Information Sciences 232 (2013) 27–43

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Efficient processing of requests with network coding in on-demand data broadcast environments

Jun Chen [a,*], Victor C.S. Lee [b], Kai Liu [b], G.G.M.N. Ali [b], Edward Chan [b]

[a] School of Information Management, Wuhan University, Wuhan, Hubei, China
[b] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 16 September 2011
Received in revised form 27 December 2012
Accepted 31 December 2012
Available online 23 January 2013

Keywords:
Network coding
Mobile computing
On-demand broadcast
Data scheduling

ABSTRACT

On-demand broadcast is an effective wireless data dissemination technique to enhance system scalability and the ability to handle dynamic user access patterns. In traditional on-demand broadcast, only one data item can be retrieved by mobile clients during the course of each broadcast, which limits bandwidth utilization and throughput. In this paper, we consider data broadcast with network coding in on-demand broadcast environments. We analyze the coding problem in on-demand broadcast and transform it into the problem of finding the maximum clique in graph theory. Based on our analysis, we first propose a new coding strategy called AC, which exploits the cached information related to clients and data items requested by them, to implement a flexible coding mechanism. Then, based on AC, we propose two novel coding assisted algorithms called ADC-1 and ADC-2 which consider data scheduling, in addition to network coding. In ADC-1 data scheduling and coding are considered separately, while these two factors are fully integrated in ADC-2. The performance gain of our proposed algorithms over traditional and other coding assisted broadcast algorithms is demonstrated through simulation results. Our algorithms not only reduce request response time but also utilize broadcast channel bandwidth more efficiently.

1. Introduction

Data broadcast has attracted many academic researchers' attention as it has been increasingly used to disseminate information to large populations of mobile clients in many new mobile applications, such as location based services, where efficient data broadcast is critical to the system performance [14,31]. In general, there are two major data broadcast approaches [8,9]: (a) push-based and (b) pull-based. *Push-based broadcast* periodically broadcasts data according to a static schedule which is computed offline from clients' historical data access statistics. *Pull-based broadcast*, commonly referred to as on-demand broadcast, compiles requests in the service queue and broadcasts data based on various attributes of pending data items at the server. Push-based broadcast is efficient with applications which require a small set of data items with stable access pattern, while on-demand broadcast is more widely used for dynamic, large-scale data dissemination [15,18]. In this paper, we focus our discussion on on-demand broadcast, where *response time* is one of the most important metrics to measure the system performance [25]. In existing on-demand broadcast strategies, mobile users can retrieve only one data item from each broadcast unit. This constraint restricts full utilization of the limited broadcast bandwidth and leads to long response time to mobile clients.

* Corresponding author.
E-mail addresses: christina.cj@hotmail.com (J. Chen), csvlee@cityu.edu.hk (V.C.S. Lee), kevin.liu@cityu.edu.hk (K. Liu), gnawazali2@student.cityu.edu.hk (G.G.M.N. Ali), csedchan@cityu.edu.hk (E. Chan).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.12.044

Network coding has been proposed in recent years to improve system performance in wireless networks. Previous works show that network coding can utilize available broadcast bandwidth more efficiently to improve throughput and energy consumption in multicast communication [11,19,22,27]. Our previous work [5] addressed the coding problem in on-demand broadcast environments by proposing application of the CR-graph in wireless mesh networks [3,10] to capture the relationships among all cached and requested data items. In this work we further investigate the coding and scheduling problems in on-demand broadcast environments. Based on our analysis, we propose two novel broadcast strategies and compare their performance with existing algorithms. Our main contributions are as follows:

1. We give a probabilistic analysis of the extent to which coding can be effective in on-demand broadcast environments. Effectiveness of network coding and the necessity of adopting adaptive network coding strategy in this dynamic environment is established by the analysis.
2. We propose a coding strategy AC (Adaptive Coding) for on-demand broadcast applications, based on the CR-graph. AC removes the need for constructing the whole CR-graph G by defining a sub-graph G' of G. This helps prune the search space and reduce computation cost.
3. Two new broadcast algorithms called ADC-1 and ADC-2 (Adaptive Demand-oriented Coding) are proposed that combine the strength of data scheduling and network coding. ADC-1, which is similar to ADC proposed in [5], considers data scheduling and network coding in isolation while ADC-2 integrates coding and scheduling in an attempt to achieve better performance. It is demonstrated that ADC-2 not only serves the maximum number of clients in each broadcast unit but also significantly improves overall system performance in terms of request response time and broadcast bandwidth utilization. Both ADC-1 and ADC-2 outperform existing algorithms significantly over a wide range of settings.
4. We introduce two mechanisms in ADC-2 to make its implementation efficient and practical. First, an efficient data structure and a pruning technique are used to reduce the search space. Second, the degree of a vertex, i.e. the upper bound of its maximum clique size, is used to skip computation of the priority of most vertices in the graph, thereby saving time required for processing in search for the most rewarding data item to be broadcast.

The rest of this paper is organized as follows. Section 2 gives the background of the research area. Section 3 describes the system architecture. Section 4 analyzes the coding problem in on-demand broadcast environment. Section 5 outlines our new coding assisted algorithms, ADC-1 and ADC-2. Section 6 describes the simulation model and experimentally compares the performance of ADC-1 and ADC-2 with existing algorithms. Finally,we present our conclusions and suggested direction for future research in Section 7.

2. Related works

Data broadcast through wireless channel is a common way to disseminate information to a large population of mobile clients. In traditional data broadcast environments, data scheduling algorithms play an important role. Various scheduling algorithms have been proposed to determine the sequence of broadcast data items in on-demand broadcast environments. Dykeman and Wong [9] proposed two widely used strategies: Most-Requested-First (MRF) and Longest-Wait-First (LWF). MRF broadcasts the data item with the maximum number of pending requests (also called broadcast productivity) first. When the system load increases and the data access pattern follows the Uniform distribution, MRF has been shown to have the shortest response time. For LWF, the sum of the time that all pending requests for a data item have been waiting is computed, and the data item with the largest total waiting time is chosen for broadcasting next. When the data access pattern follows the Zipf distribution, LWF has the best performance. Although LWF outperforms other strategies in minimizing wait time, it is expensive to implement. Tan and Ooi [24] proposed an adaptive batching scheduling scheme called Maximum Queue Length with Time restriction (MQL-time). In MQL-time, a predetermined and fixed time, say ts, is assigned to each request. Requests are served based on the maximum queue length under normal conditions while requests that have been waiting for more than t s are given higher priority and are served immediately, in order to reduce the overall waiting time. Aksoy and Franklin [1] proposed a low-overhead and scalable scheduling algorithm called RxW, an approximation of the LWF algorithm wherein the number of pending requests for a data item is multiplied by the amount of time for which the oldest outstanding request for the data item has been waiting in the service queue. The data item with the largest product is chosen for broadcasting. RxW combines the strength of MRF and FCFS to provide good performance for both hot and cold items. For systems supporting time critical services, an online scheduling algorithm called SIN (Slack time Inverse Number of pending requests) [28] has been proposed. It is motivated by two existing strategies: EDF (Earliest Deadline First), which considers urgency of requests, and MRF, which focuses on broadcast productivity. According to the simulation results presented in [1,28], SIN outperforms other existing on-demand broadcast algorithms significantly.

In all the above scheduling algorithms, only one data item can be retrieved from the broadcast channel by mobile clients in each broadcast unit. This constraint limits both bandwidth utilization and throughput of broadcast systems. Recently, researchers have proposed the use of network coding to further improve performance. Yitzhak and Tomer [2] first provided insights into the coding problem in on-demand data broadcast. They pointed out that the basis of coding in on-demand broadcast is that the server needs to have the full knowledge of requested and cached data items of each client. Recently, several strategies have been proposed to apply network coding technique to data broadcast [6,20,29,30]. Dong et al. [20]