Ad Hoc Networks 36 (2016) 203–213

Contents lists available at ScienceDirect

Ad Hoc Networks

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

An efficient adjustable grid-based data replication scheme for wireless sensor networks

Tzung-Shi Chen [a], Neng-Chung Wang [b,*], Jia-Shiun Wu [c]

[a] Department of Computer Science and Information Engineering, National University of Tainan, Tainan 700, Taiwan, ROC
[b] Department of Computer Science and Information Engineering, National United University, Miao-Li 360, Taiwan, ROC
[c] Department of Electrical Engineering, National University of Tainan, Tainan 700, Taiwan, ROC

ARTICLE INFO

Article history:
Received 24 April 2014
Revised 17 April 2015
Accepted 6 July 2015
Available online 13 July 2015

Keywords:
Query
Data replication
Load balancing
Virtual grid
Wireless sensor networks

ABSTRACT

Wireless Sensor Networks (WSNs) consist of hundreds or thousands of sensor nodes connected to each other through short-distance wireless links. Because sensor nodes are energy-constrained, efficient energy use is a critical issue for WSNs. In WSNs, the energy of popular nodes may be quickly depleted when a large number of sensor nodes are interested in the same popular events, and they are frequently queried. This paper proposes a data replication scheme called adjustable data replication (ADR), which is based on a virtual grid in order to improve the lifetime of data nodes. In each grid, a head node is selected as a manager, and is responsible for receiving or transmitting packets from or to other nodes in the same virtual grid. The head nodes around data nodes or replica nodes will compute the number of requested packets in order to determine how best to build a replica node in an appropriate location. ADR repeatedly builds data replica nodes close to query nodes in order to balance the overhead and energy consumption of sensor nodes. Simulation results show that the proposed ADR scheme outperforms existing approaches.

1. Introduction

Wireless Sensor Networks (WSNs) consist of hundreds or thousands of sensor nodes connected to each other through short-distance wireless links [1]. In a WSN, energy conservation is a primary concern, since its sensor nodes operate on small batteries with limited energy, thus limiting the lifespan of the nodes. A greater number of sensor nodes allows for sensing over larger geographical regions, with greater accuracy. WSNs have been extensively used in many applications, such as environmental monitoring, vehicular ad-hoc networks and battlefield surveillance [2–5]. Sensor nodes measure ambient conditions in their surrounding environments and then transform these measurements into signals that can be processed to reveal characteristics of phenomena located in the area around the sensor nodes. Since sensor nodes have significant power constraints, energy efficient methods must be employed to prolong network lifetimes [6]. The development of a means of effectively and efficiently controlling sensor energy is thus an important and challenging task. A WSN may consist of a large number of sensor nodes for data sensing, data processing and communication components. Minimizing communication in query and sending execution can save a significant amount of energy, and help to prolong the lifetime of a network.

In WSNs, all In-Network Storage schemes proposed to date have been in Data-Centric Storage (DCS) schemes [7–11]. The DCS schemes store sensor data readings and to process a query efficiently. DCS schemes may have the same phenomenon when query requests or storage requests [12] continuously access the same area or node. The data are stored or accessed by every kind of routing protocol [13,14].

* Corresponding author. Tel.: +886 37 381899.
E-mail addresses: chents@mail.nutn.edu.tw (T.-S. Chen), ncwang@nuu.edu.tw (N.-C. Wang), mcnknight@gmail.com (J.-S. Wu).

http://dx.doi.org/10.1016/j.adhoc.2015.07.002
1570-8705/

In real-world application, data in WSNs cannot be uniformly distributed. This study deployed sensor nodes in a disaster environment, and used sensor nodes for data storage, those sensor nodes are also called data nodes. Every sensor node in a WSN can be represented as a query node. When the data node receives query packets from query nodes, it sends data to those nodes. By using the DCS mechanism, nodes can determine where required data is. However, when the query frequency is larger than the storage frequency, the data node must expend more energy on replying the query data. This kind of increased query request frequency may result in the rapid depletion of the data node's power. This kind of the phenomenon is called the query hotspots problem.

This paper proposes a data replication scheme called Adjustable Data Replication (ADR). This scheme uses a virtual grid to mitigate traffic overloading in hotspots. When the data traffic rate of a data node increases, the node will set itself as a hotspot, and collect statistical data on the query frequency from neighboring nodes. The data node will then send a replication packet to neighboring nodes, and the replicated data will be the same as the original data. The neighboring nodes will receive the original data, and are called replica nodes. When replica nodes begin to receive fewer queries, this scheme will dynamically remove the replica nodes. ADR is thus able to use a few replica nodes to maintain a routing tree, and conserve node energy in a WSN.

The contributions of the proposed scheme are given below:

1. The ADR scheme uses a few replica nodes to maintain a routing tree reducing energy consumption of data nodes.
2. The ADR scheme uses a virtual grid to mitigate traffic overloading in hotspots to balance the overload and energy consumption of sensor nodes.
3. The ADR scheme has good performance in terms of the metrics of the number of replica nodes, the number of packets, the energy consumption, and the query latency.

The remainder of this paper is organized as follows. Section 2 will review related work. Assumptions on environment and the network model are introduced in Section 3. Section 4 will describe the proposed data replication scheme for WSNs. Simulation results are described in Section 5. In Section 6, conclusions are drawn.

2. Related work

In the DCS routing protocol, nodes requesting data may encounter an over-emphasis on data storage and data replication [8,11]. Sensor nodes may forward larger data to the next hop, while sensor nodes are positioned close to the sink or the large query target. When sensor nodes constantly forward data to a static sink which uses their nominal communication range, the energy sink-hole problem will occur at the sensor nodes nearby the sink [15,16]. A number of studies to date have proposed various solutions to this problem.

Cheng proposed a Grid-based Data Replication algorithm (GDR) [17], used to mitigate the query overload of a target area. GDR is a strategy that reduces query hotspots based on a virtual grid structure. In GDR, each grid has a head node that is responsible for receiving or transmitting packets from or to other nodes in the same virtual grid. The head nodes around data nodes or replica nodes will compute the number of requested packets in order to determine data replication. When the average query frequency is greater than the overload threshold, data nodes will produce a replication, and place it in head node.

Ishihara and Suda proposed a novel replica arrangement scheme called Dynamic Replica Arrangement on Concentric Circular Arcs (DRACA) [18]. DRACA is based on concentric circles, and builds new replica nodes in fixed target positions. In DRACA, replica nodes of a data item are arranged so that they are placed on nodes at closer locations from where queries are frequently sent. Replica nodes make nodes on the same circular arc centering the hashed location of the key, which has a pointer to indicate the position of the replica node. On the way to the hashed location, a query message may arrive at a node with a pointer pointing to a replica node. The node then forwards the query message to the location of that replica node. If the replica node receives the query message, it sends a reply message to the query node. If there are replica nodes and pointers near a query node, the communication cost required for sending the query and reply message is small. In DRACA, replica nodes are arranged according to the geographical distribution of the frequency of queries. New replica nodes and pointers are arranged based on whether and where the frequency of queries increases.

The basic concept of DRACA is to adaptively arrange replica nodes at positions close to query nodes that frequently send queries. The basic idea of GDR is that the first replica node can distribute almost half of the query frequency. However, DRACA and GDR have a major drawback, in that the replica nodes cannot be dynamically removed. DRACA must first remove the outermost nodes until the target node is at the outermost position, and only then can the target node be removed. GDR is also very restrictive in this sense, because the back produced replica node is based on the boundary of the former replication. When removing replica nodes, GDR must first remove the tail-end replica node in order to maintain the entire system. However, end replica nodes are often more significantly responsible for dispersed flow and reducing the distance between replica and query nodes.

3. Assumptions and network model

In this section, some basic assumptions are first introduced, and the proposed network model is presented.

3.1. Assumptions

This section presents the basic design of the proposed method with the following basic assumptions: the monitoring field is assumed to consist of a large number of homogeneous sensor nodes which communicate with each other through short-range radio signals. Each sensor node is aware of its own location, for example, through Global Positioning System (GPS) signals [19], or through techniques such as [20,21]. In the proposed scheme, each sensor node stores data, and the query methods are based on the Geographic Hash Table (GHT) for data-centric storage [11]. All packets forwarded by sensor nodes use Greedy Perimeter Stateless Routing for wireless networks (GPSR) [22]. Mobile issuer