evaluate two performances: the average waiting time (AWT) and the success rate of the queries. The average waiting time is defined as the average elapsed time between query generation and its fulfillment. The success rate is defined as the ratio of the number of queries being successfully responded within a time-out period to the total number of queries the clients issued. The time-out period is the maximum time allowed for a query to be responded. In our experiments, only the evaluation of the success rate considers the limit of time-out, whereas the waiting time of each query is used in the statistical evaluation of the AWT no matter whether it is beyond the time-out or not.

On the other hand, some related approaches are introduced into the experiments to compare with the proposed HWB approach, which are the random WB approach, the push/pull approach, the push/w approach, the pure pull approach, and the pure push approach.

The communication mode of the random WB approach is the same as the HWB approach. Both of them hold the base station cache and offer the three data delivery ways: the main channel, the on-demand sub channel and the on-demand wireless channel. The only difference is the selection method of query processing. The random WB approach just adopts one way randomly. By comparing the HWB approach with the random WB approach, the association between the HWB model and the HWB processing can be clearly clarified.

Moreover, we compare the HWB approach with two conventional hybrid data delivery methods. The push/pull approach provides push-based and pull-based broadcast by using the main channel and the on-demand sub channel; while the push/w approach utilizes the main channel for push-based broadcast, together with the point-to-point on-demand wireless channel. They both select the better way with a shorter waiting time, from the two possible data delivery ways, to respond to the client query.

In addition, to make comparisons, the pure pull approach provides the pure pull-based broadcast by using the on-demand sub channel, and the pure push approach provides the pure push-based broadcast by using the main channel.

The primary parameters and their default values used in our experiments are presented in **Table 1**. The number of data items in the database is 10,000; the number of the base stations is 10. The bandwidth for the main channel, the sub channel, and the wireless channel are assumed as 100 Mbps, 10 Mbps, and 1 Mbps, respectively, due to the satellite communication system with a high data rate above 100 Mbps having been developed, meanwhile the data rate of the 3G mobile service being

Table 1 Parameter settings.

Parameters	Values
Database Size [Data Items]	10,000
Data Item Size [KB]	100
Number of Base Stations	10
Cache Size of BS [Data Items]	200
Main Channel Bandwidth [Mbps]	100
Sub Channel Bandwidth [Mbps]	10
Wireless Bandwidth [Mbps]	1
Evaluation Time [Time Slot]	100,000
Query Interval [ms]	50∼2,100
Data Group Size [Data Items]	1000
Skewness for Zipf Distribution	0/0.95
Time Out [s]	30

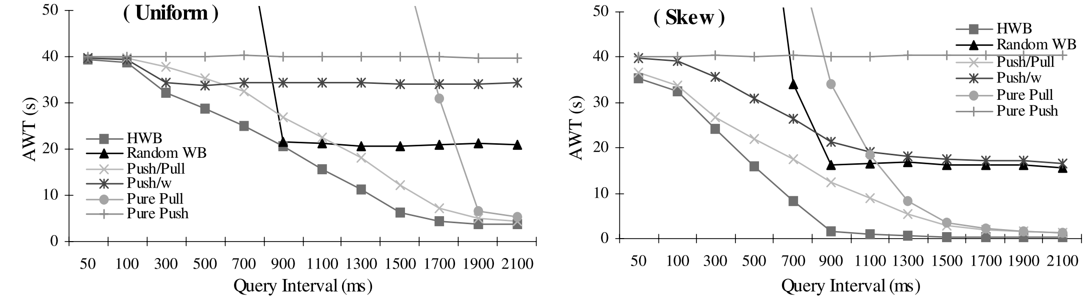

Fig. 3 Waiting time vs. query interval.

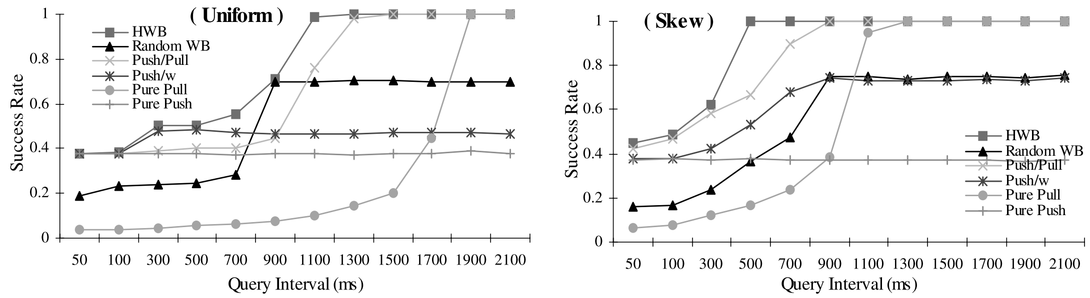

Fig. 4 Success rate vs. query interval.

a few hundred Kbps and several Mbps. Each experiment runs 100,000 time slots, where one time slot is defined as the time that a single data item is broadcast through the main channel.

4.1 Impact of Query Frequency

In the first experiment, we evaluate the system performance under different workloads, by varying average query intervals from 50 ms to 2,100 ms. The results are shown in **Fig. 3** and **Fig. 4**. Both the average waiting time and the success rate upgrades as the query interval increases for almost all the approaches due to the decreasing workload. Only the pure push approach has no change, because it just provides the broadcast via the main channel which can be shared by an arbitrary number of clients.

Observing the whole process of the varying query interval, the proposed HWB approach always outperforms the other approaches, of which the average waiting time is the shortest and the success rate is the highest. In the end of the curve, the difference between the HWB approach and the push/pull approach as well as the pure pull approach becomes very small. The reason for these behaviors is that the number of queries for the whole system gets quite small when the query interval is very long. In this case, the sub channel is dominantly used for the above approaches to respond to the requests.

In addition, the performance of the random WB approach is much lower than that of the HWB approach, even though they have the same communication modes. This is because the query processing is different. The HWB approach only selects a data delivery way with the shortest waiting time; in contrast, the random WB approach not only may select the shortest one, but also maybe the longest one, due to the random choice.

Moreover, comparing the two different access patterns under the varying query interval, almost all the approaches perform better under the skewed access rather than the uniform access, except for the pure push approach. The best account is that for the pure push approach the whole data items of the database should be broadcast in both cases. In contrast, for the on demand sub channel, which can be shared by a large number of clients, the length of the queue dramatically decreases as the queries get skewed. Additionally, the skewed access is also helpful to increase the cache hits of the base station. As a result, the proposed HWB approach performs even better under the skewed access, by benefiting from the above mentioned two merits to gain the best performance.

4.2 Impact of Database Size

In the next experiment, we examine the influence of the number of data items in the

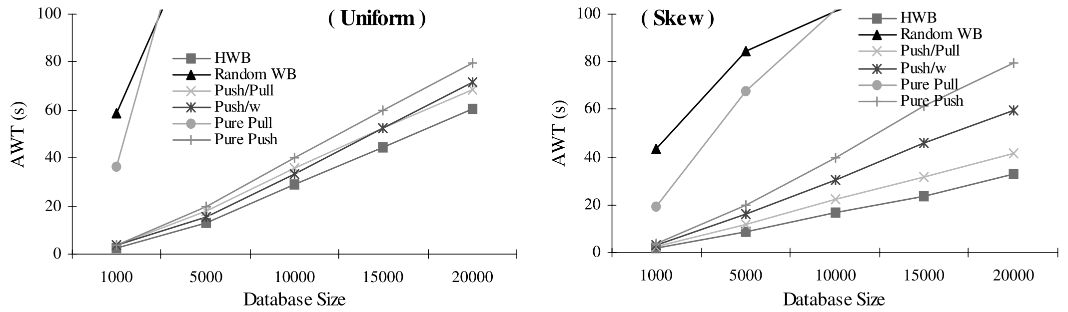

Fig. 5 Waiting time vs. number of data items in database.

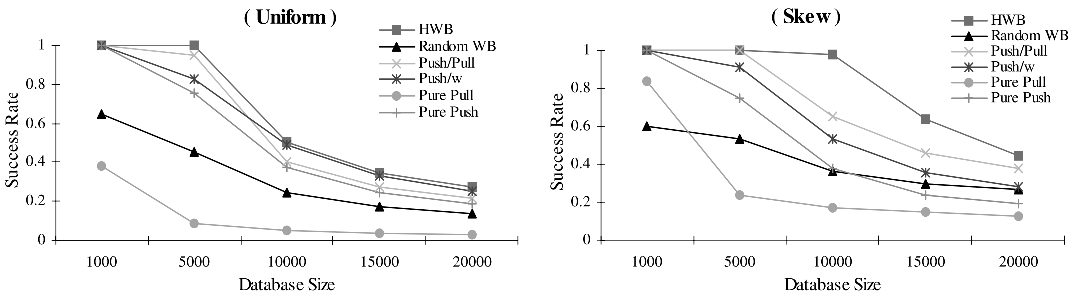

Fig. 6 Success rate vs. number of data items in database.

database. As shown in **Fig. 5** and **Fig. 6**, the performances for all the approaches decline as database size increases, due to the increasing transmission time for more data items. However, the declining degree for the HWB approach is the smallest, i.e., the HWB approach all the time has the shortest average waiting time and the highest success rate.

Figure 5 shows that the average waiting times for the most of approaches are approximately proportional to the number of data items, but at different slopes. Ranking from the smallest gradient under the skewed access, they are HWB approach, push/pull approach, push/w approach, and pure push approach. The smaller the gradient, the shorter the average waiting time they get. The gradient of each curve clearly reveals its performance, which relies on its communication mode. The pure push approach sequentially broadcasts each data item of the database with the main channel. Based on the broadcast of the main channel, the other approaches add one or two channels to improve performance. The push/w approach adds the on-demand wireless channel, while the push/pull approach adds the on-demand sub channel, in which the latter performs better than the former since the on-demand broadcast can make more efficient use of the bandwidth to respond to the skewed access. Moreover, the HWB approach adds both the on-demand wireless channel and the on-demand sub channel; hence, it outperforms the other approaches with the smallest slope.

Meanwhile, all approaches improve performance under the skewed access. The reason is that, under the uniform query, the queue of the sub channel or the wireless channel increases markedly as the database size gets larger, which drops the performance sharply. However, under the skewed access, there is a relatively slight increase for the sub channel queue or the wireless channel queue. It is because in this case the increasing number of data items via the on-demand sub channel can be shared by all the clients with a similar access interest, additionally, the skewed access is also helpful to increase the cache hits of the base station, and to some extend increase the possibility of the selection of the wireless channel. Furthermore, the HWB approach performs even better under the skewed access, owning to its optimal combination of the sub channel as well as the wireless channel with the main channel.

4.3 Impact of Access Pattern

This experiment evaluates client access patterns, by varying the skewness parameter θ of Zipf distribution from zero to one.

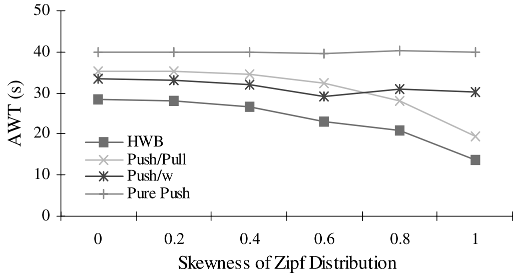

Fig. 7 Waiting time vs. access pattern.

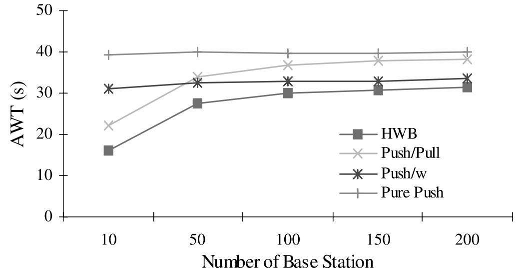

Fig. 9 Waiting time vs. number of BS.

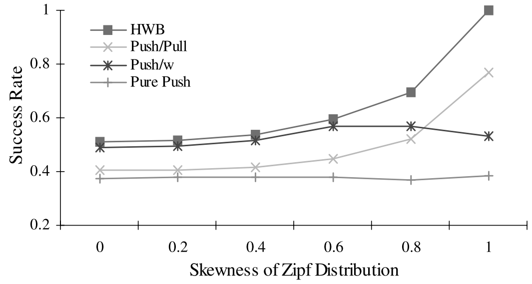

Fig. 8 Success rate vs. access pattern.

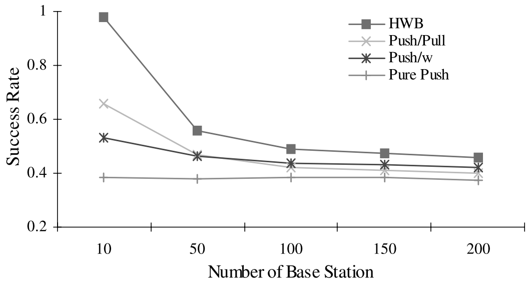

Fig. 10 Success rate vs. number of BS.

As shown in **Fig. 7** and **Fig. 8**, the skewed access patterns ($\theta > 0$), achieves much more performance improvement than the uniform access. Moreover, the performances for all the hybrid approaches upgrade as the skewness increases, since clients' queries become increasingly skewed, and the skewed access is helpful to make more efficient utilization of the shared sub channel, as well as improve the cache hits of the base station. However, when $\theta > 0.6$, the performance of the push/w approach cannot improve any more, but to some extend decreases. The reason is that in this case although the clients' queries become rather skewed and the most of the requested items are kept in the base station cache, the responses via the point-to-point wireless channel are limited by the narrowband bandwidth. In contrast, there is a large performance improvement for the push/pull approach and the HWB approach. Especially, across the entire region of the evaluation, the HWB approach has the best performance among all the approaches by benefiting from the optimal combination of the three data delivery channels.

Mobile users normally have a biased access interest, and thus we pay more attention on the skewed access pattern. The following experiments are performed under the skewed access with the default skewness $\theta = 0.95$. Additionally, we focus on comparing the proposed HWB approach with the other two hybrid data dissemination methods, i.e., the push/pull approach and the push/w approach. To provide a comparison baseline, the pure push approach is also included in the figures.

4.4 Impact of BS Scalability

To examine the scalability of the HWB system, we vary the number of the base station from 10 to 200. In this set of experiments, the database size is 10,000, and thus the maximum size of data group is 1000 for 10 base stations, and the minimum size of the data group is 50 for 200 base stations, which is enough to evaluate the system scalability. In addition, each base station has a same query frequency with the average query interval of 500 ms. Accordingly, the query load of the system increases as the base stations increase.

As shown in **Fig. 9** and **Fig. 10**, the performances for all hybrid approaches decline as the number of the base station increases. The push/pull approach even performs worse than the push/w approach when the base stations are more than 50. The reason is that as the base stations increase largely, the query load of the system increases markedly, moreover, the amount of the heterogeneous access also gets larger. As a result, the only one on-demand sub channel cannot effective serve so many requests with a dissimilar local access interest; however, it can be better served by the increasing num-

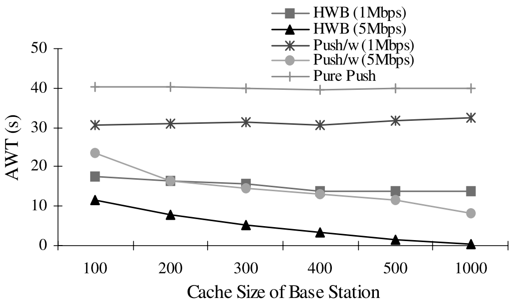

Fig. 11 Waiting time vs. cache size.

ber of the wireless channel of the base station. On the other hand, the performances drop a lot as the base stations increase though; the HWB approach still performs the best, which reveals that the HWB approach has a good scalability.

Additionally, we conduct experiments to evaluate the impact of the simultaneous queries. The result show that the performance of the HWB approach falls a little as the number of simultaneous queries increases, but it still outperforms other approaches. Due to the limitation of space, the simulation results are not shown in this paper.

4.5 Impact of BS Cache Size

By varying the cache size of the base station from 100 to 1,000, we compare the HWB approach with the push/w approach under two bandwidths of the wireless channel set to 1 Mbps and 5 Mbps, respectively. **Figure 11** shows that the performance of the push/w approach with 1 Mbps wireless channel has a slight decrease as the cache size gets larger, but there is a much improvement under 5 Mbps bandwidth. The reason is that the cache hits increases with the cache size, but the narrow-band bandwidth limits the effective use of the wireless channel. On the other hand, as the increase of the cache size, the HWB approach with 5 Mbps wireless channel has the greatest performance improvement, meanwhile there is also a slight increasing under 1 Mbps wireless channel. All these behaviors reveal that the HWB approach has a more flexible processing by taking advantage of the three complementary data delivery ways.

4.6 Impact of Pull Bandwidth

The evaluation about pull bandwidth is examined by varying the bandwidth of the sub channel and the wireless channel, respectively.

As indicated in **Fig. 12**, the waiting times for the HWB approach, the push/pull approach and the pure pull approach decrease as the

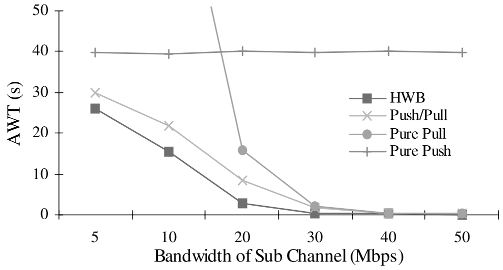

Fig. 12 Waiting time vs. bandwidth of sub channel.

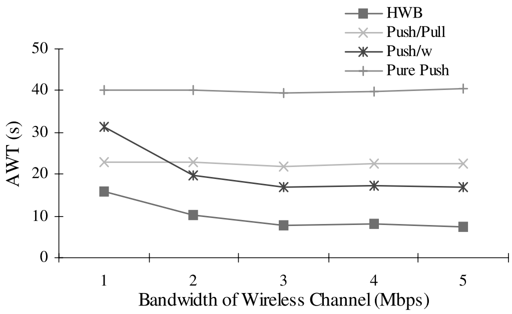

Fig. 13 Waiting time vs. bandwidth of wireless.

bandwidth of the sub channel increases, and the HWB approach always performs the best. However, when the bandwidth is larger than 30 Mbps, the performance difference among these approaches is very little. This is because in this case the bandwidth of the sub channel is huge enough to respond to almost all the requests.

On the other hand, **Fig. 13** shows that the average waiting times of the HWB approach and the push/w approach improve as the bandwidth of the wireless channel increases. When the bandwidth of the wireless channel is larger than 2 Mbps, the push/w approach even performs better than the push/pull approach, but its performance is still far lower than the HWB approach. The reason is that each base station holds its own cache and utilizes the wireless channel to serve the local clients, whereas the sub channel is only one, which is shared by all the clients in a large scope. By making effective use of the on-demand wireless channel and the on-demand sub channel, the HWB approach has the best performance.

4.7 Summary of Experiments

We conclude the above experimental results that the proposed HWB approach outperforms the other approaches with the shortest average waiting time and the highest success rate under the skewed access pattern as well as uniform access pattern. It is clear that the skewed access pattern dramatically improves the efficient utilization of the sub channel, and also increases the cache hits of the base station. Hence, the HWB approach performs even better by benefiting from the optimal combination of the on-demand sub channel as well as the on-demand wireless channel with the push-based main channel. Moreover, the HWB approach performs the best, even when the system workload increases greatly; meanwhile the HWB approach has a good scalability. Additionally, 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, cannot be separated. It is also indicated that the push/pull approach and the push/w approach always perform better than the pure push approach and the pure pull approach; however, both of them perform much more poorly than the HWB approach.

5. Related Work

The idea of employing broadcast for information dissemination has been studied for more than a decade. Some studies focused on the push-based broadcast [1),2)] or the pull based broadcast [3),13)]. Some other studies enhancing the push-based broadcast, interleaved the on-demand responses on the broadcast channel, and discussed the trade off between the push-based broadcast and the pull-based broadcast [4),5),7),14)]. In these approaches, similarly to our evaluated push/pull approach, clients explicitly request data through the point-to-point uplink, and the responses are transmitted through the push-based broadcast or the on-demand broadcast. Additionally, in contrast to adopting a shared on-demand broadcast, several researches integrated the push-based broadcast with the on-demand wireless communication by using a set of point-to-point on-demand channel [8),11)]. To some extend, their mechanism is similar to our examined push/w approach.

These researches confirmed that integrating push-based broadcast with on-demand broadcast or with on-demand wireless communication both achieved a considerable performance improvement. This result was also confirmed in our experiments. However, as the recent advances in wireless technologies and mobile devices, it is necessary to construct a more effective hybrid data delivery model to further improve the performance of the information dissemination.

Some studies discussed constructing a hybrid network based on the wireless data broadcast [6),8),11),15)]. They normally assumed that the base station takes charge of the two channels respectively for the push-based broadcast and the pull-based unicast communication. Their assumed hybrid system is completely different with the emerging mobile digital broadcasting, in which a broadcast server broadcasts information to the clients in a large scope. Moreover, in their assumed hybrid network, the two data delivery ways have the same service range, i.e., both limited in the service region of the base station. Additionally, only one kind of pull-based data dissemination, i.e., the on-demand point-to-point wireless communication, is integrated with the push-based broadcast without a shared on-demand broadcast. Therefore, it is hard to make an efficient use of the broadcast to provide a flexible information service. For example, if the clients in a different region have a common interest on some popular information, like the report of the president election or the sports scores of Olympic Games, each base station has to individually broadcast the same information.

To our best knowledge, there is no study constructing a hybrid data delivery by effectively taking advantage of the most recent emerging mobile/wireless communication. Also there is no study considering the combination of the push-based broadcast, pull-based broadcast, and pull-based wireless communication to provide a flexible and complementary service in different service ranges.

6. Conclusion

In this paper, we put forward a new hybrid data delivery model, i.e., Hybrid Wireless Broadcast (HWB) model, by taking advantage of the recent advances of the wireless technologies and mobile terminals. The proposed HWB model combines push-based and pull-based broadcast and pull-based point-to-point wireless communication, and provides a flexible and complementary information service in different bandwidths and service ranges to greatly improve the responsibility, scalability, and efficiency of the system. Based on the HWB model, furthermore, we proposed an effective query processing method, and developed a simulation model for examining the performance of the proposed approach. The experimental results confirmed that the HWB approach significantly enhances the performance of the system, and has a good scalability.

Based on the effective HWB model, there is a lot of interesting work to be done in the near future. To explore a more significant query processing, the optimal selection among the three data delivery ways of the HWB model will not only consider the waiting time, but also take into account the type of transmitted data and the response cost of each data delivery way. In addition, we will investigate an effective broadcast scheduling and cache management of the HWB model, and construct an integrated control of scheduling and caching. Furthermore, we like to extend our study on the dynamic control of the HWB system, and the issue of data update.

Acknowledgments This research was supported in part by Grant-in-Aid for Scientific Research (B) (15300033) and (18049050) 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) Acharya, S., Alonso, R., Franklin, M. and Zdonik, S.: Broadcast disks: Data management for asymmetric communication environments, *Proc. ACM SIGMOD'95*, pp.199–210 (1995).
2) Acharya, S., Franklin, M. and Zdonik, S.: Dissemination-based data delivery using broadcast disks, *IEEE Personal Communications*, Vol.2, No.6, pp.50–60 (1995).
3) Aksoy, D. and Franklin, M.: RxW: A scheduling approach for large-scale on-demand data broadcast, *IEEE/ACM Trans. Networking*, Vol.7, No.6, pp.846–860 (1999).
4) Acharya, S., Franklin, M. and Zdonik, S.: Balancing push and pull for data broadcast, *Proc. ACM SIGMOD'97*, pp.183–194 (1997).
5) Beaver, J., Chrysanthis, P.K., Pruhs, K. and Liberatore, V.: To broadcast push or not and what?, *Proc. MDM'06*, pp.40–45 (2006).
6) Buchholz, S., Schill, A. and Ziegert, T.: A simulation study of update techniques for cyclic data broadcast, *Proc. MSWiM*, pp.115–122 (2001).
7) Guo, Y., Pinotti, M.C. and Das, S.K.: A new hybrid broadcast scheduling algorithm for asymmetric communication systems, *ACM SIGMOBILE Mobile Computing and Communications Review*, Vol.5, No.3, pp.39–54 (2001).
8) Hu, Q.L., Lee, D.L. and Lee, W.-C.: Performance evaluation of a wireless hierarchical data dissemination system, *Proc. Mobicom'99*, pp.163–173 (1999).
9) Imielinski, T., Viswanathan, S. and Badrinath, B.: Energy efficient indexing on air, *Proc. ACM SIGMOD'94*, pp.25–36 (1994).
10) Jung, I., You, Y., Lee, J. and Kim, K.: Broadcasting and caching policies for location-dependent queries in urban areas, *Proc. WMC'02*, pp.54–60 (2002).
11) Lin, C.W., Hu, H. and Lee, D.L.: Adaptive realtime bandwidth allocation for wireless data delivery, *Wireless Networks*, Vol.10, pp.103–120 (2004).
12) Ren, Q. and Dunham, M.H.: Using semantic caching to manage location dependent data in mobile computing, *Proc. Mobicom 2000*, pp.210–221 (2000).
13) Sun, W., Shi, W., Shi, B. and Yu, Y.: A cost-efficient scheduling algorithm of on-demand broadcasts, *Wireless Networks*, Vol.9, pp.239–247 (2003).
14) Stathatos, K., Roussopoulos, N. and Baras, J.S.: Adaptive data broadcast in hybrid networks, *Proc. VLDB'97*, pp.326–335 (1997).
15) Vlajic, N., Charalambous, C.D. and Makrakis, D.: Generalized wireless data broadcast system analysis: multiobjective performance measures, *Proc. IASTED*, pp.229–234 (2002).
16) Zheng, B. and Lee, D.L.: Information dissemination via wireless broadcast, *Comm. ACM*, Vol.48, No.5, pp.105–110 (2005).
17) Zheng, B., Lee, W.C. and Lee, D.L.: On semantic caching and query scheduling for mobile nearest-neighbor search, *Wireless Networks*, Vol.10, pp.653–664 (2004).
18) Zhang, J. and Gruenwald, L.: Prioritized sequencing for efficient query on broadcast geographical information in mobile computing, *Proc. GIS'02*, pp.88–93 (2002).
19) http://www.au.kddi.com/seihin/kinobetsu/
20) http://www.nttdocomo.co.jp/product/
21) http://www.vodafone.jp/japanese/products/
22) http://www.mobaho.com/index.html

(Received June 19, 2006)
(Accepted August 9, 2006)

(Editor in Charge: *You Shiraishi*)

Jing Cai is currently pursuing Ph.D. in Graduate School of Information Science and Technology, Osaka University, Japan. She received B.E. and M.E. degrees from Dalian Maritime University and Wuhan University of Technology, China, in 1990 and 1993, repectively. Her research interests include broadcasting computing, information system access control, and e-Learning. She is a student member of IPSJ and DBSJ.

Tsutomu Terada is a lecturer at the Cybermedia Center of Osaka University. He received B.E., M.E., and Ph.D. degrees from Osaka University in 1997, 1999, and 2003, respectively. Since 2005 he is Assistant Professor at the same center. Dr. Terada has previously worked on active database technologies, and he is currently investigating the application of active database techniques in the context of smart object systems. He is a member of IEEE, IEICE, and IPSJ.

Takahiro Hara received his B.E., M.E., and D.E. degrees in Information Systems Engineering from Osaka University, Osaka, Japan, in 1995, 1997, and 2000, respectively. Currently, he is a Associate Professor of the Department of Multimedia Engineering, Osaka University. His research interests include distributed database systems in advanced computer networks such as high-speed networks and mobile computing environments. Dr. Hara is a member of four learned societies, including ACM, IEEE, IEICE, and IPSJ.

Shojiro Nishio received his B.E., M.E., and Dr.E. degrees from Kyoto University, Japan, in 1975, 1977 and 1980, respectively. From 1980 to 1988 he was with the Department of Applied Mathematics and Physics of Kyoto University. In October 1988, he joined the faculty of the Department of Information and Computer Sciences, Osaka University, Japan. In August 1992, he became a full professor in the Department of Information Systems Engineering of Osaka University. Since April 2002, he has been a full professor in the Department of Multimedia Engineering of Osaka University. From April 2000 to July 2003, he served as the founding director of Cybermedia Center of Osaka University. Since August 2003, he has been serving as the dean of the Graduate School of Information Science and Technology of Osaka University. His current research interests include database systems and multimedia systems. Dr. Nishio has served on the Editorial Board of *IEEE Transaction on Knowledge and Data Engineering*, *ACM Transactions on Internet Technology*, and is currently involved in the editorial board of *Data and Knowledge Engineering*. He is a fellow of IEICE and IPSJ, and he is a member of eight learned societies, including ACM and IEEE.