Computer Networks 52 (2008) 3017–3027

Contents lists available at ScienceDirect

Computer Networks

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

# Scalability study of cache access mechanisms in multiple-cell wireless networks

Hui Chen [a], Yang Xiao [b,*], Susan V. Vrbsky [b]

[a] Department of Math and Computer Science, Virginia State University, Petersburg, VA 23806, United States
[b] Department of Computer Science, The University of Alabama, 101 Houser Hall, Tuscaloosa, AL 35487-0290, United States

ARTICLE INFO

*Article history:*
Received 25 September 2007
Received in revised form 18 June 2008
Accepted 25 June 2008
Available online 11 July 2008

Responsible Editor: C. Westphal

*Keywords:*
Cache access algorithms
Invalidation reports
Poll-each-read
Scalability
Multi-cell networks
Communication cost
Call-back
Lease

ABSTRACT

Strong-consistent cache access algorithms have been proposed and studied for wireless networks. In this paper, we carry out a comprehensive study to compare four strong-consistent cache access algorithms including the invalidation report (IR), poll-each-read (PER), call-back (CB), and Lease schemes in multiple-cell wireless/mobile networks. The purpose of this study is to examine the scalability of the IR schemes. To the best of our knowledge, this is the first such study. We focus on the scalability issue of these four fundamental strong-consistent schemes in terms of network transmission costs regarding network size, database size, subscription ratio, and network traffic, through extensive computer simulations. Our results show that: (1) the IR schemes do not perform well in a multiple-cell wireless/mobile networks with a large update rate, database size, and IR window size, and with a small subscription ratio; (2) the IR schemes do scale up well with the IR period and the access rate; however, while good performance can be obtained when the IR period is large, this implies a large access latency; (3) the PER, the CB and the Lease schemes perform well with a large update rate, small subscription ratio, and large database size.



## 1. Introduction

Wireless links are the scarcest resource in wireless networks. Reducing usage of wireless links often means consuming less battery power at wireless clients. Battery power at wireless clients are scarce resources as well. Therefore, research in wireless caching mostly studies client caching, which can reduce wireless link bandwidth usage, battery power consumption and access latency. A cache mechanism consists of two basic procedures, i.e., the cache access algorithm and cache replacement policies.

Cache access algorithms describe how clients and servers exchange messages and maintain the consistency between the cached copies at clients and the original copies at servers. They are widely used in distributed systems for improve system performance, especially, access latency in wired networks. For example, Coda is a distributed file system that uses a cache access algorithm called call-back to improve access latency [22]. Readers can also find a number of cache access algorithms that primarily aim at improving access latency for the web in [23]. Compared with wired networks, wireless networks are more constrained with bandwidth, client battery power and link quality. Fundamental cache access algorithms initially proposed for wired networks including call-back (CB), poll-each-read (PER), Lease, and time-to-live are studied in [9,17,24] for wireless networks. Invalidation report schemes were initially proposed in [1] and were designed with consideration of frequent disconnections of wireless clients. Extensions of the invalidation report schemes are well-studied in literature, such as [2,3,5–8,10–12,14,20,21]. These extensions address several issues, e.g., access latency, invalidation report overhead, asymmetry of uplink and downlink, as well as erroneous of wireless links. However, the invalidation report schemes require broadcasting channels which are not readily available in

* Corresponding author. Tel.: +1 205 348 4038.
*E-mail addresses:* huichen@ieee.org (H. Chen), yangxiao@ieee.org (Y. Xiao), vrbsky@cs.ua.edu (S.V. Vrbsky).

1389-1286/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2008.06.018

every wireless network, such as personal communications networks [9,17].

Cache access algorithms previously discussed follow a client-server model, where a server has a database and a client requests data items stored in the database and duplicates the received data items in the client cache. Recently, studies [16,25–27] have expanded the cache access algorithms to ad hoc networks and peer-to-peer networks. Caching in these scenarios is often referred to as "cooperative caching" since wireless nodes can be either servers or clients and they work cooperatively. In fact, in those caching mechanisms, the roles of servers and clients remain. In cooperative caching, a client often needs a few hops to reach the server and vice versa. These caching mechanisms often focus on issues such as cache placement and architecture [16,25–27] in the network while the underlying cache access algorithms remain the same as those mentioned above.

Some cache access mechanisms [1–19] maintain so-called strong consistency between data item copies at clients and those at servers, i.e., only the latest copy of a data item is regarded as valid and can be used by a client. On the contrary, some cache access mechanisms [1,7] only maintain a certain degree of weak consistency between data item copies at clients and those at servers, i.e., a client can use its cached copy of a data item even if the data item has been updated at the server. Whether or not strong consistency or weak consistency is required depends on applications and users.

A replacement policy describes what data items need to be evicted from the cache when there is no available cache space for storing a copy of the newly accessed data item. Replacement policies are important to the effectiveness of cache mechanisms. A well-designed replacement policy can significantly improve system performance. In this paper, we are interested in cache access algorithms only and interested readers are referred to [19] for more references on replacement policies.

We study only strong-consistent cache access algorithms. Four fundamental strong-consistent cache access algorithms/schemes have been proposed and widely studied. They are the invalidation report (IR) [1–3,5–8,10–12,20,21], the PER [5], the CB [5], and the Lease schemes [4,6]. However, to the best of our knowledge, two issues have not been addressed when strong consistency is required.

(1) Most of these cache access schemes were studied in single cell wireless networks. Though multiple-cell wireless networks are studied in [11,21], the schemes proposed therein can not guarantee strong consistency. How strong-consistent cache access schemes scale to wireless networks with multiple-cells, e.g., a personal communications service network (PCS), needs further investigation.
(2) The IR-based schemes have been well-studied. It is assumed the IR-based schemes can better use the broadcasting capability of wireless communications. The IR-based schemes have long been considered good schemes for wireless networks. Many schemes have been proposed [1–3,5–8,10–12,20,21]. However, we have not seen the IR-based schemes compared with the PER, the CB, and the Lease schemes. It is important to compare them in terms of different parameters, such as remote database size, network size and traffic load.

Design of distributed cache schemes, like wireless caching, is a challenge [15,16]. Many factors have to be considered, such as (1) narrow wireless bandwidth; (2) mobility of clients; (3) limited battery power of mobile clients; (4) frequently-disconnected mobile clients and poor channel quality; (5) network routing and topology; (6) scalability of cache schemes with database size, network size, user population and network traffic load.

The invalidation report scheme is typically employed in wireless/mobile networks with little concern regarding its scalability. However, the results we present in this study demonstrate the need to consider alternative cache access strategies for wireless/mobile networks.

This paper focuses on the scalability of cache mechanisms, and compares the performances of the IR with the three fundamental cache access algorithms. Since the comparison among the PER, the CB and the Lease have been studied previously, we will not provide an extensive discussion here of their differences. Section 2 gives a brief review of the IR, the PER, the CB and the Lease schemes. Section 3 presents the performance metrics and the evaluation assumptions. Section 4 compares these schemes. Section V concludes the paper with a summary.

## 2. Cache schemes

Data are stored at the server database where data updates may happen. Data access is initiated at a client. Each client has a cache to store copies of accessed data. In a strong-consistent cache scheme, a data update makes the client cached copy invalid and the copy can not be used by the client any more. This section introduces the IR, the PER, the CB, and the Lease schemes. Each of these four strong consistent cache access algorithms has many variants. Only one representative variant of each scheme will be studied in this paper. We will show the message flows of such representative variants.

### 2.1. Time stamp (TS) invalidation report scheme

In the invalidation report (IR) schemes [1], invalidation reports are sent from the server to clients periodically or aperiodically. An invalidation report contains sufficient information about updates on data items that happened in a past period of time. Any access request from a client will not be processed until the client receives the invalidation report. The client can then use the report to examine whether or not a cached data item is still valid. Many variants of the strong-consistent IR schemes have been proposed [2,3,5–8,10,12,20]. They studied several related major issues such as the sizes of IR, energy-savings, disconnections of mobile terminals from networks, access latency, as well as communication cost. These approaches are to adjust broadcasting intervals, organize contents of