Effective Content Organization and Retrieval within Point-to-multipoint Channels towards Mobile Grid IPTV

Shaohua Liu, Junsheng Yu
Beijing University of Posts and Telecommunications
Beijing 100876, China
liushaohua@ieee.org

Yinglong Ma, Lin Zuo,Qiang Liu
Institute of Software,
Chinese Academy of Sciences
Beijing 100080, China

Peng Gao
China Mobile Group Design Institute Co., Ltd
Beijing 100080, China

ABSTRACT
Abundant information in the Grid is being delivered to enormous destinations among which the mobile users maybe the mainstream in near future. Accordingly, the bottleneck exists at the last mile from the Grid backbone to those mobile terminals where wireless environment places limitations on the rate and amount of communication. Point-to-multipoint (PTM) communication is an effective solution to this limitation. This paper develops the PTM parallel channels content organization mechanism to support timely and reliable access to the common interested information on the Mobile Grid, which has great practical value in the mobile IPTV application scenario. Moreover, we systemically study the parallel channels content organization for non-concurrent and concurrent content retrieval, and then contribute better content organization and retrieving algorithms which fully utilized the concurrent parallel content retrieval capability. The algorithms proposed could deliver information in a very high-performance way to larger user groups so as to achieve shorter response time and less network latency, both for the source-side and for the destination-side. We demonstrate the effectiveness of related mechanisms using a number of examples and some performance experiments.

Keywords
IPTV, Mobile Grid, Point-to-multipoint (PTM) communication.

1. INTRODUCTION
As the information superhighway, the Grid contains countless needed information and is capable of delivering them at extremely high performance and high throughput [1]. Meanwhile, we have also seen the Grid being extended to the Mobile-computing environment, which leads to the concept of Mobile Grid [2]. However, in such scenario, the abundant information in the Grid may not be available at the mobile users' fingertips, by any device, at anytime, at anywhere. The real bottleneck is at the last mile where wireless environment places limitations on the rate and amount of communication [3].

GridNets 2007, October 17–19, 2007, Lyon, France.

DOI 10.4108/gridnets.2007.2244

Almost all existing mobile communicational applications such as telephony, messaging and on-demand streaming and download services are based on point-to-point (PTP) communication.

Nowadays, Mobile IPTV that allows users to watch TV or play other program on their mobile terminals is becoming one of the highest priority markets. But PTP would quickly exhaust the networks bandwidth resources, which is clearly infeasible.

Point-to-multipoint (PTM) communication where data packets are simultaneously transmitted from a single source to multiple destinations is potential solution to this limitation. In fact, many mobile users are directed towards common interested information, e.g. hot spot news and sports games videos. Furthermore, the number of mobile users soliciting such simple and similar requests is usually massive. In this case, PTM is very useful.

Broadcast and multicast are synonyms for PTM communication. Broadcast refers to the ability to deliver content to all users. Known examples are traditional radio and TV. Multicast refers to services that are solely delivered to a group of users who have common interests in some information, for example, sports, news, cartoons and so on. A multicast-enabled network ensures that the content is solely distributed over those links that are serving receivers that belong to the corresponding multicast group [4].

This paper developed the PTM parallel channels content organization and retrieval mechanisms so as to support timely and reliable access to the common interested information on the mobile Grid, in particular for the mobile IPTV application.

The algorithms proposed can deliver information, contents and services to larger user groups in a high efficient way. The benefits include scaling up as the number of users increases, eliminating the need to multiplex the bandwidth among users accessing the air channel, and achieving shorter response time, lessening power consumption, and so on.

2. BACKGROUND
The Grid has already been successfully used in many demanding applications where huge amounts of data have been stored, processed and distributed [5]. Such killer applications coined the concept of the Grid within the scientific community. The amount of potential Grid users is really enormous. In particular, mobile users who travel and work only seldom at their offices might be the future mainstream users of the Grid.

The Mobile Computing is a generic term describing the application of small, portable, and wireless computing and communication devices. It focuses on the requirement of providing access to information, communications and services everywhere, anytime and by any available means [6].

Mobile Grid, in relevance to both the Mobile Computing and the Grid, inherits from the Grid, meanwhile has the additional feature of supporting mobile users and resources in a seamless, transparent, secure and efficient way. Wireless devices with currently limited resources such as low processing power, finite battery life and constrained storage space would benefit from the opportunity of connecting to the Mobile Grid network.

Market research shows a desire of mobile users to watch TV while on the move. An initial service is available on some mobile networks including EDGE and 3G. Furthermore, IPTV became a new-coined killer application to which Mobile Network Operators (MNOs) also resort so as to accelerate 3G adoption.

In 2006, ITU defined IPTV as "multimedia services such as television/video/audio/text/graphics/data delivered over IP-based networks managed to provide the required level of QoS/QoE, security, interactivity and reliability".

Moreover, Mobile IPTV obviously allows users to watch TV on their mobile terminals where the contents may be distributed by the Mobile Grid. In this case, at the last mile, information is delivered by the air channel. The challenge is to meet the consumers' high expectations, such as quick and easy access to useful information and available when and where they need it.

In fact, there exist some similar applications. In Europe, some operators have launched sports such as FIFA 2006 information services that push short video clips of game highlights via MMS. In Japan, mobile users who subscribe to KDDI's EZChannel multimedia service would receive various kinds of content in their terminals. Thus far, Vodafone, TIM, Three, and Sprint etc. have each launched mobile IPTV services. The content, in some cases, is taken from existing living broadcasting TV channels; in other cases, offered specially edited versions of programs.

However, most mobile IPTV services at present are delivered over PTP. Consequently, a content server that delivers content to several users at a time must establish and maintain a separate PTP connection for each recipient. This approach works well for low to moderate numbers of subscribers but does not scale well as the number increases. Furthermore, the service would generate a tremendous amount of outbound traffic when many subscribers use it at the same time. The wireless link can also easily become a bottleneck if numerous recipients are located in the same cell [7].

Imagine the 29th Olympic Games in 2008 to which Beijing city promised to provide anyone by any device with rich information at anytime and anywhere. Maybe at that time, a great number of fans sitting in a soccer stadium use their mobile phones to monitor parallel games, in that way they currently use transistor radios. In this case, the use of PTP would be very inefficient. Therefore, there is a clear need for new PTM telecommunication bearers that can support broadcast/multicast services more efficiently.

3GPP and 3GPP2 addressed broadcast/multicast services in GSM/WCDMA and CDMA2000 respectively. In 3GPP, the work item is called Multimedia Broadcast and Multicast Service (MBMS) [8], [9]. In 3GPP2 it is called Broadcast and Multicast Service (BCMCS). 3GPP MBMS and 3GPP2 BCMCS have many commonalities. Meanwhile in China, TDSCDMA has also integrated similar MBMS mechanism into existing standards.

MBMS and BCMCS will boost the capacity of existing services. For instance, WCDMA could be set up with a mix of 64kbps news channels and 128kbps sports channels with an aggregated channel bandwidth of 2.5Mbps on a cell carrier. This enhanced capacity would stimulate the development of new, mobile, mass-media services on the Mobile Grid environment [10].

3. RELATED WORK

In the Mobile Grid IPTV scenario, the characteristics of mobile devices and limitations of wireless communication technology pose challenges on PTM strategy as well as methods to retrieve data. Desire to have timely access to information is compromised by the network latency and power/memory consumption.

Major issues of concern include reducing power consumption, reducing access latency, and disseminating relevant data of interest to the user group. User profiling, monitoring the access patterns, application of indexing, broadcasting over parallel channels, data distribution and replication strategy, conflict resolution, and scheduling of data retrieval are solutions to these issues that have advanced in the literature. Hurson et al. gave a quite complete review and analysis on those solutions [11].

The literature addressed some methods to determine and generate data contents on one or more broadcast channels. Some methods are classified as push-based techniques, where the users are passive listeners and the server assumes a prior client access patterns that does not change during the course of broadcast. Based on this assumption, the server repeatedly broadcast a set of selected data items [12]-[15]. Alternatively, the dynamic and adaptive methods, also called pull-based methods, assume the clients' access patterns are either unknown in advance or changes relatively frequently during the course of broadcast. Finally, hybrid schemes assume a back channel for users to explicitly request data items that are not included in the standard broadcast cycle. As a result, the push-based and pull-based data are interleaved on a broadcast cycle [16]-[18].

This paper adopts the push-based methods in which content organization, as a means of reducing access latency, has been researched intensively in the past. Whether the physical storage medium is a flat memory [12] or a disk rack structure [14], an appropriate data placement algorithm should attempt to detect data locality and cluster related data close to one another.

Flat memory way is the basic schema to distribute data on single channel. All information units are statically and sequentially delivered to the users of the network [12]. Users retrieve information units from the channels when they are available. There are no requests made by the users to the server. Clearly, wait time for information units is relatively long and proportional to the half of the total content length on the average.

Disk rack structure could reduce the waiting time for retrieving hot spot information [14]. Instead of using one spinning flat memory to deliver all information, it distributes information among multiple disks with varying sizes and spinning speeds. Information units available on faster spinning disks get mapped more frequently than those available on slower disks. The disks create a "memory hierarchy" for the information units.

Peng and Chen proposed the hierarchical allocation that extended the disks rack structure to parallel channels with different speeds [15]. Besides, they used the access frequency of an information unit as a means to allocate the unit, i.e., hot information units are allocated to the faster channel. Given a set of information units with different access frequencies and several channels of different speeds, the units are heuristically allocated to different channels in order to minimize the average expected delay of all units.

On the other hand, in order to reduce the overall response time and the amount of power consumed in the retrieval process at the mobile device, the index can be used to point to the location or possible availability of an information item on the channels and allowing the mobile device to predict the arrival time of the information requested. This enables the mobile device to switch its operational mode into an energy-saving mode and minimize energy consumption. Of course, the advantages of indexing come at the expense of computational overhead and increased length of the content, which increases the response time. Currently, there are mainly two categories of indexing techniques, namely, signature-based and tree-based indexing.

The signature-based indexing is to add a control part to the contents of an information frame. That is to say, a hash function is applied to the contents of the information frame to generating a bit vector that is superimposed on the data frame. As a result, signature partially reflects the data content of a frame. In short, this technique creates a set of signatures for data frames and interleaves them with their associated data frame. Different allocations of signatures have been studied, including single signature, integrated signature, and multi-level signature [19].

An index is usually organized as a tree in which the lowest level of the tree points to the location of the information frames. To reduce the mobile device active time and consequently to reduce the power consumption, the index frames are usually replicated and interleaved with the data frames. Two index replication schemas, i.e., distributed indexing and (1, m) indexing, have been studied. In distributed indexing, the index is partitioned and interleaved in the cycle. Each part of the index is followed by its corresponding data frames. In (1, m) indexing the entire index is interleaved m times during the cycle, that is, the whole index is given before every 1/m fraction of the cycle [3].

In general, the tree-based indexing schemas are more suitable for our application domains where the user request is directed towards interrelated information units clustered on the channels. Furthermore, tree-based indexing schemas relative to signature-based indexing schemes are more suitable in reducing the overall power consumption. This is due to the fact that a tree-based indexing provides global information regarding the physical location of the data frames on the channel.

4. PARALLEL CHANNELS CONTENTS ORGANIZATION AND RETRIEVAL

4.1 PTM Communication Model

In the Mobile Grid IPTV scenario, the information to be delivered is of multimedia or hypertext nature published from an information source subscribed by multiple mobile users.

The PTM communication complies with following assumptions:

(1) Information content comprises calculable information units.

(2) Different unit can have different size, i.e. content length.

(3) Dependencies can exist among those units [20]. *For example, the temporal relationship between units maybe exist, requiring one unit must be prior to the other, but parallel to another.*

(4) The information is supplied by one source and read by multiple receivers in general.

(5) Information can be delivered on one or more channels.

(6) PTM communication is performed in a cyclic manner usually.

(7) Information units might be replicated within a PTM communication cycle.

(8) Information units retrieved can be cached, so as to reconstruct the relationship with other units. But the cache size is limited, so we should also reduce the use of cache.

(9) Information receivers can retrieve information from one or more channels at the same time. This assumption is based on the spread-spectrum communication nature in current 3G mobile communications, i.e. Code Division Multiple Access (CDMA), such as in TDSCDMA/WCDMA/CDMA2000. As compared with the increase in telecommunication network bandwidth, we think the mobile devices' memory and computational capability, relative to parallel information retrieval from several channels, will not be a problem in near future. This is a very important assumption that distinguishes our approaches from previous.

(10) The updates of information only reflected at the following cycle. This constraint enforces the integrity among the information units within a PTM communication cycle so as to avoid the inconsistency either between the original unit and its replica or other units who are dependent on it within a cycle.

4.2 Data Structure for Content Expression

Here, we adopt the Directed Acyclic Graph (DAG) to express the information and the dependency, as illustrated in Figure 1.

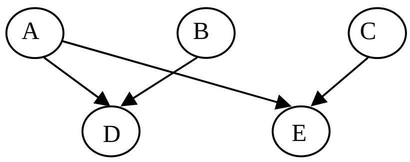

Figure 1. Directed Acyclic Graph (DAG) Data Structure.

In this data structure, all information units are represented as the vertexes, and directed edges are the dependency among information units. For example, the direction can be the symbol of temporary relationship. Previous work showed that such clustering could improve the response time [21].

4.3 Contents Organization Mechanism

Delivery information along parallel channels can reduce the content length than placement on a single channel. This goal is a little similar to typical scheduling tasks in a multiprocessor environment that is a classic NP-hard problem.

In order to organize content effectively, three kinds of heuristic-based static scheduling algorithms would be discussed here.

4.3.1 Longest Information Unit First

First of all, content organization mechanism could rely on a simple and localized intuition by giving priority to longest information units so as to achieve better load balance because they would consume more resources to deliver and retrieve.

For instance, consider a two channel allocation and three units A, B, and C. Further assume the following relationships among the length of the information units B > C > A. Figure 2 (a) shows a random allocation, whereas, Figure 2 (b) shows the allocation based on the aforementioned heuristic. As can be seen, longest information unit first achieves better load balancing.

Formally, the first heuristic algorithm is given in Algorithm.1 that is equivalent to typical "largest object first" algorithm in which all units is modeled as objects [22].

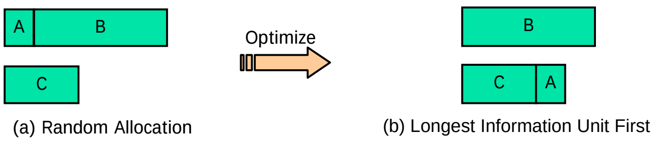

Figure 2. Two Channels Content Organization Example.

Algorithm. 1 (Longest Information Unit First)

Initially, all information units are expressed as a DAG.

Recursively, the longest content length node with in-degree of zero (root node) is chosen and assigned to the least loaded channel. The assigned node along with all of its out-edges are eliminated from the DAG.

Figure 3 is an example of Algorithm.1 execution. Because the length of information unit B is the longest, the B is allocated at first and the directed edge from B to D is eliminated. Then the length of C is longer than A, so C is allocated before A. Therefore, we would get the Figure 2(b) allocation result according to algorithmic partly execution in Figure 3. The time complexity of Algorithm.1 is $O(n^2)$ where n is the number of information units.

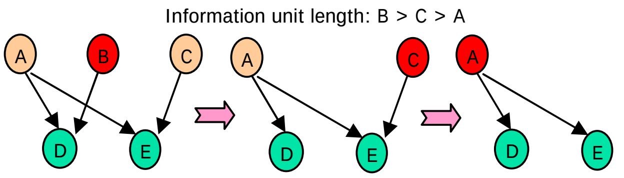

Figure 3. Execution Example of Algorithm.1.

4.3.2 Related Information Units Clustering

Another intuition is that related information units should be clustered. Here the clustering is based on the notions of critical node and critical path. *A critical node is a node that has a child with an in-degree greater than 1. A critical path is defined as the longest sequence of dependent node accessed serially.* The critical path is often defined based on the weights assigned to each node.

Consider the sub-DAG shown in previous Figure 1, A, B, and C are all critical nodes. A has precedence over two nodes whereas B and C have precedence over one node each. Assume that there are two parallel channels. Regardless of the length of information unit A, B, and C, allocating A first results in a better load balancing. Hence, we should allocate a critical node with the highest number of children with in-degree greater than one (In-degree > 1) first.

Then consider the allocation of children B and C as shown in Figure 4. Assume that at the starting point only one channel is available. As can be seen, allocating B first as illustrated in Figure 4 (b) produces better load balancing than the allocation in Figure 4 (a). Therefore, we should allocate nodes with the highest number of child with in-degree = 1 first because this could free up more nodes to be allocated in parallel.

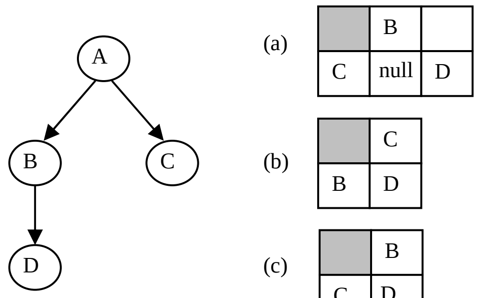

Figure 4. Related information units allocation.

In summary, the effect of the number of the incoming and outgoing edges should be involved in the weight of a node.

Eventually, in our practical applications, the weight W of each node is calculated based on following factors: *the maximum weight of its children, frequency of access, length of the information unit on this node, total length of all information units who are the parents of this node, the number of children with in-degree = 1, the number of children with in-degree >1, the number of the incoming edges, the number of the outgoing edges.*

Furthermore, the calculation of weight should be started at the leaf nodes and then traversed in a breadth-first manner.

Accordingly, a new algorithm that uses a more sophisticated set of heuristics in determining the weight of the node is presented in Algorithm.2. The main idea is to cluster related units, which is similar with existing "clustering critical-path" algorithms [22].

Algorithm.2 (Related Information Units Clustering)

Initially, all information units are expressed as a DAG.

Assign weight to every information unit node in the DAG.

Repeat until all the nodes have been processed

Select the free node N with the largest weight

If all parents of N are fully allocated on the channels

Place it on the currently least-loaded channel

Else

Fill up the least-loaded channels with nulls up to the end of the last allocated parent of N then place N on it.

The third line *repeat* loop implements the critical path heuristics by selecting the free node with the largest weight and placing it on the least-loaded channel, after all of its parents have been fully allocated. Hence, the allocation in Figure 4(c) cannot be reached.

The critical path in this algorithm itself is actually the *repeat* loop, which has to be repeated n times where n refers to the number of nodes. Thus, the overall running time of Algorithm.2 is also $O(n^2)$.

4.3.3 Performance Improvement for Concurrent Retrieval

Here, let's recall the parallel content retrieval in the PTM communication model in previous section, which said information receivers could retrieve information from one or more channels at the same time and the information units retrieved could be cached so as to reconstruct the relationship with other units later. That means the allocation in Figure 4(c) is acceptable in this case.

Thus, in fact, we could eliminate the dependency on the parents' allocation from above Algorithm.2 procedure. It leads to following performance-improved Algorithm.3, which is exactly an important contribution in this paper.

Algorithm.3 (Parallel Channels Organization for Concurrent Content Retrieval from Multiple Channels at the Same Time)

Initially, all information units are expressed as a DAG.
Assign weight to every information unit node in the DAG.
Repeat until all the nodes have been processed
  Select the free node N with the largest weight
  Place it on the currently least-loaded channel

Obviously, the new Algorithm.3 is faster than original Algorithm.2, although its time complexity is still $O(n^2)$.

4.4 Retrieving Contents within PTM Channels with Index

Due to the power limitation of the mobile devices, in retrieving information from the air channel, we could reduce the amount of power consumption by employing some indexing technique.

Indexing is normally implemented via a multi-level tree. Due to the sequential nature of the air channel, the allocation of nodes of a multi-level tree has to follow the navigational path used to traverse the tree starting at the root. Hence, an ordering schema should be used to sequentially map the nodes on the air channel.

In the presence of an indexing mechanism one could use the following procedure to retrieve content from the air channels.

Initially, the mobile device must be in active operational mode and tune into the channel to determine when the next index tree is sent. As soon as the mobile device retrieves the offset of the next index, its operational mode could change to doze mode. Once the index tree arrives, the mobile device must be in active mode again so as to access the index and determine the offset for the requested information. After that, it could switch to doze mode again. Finally, when the requested information arrives, the mobile device would change its operational mode to active mode and retrieve them. When the content is downloaded, the mobile device changes to doze mode again.

In our research, we construct only one multi-level tree representing an index for all information units. Thereby, any query requesting one or more information units needs just navigator this common tree.

In addition, we adopt the (1, m) indexing schema. In other words, the same whole indexing tree for all content is given before every $1/m$ fraction of the cycle.

In traditional telecommunication applications, even with the indexing for parallel channels, the mobile device can only retrieving content from one of these channels at any time. This mechanism conforms to Algorithm.4 that is discussed in [11].

Algorithm.4 (Retrieving Required Contents from One of Parallel Channels with Indexing)

Switch to active mode
Probe onto channel and retrieve offset to the index
Switch to doze mode
Do {
  Reach the index
  Switch to active mode
  Retrieve the index to know the location of required content
  Switch to doze mode
  Do {
    Reach the next location including required content
    Switch to active mode
    Tune into one channel including required content
    Retrieve the required content
    Switch to doze mode
  } While each required content is retrieved within this PTM cycle
} While there are un-accessed information units because of overlapping between information units on different channels

---

However, in our mobile Grid IPTV solution, we designed a more effective retrieving method as Algorithm.5 that is another important contribution in this paper.

Although this algorithm seems very simple and naive, but it fully utilizes the capability for simultaneously retrieving from parallel air channels. As this method reduces the expensive of time for frequently tuning-into different channel, the retrieving operation based on such Algorithm.4 would obviously outperform traditional implementation based on just retrieving one of several parallel channels each time.

Algorithm.5 (Simultaneously Retrieving Required Contents from Parallel Channels with Indexing)

Switch to active mode
Probe onto channel and retrieve offset to the index
Switch to doze mode
Do {
Reach the index
Switch to active mode
Retrieve the index to know the location of required content
Switch to doze mode
Do {
Reach the next location including required content
Switch to active mode
Retrieve the required content
Switch to doze mode
} While either each required content is retrieved or the end of this cycle is reached
} Until all required contents are retrieved

5. PERFORMANCE EVALUATION

5.1 Design of the Experiment

To illustrate the effectiveness of the algorithms across an unbiased test bed, the degree of connectivity among the nodes in the DAG was randomly varied. In total, 10,000 different DAGs were generated. In these DAGs, the out-degrees of nodes were randomly determined within the range from zero to three. In addition, the weight of each node was calculated according to a set of rules which realizes those factors listed in last section. But, it should also be noticed that the access frequency of each information unit is assumed to be equal in these experiments.

On the other hand, we tested the content retrieving with and without indexing so as to show the effectiveness of the retrieving algorithm, i.e., Algorithm.4 and Algorithm.5. In such kind of simulation, we designed the indexing tree with max fan-out 6, and then laid 1,000 information units with the same size on those PTM channels. Finally, 10% information units were required.

5.2 Performance on the content organization

Above all, we should evaluate the expense of time for content organization on the source-side.

Figure 5 shows the effect on time expense for content organization as the number of information units increases.

As can be seen, Algorithm.3 outperformed Algorithm.2, although Algorithm.3 would spend a little more time than Algorithm.1, the simplest intuition.

Furthermore, Algorithm.3 would bring about higher performance on the destination-side for content retrieval than Algorithm.1 and Algorithm.2, which will be discussed in the next section.

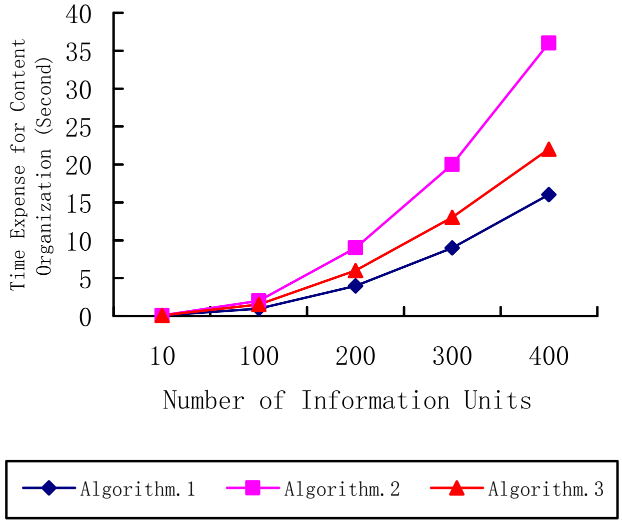

Figure 5. Time expense for content organization on the source-side

5.3 Performance on the content retrieval

Next, we will evaluate the performance of the proposed algorithms by simulating and measuring the average response time per information unit retrieval by mobile end-users.

5.3.1 Number of Channels Effect Analysis

Figure 6 shows the effect of varying the number of channels on the average response time per information unit.

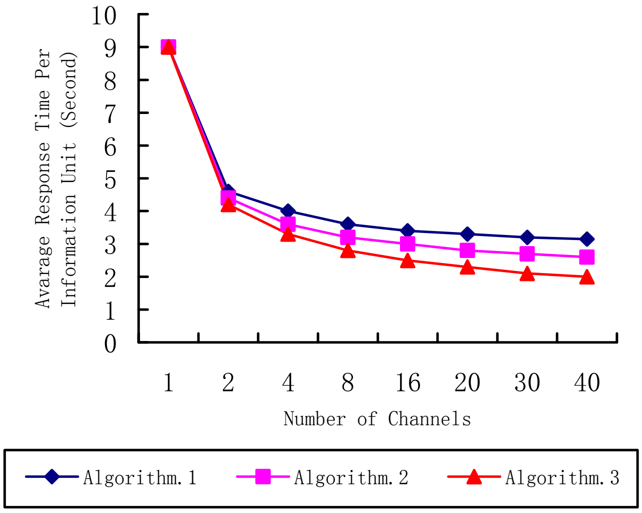

Figure 6. Number of Channels Effect.

We found that increasing the number of channels resulted in a better response time for all of three algorithms.

However, this improvement flattened as the number of channels increased above a certain threshold, because additional parallism provided by the additional number of channels did not match the number of free nodes available to be allocated, simultaneously.

In addition, Algorithm.2 outperformed the Algorithm.1 because Algorithm.2 heuristics attempt to smooth the distribution of the information units among the parallel channels while clustering the related units.

Moreover, Algorithm.3 outperformed the Algorithm.2 because the content is organized more compactly so that the total content length becomes shorter, which leads to less latency on the information units retrieval.

5.3.2 Out-degree Distribution Effect Analysis

Figure 7 depicts the effect of modifying the distribution of the out-degrees of the nodes on the average response time per information unit. The distribution of the out-degrees is shown in terms of a distribution vector composed of four entries. Each entry corresponds to the relative percentage of nodes with a certain out-degree. For example, a vector [40%:30%:20%:10%] represents the fact that within a DAG, 40% of nodes have no out-degrees, 30% have out-degrees of 1, 20% have out-degree of 2, and 10% have out-degree of 3.

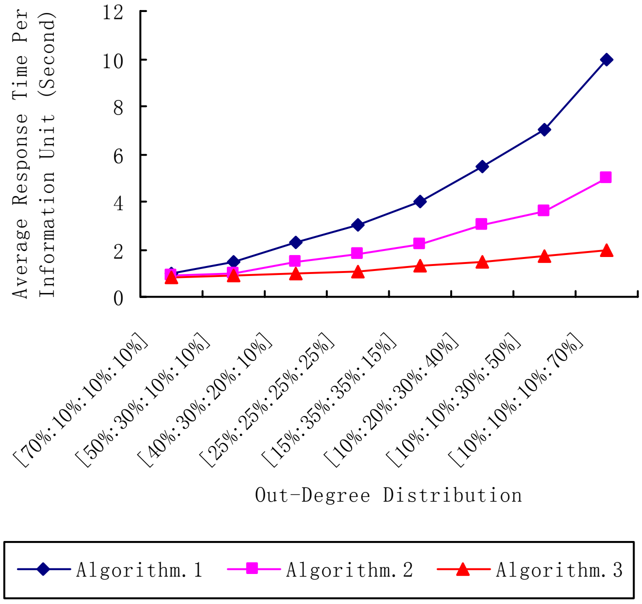

Figure 7. Out-degree Distribution Effect.

In general, Algorithm.2 outperformed Algorithm.1, and Algorithm.3 performed better than either Algorithm.1 or Algorithm.2.

As can be seen, when the out-degree distribution is biased to include nodes with larger out-degrees (i.e. making the DAG denser), the Algorithm.1 performance degrades at a much faster rate than the Algorithm.2 and Algorithm.3. This is due to the fact that such bias introduces more critical nodes and a larger number of children per node. Algorithm.2 is implicitly capable of handling such cases, so its performance is better than Algorithm.1. As to as less average response time per unit as possible, because of its imperceptible performance degradation, Algorithm.3 is more stable than Algorithm.1 and Algorithm.2.

5.4 Performance with and without indexing

To study the effectiveness of indexing, we measured the performance, in terms of response time for retrieving required contents, with and without indexing.

Figure 8 shows the simulation result. From this figure, it can be seen that in the case of no indexing, the response time was almost constant and independent of the number of channels in that without any indexing mechanism in place, the mobile device has to scan every information unit in sequence on each channel until all required content are acquired.

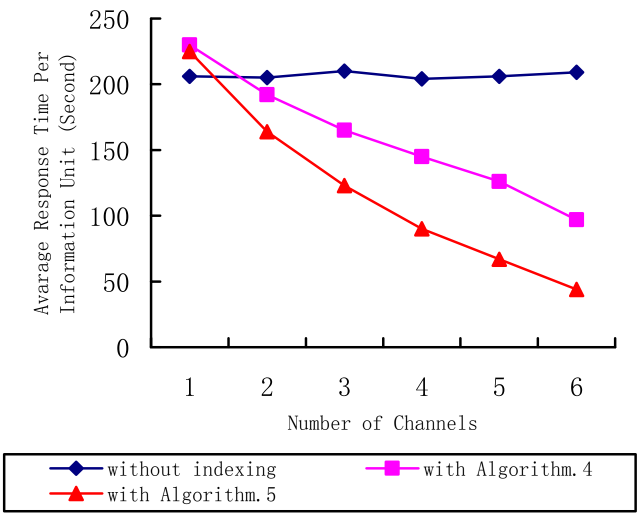

Figure 8. Response time for required contents retrieval with and without indexing.

When the indexing is used, the response time lessens as the number of channels increases in general. It should be noticed that for a single channel environment, the response time without indexing is less than with any indexing schema. That is because placing an indexing tree along one air channel contributes to extra storage overhead and thus longer response time.

Moreover, the Algorithm.5 proposed for the mobile Grid IPTV application indeed performs better than traditional method like Algorithm.4 mainly because of reduced expensive of time for frequently tuning-into different channel.

For the mobile device, the active time is proportional to the number of index and information units to be retrieved. The mobile device retrieving PTM contents without an index has to be active all the time in order to scan contents from parallel air channels until all required contents are accessed. This simply means that the active time is the same as the response time. In other words, the active time for retrieving without indexing remains almost constant and independent of the number of air channels, either. Once we use the index to lessen the response time, we are also saving precious energy resources of mobile devices.

Therefore, indexing could bring about shorter response time and lower energy consumption for content retrieving on the mobile devices.

6. CONCLUSION

To achieve timely access to PTM communication in the Mobile Grid IPTV Scenario, this paper focused on the proper mapping of information units on multiple parallel channels.

The goal was to find the most appropriate allocation schema that would balance the overall load on parallel channels and cluster related information units, so as to provide the minimum overall time both for content organization on the source-side and for content retrieval on the destination-side.

Applying naive Algorithm.1 (longest information unit first) intuition showed an improvement in load balancing in contrast with the random allocation. However, it was proved short in solving the clustering requirement.

Algorithm.2 (related information units clustering) was presented in order to compensate this shortcoming. Relying on the critical path paradigm, Algorithm.2 assumed several heuristics and showed better performance. Finally, reducing the overall response time could also minimize the amount of energy consumption at the mobile devices.

Although the dedicated Algorithm.2 heuristic satisfied the clustering requirement, it ignored the parallel content retrieval capability hide behind the spread-spectrum nature in current 3G telecommunications.

Accordingly, Algorithm.3 improved Algorithm.2 by utilizing the capability of concurrent parallel content retrieval from multiple channels at the same time, which can be given by the 3G spread-spectrum communications, and finally achieved better performance both on the source-side and on the destination-side.

The techniques in this paper, especially the Algorithm.3, can improve the mobile GRID IPTV applications meanwhile boost the 3G communication. Indeed, it can also be applied in the broadcast/multicast group communication for the Grid, Internet and some other networks environments.

Technically, mobile users have to scan the channels to locate its desired content. In an effort to reduce such scanning time, providing an effect indexing directory for the information units on the channels could be proved to be beneficial in minimizing resource usage and power consumption. Therefore, the more effective indexing trick proposed in this paper, i.e. Algorithm.5 outperformed traditional method like Algorithm.4, is another contribution to the mobile Grid IPTV application scenario.

7. ACKNOWLEDGMENTS

The authors would like to thank the experts within the "Europe-China Grid InterNetworking (EC-GIN)" project consortium for their encouragement and comments, especially thank those partners in China for their collaborative work. This research is supported by the European Union through the FP6-2006-IST-045256 project EC-GIN, the National Natural Science Foundation of China under Grant No.60703036, and the National High-Tech Research and Development (863) Program of China under Grant No. 2006AA01Z163.

8. REFERENCES

[1] Foster, I. "The GRID: Computing without bounds". *Scientific American*. Vol 228, Iss 4. April 2003.

[2] Park S.-M. Ko Y.-B., and Kim J.-H., "Disconnected Operation Service in Mobile Grid Computing," *In Proceedings of International Conference on Service Oriented Computing* (ICSOC'2003), Trento, Italy, December 2003.

[3] Imielinski T., Viswanathan S., Badrinath B. R., "Data on air: Organization and access", *IEEE Trans. Knowledge and Data Engrg*. 9(3), 1997, pp. 353-372.

[4] Xylomenos G. "Group Management for the Multimedia Broadcast/Multicast Service". *In Proceedings of the 14th IST Mobile Summit*, 2005.

[5] Theotokis, S and Spinellis, D. "A survey of peer-to-peer content distribution technologies". *ACM Computing Surveys*. 36(4). December 2004.

[6] Chlamtac I., Lin Y.-B., "Mobile computing: When mobility meets computation", *IEEE Trans. Comput*. 46(3), 1997, pp. 257-259.

[7] Weiser M., "Some computer science issues in ubiquitous computing", *Communication of ACM*, 36(7): 75-84, 1993.

[8] Third-generation Partnership Project (3GPP), *3GPP Technical Specification TS 26.346*, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codes", V6.7.0, December 2006.

[9] Third-generation Partnership Project (3GPP), *3GPP Technical Specification TS 23.246*, "Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description", V6.11.1, December 2006.

[10] Parkvall S, Englund E, Lundevall M, Torsner J. "Evolving 3G mobile systems: broadband and broadcast services in WCDMA". *IEEE Communications*, 44(2) , 2006, pp. 68-74

[11] Hurson A. R., Jiao, Y., and Shirazi B. A., "Broadcasting a Means to Disseminate Public Data in a Wireless Environment – Issues and Solutions", *Advances in Computers*, Vol. 67, Academic Press, 2006.

[12] Acharya S., Franklin M., Zdonik S., "Dissemination-based data delivery using broadcast disks", *IEEE Personal Commun*.2(6), 1995.

[13] Vaidya N.. Harneed S., "Scheduling data broadcast in asymmetric communication environments", *Wireless Networks*, 5(1999) pp. 171-182.

[14] Cetintemel U., Franklin M., Giles L., "Self-adaptive user profiles for large-scale data delivery", *in Proceedings of ICDE*, 2000, pp. 622-633.

[15] Peng W.-C., Chen M.-S., "Dynamic generation of data broadcasting programs for a broadcast disk array in a mobile computing environment", *In ACM International Conference on Information and Knowledge Management*, 2000, pp. 38-45.

[16] Stathatos K., Roussopoulos N., Baras J., "Adaptive data broadcast in hybrid networks", *in Proceedings of VLDB*, 1997.

[17] Oh J., Hua K., Prabhakara K., "A new broadcasting technique for an adaptive hybrid data delivery in wireless mobile network environment", *in proceedings of IEEE Conference on Performance, Computing, and Communications*, 2000, pp. 361-367.

[18] Bhide M., Deolasee P., Katkar A., Panchbudhe A. Ramamritharm K., Shenoy P., "Adaptive push-pull: Disseminating dynamic web data", *IEEE Trans. Comput*. 51(6), 2002, pp. 652-668.

[19] Banerjee J., Kim W., Kim S.-J. Garza J. F., "Clustering a DAG for CAD databases", *IEEE Transaction on Software Engineering*. 14(11), 1988, pp. 1684-1699.

[20] Liberatore V., "Multicast scheduling for list requests", *in Proceedings of IEEE INFOCOM 2002*, June 2002, pp. 1129-1137.

[21] Hu Q.L., Lee D.L., "Power conserving and access efficient indexes for wireless comuping", in: Tanaka K., Ghandeharizadeh S., Kambayashi Y. (Eds.), *Information Organization and Databases*, Kluwer Academic Publications, Boston, MA, 2000, pp. 249-264.

[22] Hurson A.R., Chehadeh C.Y., hannan J., "Object organization on parallel broadcast channels in a global information sharing environment", *in IEEE Conference on Performance, Computing, and Communications*, 2000, pp. 347-353.