Computer Networks 59 (2014) 17–32

Contents lists available at ScienceDirect

Computer Networks

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

On the optimal randomized clustering in distributed sensor networks

Ali Dabirmoghaddam [a,*], Majid Ghaderi [b], Carey Williamson [b]

[a] Department of Computer Engineering, University of California, Santa Cruz, CA 95064, USA
[b] Department of Computer Science, University of Calgary, AB T2N 1N4, Canada

ARTICLE INFO

Article history:
Received 30 May 2013
Received in revised form 6 October 2013
Accepted 7 December 2013
Available online 12 December 2013

Keywords:
Energy-efficiency
Clustering
Data correlation
Data compression
Wireless sensor networks

ABSTRACT

Cluster-based data gathering is widely used in wireless sensor networks, primarily to overcome scalability issues. While clustering is not the most efficient means of gathering data, many clustering algorithms have attempted to provide energy efficiency as well. In this paper, we first demonstrate that the general problem of optimal clustering with arbitrary cluster-head selection is NP-hard. Next, we focus on randomized clustering in which sensor nodes form clusters in a distributed manner using a probabilistic cluster-head selection process. In order to find tractable and efficient solutions, we develop a mathematical framework that carefully captures the interplay between clustering and data correlation in the network. We further generalize this model to allow heterogeneous-sized clusters in different regions of the network. According to this model, we observe that clusters tend to become larger further from the sink. We also present simulation results to quantify the energy savings of joint clustering and compression. The results demonstrate that: (1) optimal selection of cluster sizes with respect to the correlation among sensor data has a significant impact on energy consumption of the network and (2) while non-uniform clustering slightly improves the energy efficiency of the network, simple uniform clustering is remarkably efficient and provides comparable results for energy savings.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

A *Wireless Sensor Network* (WSN) is formed by a large collection of cooperative micro-electronic sensing devices that are equipped with wireless communication capability. These autonomous self-configurable networks have given rise to many types of applications, from disaster management to home automation, and from health control to military missions [1]. Some WSN applications require dense deployment of sensor nodes in harsh and remote environments where human access is impossible or inadvisable. Such networks typically have many nodes, because of their geographic size, as well as the need for robustness to node failures. Such deployments require efficient architectures that can easily scale with network size without significant loss in performance.

Clustering is a well-established technique that has primarily been adopted to address scalability issues in WSNs [2]. With clustering, sensor nodes are grouped into small disjoint sets that are coordinated by one of the cluster members known as *Cluster-Head* (CH). The CH is in charge of managing the internal activities of the cluster, such as scheduling nodes for intermittent subject monitoring and data transmission.

Apart from providing a scalable structure, another advantage that clustering can offer is local data compression. Since in most applications, sensor nodes are deployed densely within the environment, significant redundancy is likely to be present among the readings from adjacent sensors. For instance, in a camera sensor network, the same event may be detected by multiple camera sensors in a local neighborhood [3]. Likewise, for temperature

* Corresponding author. Tel.: +1 831 459 5436.
E-mail addresses: alid@soe.ucsc.edu (A. Dabirmoghaddam), mghaderi@ucalgary.ca (M. Ghaderi), carey@cpsc.ucalgary.ca (C. Williamson).

1389-1286/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bjp.2013.12.008

monitoring, measurements reported by proximally-located sensors are likely to be very similar. This dependence can be exploited to eliminate redundancies and reduce the volume of data transmitted in a WSN.

In a cluster-based sensor network, individual sensors transmit their observations to their corresponding CH. The CH compresses the whole cluster data and transmits a representative condensed message (subject to some tolerable distortion level) to the *sink* (the designated fusion center). In this sense, cluster-based data gathering schemes can construct a hierarchy of nodes in multiple levels to route the data from sources to the sink. The most trivial implementation includes a bi-level structure comprising cluster members and CHs. In a similar fashion, CHs can form tier-2 super-clusters whose members are tier-1 CHs and one of them may act as a tier-2 CH as well. Following this strategy, data compression can be performed in multiple levels. However, as we shall see later, with spatial data correlation, the dependency between observations rapidly decays with their geographical distance. As a result, the amount of reduction in message size by applying more levels in the hierarchy would be negligible. Therefore, in this paper, we focus on a bi-level hierarchy in which data compression is only performed at the CH level. The model we develop, however, can be extended to multi-level networks as well.

We should emphasize that *cluster-based data gathering* and *correlated data gathering* have both been extensively studied in the past, though separately. Specifically, the joint problem of *optimal clustering* and *correlated data gathering* is not fully addressed in the existing literature. Once again, it is noteworthy to highlight that clustering is essentially adopted as a means to achieve scalability in large WSNs and in that sense, is not intended to serve as the most efficient method of data gathering in WSNs with correlated data. Besides, as we shall show later, the problem of optimal clustering for minimizing network energy consumption is computationally intractable. Still, viable frameworks can be constructed and optimized to generate clusters that provide maximum energy efficiency while enabling scalability, as well.

There have been a number of works that studied optimal (energy-efficient) clustering, but ignored the effect of data correlation and compression on optimal cluster sizing [4–10]. A pioneering example of energy-aware clustering protocols is LEACH [4] in which each node has a pre-determined chance of becoming CH based on some probability function. The basic idea of LEACH was quickly adopted and extended in many different directions by the research community. EEHC [6], MOCA [9] and GESC [10] for instance are randomized clustering protocols which are based on a similar foundation as LEACH. In all such works, although data fusion is performed to reduce the size of communicated data in the network, no notion of data compression is taken into account while forming the clusters.

On the other hand, some researchers considered optimal data compression in WSNs without explicitly focusing on the clustering aspect of the problem [11–14]. A seminal analysis of energy-efficient correlated data gathering is presented by Cristescu et al. [11]. In that work, the authors consider Slepian–Wolf Coding (SWC) [15], a well-known method of distributed source coding, for which establishing the routing tree is easy, yet the data coding is complex and requires global network knowledge for optimal implementation. The authors prove that joint optimization of rate allocation and transmission structure in distributed networks is NP-complete. Aside from energy-conservation, efficient data gathering has also been investigated from other perspectives, such as minimizing latency (*e.g.*, GroCoca [16]) or improving throughput and scalability (*e.g.*, SelectCast [17] and DDA [18]).

There are only a few sporadic works that study optimal clustering in the presence of data correlation [19–22]. For instance, [19,20] model and analyze various configurations of a simple linear network topology and formulate the optimal cluster size with respect to the number of locally similar observations. Due to the complexities of modeling the joint data compression in correlated data fields, the authors make some simplifying assumptions, *e.g.*, trivial network topologies (linear or grid) and *fixed rate* of data reduction per source after compression, that inevitably influence the reliability and accuracy of the outcomes under realistic situations.

Furthermore, a de facto approach sought after by researchers studying clustering with data compression (*e.g.*, LEACH [4], EEHC [6], NOLBC [20] and MOCA [9] to name a few) attempts to find a *globally optimal* cluster size that minimizes the total network energy consumption. In all such works, for simplicity of model and analysis, the problem has intentionally been restricted to find a *uniform* clustering pattern that results in clusters that contain, on average, the same number of nodes. However, the fundamental question being overlooked here is whether uniform clustering is optimal for total energy consumption. In fact, although all foregoing proposals result in some form of energy-efficient topology, their methodology for tackling the problem inherently lacks the flexibility to form independently-sized clusters in different areas of the network. This paper challenges the existing belief by introducing a comprehensive model that collectively considers the joint impact of all important network attributes in forming clusters.

In particular, in a precursor study [23], we demonstrated that for a simple single-cluster network model, the optimal size of the cluster is directly proportional to its geographical distance from the sink. Such a proposition intuitively promotes a *non-uniform* clustering strategy with larger clusters at further distances from the sink. In this work, for the first time, we examine the foregoing hypothesis under more realistic conditions and establish that although the optimal cluster size grows with the distance from the sink, in practice, uniform clustering – if carefully done – can perform reasonably close to any optimal non-uniform clustering scheme. We verify such unexpected behavior using both mathematical analysis and simulation validation throughout this paper.

In short, our contributions in this paper can be summarized as follows:

- We provide a formal and general definition of the problem of optimal clustering in distributed sensor networks with arbitrary CH selection and prove that this problem is NP-hard.

Download English Version:

https://daneshyari.com/en/article/451056

Download Persian Version:

https://daneshyari.com/article/451056

Daneshyari.com