DATA ORGANIZATION FOR DATA BROADCASTING IN MOBILE COMPUTING

by

ROZAIDA BINTI MOHD DARUS

Thesis submitted in fulfillment of the requirements for the degree of Master of Science

NOVEMBER 2006

ACKNOWLEDGEMENTS

Alhamdulillah thank to Almighty Allah that gave me the guidance and patience to finish this thesis. Firstly, I would like to express my deepest gratitude and appreciation to my supervisor, Dr. Putra Sumari for his invaluable help and guidance. His moral support and encouragement throughout the period of conducting this project has kept me to continue the project till the end.

My special thanks go to Department of Civil Services (JPA), for having selected me to pursue my studies under Scholarship of Training in Services, which provides me with the financial security during the period of my post graduate study.

Last but not least, a special thank you to my parents, my husband, Khairil Azhar Kassim and also my three lovely kids, Iman Hakimi, Iman Hairie and Iman Haris for their support and understanding.

TABLE OF CONTENTS

CHAPTER 6 - CONCLUSIONS AND FUTURE WORK

LIST OF TABLES

LIST OF FIGURES

PENYUSUNAN DATA BAGI PENYIARAN DALAM PENGGUNAAN TEKNOLOGI TANPA WAYAR

ABSTRAK

Peningkatan penggunaan di dalam teknologi tanpa wayar, membenarkan data atau maklumat dicapai oleh pengguna pada bila-bila masa dan di mana sahaja. Melalui teknik penyiaran, keupayaan pelayan penyiaran merupakan komponen yang penting dalam penyebaran maklumat. Data disiarkan ke udara contohnya, melalui satelit sementara pengguna akan menunggu dan terus mencapai data yang sampai kepadanya.

Terdapat berbagai pendekatan digunakan dalam penyiaran maklumat di mana salah satu daripadanya adalah melalui kaedah cakera peyiaran (Chang & Yang, 2000; Acharya, 1998). Algoritma yang dihasilkan daripada kaedah cakera penyiaran ini adalah tertumpu kepada penyusunan data secara berjujukan. Data-data akan disusun mengikut kepentingan serta kepopularan data yang diperlukan oleh pengguna-pengguna teknologi tanpa wayar ini. Data-data yang mempunyai permintaan yang tinggi akan dimasukkan ke dalam cakera yang panas manakala data-data yang kurang permintaan akan diletakkan ke dalam cakera yang kurang panas atau dikenali sebagai cakera sejuk. Berdasarkan kepada pendekatan yang digunakan dalam cakera penyiaran ini, wujud slot-slot kosong dalam persekitan penyiaran tersebut. Slot-slot kosong tersebut juga dikenali sebagai slot yang tidak berfungsi. Slot kosong yang wujud ini menyebabkan pembaziran terhadap lebar jalur yang digunakan dalam persekitaran penyiaran. Ini akan mengakibatkan masa menunggu untuk mencapai sesuatu data yang diperlukan adalah agak lama. Oleh itu satu pendekatan baru telah diperkenalkan bagi mengatasi masalah slot kosong ini. Slot-slot yang kosong akan dihapuskan digantikan dengan slot-slot yang telah disusun di dalam cakera mengikut

kepentingan capaiannya. Daripada kajian kes serta simulasi yang digunakan, didapati pendekatan baru yang diperkenalkan ini dapat membantu mengoptimumkan penggunaan jalur lebar, di samping mengurangkan masa menunggu dalam mencapai data oleh pengguna.

DATA ORGANIZATION FOR DATA BROADCASTING IN MOBILE COMPUTING

ABSTRACT

The advances in mobile devices and wireless communication techniques have enabled anywhere, anytime data access. In Broadcast Disks, one of the important components of this environment is the server with broadcasting capabilities. Data are fetched from the server and being broadcasted to the air. Mobile users keep listening to the air and catch the data that they maybe interested.

There are many strategies proposed for efficient broadcast delivery such as disk broadcasting (Chang & Yang, 2002; Acharya, 1998). Disk broadcasting approach has developed an algorithm which arrange the sequence order of data to be broadcasted. This generate the sequence of data with the policy where data that frequently accessed are called "hot" and less frequently access are called "cold". However, based on disk broadcasting approach, some broadcast sloted in the broadcast are empty in which we referred as unused slots. The unused slots are considered as a waste of bandwidth thus increasing the client's waiting time. Therefore, we propose a new data organization for the server which is to eliminate the unused slots in the broadcast and will fill them with data selected according to their popularity. We show by case study and simulation that our data organization scheme has resulted bandwidth utilization and reducing the clients' average waiting time.

CHAPTER 1
INTRODUCTION

1.1 Background

The emergence of powerful portable computers, along with advances in wireless communication technologies, has made mobile computing a reality. In the evolving field of mobile computing there is a growing concern to provide mobile users with timely access to a large amount of information (Barbara, 1999; Dunham, 1998) in which known as mobile information system. Several existing mobile information system applications such as whether broadcast system, highway condition monitoring system, traffic direction system, news and stock queries information system, are some of examples in this field. In this kind of applications, a large amount of data has to be broadcasted to a large number of mobile clients. The server in which equips with broadcasting facilities is responsible to provide data to a large number of mobile clients when requested. This is indeed a tricky task mainly due to the fact that wireless has a very limited bandwidth. Bandwidth is a very valuable resource in mobile broadcasting. Idle bandwidth in which refer to the channel with no data is considered a waste and should be avoided. Fully utilize the bandwidth resources is an important criteria in this field (Alonso & Korth, 1992).

Another important criterion in mobile information system is the average waiting time for clients to receive information from the server. Information must arrive at the clients mobile devices within acceptable delay (Acharya, Franklin & Zdonik, 1996). A long waiting time would make the mobile information become unattractive. The issues of bandwidth utilization and waiting time can be improved by having a good data organization in the server and its buffer. Data in server site should be organized in intelligent manner so that they can be grouped together to fully utilize the bandwidth capacity. Furthermore the organization of data in term of its order and duplication can influence the client waiting time. Therefore this thesis has concern itself with the data organization within the server site in mobile broadcasting infrastructure to improve the bandwidth utilization and the client's average waiting time.

In this chapter, we first elaborated the broadcasting system for mobile information system specifically the infrastructure and components. We narrow the elaboration into the issue of bandwidth utilization and client's average waiting time. Finally, follow by thesis objectives, contribution and organization.

1.2 Data Broadcasting

Mobile information system can be characterized by massive numbers of mobile users requested for information (Xuan, Sen, Gonzalez, Fernandez & Ramamritham, 1997). Figure 1.1 illustrates a scenario of mobile information system. The system consists of server, transmitter, broadcast channel and mobile clients.

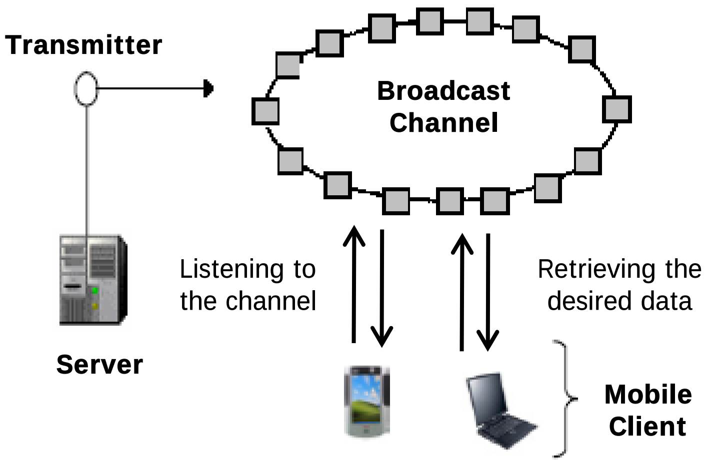


Figure 1.1: Data broadcasting components

The server, in which data are resided are being broadcasted to the mobile clients. The transmitter is simply kind of broadcasting equipment that turning the data into signal form and send to the broadcast channel. Data fly within broadcast channel and waiting to be caught and filtered by clients.

Asymmetric communication is one of the well-known technologies for broadcasting strategy. Asymmetry in the communication means that the bandwidth in the downstream (server to clients) is much greater than that in the upstream (client to server) direction. In other word it is almost a one way direction of data from the server to client. The server will continuously and repeatedly broadcast in cyclic manner to client community. There are several advantages of having asymmetric in communication. First, is scalability: it is independent of the number of users the server is serving (Chang & Yang 2002; Acharya, 1998). In the simplest scenario, given an indication of the data item that are desired by each client listening to the broadcast, the server would simply accumulate those requests and broadcast the resulting set of data items cyclically. This solution is efficient since it eliminates the need to send the same item multiple times when each item is used by many clients.

Second, it allows the server to broadcast different items with differing frequency (Barbara, 1999; Acharya, Franklin, Zdonik & Alonso, 1995). In other words the server can emphasize to broadcast the most popular items and de-emphasize the less popular ones. Third, saving power consumption of clients' device, that is by allowing transmitting request to server can degrade the power usage.

1.3 Bandwidth Utilization

Wireless networks are more expensive, offer less bandwidth, and are less reliable than wire line networks. The effective bandwidth of wireless networks is only a fraction of the bandwidth that is available in wired networks. The current ATM (Asynchronous Transfer Mode) standards are designed to yield a bandwidth of up to 622 Mbps. This bandwidth is projected to go up to gigabits (Partridge, 1993). Even the widely used Ethernet has a bandwidth of 10 Mbps. Compared to the wired bandwidth; the bandwidth of a wireless channel is very limited. It varies from 1.2 Kbps for slow paging channels, through 19.2 Kbps (characteristic of cellular proposals like Cellular Digital Packet Data (CDPD)) to about 2 Mbps of the wireless LAN. Therefore, bandwidth utilization is of vital importance in a wireless network.

Such applications typically involve a small number of servers and a much larger number of clients with similar interests. Examples include stock trading, electronic commerce applications, such as auction and electronic tendering, and traffic control information systems. The number of clients can grow dramatically as more clients can join and monitor the broadcast channel at anytime they want. Thus, data need to properly organize to cater to the needs of the client within a limited bandwidth. There have been many strategies proposed for an intelligent data organization for broadcast delivery (Jing, Helal & Elmagarmid, 1999).

Among those works is disk broadcasting method known as broadcast disk scheme. Based on the broadcast disk scheme, different hierarchies of slots are formed. The highest levels of hierarchy will be filled by hot data/pages and to be broadcasted more frequent whereas low hierarchy will be filled by cold data/pages and to be broadcasted less frequent. However, based on disk broadcasting approach, at a certain condition some slots in the hierarchy are empty in which we referred as unused slot. The unused slots are considered as a waste of bandwidth. As bandwidth is an expensive resources therefore to fully utilize the bandwidth particularly the empty slots is important.

1.4 Clients Waiting Time

A client waiting time is referred as the time of waiting between once clients' submitted request and requests being served to clients. Clients waiting time is important criteria that make the mobile information system usable and reliable. The broadcasting system is an asymmetric type of delivery in which data are broadcasted in one way direction that is only from the server to the clients with no opposite direction. The server keeps repeatedly broadcast data to the channel in cyclic manner. Once the pages are being broadcasted from the server, the clients listen to those data and downloaded the one they interest (Acharya, Franklin, Zdonik & Alonso, 1995). This approach has resulted that client has to wait (listen mode) for some time until data are passed to them. Common approach in reducing waiting time is to know in advance the pattern of clients catching data where, what those data are most catched by clients and what data are less catched by clients then by broadcasting those most wanted compared to those less wanted will reduced the clients' average waiting time.

One approach in tackling waiting time issue is to have good data organization for data broadcasting. Data are organized into a proper sequence prior being broadcasted. The existence of hot data in the sequence must ascend in term of number to those cold data. As more hot data are in the sequence then more clients can catch. Disk broadcasting has fulfill this concept in which data hierarchy is established where the higher level contain slots that fill by most wanted data and being broadcast frequently whereas the low hierarchy is the opposite way. As mention in previous section there are empty slots occurred in the hierarchy. These empty slots if filled by most wanted data will reduce the clients' average waiting time.

1.5 Thesis Objectives

In mobile computing environment, the available bandwidth is quite low, user devices are typically portable computers with less processor speed, memory size, disks capacity and battery lifetime. One should consider these constrains particularly in our work we concentrated on bandwidth issue in order to apply broadcast technology to mobile computing.

The objectives of this research are:

i. To utilize the bandwidth of wireless channel particularly from the server to client. Bandwidth is an expensive resource in mobile information system. It should be fully utilize for maximum performance.

ii. To reduce the client average waiting time in mobile information environment. Average waiting time is an important aspect in mobile information system as it makes the field attractive.

1.6 Thesis Contribution

In this thesis we present a new data organization for the broadcast server. The proposed scheme is an enhancement of existing disk broadcasting scheme. The main contribution of this thesis is as new data organization that eliminates the unused slots introduced by broadcast scheme. Our data organization scheme has resulted bandwidth optimization and reducing the clients average waiting time.

We have in order to cater the needs of the clients by making an effective use of the low bandwidth; we proposed a new scheme of data organization for mobile computing. The data are not always homogeneous and clients sometime are more interested in particular data elements. Therefore some data, more frequently accessed, are called “hot” and the other data, less frequently accessed, are called “cold”. To deal with this kind of data the idea of Broadcast Disks was introduced (Barbara, 1999; Acharya, 1998). Here, the broadcast organized as a set of disks with different speeds. “Hot” data are placed on the fast disk and the “cold” data are placed on the slow disk. However, based on disk broadcasting method, some broadcast slots in broadcast may be unused, which result in the waste of bandwidth thus increase the waiting time.

Therefore, if most of the data that client needs are “hot”, the client waiting time will reduce. Our proposed method is to eliminate the unused slots that occur in the broadcast and fill them with data selected according to their group of popularity. Besides, we experimentally show that our proposed method utilizes the bandwidth efficiently and reduces the waiting time of users than the disk broadcasting method.

1.7 Thesis Organization

This thesis is organized into six chapters. The contents are arranged such that each previous chapter provides a basic idea to further proceed to the next chapter.

Chapter 1 introduces the general concept and benefits of data broadcasting in mobile information system. Then we look at the issues of bandwidth utilization and client average waiting time. Finally we outline the thesis objectives and it contribution and also thesis organization.

In Chapter 2, we present a literature survey of research work closely related to this thesis. This chapter introduces basic background knowledge of organizing data broadcasting in mobile computing. We discuss the characteristics of broadcast data delivery model and survey the broadcasting scheme.

Chapter 3 describes the existing work called Broadcast Disks system. This chapter elaborates in detail the Broadcast Disks environment, concept and its data broadcasting algorithm. This broadcasting algorithm is responsible to organize data item into sequence order before broadcast to clients. We discuss on two main issues arise from Broadcast Disk method which are bandwidth utilization and client waiting time.

Chapter 4 presents our proposed methodology and approach of data organization for mobile broadcasting. The proposed work is an enhancement of disk broadcasting method known as 'Broadcast Disks' architecture. Based on disk broadcasting approach, some broadcast slots may be unused refer as empty slot phenomena which result in the waste of bandwidth and the increase of client access time. In this chapter, we present a new scheme that solves the empty slots phenomenon.

In Chapter 5, we present the performance of our proposed scheme for broadcast disk in mobile environment. This evaluation is done by comparing our proposed method and disk broadcasting method which is divided into case study and simulation. The case study is performed to measure the percentage of empty slots of both methods while the simulation is carried out due to the performance of client's average waiting time in the broadcast.

In Chapter 6, we summarize the scope of works along with the conclusion or our research work; the discussions of the results we obtain in the previous chapter are presented. Finally, a discussion and suggestion for future work pertaining to this research is proposed.

CHAPTER 2
SURVEY OF RELATED STUDIES

This chapter introduces basic background knowledge of organizing data broadcasting in mobile computing. We discuss the characteristics of broadcast data delivery model and survey the broadcasting scheme.

2.1 Broadcast Delivery

Broadcast delivery is used in most wireless systems to disseminate data items to all mobile elements inside cell. This form of data delivery is different from the one in traditional client/server systems. In this section, we look at mobile broadcasting characteristics.

2.1.1 Asymmetry Communication

In developing traditional computing systems, designers have made an implicit assumption about the nature of the communication medium. It has always been assumed that the communication capacity (bandwidth) available to each machine to receive data from the network (downstream channel). This logically is followed because machines on the network typically shared the same physical medium (wire) for both the uplink and downlink communication.

However, recently a number of new network technologies have come to the fore which has the property of asymmetric. In a client-server mobile framework, bandwidth asymmetry occurs in an environment in which the effective bandwidth from the servers to the client on the downstream channel is significantly higher than the bandwidth available in the reverse direction on the upstream connection. Examples of such environments include satellite and cable television networks. Asymmetry can also arise for reasons other than bandwidth. The various types of asymmetric communication that are possible are as follows:

i. Network Asymmetry

In many networks the bandwidth of the communication channel from the server to the clients is much greater than the bandwidth in the opposite direction. For example, satellite networks often require the use of a wired channel (such as phone network) for the uplink channel. Table 2.1 lists some asymmetric technologies and the bandwidths they provide in both directions. As can be seen in the table, there is at least an order of magnitude difference between the downstream and the upstream bandwidth.

Table 2.1: Some Examples of Bandwidth Asymmetry

Network Type	Example Technology	Downstream Bandwidth	Upstream Bandwidth
Satellite	DirectPC	400kbps	56.6 Kbps (Via Telephone)
Cable Television	Cable Modems	10-30 Mbps	128 Kbps (Shared)
Telephone	ADSL, VDSL	2-6 Mbps	9.6-640 Kbps
Wireless	WiFi, 3G	1-10 Mbps	< 10 Mbps (Shared)

ii. Slow Uplink

Such systems are designed to deliver data from a few servers to a large number of mobile clients, but there is significantly more “downlink” bandwidth from servers to clients than in the opposite or “uplink” direction. Also, clients have low power reserves, while servers have plenty of power.

iii. More Client, Less server

A system with a small number of servers with larger number of clients also results in asymmetry. For example, since a server has fixed resources and data processing capacity must be divided among all of clients faces such an asymmetry.

iv. Updates and new information

Asymmetry can also arise in an environment where newly created items or updates existing data items must be disseminated to clients. In such cases, there is a natural (asymmetric) flow of data in the downstream direction.

As mentioned above, asymmetry can arise not only due to the properties of the communications network, but can arise even in a symmetric environment due to the nature of the data flow in the application.

2.1.2 Pull / Push-based Broadcast

Basically, there are two basic architectures for a broadcast delivery system: pull-based broadcast and push-based broadcast. In push-based broadcast delivery as Figure 2.1, the data items are sent from the server to the clients without requiring a specific requires from the clients. The clients just listen on the common channel and filter the data items they need. In this case, the profile of the clients which describe clients access pattern are sent beforehand or the client interests are defined or already known to the server. The server usually broadcast data items periodically.

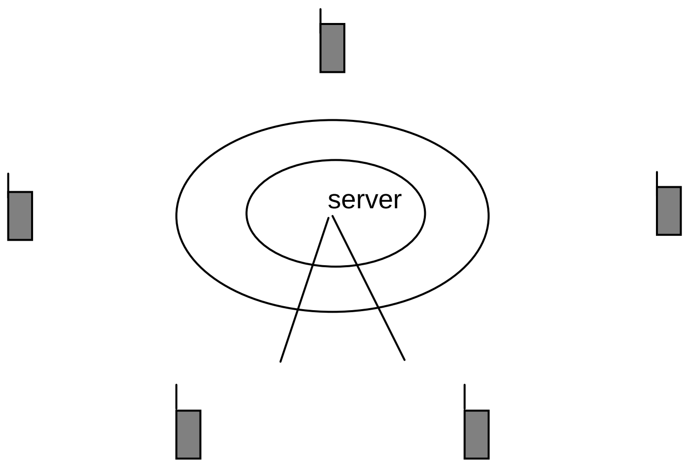


Figure 2.1: The Push-based Broadcast Delivery

In pull-based broadcast delivery as in Figure 2.2, the clients use their uplink channel to make requests and then listen on broadcast. The server groups the responses and broadcasts them on downlink channel. This kind of data delivery is also called on-demand broadcast, since it is demand-driven. One example is large-scale on-demand broadcast using satellite for broadcasting server's data items (Franklin & Zdonik, 1998)].

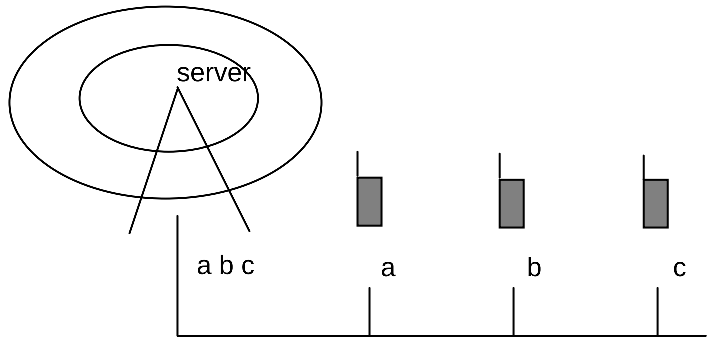


Figure 2.2: The Pull-based Broadcast Delivery

By combining these two broadcast data delivery architectures, we obtain a variation. The data items which are commonly well-known interests to the clients are sent for the server to clients on the downlink channel without requiring any specific request. In the mean time, the server also sends some specific responses on the broadcast channel which are requested specifically by the clients via backchannels. In other words, the data items are only interests of particular clients on the broadcast channel. In (Stathatos, Rousopoulos & Baras, 1997) another slightly different model is introduced. In this model, the broadcast delivery is used for the push program but the pull responses will go only to the client that made the request.

The data delivery in a traditional client/server system is pull-based where a client makes an explicit request to the server. The server receives the request, locates the information of interest and returns it to the client. And in most case point-to-point connection is employed, the communication between the server and a client is dedicated. However, in mobile broadcasting it mainly suffers two problems:

i. Backchannel congestion

In many environments, the backchannel (channel from the client to the server) is limited in bandwidth and can easily become a bottleneck.

ii. Server saturation

The server must be continually interrupted to react to the client backchannel requests. Thus as the number of clients increase, eventually the server will become saturated.

In contrast, mobile broadcasting data delivery is usually push-based. In push-based data delivery, the server repetitively broadcasts data to a client population without a specific request. Clients monitor the broadcast and filter what they need. It has three fold advantages which are:

i. Efficiency

Data is transferred only as need as when new data is created or if updates take place. Thus client and server processing are minimized which can eventually lead to better performance.

ii. Scalability

It is independent of the number of users the system is serving (Delis & Rousopoulos, 1992). Moreover, once data items are broadcast, the entire clients in the coverage receive them simultaneously. Thus the client's access time is reduced.

iii. Lower bandwidth utilization

By avoiding unnecessary data transfer, both the uplink and downlink bandwidth is saved. This is particularly useful for asymmetric environments.

2.1.3 Broadcast Program

Once a broadcast program is set up, any number of clients can join the system without impacting the performance of other clients. From the server's viewpoint there are two options for the "programs" that can be generated which are:

i. Periodic Broadcast Program

In the periodic case, the server would broadcast the same program repeatedly in a cyclic fashion. An example of periodic dissemination is a stock market monitor which cycles though all the stock values. Periodic program has the advantage of allowing clients to disconnect for certain periods and still not miss out items (since they will be broadcasted again after the client is reconnected). The different styles of periodic broadcast are as follows:

- **Static or Dynamic**

  A static broadcast is where the schedule of program is fixed, and even though the contents of a program can change with time, direct feedback from the users is not supported. An example of a static program is a traditional weather information service. In such a service, the contents of program change between broadcast in order to reflect the latest climate conditions. In contrast, in dynamic broadcast, both the schedule of programs and its contents can change and there exists limited support to handle user's requests.

- **Regular or Irregular**

  A program is defined to be regular if the inter-arrival time between two consecutive occurrences of a page is always the same otherwise it is irregular.

ii. Aperiodic Broadcast Program

Television broadcast, on the other hand is an example of aperiodic dissemination since there does not exist any repeatability in the schedule. However, there are some programs which follow a periodic schedule. For example, the nightly news programs are usually broadcast at the same time each night.

2.2 Flat Broadcast

In the simplest scenario, given an indication of the data items that are desired by each client listening to the broadcast, the server would simply take the union of the requests and broadcast the resulting set of data items cyclically. Figure 2.3 illustrates this scheme, where 6 pages (indicate by A, B, C, D, E and F) are broadcast cyclically in a wireless environment.

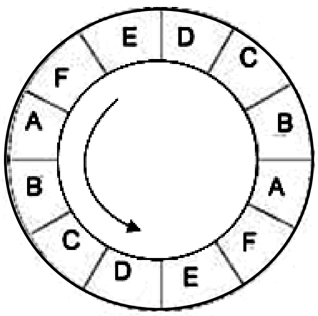


Figure 2-3: A Flat Data Organization

When an application running on a client needs a data items, it first attempts to retrieve the item from its local memory of disk (if it has). If the desired item is not found, then the client monitors the broadcast channel and waits for the item to arrive. In this scenario, the expected waiting time for an item on the broadcast is same for all items (namely, half a broadcast period on the average) regardless of their relative importance to the clients. This “flat” approach has been adopted in earlier work on broadcast-based database systems such as Datacycle.

2.3 Datacycle

The Datacycle project (Herman, Gopal, Lee & Weinrib, 1994; Herman et. al., 1987) at Bellcore was the first push-based approach in building a database system. Datacycle broadcast data using a flat disk approach (Bowen, 1992). Datacycle provided all the functionality of a traditional DBMS including query processing, transaction management and also recovery. In Datacycle, a database circulates on a high bandwidth network (140 Mbps) and users query this data by filtering relevant information using a special massively parallel transceiver capable of filtering up to 2,000,000 predicates a second. The main differences between wireless broadcasting considered in this paper and broadcasting considered in the Datacycle architecture are as follows:

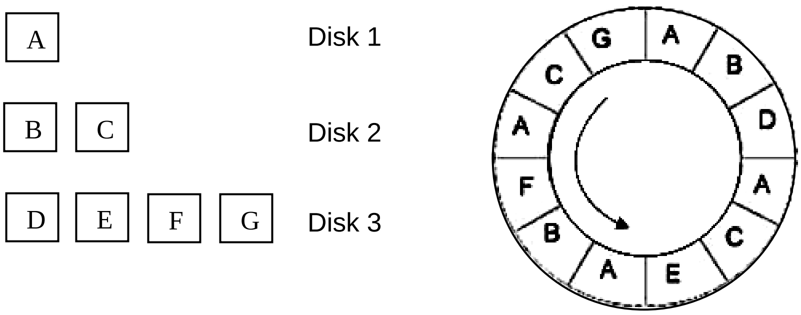
i. Power conservation is of no concern in the Datacycle architecture, while it is a major physical requirement for wireless broadcasting.

iii. The wireless bandwidth is much lower than the bandwidth assumed in the Datacycle architecture. Hence, bandwidth utilization is a major concern in data broadcasting.

2.4 Broadcast Disks

The flat organization assumes that every data item is accessed uniformly, that is, each data items are broadcast with fix interval in cyclic basic. This is rarely true in real-life. Typically, some portions of data are more important and are more common needed by clients whereas others are not. This phenomenon is referred as hot and cold data respectively become the motivation of the Broadcast Disks architecture. In other words, data items of greater interest or greater importance should be broadcast more often than items for which there is very little demand. As important data being broadcasted more frequently, clients waiting time can reduced.

Broadcast Disk architecture organizes data into hierarchy in which simulates multiple disks scenario in the way that the important or highly demanded data items (hot) are placed in the “fast spinning disks” whereas the less important or low-demanded data items (cold) are in the “slow spinning disks”.



(a) (b)

Figure 2.4: Broadcast Disks

Figure 2.4 shows an example of Broadcast disk architecture. In this example as Figure 2.4(a), data are divided into 12 “pages”, which are fix sized independent units. These 12 pages are assigned to 3 different “disk” based on their popularity (access probabilities). The first disk contains most hot data. The second disk contains less hot data where as the last disk contains cold data. In the example, by assuming the first disk (data A) are to be broadcast twice as often as those in the second disk (data B and C) and four times as often as those of the last disk (data D,E,F and G). Therefore as seen in Figure 2.4(b), we see that data A is being broadcast more than others.

Consequently, in broadcast disk architecture, the expected delay in accessing the hot pages can be considerably shorter in comparison to a flat disk. However, the delay in accessing colder items is longer than with a flat disk, introducing an interesting tradeoff. A good broadcast program should be periodic, has fixed inter-arrival times for repeated occurrences of a page, and allocates bandwidth to pages in accordance with their access probabilities (Acharya, Franklin & Zdonik, 1995).

2.5 Complementary Approach on Data Organization

In data broadcasting, any number of clients can monitor the broadcast channel. Data need to be properly organized to cater the needs of the client (Acharya, 1998; Imielinski & Knorth, 1996). Obviously, broadcasting irrelevant data items increases client waiting time and hence, deteriorates the efficiency of a broadcast system.

Based on Broadcast Disks, the server can construct a memory hierarchy in which the highest level contains a few items and broadcasts them with high frequency while subsequent levels contain more and more items and broadcast them with less and less frequency. In disk broadcasting by Acharya (1998), some broadcast slots may be unused, which results in the waste of bandwidth and increase of access time. However, the purposed of complementary approach by Chang and Yang (2002), in organizing data for broadcasting is to solve the empty slots problem that occurs in the disk broadcasting approach.

Figure 2.5 shows an example of a broadcast program generated by disk broadcasting approach, in which several empty slots can occur, consists of 6 minor cycles, and has a period of 24 slots with 8 empty slots. Based on the disk broadcasting method above, in complementary approach, the broadcast program generates a small number of slots in one broadcast cycle and shorter access time.

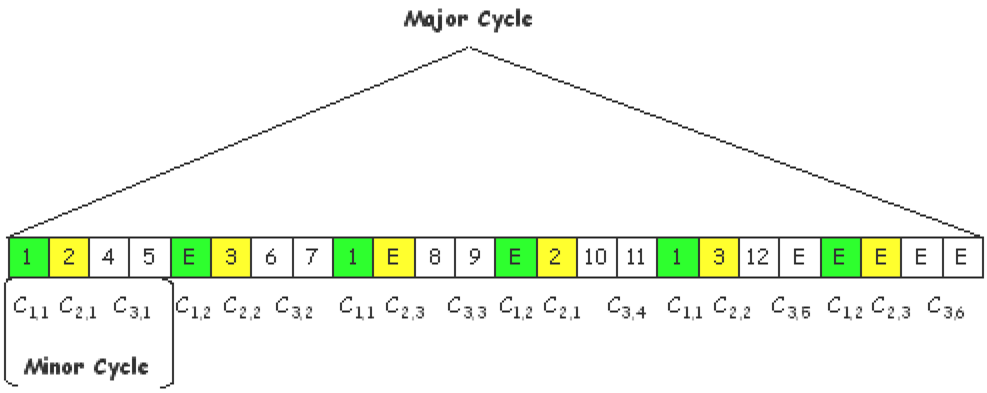
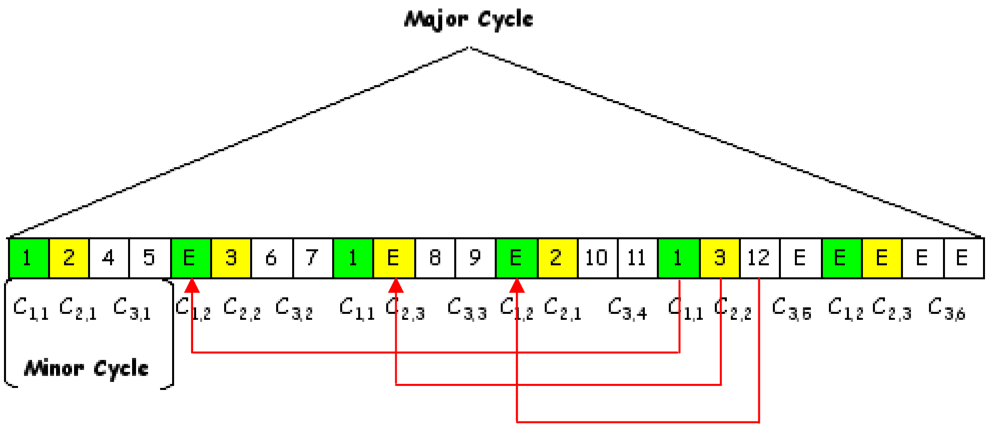
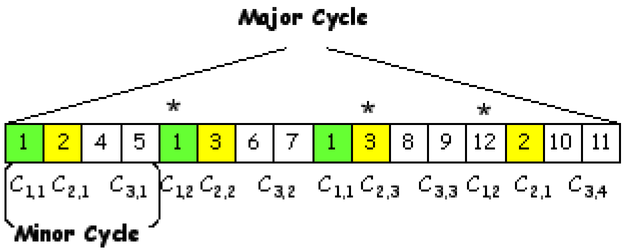


Figure 2.5: A broadcast program with 8 empty slots based on disk broadcasting method

The outline of broadcast data can be determined as shown in Figure 2.6(a). Pages which are located near the end of broadcast cycle are moved to those empty slots which occur before those pages as shown in Figure 2.6(b) where, the "*"symbol denotes those moved pages. Finally, the total number of slots in a broadcast cycle is equal to the one from disk broadcasting method minus the number of empty slots which is 16 slots with no empty slots. It generates a small number of slots in one broadcast cycle and shorter mean access time than disk broadcasting method. The drawback of this approach is that the broadcast cycle is not fully utilize by shorten the broadcast size as the empty slots can be replaced with other popular pages to clients.



(a)



(b)

Figure 2.6: A broadcast program based on the complementary approach:
(a) Deciding the outline; (b) The result after the complementary approach

2.6 Hierarchal Caching Management

With the ever growing performance gap between memory and disk as well as rapidly improving CPU performance, efficient cache management is becoming an increasingly important consideration in system design. Caching data items from the broadcast can be used to reduce delay for accessing data, to reduce workload of channels and to make disconnection from the server transparent to the user somehow. In hierarchal caching, one level of cache is lower than another if it is further from client. A first-level cache is at the highest level, residing in a client.

Only a first-level cache is exposed to the original locality and has the highest potential to exploit it. A stream of access requests from a client, as seen by a lower-level cache, has been filtered by the higher-level caches. Williamson (2002) provides a more in-depth explication of this filtering effect and some of its implications. Another challenge in hierarchical caching is to eliminate redundancy which caching of the same block at more than one level on its retrieval route. Because of redundancy, for some access pattern, the effective aggregate size of a multilevel cache may be no larger than that of the single level of cache with largest size. There may be multiple clients at the first level in the hierarchical cache system. In such a system, each client, along with its lower levels of caches, exhibits the weakened locality and redundant caching problems.

2.7 Indexing

The major drawback of Broadcast Disk scheme is that it requires clients continually listen to the broadcast, which makes it infeasible or inefficient for some mobile clients machines. The reason is mobile clients usually suffer from disconnection or weak connection due to limited battery power. Indexing scheme is proposed to overcome this issue. The idea is to set mobile clients always in doze mode and turn to wake up mode only when receiving data. Having mobile client in wake up mode when only receiving data can save power consumption and hence can prolong the life. Indexing scheme lets the servers to interleave the broadcasts data and its directory (index) during the broadcasting as shown in Figure 2.7. Clients use the index to determine of how long to “doze” before the data of interest arrives on the broadcast (Herman, 1987). Use of the index provides a means to allow the client to disconnect and “wake up” at appropriate time to pick up data of interest. The drawback of this solution is that broadcast cycles are lengthened due to additional indexing information.

(Imielinski, Visvanathan & Bandrinanth, 1997) provides detailed description of several indexing algorithms and their analysis.

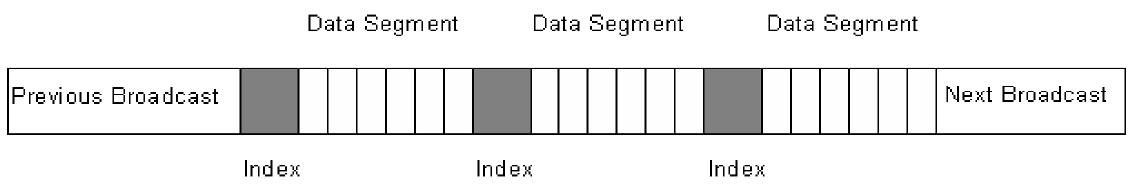


Figure 2.7: Index broadcasting

2.8 Hybrid Broadcast

Hybrid architecture was first investigated by Wong (1988). The model is shown in Figure 2.8. In that model, items are classified as either frequently requested (f-request) or infrequently requested (i-request). It is assumed that clients know which items are f-requests and which are i-requests. The model services f-request is using a broadcast cycle and i-request is on-demand request. In the downlink scheduling, the server transmits of f-requested items (according to a broadcast program), followed by the transmission of the first item in i-request queue (if at least one such request is waiting).

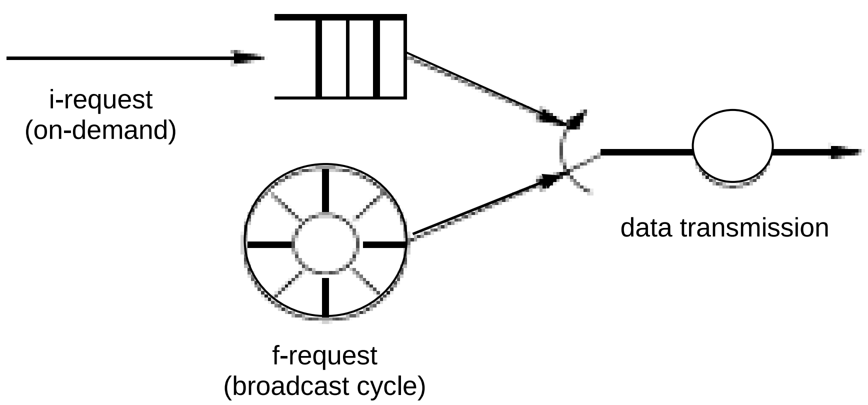


Figure 2.8: Architecture of Hybrid Broadcast