Ad Hoc Networks 24 (2015) 29–45

Contents lists available at ScienceDirect

Ad Hoc Networks

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

Data similarity aware dynamic node clustering in wireless sensor networks

Fernando Gielow [a], Gentian Jakllari [b], Michele Nogueira [a], Aldri Santos [a,*]

[a] Department of Informatics, Federal University of Paraná, Brazil
[b] Dept. of Telecommunications and Networks, University of Toulouse, France

ARTICLE INFO

Article history:
Received 26 November 2013
Received in revised form 31 May 2014
Accepted 24 July 2014
Available online 7 August 2014

Keywords:
WSNs
Data similarity
Clustering
Protocols
Bio-inspired

ABSTRACT

Wireless Sensor Networks (WSNs) have been used by several kinds of urban and nature monitoring applications as an important interface between physical and computational environments. Node clustering is a common technique to organize data traffic, reduce communication overhead and enable better network traffic management, improving scalability and energy efficiency. Although current clustering protocols treat various kinds of dynamicity in the network, such as mobility or cluster-head rotations, few solutions consider the readings similarity, which could provide benefits in terms of better use of compression techniques and reactive detection of anomalous events. For maintaining similarity aware clusters, the synchronization of the cluster's average reading would allow a distributed and adaptive operation. In this article, we propose an architecture for dynamic and distributed data-aware clustering, and the Dynamic Data-aware Firefly-based Clustering (DDFC) protocol to handle spatial similarity between node readings. The DDFC operation takes into account the biological principles of fireflies to ensure distributed synchronization of the clusters' similar readings aggregations. DDFC was compared to other protocols and the results demonstrated its capability of maintaining synchronized cluster readings aggregations, thereby enabling nodes to be dynamically clustered according to their readings.

1. Introduction

Even though sensors are increasingly common in practical applications, most of them are primitive, when taking only singular and individual data interpretations into account, without establishing further relations between readings. The most usual relations between data readings are referred to as spatial and temporal relations [1]. For example, for readings such as temperature, humidity and lighting sensor readings are likely similar when taken in regions near each other, due to their **spatial relation**. Similarly, successive readings in a single localization tend to vary gradually due to the **temporal relation**.

When exploring and analyzing data readings collectively one could leverage possible relations in the data readings for building more robust applications. In an urban environment, collectively interpreted data can enable streets traffic analysis so that optimal routes can be determined; spatial patterns of temperature readings can be analyzed for locating heat islands for driving improvements in urban planning; audio readings would determine the level of auditive pollution or even map the sound propagation in the environment, etc.

Wireless Sensor Networks (WSNs) have not reached their maximum potential in term of data collection [2]. They have been used as a **communication interface**

* Corresponding author. Tel.: +55 41 3361 3412; fax: +55 41 3361 3205.

E-mail addresses: fhgielow@inf.ufpr.br (F. Gielow), Gentian.Jakllari@irit.fr (G. Jakllari), michele@inf.ufpr.br (M. Nogueira), aldri@inf.ufpr.br (A. Santos).

http://dx.doi.org/10.1016/j.adhoc.2014.07.008

between the physical and computational mediums, formed virtually by many data sets [3]. Hence, they are essential for the Cyber-Physical Systems [4], which rely heavily on an interface between the computation and physical environments, and for the advent of the Internet of Things [5], which integrate several kinds of heterogeneous devices that need environment information [6]. Among techniques that aim to provide better network scalability, clustering techniques organize nodes in WSNs into hierarchical logical groups (clusters), allowing data aggregation and organization of the traffic in the network [7]. Further, through the use of clustering techniques, an organization that keeps nodes with similar readings grouped together could bring advantages such as much more efficient data aggregation. Moreover, such similarity aware clusters enhance the capability of detection of anomalous events.

WSNs are dynamic in terms of topology, routes and positioning of the nodes. Thus, clustering mechanisms should be **adaptive** and **reconfigurable**. Nevertheless, little research in protocols that handle simultaneously the correlation and variation of data has been developed up to the moment. Clustering protocols for WSNs may have several objectives: some aim to handle the dynamicity due to mobility [8,9], others try, sometime periodically, to recreate entire cluster hierarchies [10]. However, few protocols consider spatial data similarity [1,11] and even fewer support the dynamic nature of data using a dynamic clustering approach [12]. Thus, the lack of suitable protocols for handling such data highlights the need for a data aware protocol that creates clusters of nodes with **similar readings** in an adaptive and dynamic way.

There are some design difficulties to be handled to provide an adaptive operation that continuously keeps nodes with similar readings grouped together. A readings similarity aware protocol for WSNs should ideally operate in a distributed and self-organizable fashion, avoiding coordination from the sink and complete re-clustering operations. Such characteristics are commonly found in biological systems, whose principles have often inspired distributed networking algorithms [13]. Although biological algorithms have inspired clock synchronization mechanisms [14], they pose an unexplored potential in other kinds of synchronization tasks [15].

In this work, in contrast to our work in [16], where the proposal was still static lacking an Adaptive Agent, we propose a conceptual architecture for dynamic and distributed data-aware clustering, and a logical organization protocol, named DDFC (***D**ynamic **D**ata-aware **F**irefly-based **C**lustering*), that considers spatial data similarity in dynamic environments. The protocol, utilizing the biological principles of fireflies, groups nodes with similar readings. DDFC synchronizes similar reading aggregations in clusters, supporting their **dynamic maintenance** and **internal routing**, thereby enabling an easy detection of nodes which should be clustered together. DDFC acts between the link and network layers, making use of link layer broadcasts to establish logical clusters and perform intra cluster routing. Hence, the network layer uses the clusters created by DDFC, routing messages between the cluster-heads and the sink. Such data similarity aware clusters enable several kinds of applications in the real world. For instance, with seismic data similarity information, patterns can point to eruptions with some weeks of antecedence [17]; with pollution data similarity analysis, water quality could be estimated and this information could be used for identifying possible areas of contamination and emission.

To assess DDFC's general characteristics and capacity of grouping nodes together, simulations were conducted on the Network Simulator, version 3. Using data readings collected from a real environment, DDFC was compared to a variant and another protocol in terms of cluster-heads stability, readings similarity of nodes clustered together and inconsistent routes. Results obtained prove the efficiency of DDFC in keeping nodes with similar data clustered together and in electing adequate cluster-heads.

Our main contributions consist of (i) the usage of the biological principles of fireflies to synchronize atemporal data, different from traditional approaches that employ fireflies to synchronize exclusively temporal based operations or clocks; (ii) a readings similarity aware clustering protocol which differs from other solutions that focus on more static clusters and dynamic indexing, while DDFC focuses on creating and maintaining the clusters dynamically without considering index based network queries; (iii) a thorough evaluation of DDFC through simulations where it is compared to the best scheme available in the literature; the results demonstrate that our scheme improves its performance by being more stable and by decreasing the number of invalid routes.

The rest of the paper is organized as follows. Section 2 presents the related work. Section 3 presents the principles of fireflies synchronization. Section 4 describes an overview of the data similarity concepts. Section 5 details the DDFC protocol. Section 6 shows its performance evaluation. Finally, Section 7 presents the conclusions and ideas for future work.

2. Related work

WSNs are dynamic in multiple dimensions, such as topology, routes and node locations. Hence, to support the WSN operation, clustering protocols should be adaptive and reconfigurable [18]. Although there are many solutions in the literature for handling several kinds of dynamicities, none performs **data-aware** clustering in a **dynamic** and **scalable** way.

Some protocols handle the dynamicity due to **mobility** by maintaining clusters while nodes arbitrarily transit through the network. The $SPRP_G$ protocol [8], for instance, establishes a spanning tree though a recursive process, in order to establish cluster-heads and gateways to connect the clusters inside the tree. KHOPCA [9] operates proactively through a simple set of rules that defines clusters with variable k-hops. Those rules consider and manipulate a score system, considering a node's neighbors' scores to calculate its own score.

Other approaches support dynamicity through **cluster recreation**, whether periodic or reactive. DCRR [10] considers that clusters are relevant only when there is an event detection, thus being created only on such occasions, while supporting that, continuously maintaining a cluster