Researching on Speculating Algorithm Based on End-To-End Date in Wireless Sensor Networks

Ji Wenli[a], li Xiangjun, Jiang Meiying

School of Communication Engineering, Xi'an Univ. of Science and Technology, Xi'an, P.R.China

[a] wenlixa@gmail.com

Abstract. In this paper, we focus on how to gain network topology information of wireless sensor network (WSN) with higher accuracy. It is very important for Network planning and management. We propose a topology identification algorithm based on data fusion system in WSN. Firstly, based on the information of packet delay and/or packet loss, the algorithm got the collections of approximate ancestors of each node, according to the classification inference algorithm of graph theory. Secondly, it identifies parent-child relationship of the nodes by calculating the Hamming distance between the current node and approximate ancestors' nodes; and infers the topology of the network layer by layer. The proposed algorithm doesn't require support from internal nodes. It employs end-to-end measurements and does not incur any additional burden on the network. NS2 simulation results show the high accuracy and efficiency of the proposed algorithm.

Introduction

The Wireless Sensor Network (WSN) consists of a large number of intelligent tiny Sensor nodes in monitoring area, which forms a multi-hop and self-organizing wireless networks. It serves the vital purpose of sense and data acquisition and processing [1]. With the propagation of WSN in the machinery and military field, how to measure network performance metrics and system performance evaluation is coming focus of study. To better do network maintenance and optimization of network performance, subscribing to several real time condition parameter of network must be made. Network topology is one of the most important methods in a WSN's performance [2]. But it is more challenging than the traditional computer network due to their own network dynamic characteristics and capacity constraint and so on. To sum up, the existing topology inference methods all depend on the collaboration among the internal nodes, and increase the burden on the network. Especially for the autonomous system which is not open to outside, it's difficult for measurements within the network to achieve the collaboration and information exchange of internal nodes, and can not guarantee the accuracy of measurement.

Currently, a new network measurement has been proposed internationally, which is called Network Tomography (NT).Network tomography (NT) introduces the medical computed tomography (CT) into the network measurement. It analyzes and infers the network performance and topology of the internal according to the measurement data of network boundary nodes. Using this technology can speculate network logical topology by End-to-End measurement, through the analysis of mutuality of sensor without the need for the collaboration of network internal node [3].

As to the analysis above, in this paper, we proposed a topology identification algorithm based on Network Tomography in sensor network. Firstly, using the information of packet delay and/or packet loss, relevance of sensors is been determined. Algorithm excludes part of non-ancestor node of each node, and then gets the collections of approximate ancestors of each node. Secondly, by calculating the hamming distance between the current node and approximate ancestors nodes, the nodes with parent-child relationship can be identified between the current node and approximate ancestors. The nodes which has same parent node are brother of one another. By this way, each node can be inferred from the parent node, so that the entire network topology can be estimated. Received data of sink node is only used to infer network topology as a useful data, and did not incur any additional burden on the network

System Structure

Because bandwidth and energy of sensors network are limited, for mass deployment of wireless network to become a reality, there are large number of data redundancy. The method of data aggregation is often adopted in the process of collecting data, it sometimes called data fusion.

Data aggregation. In wireless sensor network, some nodes of network can transmit the collected data to node sink through an inverted converge tree. As shown in figure 1, node 0 is a sink node, which transfer collected data through network to task management node. When nodes have collected the data transmitted by all the child nodes or waited for timer overtime, transmit perception data of all child nodes and their own to the next hop node, at the same time the information on nodes aggregated in aggregation data is also transponder to the next hop node with aggregation data. Boolean variable is described whether perception data reach sink node during data fusion. And in each turn data collection, perception data of internal node will reach sink node, which can be described as vectors [4]. Calculation and storage resources of the node are added during data fusion, but it can reduce the traffic so as to save node energy.
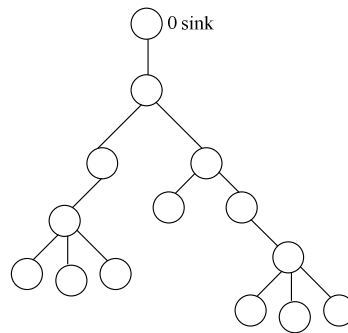

Figure 1. An inverted converge tree

Lost package model. A handstand aggregation tree can be shown, which is used to describe the logic network topology formed in data aggregation process, as shown in figure 1. If a link packet loss process is an independent Bernoulli process, the probability of date successfully transmitted from node to sink node is defined as $P1$, and the probability of packet loss is s defined as $1-P1$. Assume that network T contain n nodes, where node sink which is expressed by s, is the root node of T. The mess starts here. $<i, j>$ is been used to indicate the link from node i to node j, or j is the ancestors of i. data flow through handstand gathered tree can be described as $Z = (z_{i,j}), i \in d(j), j \in V$,which is a random process. $z_{i,j} = 1$ Shows data sent from node i successfully reaches the node j. Boolean variable x_i is described whether perception data of this round node i reach sink node. $x_i = 1$ Suggest perception data of node i reach sink node in each turn data collection. Sink node know which data arrives after completion of a preceding operating cycle. In n one-round data collection, in sink node a sequence of every node, whose length is N, will be maintenance, written as:

$$X_i = \{x_i^{(m)}, 1 \le m \le N\} \tag{1}$$

Hamming distance. In the above basis of lost model ,Hamming Distance is a simple and effective method to show the similarity between nodes . Hamming Distance is named after Richard Hamming. In information theory, the Hamming distance is the number of positions in two strings of equal length for which the corresponding elements are different

   In the network topology based on NT technology that introduced into the algorithm hamming distance, called the two nodes received a message of the sequence Hamming distance. Hamming distance of Node i and j is defined as, in N wheel data collection, the convergence of the two nodes in receiving node message lost or reception of 0-1 sequences, and the same round the receipt or lose the number of different, the formula is as follows:

$$H_d(i,j) = \sum_{m=1}^{N} x_i^{(m)} \oplus x_j^{(m)} \tag{2}$$

Topology Speculate Algorithm Analyze

The current network topology inference employs a property on the whole, and different parameters also adapt to different network environment. For example, the method merely depends on package loss rate is preferable when the network load is high. The proposed algorithm excludes part of non-ancestor node of each node, and then gets the set of approximate ancestors of each node, according to the classification inference algorithm of graph theory and the information of packet delay and/or packet loss. And then in the same division, the calculation of the current node i and approximate ancestors set node of the Hamming distance, can be used to infer the current node's father node written as j. Thus the whole the topology of the network may be inferred.

Network node layered. According to the analysis of lost package model above, in each turn of the process of data collection, each node can wait for a short time in order to send a new message that contains itself sensor-data and another data receiving all child nodes to the next hop node by protocol. There are ultimately message information of all nodes in the sink node, which contains all child node information and hop information. If I is behind j and $X^m_i=1$ and $X^m_j=1$ in the link of $R_i=\{ X^m_{i,1}=<m<=n \}$, this makes it impossible to determine that j is the ancestors of i. But If i is behind j and $X^m_j=0$ and $X^m_i=1$, we can confirm that j is not the ancestors of i. In this way, each node can get a similar set of ancestors written as P.

Network topology inference. For these approximate ancestors, by defining the hamming distance between the nodes it can confirm that the hamming distance between parent node and child node is much less than the one between the nodes without the parent-child relationship [5]. It was upon this principle that the current node's parent - j will be recognized, then register node j from approximate ancestors node. In the rest of the nodes j node is current node, the same thing is done until farther node of all node is recognized. Note that the nodes with the same parents are brother.

Algorithm description. For simplicity we assume that the node set is usually expressed as U, data sequence of observation is expressed as X. The proposed algorithm is described as follows:

   Step 1: Calculate the collections of approximate ancestor set p of each node based on the packet loss information.

   Step 2: In node set U, choose node i. In the approximate ancestor set of i, calculate the Hamming distance between node i and other node. Select the node j, which has the smallest hamming distance with the node i, as the parent node of i. And then the node j is marked in P.

   Step 3: The hamming distance between node j and any node of node set p is calculated. Select the node k, which has the smallest hamming distance with the node j, as the parent node of j. And then the node k is marked in P. Repeat step 2 until node set p is all marked. The node i is removed from node set U.

   Step 4: Choose any node from node set U, and repeat the above step 2 and step 3 until the node set U is empty. Note that the nodes with the same parents are brother. Finally, the logical topology of sensor network is inferred.

The Simulation Experiment and Analysis

Using NS2 simulation tool, the proposed algorithm was verified effectively in two networks, which have 20 nodes and 200 nodes, respectively. By collecting data in accordance with the sequence of data transmission, the inference algorithm was realized and the collected data were analyzed using Matlab-software. On host computer with the frequency of 2.8GHZ, a simulation program was run, and the time of inference from beginning to end inference was measured, as shown in the table 1.

Table 1:   Inference running time of the algorithm

The scale of network	Data collection rounds/time	Time/s
20	60	2.0
200	180	4.2

It is also clear from table 1 that it needs less 5 seconds to infer 200-node sensor network, which show that the proposed algorithm is very efficient. In 200-node of the network topology, when data gathered round up to control 180 times, the sensor network topology is inferred correctly. The bigger the network needs more data collection rounds to infer its network topology correctly. As the network scale increases, the variation of execution time of inference algorithm is smaller, which shows that the algorithm do not increase in proportion with the network size

Summary

Mathematical theory from the correlation of the nodes in the network nodes can reflect the characteristics of share links, and combining with the node of the Hamming distance. This paper proposed a new topology algorithm. Through simulation, the conclusion is that this algorithm can infer the topology of the network accurately by network tomography technology.

Acknowledgment

References

[1] Jennifer Y,Biswanath M,Dipak G .Wireless sensor networks survey[J].Computer Networks,2008, 52(12) :2292-2330．

[2] zhang zhiyong,hu guangmin.Novel topology inference algorithm for wireless sensor network [J].Journal of computer Application,2010,30(3):733-744

[3] zhao honghua,chen ming.Topolpgy inference based on tomography technology[J].Computer engineering,2009,35(2):91-94

[4] Zhao T, Cai W D, LI Y J. MP IDA: A sensor network topology inference algorithm [C] .2007 ternational Conference on Computational Intelligence and Security. Washington, DC: IEEE Computer Society, 2007: 451- 455

[5] Zhao Tao, Cai Wandong, Li Huixian.,Sensor network level—topology inference based on Hamming distance.[J] J.Huazhong Univ.of Sci. & .Tech.(Natural Science Edition). 2008(10),V01.36 NO.10:71-76

[6] ZHANG Zhi-yong,HU Guang-min. Novel topology inference algorithm for wireless sensor network.[J] Journal of Computer Applications,30(3):733-736. (In Chinese)

[7] Yongyi Mao, Kschischang F.R., Baochun Li, Pasupathy S. A factor graph approach to link loss monitoring in wireless sensor networks.IEEE Journalon Selected Areas in Communications, vol.23, no.4, Apr 2005:820-829.

[8] Yongjun Li,Wandong Cai.Wireless Sensor Network Topology Identification based on Data aggregation [J], Journal of Computational Information systems, 2007,456-458.

[9] Zhao Tao, Cai Wandong, Li Yongjun. Using end-to-end data to infer sensor network topology[C]. In Proc. of IEEE international symposium on signal processing and information technology 2007, 504-508. .(in Chinese)