Mining Sensors in Wireless Sensor Networks

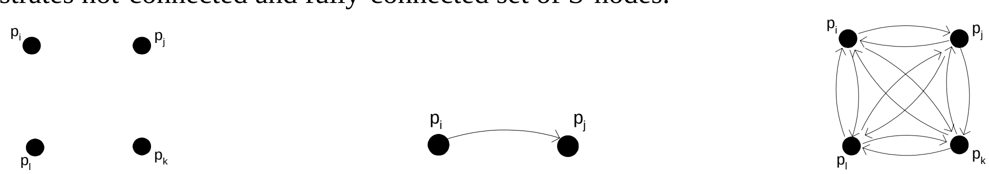
Vijay Kumar
Computer Science
University of Missouri-Kansas City
kumarv@umkc.edu

Problem Statement: To find as quick as possible a subset of sensor nodes satisfying a set of desired criteria in a highly populated sensor network through data mining approach. Current approach is inadequate to handle the mining of massive sensor network.

Justification: A sensor is commonly viewed as a programmable, low-cost, low-power, functional tiny mobile or stationary device which usually has a much shorter working life span [1]. The current sate of the deployment of sensor nodes (S-nodes) seems to be ubiquitous. Starting from small objects such as toys to very large objects (and systems) such as national highway infrastructure has large number of embedded S-nodes. In fact any kind of automation relies heavily on a set of programmed S-nodes. For example, programmed set of S-nodes drive fire-detector, distance sensor automatically flushes toilet, immersive sensors monitor human and animal movements, etc. In fact, there are very few units which do not use S-nodes to perform their functions. We argue that since a sensor is programmed to do a specific task, it has its own semantics. Thus, with its commonly accepted functional properties and its semantics, we regard a sensor also as a data point in an information space. This allows us to represent a sensor with a database relation and create a sensor database with sensor nodes (S-nodes) as data items. We try to establish that a S-node and a database data item share some important common properties and as a result the former can be processed similar to a data item. This model presents us a suitable platform, similar to a database, for the applying data mining technique for sensor node identification.

Definition 1: Embedded Sensor Data Space (ESDS) is a countably infinite set of uniquely programmed network of S-nodes. Thus, $ESDS = \{s_1, s_2, ..., s_\infty\}$ where s_i ($i = 1, 2, ..., \infty$) are programmed S-nodes and s_i and s_j ($i \neq j$) are fully connected. As a result, we regard *ESDS as a database of S-nodes and with embedded semantics*.

Definition 2: A pair of S-nodes is fully connected if they are the nodes of a directed cyclic graph. Thus, in an ESDS every pair of S-nodes has a cycle which is essential for ESDS to captures data of its environment and dispatches it to other S-nodes through routers [2, 3, 6, 7, 8, 9, 11]. Figure 1 illustrates not-connected and fully-connected set of S-nodes.



None-connected One-way connected Fully connected

Figure 1. None and Full connectivity

Definition 3: An Embedded Sensor Net (ESN) is a fully connected wired or wireless or both network that implements ESDS. Thus, $ESN = \{S, L\}$ where $S = \{s_1, s_2, ..., s_\infty\}$; s_i ($i = 1, 2, ..., \infty$) are programmed S-nodes and $L = \{l_1, l_2, ..., l_m\}$ where l_i is a pair of directed arcs between a pair of S-nodes creating a cycled.

Definition 4: A S-node is a mobile or stationary processing unit which is a pair <*C, A*>; where C is a set of its processing and communication capabilities and set of attributes $A = \{a_1, a_2, \ldots, a_n\}$; where a_i a property such as size, location of deployment, type, etc.

Sensor mining: It is highly realistic to assume a huge ESDS with millions of S-nodes each with their own semantics. Figure 2, illustrates an ESS, with a large number of S-nodes with varying capability and properties. S-nodes are fully connected with wired and wireless connections. This setup must be managed efficiently and to do so the criteria (a) all S-nodes must be active, (b) wireless and wired connection among them must be up, (c) any kind of failure and failure to a S-node, software system, and communication must be repaired immediately, and (d) any require software update must be done without delay.

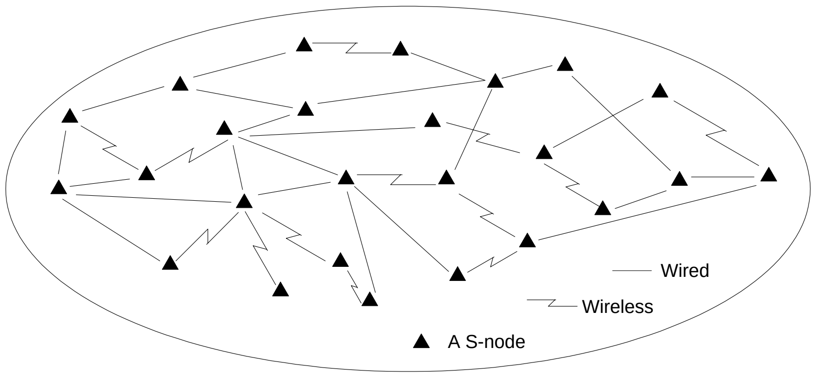


Figure 2. An ESDS with a large number of S-nodes.

To meet the set criteria listed above we must have the following information. We want to apply data mining techniques to discover (a) what set of S-nodes needs software upgrade, (b) what set of S-nodes will need battery replacement at the end of the month, (c) what are the geographical locations for S-nodes that monitors the temperature of the building, and so on.

Identification of S-nodes at present is done by probing individual node. Each S-node is fitted with some notification device through which the node announces its coordinates and its status. This approach works but it has the following problems:

1. A dead S-node cannot announce its location.
2. If every S-node announces its location, then this is likely to create congestion.
3. The available channels, especially wireless, cannot support such communication traffic.
4. Identification of a central S-node which will listen to these announcements.

These problems motivated us to exploit the power of conventional data mining to find the desired set of S-nodes. We argue that a need-based mining in ESN is what we need to manage it efficiently and to trace an individual sensor for monitoring its functionality, repairing dead sensors, reprogramming a set of sensors, and so on.

1. Vijay Kumar, "Sensor: The Atomic Computing Particle", *ACM Sigmod Record*, December 2003.