Data Caching in Mobile Ad-Hoc Networks

Narottam Chand
Indian Institute of Technology Roorkee, India

R. C. Joshi
Indian Institute of Technology Roorkee, India

Manoj Misra
Indian Institute of Technology Roorkee, India

INTRODUCTION

Mobile wireless networks allow a more flexible communication structure than traditional networks. Wireless communication enables information transfer among a network of disconnected, and often mobile, users. Popular wireless networks such as mobile phone networks and wireless local area networks (LANs), are traditionally infrastructure based—that is, base stations (BSs), access points (APs), and servers are deployed before the network can be used. A mobile ad hoc network (MANET) consists of a group of mobile hosts that may communicate with each other without fixed wireless infrastructure. In contrast to conventional cellular systems, there is no master-slave relationship between nodes, such as base station to mobile users in ad-hoc networks. Communication between nodes can be supported by direct connection or multi-hop relays. The nodes have the responsibility of self-organizing so that the network is robust to the variations in network topology due to node mobility as well as the fluctuations of the signal quality in the wireless environment. All of these guarantee anywhere and anytime communication. Recently, mobile ad-hoc networks have been receiving increasing attention in both commercial and military applications.

The dynamic and self-organizing nature of ad-hoc networks makes them particularly useful in situations where rapid network deployments are required or it is prohibitively costly to deploy and manage network infrastructure. Some example applications include:

- attendees in a conference room sharing documents and other information via their laptops and PDAs (personal digital assistants);
- armed forces creating a tactical network in unfamiliar territory for communications and distribution of situational awareness information;
- small sensor devices located in animals and other strategic locations that collectively monitor habitats and environmental conditions; and
- emergency services communicating in a disaster area and sharing video updates of specific locations among workers in the field, and back to headquarters.

Unfortunately, the ad-hoc nature that makes these networks attractive also introduces many complex communication problems. From a communications perspective, the main characteristics of ad-hoc networks include:

1. lack of pre-configuration, meaning network configuration and management must be automatic and dynamic;
2. node mobility, resulting in constantly changing network topologies;
3. multi-hop routing;
4. resource-limited devices, for example, laptops, PDAs, and mobile phones have power and CPU processing constraints;
5. resource-limited wireless communications, for example, a few kilobits per second per node; and
6. potentially large networks, for example, a network of sensors may comprise thousands or even tens of thousands of mobile nodes.

A key research challenge in ad-hoc networks is to increase the efficiency of data transfer, while handling the harsh environmental conditions such as energy constraints and highly mobile devices. Presently, most of the researches in ad-hoc networks focus on the development of dynamic routing protocols that can improve the connectivity among mobile nodes which are connected to each other by one-hop/multi-hop links. Although routing is an important issue in ad-hoc networks, other issues such as information/data access are also very important since the ultimate goal of using such networks is to provide information access to mobile nodes. One of the most attractive techniques that improves data availability is caching. In general, caching results in:

1. enhanced QoS at the nodes—lower jitter, latency, and packet loss;

2. reduced network bandwidth consumption; and
3. reduced data server/source workload.

In addition, reduction in bandwidth consumption infers that a properly implemented caching architecture for ad-hoc network can potentially improve battery life in mobile nodes.

BACKGROUND

Caching has been proved to be an important technique for improving the data retrieval performance in mobile environments (Chand, Joshi, & Misra, 2004, 2005; Cao, 2002, 2003). With caching, the data access delay is reduced since data access requests can be served from the local cache, thereby obviating the need for data transmission over the scarce wireless links. However, caching techniques used in one-hop mobile environments (i.e., cellular networks) may not be applicable to multi-hop mobile environments since the data or request may need to go through multiple hops. As mobile clients in ad-hoc networks may have similar tasks and share common interest, cooperative caching, which allows the sharing and coordination of cached data among multiple clients can be used to reduce the bandwidth and power consumption.

To date there are some works in literature on cooperative caching in ad-hoc networks, such as consistency (Yin & Cao, 2004; Cao, Yin, & Das, 2004), placement (Zhang, Yin, & Cao, 2004; Nuggehalli, Srinivasan, & Chiasserini, 2003; Papadopouli & Schulzrinne, 2001), discovery (Takaaki & Aida, 2003), and proxy caching (Lau, Kumar, & Venkatesh, 2002; Friedman, Gradinariu, & Simon, 2004; Sailhan & Issarny, 2003; Lim, Lee, Cao, & Das, 2004, 2006). However, efficient cache replacement is not considered yet.

Cache management in mobile ad-hoc networks, in general, includes the following issues to be addressed:

1. The cache discovery algorithm that is used to efficiently discover, select, and deliver the requested data item(s) from neighboring nodes. In a cooperative architecture, the order of looking for an item follows local cache to neighboring nodes, and then to the original server.
2. The design of cache replacement algorithm—when the cache space is sufficient for storing one new item, the node places the item in the cache. Otherwise, the possibility of replacing other cached item(s) with the new item is considered.
3. Cache admission control—this is to decide what data items can be cached to improve the performance of the caching system.
4. The cache consistency algorithm, which ensures that updates are propagated to the copies elsewhere, and no stale data items are present.

In this article, a *Zone Cooperative (ZC)* cache is proposed for mobile ad-hoc networks. Mobile nodes belonging to the neighborhood (zone) of a given node form a cooperative cache system for this node since the cost for communication with them is low both in terms of energy consumption and message exchanges. In ZC caching, each mobile node has a cache to store the frequently accessed data items. The cache at a node is a nonvolatile memory such as hard disk. The data items in the cache satisfy not only the node's own requests, but also the data requests passing through it from other nodes. For a data miss in the local cache, the node first searches the data in its cooperation zone before forwarding the request to the next node that lies on a path towards server. The caching scheme includes a discovery process and a cache management technique. The proposed cache discovery algorithm ensures that requested data are returned from the nearest node or server. As a part of cache management, cache admission control, Least Utility Value (LUV)-based replacement policy, and cache consistency technique are developed. The admission control prevents high data replication by enforcing a minimum distance between the same data item, while the replacement policy helps in improving the cache hit ratio and accessibility. Cache consistency ensures that clients only access valid states of the data.

SYSTEM ENVIRONMENT

The system environment is assumed to be an ad-hoc network where a mobile host accesses data items held as originals by other mobile hosts. A mobile host that holds the original value of a data item is called data server. A data source may be connected to the wired network. A data request initiated by a host is forwarded hop-by-hop along the routing path until it reaches the data source and then the data source sends back the requested data. Each mobile host maintains local cache in its hard disk. To reduce the bandwidth consumption and query latency, the number of hops between the data source/cache and the requester should be as small as possible. Most mobile hosts, however, do not have sufficient cache storage, and hence the caching strategy is to be devised efficiently. In this system environment, we also make the following assumptions:

- Assign a unique host identifier to each mobile host in the system. The system has a total of M hosts, and MH_i $(1 \le i \le M)$ is a host identifier. Each host moves freely.
- We assign a unique data identifier to each data item located in the system. The set of all data items is denoted by $D = \{d_1, d_2, ..., d_N\}$, where N is the total number of data items and d_j $(1 \le j \le N)$ is a data identifier. D_i denotes the actual data of the item with id d_i. Size of