Computer Networks 57 (2013) 3654–3672

Contents lists available at ScienceDirect

Computer Networks

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

SCALAR: Scalable data lookup and replication protocol for mobile ad hoc networks

Emre Atsan [a,b], Öznur Özkasap [b,*]

[a] School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[b] Department of Computer Engineering, Koc University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 6 April 2012
Received in revised form 8 August 2013
Accepted 12 August 2013
Available online 28 August 2013

Keywords:
Data replication
Data lookup
Scalability
Virtual backbone
Connected dominating set
Mobile ad hoc networks

ABSTRACT

Data replication, as an essential service for MANETs, is used to increase data availability by creating local or nearly located copies of frequently used items, reduce communication overhead, achieve fault-tolerance and load balancing. Data replication protocols proposed for MANETs are often prone to scalability problems due to their definitions or underlying routing protocols they are based on. In particular, they exhibit poor performance when the network size is scaled up. However, scalability is an important criterion for several MANET applications. We propose a scalable and reactive data replication approach, named SCALAR, combined with a low-cost data lookup protocol. SCALAR is a virtual backbone based solution, in which the network nodes construct a connected dominating set based on network topology graph. To the best of our knowledge, SCALAR is the first work applying virtual backbone structure to operate a data lookup and replication process in MANETs. Theoretical message-complexity analysis of the proposed protocols is given. Extensive simulations are performed to analyze and compare the behavior of SCALAR, and it is shown to outperform the other solutions in terms of data accessibility, message overhead and query deepness. It is also demonstrated as an efficient solution for high-density, high-load, large-scale mobile ad hoc networks.

1. Introduction

A mobile ad hoc network (MANET) is a self-organizing, infrastructureless, dynamic wireless network of autonomous mobile nodes. The network is ad hoc since there is no fixed and known network structure that every other node forwards data. Each node is not only an end system, but also a router in the network's multi-hop communication structure. MANETs are adaptive networks, which are reconstructed in the case of network changes due to mobility. Possible application areas are military communication systems, personal area networks, wireless peer-to-peer networks, and vehicular ad hoc networks.

In MANET research, most of the effort has focused on the design of routing protocols that aim to find multi-hop paths between two nodes. Besides routing, accessing to a remote data is equally important since the goal of an ad hoc network structure may be to provide the necessary data items to requester nodes. Different than static, infrastructure-based conventional networks, locating and accessing the remote data (*data lookup*) is a challenging problem in ad hoc environments. In this case, mobile users need to learn the availability of data items in an ad hoc manner without the help of any central server. Moreover, due to the unpredictable mobility behaviors of nodes, a rapid change in the MANET topology is presumable. This change can result in partitioning, dividing the network into isolated sub-networks. Thus, data availability in MANETs is lower than static networks.

One possible solution to this problem can be *replication* of popular data items in the network. In conventional distributed systems, replication not only increases availabil-

* Corresponding author. Tel.: +90 212 338 1584; fax: +90 212 338 1548.
E-mail address: oozkasap@ku.edu.tr (Ö. Özkasap).

1389-1286/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2013.08.012

ity, but also helps for load balancing and fault-tolerance. Also, in geographically widely dispersed systems, having a copy at a nearby location can solve most of the communication latency related problems. On the other hand, it should be taken into consideration that it may not be possible to replicate every data item at each network node due to limited resources or power considerations of the nodes. Therefore, it is important to define good replication criteria and rules that can select the most appropriate data items and best hosts for replication. Another possible solution to the problem of low data availability, or accessibility in mobile networks is *cooperative caching* techniques [1–4]. Cooperative caching for MANETs is the coordination of several nodes to share a cached data in an efficient way for all. Caching-based systems let nodes to cache data or path to a data item in order to increase data availability and achieve lower query delays. Caching reduces communication cost of the system, which consequently results in the reduction of bandwidth and energy usage. In general, caching approaches aim to achieve that neighboring nodes cache different data items. This will increase the data diversity in a region and as a result, more data will be accessible in the network [1]. Furthermore, caching and replication schemes help to establish load balancing among nodes.

Accessing a specific data item in MANETs is an important problem and solutions to this problem can give better results when they are supported with a good data replication or caching approach. On the other hand, possible solutions to this problem can be even more complex if the size of the network increases; which means a possible solution for a meso-scale network can be inefficient for large scale scenarios. Existing protocols for data lookup and replication in MANETs do not consider scalability. In [5], three replica allocation schemes are proposed in which the main parameter for replication decision is the *access frequencies* of data items. However, that study does not consider large-scale networks. Other replication schemes, such as [6–9] also do not address the scalability problem by definition. These solutions can give unexpected results as the system size scales up, and need to utilize additional mechanisms for improvement. We observe that, most of the time, this is because of the dependency to underlying routing protocols which are shown to have scalability problems [10]. In some cases, this performance loss in large scale networks is due to proposed solution's large control message overhead. In small or meso-scale networks the message overhead may not be detrimental, but it affects the performance of the solution in large-scale scenarios.

In this study, we propose a SCAlable data Lookup And Replication framework, called *SCALAR*, which does not depend on the existence of an underlying MANET routing protocol. Our motivation is that a scalable and efficient solution for data lookup and replication in MANETS is essential for several applications. SCALAR consists of three parts: virtual backbone construction, data lookup and reactive replication. It dynamically constructs a virtual backbone structure among the mobile nodes, which is based on an approximation of minimum connected dominating set (CDS) problem in graph theory. CDS-based approach is chosen to minimize the number of nodes in the network involved in data lookup and replication process. Then, our data lookup protocol takes advantage of the dominating set behavior of constructed virtual backbone for low-cost data requests. Lastly, we extend the data lookup protocol with a novel data replication approach, which aims to replicate frequently accessed data items from far away places.

Contributions SCALAR is proposed as a fully distributed algorithm, which operates in a peer-to-peer (P2P) fashion. Different from other data lookup and replication approaches existing in the literature, SCALAR is a complete solution, in which disseminating request packets does not require an underlying routing protocol. Furthermore, SCALAR specifically aims at operating efficiently in large scale ad hoc networks different from other data replication or caching solutions developed for MANETs. To the best of our knowledge, using a virtual backbone structure in order to operate a data lookup and replication process has not been investigated in the literature before. Performance results show that our proposed system outperforms the straightforward solutions for data lookup in MANETs when the scale of the network is large. It is demonstrated that for increasing number of nodes in the network, message complexity of our solution has an acceptable bound. Besides, simulation results present that SCALAR is an efficient solution for high density networks, as well as large scale ones.

The rest of this paper is organized as follows: We review the related work in Section 2. Section 3 presents our SCALAR protocol. Experimental preliminaries are given in Section 4. Performance results and analysis of SCALAR are discussed in Section 5. Comparative results for application scenarios are presented in Section 6. Finally, Section 7 concludes and gives future directions.

2. Related work

In this section, we first review prior studies on data replication and cooperative caching techniques for MANETs, and then demonstrate a feature comparison of these techniques. We also give an overview of controlled-flooding techniques for MANETs.

2.1. Data replication techniques

Data replication is used to avoid data losses in case of unpredictable disconnections of mobile nodes by increasing system wide data availability [11]. Also, replication improves the efficiency by decreasing the number of hops that a data item is transmitted from source to destination. Recent surveys on MANET replication techniques classify protocols according to issues of energy consumption, scalability, and network partitioning [12,13].

One of the well-known schemes for data replication in MANETs is proposed by Hara and Madria in [5] in which the main parameter for replication decision is the *access frequencies* of data items, and each method proposed for replica allocation has been studied for both update-less and randomly updated systems. The following three replica allocation methods are offered for network environments without any update mechanism for data items:

SAF (Static Access Frequency) method: In SAF, each mobile host gives the replication decision based on only its