Journal of Network and Computer Applications 40 (2014) 1–11

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

An energy efficient hierarchical clustering index tree for facilitating time-correlated region queries in the Internet of Things

Jine Tang [a], ZhangBing Zhou [a,b,*], Jianwei Niu [c], Qun Wang [a]

[a] China University of Geosciences (Beijing), China
[b] Institut Mines TELECOM/TELECOM SudParis, France
[c] Beihang University, China

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form
8 May 2013
Accepted 24 July 2013
Available online 29 August 2013

Keywords:
Internet of Things
ECH-tree
Time-correlated region queries
Grid cell clustering

ABSTRACT

In the Internet of Things, smart things communicate with each other, and sensed data are aggregated and queried to satisfy certain requests of end-users. When a region of interest requires to be monitored continuously, the strategy that each query is to be executed independently through gathering sensed data of target sub-regions may not be energy efficient, since the values reported by sensors may have no significant difference in proximate sensing time-slots in some applications. To mitigate the energy consumption in this context, in this paper we firstly divide the region with sensor nodes evenly into grid cells, and propose an energy-efficiency hierarchical clustering index tree to organize these grid cells. Then, we develop a time-correlated region query technique for answering continuous queries. Generally, sensor nodes report their values to the base station at the beginning, and report their values only when these values are changed significantly with respect to those reported previously. Queries are answered through assembling the values of interested sensors saved in the base station. Theoretical analysis and experimental results show that our proposal is energy efficient compared with traditional techniques.

1. Introduction

The *I*nternet *o*f *T*hings (*IoT*) (Bandyopadhyay and Sen, 2011; Zeng et al., 2011) has envisioned the interaction and seamless integration of smart things. In this context, smart things communicate with each other, aggregate data in-network, and compose physical or virtual devices into new value-added services for satisfying certain requirements of end-users (Karmouch and Nayak, 2012). Communication is conducted between (i) people and things, and (ii) things themselves (Miorandi et al., 2012; Guinard et al., 2010). Data aggregation (Zhang et al., 2010) has been widely recognized as an efficient method to reduce energy consumption in wireless sensor networks, which can support a wide range of real applications. For instance, when required to monitor (continuously) a region for certain purposes, a large amount of (wireless) sensor nodes are deployed in this region of interest (Wang et al., 2008; AguiarAmaral et al., 2012; Zhang et al., 2013). Sensed data are collected, aggregated, and forwarded to the base station for answering certain queries of interest. Sensed data collected by an individual sensor node may present high *temporal correlation* when its surrounding environment remains relatively stable, which indicates that two successive values may have no significant difference. In this kind of applications, it is required to avoid reporting sensor data without updates, in order to reduce energy consumption.

To mitigate this issue, several rectangle based index methods are developed to perform spatial range queries (Coman et al., 2005; Ai et al., 2009) efficiently in sensor networks. In Soheili et al. (2005), an index tree is built, where the *M*inimum *B*ounding *R*ectangle (*MBR*) of a parent node is derived by merging corresponding child nodes. The sub-region that a node of this index tree occupies is adjusted, for ensuring that upper level regions can have less dead space. A semantic routing protocol (Pathan and Hong, 2007) stores the child node range in the parent node, and forwards query messages through comparing the parent node range with the query range. In Eo et al. (2006), the network structure is organized into a fully distributed spatial index tree, like a routing tree, for distributing sensor nodes into appropriate *MBRs* according to the parent–child proximity relationship. Generally, these index trees can eliminate sensor nodes outside the range of a certain query from the upper levels leveraging the tree hierarchy at the beginning. However, the index tree construction is energy inefficient, since the upper level *MBRs* are formed leveraging the parent–child relations of sensors, and these relations are derived via broadcasting messages. Besides, dynamic index tree adjustment is energy inefficient in most cases.

To support efficient data collection and aggregation, a novel *E*ner*G*y-e*F*ficiency *H*ierarchical *C*lustering index *tree* (*ECH*-tree)

* Corresponding author.
E-mail addresses: tangjine2008@163.com (J. Tang), zbzhou@cugb.edu.cn (Z. Zhou), niujianwei@buaa.edu.cn (J. Niu), qunw@cugb.edu.cn (Q. Wang).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.07.009

based on *grid cell* (Peng and Chen, 2013; Han et al., 2013) clustering is proposed in this paper. Our technique is as follows. First, the whole region of the sensor network is divided evenly into grid cells. Then multiple grid cells are clustered into one sub-region, ensuring that the energy consumption for forwarding messages between them is the smallest. Upper level sub-regions are clustered in a similar manner. This procedure iterates until a tree hierarchy, an *ECH*-tree, is formed. Generally, our *ECH*-tree has less dead space in high-level sub-regions. Leveraging the *ECH*-tree hierarchy, we propose a time-correlated region query method for answering continuous queries while mitigating the energy consumption regarding these queries. Generally, all sensor nodes are required to report their values to the base station at the beginning. They report the last values continuously, only when these last values of certain sensors are significantly different from the values reported to the base station previously. Queries are answered through assembling the values of interested sensors saved in the base station. Theoretical analysis and experimental results show that our technique is energy efficient, especially when the region is mostly interested by end-users, and queries are to be answered continually.

This paper is outlined as follows. Section 2 presents the energy model which is applied in the rest of this paper. Section 3 introduces our technique for constructing an ECH-tree in the context of skewness distribution. Section 4 develops our time-correlated region query technique for facilitating queries continually. Section 5 presents our implementation and evaluates our time-correlated region query technique. Section 6 discusses related techniques, and finally this paper is concluded in Section 7.

2. Preliminary: energy model

Table 1 presents the parameters of energy model to be used in the following sections, and this energy model was introduced in Tan et al. (2011). We assume that the sensors have the same communication radius.

The parameters E_{tx} specifying the energy consumed to transmit a k bit packet to a distance d, E_{rx} specifying the energy consumed to receive a k bit packet, and $C_{ij}(k)$ specifying the energy for transmitting a k bit packet from a node i to its neighboring node j, are computed as follows:

$$E_{tx}(k,d) = a \times k + b \times k \times d^n$$

$$E_{rx}(k) = c \times k$$

$$C_{ij}(k) = \begin{cases} a \times k + b \times k \times d_{ij}^n & \text{if } j \text{ is } BS \\ E_{tx}(k,d) + E_{rx}(k) & \\ = (a+c) \times k + b \times k \times d_{ij}^n & \text{otherwise} \end{cases}$$

3. ECH-tree construction

This section presents our ECH-tree construction technique for indexing spatial sensor nodes in the skewness distribution. Section 3.1 introduces the ECH-tree construction procedure based on grid cluster (Al-Turjmana et al., 2013), and Section 3.2 gives the weight computation between two grid cells and two cluster sub-regions.

3.1. ECH-tree generation algorithm

This section presents the construction procedure of our ECH-tree. Different from (Ai et al., 2009) whose index is based on even grid division of sensor nodes, our approach assumes the skewed distribution of nodes, which means that sensors are dense in some regions, while are sparse in some other regions. A general example of skewed distribution is shown in Fig. 1, which contains four relatively dense sub-regions.

Algorithm 1 shows the ECH-tree construction procedure when sensor nodes are distributed unevenly. Firstly, the whole region is divided evenly into grid cells, where any grid cell is a square and the side-length of each grid cell is $\sqrt{2}r$ (line 1). r is the communication radius of sensors. The communication radius of sensors is assumed the same in this research. An example of this even division is shown in Fig. 1. We choose the side-length as $\sqrt{2}r$, since according to the evaluation performed in Forster et al. (2009), the optimal cluster size is 1 hop in dense regions, and the size should be larger in sparse regions. Then, the gird cells are grouped into several clusters (lines 2–3). The new formed cluster sub-regions are clustered recursively (lines 5–8), until the root cluster is formed (line 9). Algorithm 2 details the grid-cell clustering procedure.

Line 10 selects head nodes for each sub-regions using the *LEACH* method developed by Heinzelman et al. (2002) and Tyagi and Neeraj Kumar (2013). Generally, cluster heads are selected according to the probability of optimal cluster heads decided by the network. Since being a cluster head node is much more energy intensive than being an average head node, it requires that each node takes its turn as cluster head rotatively. *LEACH* incorporates the randomized rotation of the high-energy cluster head position among sensors to avoid depleting the energy of any sensor. Therefore, the energy load of being a cluster head is evenly distributed among the nodes in a sub-region. As an example, Fig. 2 shows the head nodes of grid cells, and new formed cluster sub-regions, of our ECH-tree shown in Fig. 4. We use the notion p to indicate the selected header nodes. The specific header nodes are denoted in the corresponding grid cells and new formed cluster sub-regions in Figs. 3 and 4.

Algorithm 1. ECHTreeConstr.

Require: SN_{set}: set of sensor nodes
r: communication radius of sensors
Ensure: nd_{rt}: the root node of ECH-tree

Table 1
Parameters in energy model.

Parameter name	Descriptions
a	Energy consumption constant of the transmit electronics
b	Energy consumption constant of the transmit amplifier
c	Energy consumption constant of the receiver electronics
k	The number of bits in a packet
d	Transmission distance
r	Communication radius of sensors
E_{tx}	Energy consumed to transmit a k bit packet to a distance d
E_{rx}	Energy consumed to receive a k bit packet
$C_{ij}(k)$	Energy for transmitting a k bit packet from a node i to its neighboring node j