ARTICLE IN PRESS
YINCO:4245

Information and Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

The complexity of data aggregation in static and dynamic wireless sensor networks ☆

Quentin Bramas [a,*], Sébastien Tixeuil [a,b]

[a] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu 75005 Paris, France*
[b] *Institut Universitaire de France, France*

ARTICLE INFO

Article history:
Received 31 October 2015
Available online xxxx

Keywords:
Data aggregation
Dynamic graphs
Complexity

ABSTRACT

The contribution of this paper is threefold. First, we give tight bounds for the complexity of the problem of data aggregation in static networks. In more details, we show that the problem remains NP-complete when the graph is of degree at most three. Second, we investigate the complexity of the same problem in a dynamic network, that is, a network whose topology can evolve through time. In the case of dynamic networks, we show that the problem is NP-complete even in the case where the graph is of degree at most two. Third, we give the first lower and upper bounds for the minimum data aggregation time in a dynamic graph. We also observe that even in a well-connected evolving graphs, the optimal solution cannot be found by a distributed algorithm or by a centralized algorithm that does not know the future.

1. Introduction

The growing number of sensor nodes with sensing, computing and communication capabilities, was made possible by recent technological advances. This growth was encouraged by a variety of applications and contributes to the widespread interest in practical and theoretical aspects of wireless sensor networks. Sensor nodes should be inexpensive, small and sustainable in order to be easily deployed in a dangerous area, inside a human body or in vehicles, generally for monitoring applications. They generate data that have to be retrieved by an end-user or a base station. However, the environment and the lack of networking infrastructure do not permit direct transmission to the end user, but only transmissions between sensors that are close to one another. This raises various challenges, such as energy (sensors are battery powered) and delay efficiency (information is relevant for a short period of time only).

In a wireless sensor network (WSN), sensor nodes can communicate through a wireless ad hoc network. Then, there exists a communication link between two nodes if the Euclidean distance between them is smaller than their communication range. Since we assume all sensors to be identical, they have the same communication range and the communication graph

☆ This work was performed within the Labex SMART supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02.

* Corresponding author.
E-mail addresses: quentin.bramas@lip6.fr (Q. Bramas), sebastien.tixeuil@lip6.fr (S. Tixeuil).

http://dx.doi.org/10.1016/j.ic.2016.12.004

Please cite this article in press as: Q. Bramas, S. Tixeuil, The complexity of data aggregation in static and dynamic wireless sensor networks, Inf. Comput. (2016), http://dx.doi.org/10.1016/j.ic.2016.12.004

can be modeled as a unit-disk graph[1] [1]. Sensor nodes typically generate data from their environment, such as temperature, number of vehicles on a road, or number of passenger in a bus. The end-user, called *sink node*, wants to extract this information. To do so, a node can send its data directly to the sink node if it is located within its communication range, or, if it is far from the sink, can use intermediate nodes to relay the data to the sink node.

In this paper we investigate the problem of retrieving data from a WSN whose data transmissions are constrained by two rules: avoiding collisions, and allowing data aggregation. In more details, the time is discretized and, at each time slot, a node is able to send its data to all of its neighbors (*i.e.*, all nodes within its communication range). Now, if two or more nodes send their data in the same time slot, their common neighbors do not receive any data, due to interference. Whenever a node successfully receives data, it aggregates the data with its own and stores the result as its new data. This process ensures energy-efficiency of the protocol. Indeed, n transmissions are sufficient to retrieve the data from n sensors to the sink node (compared to possibly $\Omega(n^2)$ without the capability to aggregate data). The problem of aggregating data from all nodes in the network in a minimum amount of time slots (delay-efficiency), assuming that a node sends its data at most once (energy-efficiency) is known as the *minimum data aggregation time* (MDAT) problem [2]. A solution to this problem consists of a transmission schedule, meeting the communication constraints, with minimum duration.

In this paper we also introduce the dynamic version of the MDAT problem, where individual sensors are now mobile entities. This could model cars evolving in a smart city, medical devices in a body area network, or mobile devices monitoring an area. A WSN whose topology evolves with time is modeled as a *dynamic unit-disk graph*, *i.e.*, a sequence of static unit-disk graphs. In this setting, the communication constraints hold at each time slot, and a solution of the MDAT problem consists of a transmission schedule with minimum duration.

When sensor nodes have fixed positions, the maximum distance (in hops) from the sink node to any other node is a lower bound for the minimum data aggregation time [2]. Indeed, if no collision occurs, the data from the farthest node can be sent through a shortest path. Each avoided collision increases the duration of the schedule by one time slot. However, if we suppose the nodes are moving, avoiding collisions can intuitively have a much greater impact. Indeed, if a collision occurs and we delay the transmission of a node by one time slot, the node may not be able to transmit again (maybe the node remains isolated thereafter). In other terms, the existence of a journey (a path in a dynamic graph) from every node to the sink node is not sufficient to guarantee the existence of a collision-free schedule.

Related work The data aggregation problem we consider here was first studied by Anamalai et al. [3]. The authors assume that a fixed number of channels is available for a transmission, and a collision occurs at a receiver whenever two of its neighbors transmit on the same channel at the same time. The authors propose an algorithm that constructs a collision free convergecast tree that can also be used for broadcasting. Then, Chen et al. [2] present a well-defined model for the study of the MDAT problem in wireless sensor networks. The problem is equivalent to the convergecast problem defined by Anamalai et al. with a unique channel. In the same paper, Chen et al. prove that the problem is NP-complete, even in graphs of degree at most four (more precisely they restricted the problem to networks whose topology is a sub-graph of the grid, which cannot be considered directly as a wireless sensor network). They also gave a $(\Delta - 1)$-approximation algorithm (where Δ is the maximum node degree of the graph).

After the work of Chen et al. [2], a variety of papers proposed centralized and distributed approximation algorithms using geometric aspect of the MDAT problem to improve the data aggregation delay. Yu et al. [4] give a distributed algorithm with an upper bound at $24D + 6\Delta + 16$ (where D is the diameter, and Δ the maximum degree of the graph). Xu et al., in [5] and Ren et al. in [6] propose centralized algorithms with upper bounds at $16R + \Delta - 14$ and $16R + \Delta - 11$, respectively (where R is the radius of the graph). The best known bound is due to Nguyen et al. in [7], as they give a centralized algorithm that takes at most $12R + \Delta - 11$ time slots to aggregate all data.

Related problems such as in-network aggregation [8] focus on an orthogonal perspective. They assume that collisions are handled by the MAC layer, and aim to find routes that minimize the delay. So, those works actually differ significantly from the MDAT problem.

On the other hand, dynamic graphs have received a lot of interest recently and efforts have been done in order to standardize the underlying model [9–11]. Various problems have been studied in a distributed setting, such as designing foremost, fastest, and shortest broadcast algorithms [12,13]. For each problem, sufficient and necessary conditions on the (dynamic) graph are given. The opposite problem of data dissemination (or flooding) has also received a lot of attention in random dynamic networks. Clementi et al. [14] gave an almost tight bound for any random dynamic graphs including edge-markovian evolving graphs and geometric graphs (where nodes follow a random walk or a random waypoint). A conclusion of their work is that increasing the dynamicity of the graph (for instance by increasing the speed of the nodes) implies faster data dissemination, even if the network is at each instant very sparse (due to a short communication range). Most related to our concern are two previous attempts that consider collecting data in dynamic networks [15,16], however they use a much more powerful communication model where no collision occurs. In more details, continuous aggregation [15] assumes that data have to be aggregated, and that the result of the aggregation is then disseminated to all participating nodes. The main metric is then the delay before aggregated data is delivered to all nodes, as no particular node plays the

[1] We suppose here that the area is a two dimensional plane, but our results naturally extend to greater dimensions.