Optimal and Robust Data Aggregative Fusion in Internet of Things for Data Collection on Equipment Status

Feng Xie [a], Xiaohui Ye [b]
Naval University of Engineering, Wuhan, 430031, P.R. China
[a] *fengxie2014@163.com*, [b]*whyxh007@yahoo.com.cn*

Abstract

Internet of Things has been regarded as a promising technology and architecture for instant maintenance of equipment status in large marine ships. Data aggregation and data fusion are essential operations in sensing data collection. The energy efficiency of aggregative tree affects the lifetime of the data collection, which should maintain robustness and optimization. Current researches have not explored how to distinguish and rely on the node energy potential for tree construction. In this paper, we propose a bunch of algorithms by using random network for data aggregation. The energy efficiency is improved by our algorithms, which is justified extensively by analysis and proof.

Keywords: Data Aggregation, Data Fusion, Optimal Algorithm, Internet of Things, Robust.

1 Introduction

Internet of Things (IoT) [1-2] recently is largely applied in smart sensing of machine parameters and surrounding data. The number of equipments in marine ships usually is large, the maintenance of those equipments is difficult, and usually thus relies on the status monitoring, automatically and smartly. IoT can be utilized in this scenario by taking its advantages in equipment status monitoring. It can collect the working parameters about equipments, gather the operational surroundings about the machines, and report the fusion information to central decision server for further response.

The data fusion and data aggregation in IoT is a critical problem having been researched for many years [3-4], and it still attracts attentions. Data fusion on equipment maintenance in large marine ships is however discussed extensively, as the application of IoT on ships is still under researches. The major concern in IoT is the energy efficiency [5-9]. The energy efficiency of data fusion and data aggregation should be satisfied in the design of fusion algorithms.

Sensing data indicating the working status of the equipments usually need to aggregate for efficient transmission. It thus rises a question – how to guarantee the robustness of the data fusion and energy efficiency with optimal performance. In this paper, we tackle the problem from random network and node energy awareness. We propose an algorithm that utilizes random network to generate the aggregative tree by knowledge of node energy potentials, and make use of the energy potential to find the optimal and robust fusion tree in terms of longer lifetime.

The contribution of this paper is as follows: We make the first attempt to propose random network algorithms for data aggregation. We also formally discuss optimization and robustness problem for data aggregation in IoT.

The rest of the paper is organized as follows: the Problem Formulation section will address the normalized problem on efficient data aggregation of sensing data. The Proposed Fusion Scheme section will propose dedicated algorithms. The Analysis section will give the extensive analysis. The Conclusion section concludes the paper.

2 Problem Formulation

Definition 1: Sensing Node (SN). It is the node that generates data.

Definition 2: Random Network (RN). It is the graph consists of multiple vertexes and paths. The vertexes represent SNs. The paths represent that the communication links are available between the vertexes. The topology of the network may change.

Definition 3: Node Energy Index (NEI). The node has an index that evaluates the energy potential. The index is related to three folders as follows:

1) The data volume of the node, which is the sending data volume over the time, denoted as D.

2) The sending probability of the node, which is the sending rate over time, denoted as P.

3) The remaining energy percentage of the node, denoted as E.

Usually, NEI = K*E/(D*P), where K is a system parameter.

Definition 4: RN Energy Index (REI). It is the total count of NEI of each node in RN, which is denoted as REI.

Definition 5: The Robustness of RN (RRN). It is measured by RNEI. That is, given RN<V<D,P,E>>, the total count of NEI for each SN∈RN.

Definition 6: The Optimal Problem of RRN (OPRRN). Given RN, find the best aggregation node in RN, such that RRN is the largest.

Hereby the research problem in this paper is thus to find an efficient algorithm for solving **OPRRN**.

3 Proposed Fusion Scheme

Before we give the proposed algorithm, we discuss a straightforward algorithm for better understanding. The naïve fusion is to randomly select a node as an aggregative node, or select by round robin. The algorithm does not distinguish the node working load and energy situation. In contrast, our algorithm will make use of above heuristic information for an optimal selection.

3.1 Node Deployment by using NEI

The deployment of sensing nodes needs to consider NEI. Simply speaking, the nodes with larger NEI can be equipped with lower power supplier. The replacing period of node with larger NEI is longer than the node with less NEI.

3.2 Aggregate Node Selection

The major steps in the algorithm are as follows: 1) Count all NEI for each nodes. 2) Choose the node that has the largest NEI. The heuristics in the algorithm is the node with largest NEI will be the largest energy potential. The aggregative node will collect data from sensing nodes, so it should have more energy potential. The node selection algorithm is proposed in the following.

Algorithm 1: Greed Method for AN Selection

```
Function: SelectAN()
Input: RN
Output: AN
-----------------------------------------------------------------
CountNEI() {                      //Sub-function to count NEI
For each V[i] ∈ RN  {
D[i] <= GetDbyHistory();      //Get data volume from history data
P[i] <= GetPbyHistory();      //Get probability from history data
E[i] <= GetE();               //Get current remain energy
NEI[i] <= K*E[i]/(D[i]*P[i]);  //K is a parameter
}
}
AN <= GetLargest(NEI[i]);         //Compare NEI, Output the largest one
Output AN;
```

3.3 Aggregation Tree Generation by using NEI

The NEI information can be used in the aggregation tree generation. In the algorithm, we will select node with the largest NEI as the main branch nodes in the tree. The main branch nodes can connect all other sensing nodes. The selection procedure is greed. The tree generation algorithm is proposed in the following.

Algorithm 2: AN Selection for Tree Generation

```
Function: ContructANTree()
Input: RN
Output: ANTree
-----------------------------------------------------------------
While RN - S ≠ Φ {
AN <= SelectAN();
S <= S∪AN;                        //Extend AN Tree Node into Set S
Tree <= GenerateTree(S);          //Build the AN Tree for connecting all SNs
FLAG <= IsANTree(Tree);       // AN Tree can connect all SNs
```

```
If (FLAG) exit;
}
Output Tree;
```

3.4 Aggregation Tree Generation for Global Optimization

Above algorithm is a greed algorithm, hereby we propose an algorithm with global optimizations. The key point is we construct all possible aggregative trees and compare the REI of the trees. Finally choose the tree with the larges REI as the aggregation tree, which obviously has the most energy potential.

Algorithm 3: Tree Generation with Global Optimization

```
Function: ContructOptimalANTree()
Input: RN
Output: OptimalANTree
```

```
CountNEI();                 //Call Sub-function in Algorithm 1
For each V[i] ∈ RN{         //Recursively generate ANTree for all possible trees
ANTree[i] <= GenerateTree(S);
REI[i] <= CountREI(ANTree[i]);
}
AN <= GetLargest(REI[i]);   //Choose the largest REI
Output OptimalANTree;
```

3.5 Aggregation Tree Generation for Random Optimization

As the computation and communication for tree generation in algorithm 3 may cost too much, we finally propose a random method to shorten the tree construction and with a manageable optimization. The major heuristics in the algorithm is as follows: Select the largest NEI node as key node, and then construct sub-tree; Choose the largest NEI in the left node, and construct sub-tree, iteratively. Do it until all the nodes are in the tree.

Algorithm 4: Tree Generation with Random Optimization

```
Function: ContructRandomOptimalANTree()
Input: RN
Output: OptimalANTree
```

```
CountNEI();                       //Call Sub-function in Algorithm 1
Do {
For largest V[i] with ∈ RN{       //Generate ANSubTree
ANSubTree[i] <= GenerateSubTree(S);
}
OANTree <= CreatRandomOptimalANTree();
If (OANTree includes all SNs) exit;
```

```
    }
    Output OANTree;
```

4 Analysis

Proposition 1: Node deployment by using NEI is energy optimal and robust.

Proof. As the node with less NEI will supply more energy, the node will be robust. As the NEI of each node is measured firstly, the deployment is optimal.

Proposition 2: **Algorithm 1 is greed for energy efficiency and robustness.**

Proof. As the aggregative node is selected first from the nodes with the largest NEI, the nodes with largest energy potential is selected first. If the energy potential is larger, the robustness is better.

Proposition 3: Algorithm 2 is greed for energy efficiency and robustness.

Proof. As each aggregative node is selected by greed algorithm, the nodes in tree are also selected by greed algorithm. If the energy potential of the whole tree is larger, the robustness of the data aggregation is better.

Proposition 4: Algorithm 3 is global optimal for energy efficiency and robustness.

Proof. As each aggregative tree is evaluated for REI, the tree with the largest REI will has the best energy efficiency and best robustness in terms of the longest lifetime.

Proposition 5: Algorithm 4 has shorter construction time than algorithm 3.

Proof. The construction times in the tree in algorithm 4 is much less than that in algorithm 3, thus the construction time is shorter.

5 Conclusions

The robustness and optimal communication efficiency of data aggregation in IoT is essential for equipment maintenance in marine ships. In this paper, we make the first attempt to propose greed algorithm based on node energy evaluation. The node energy potential is distinguished for aggregative node selection and aggregative tree generation. Several key algorithms are proposed and analyzed for their robustness and energy efficiency.

References

[1] Zhu Yu; Wang Tie-ning, *Research on the Visualization of Equipment Support Based on the Technology of Internet of Things*, Proc. of 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC), pp.1352-1357, (2012)

[2] Shancang Li; G. Oikonomou; T. Tryfonas; T.M. Chen; Li Da Xu, *A Distributed Consensus Algorithm for Decision Making in Service-Oriented*

Internet of Things, IEEE Transactions on Industrial Informatics, vol.10, no.2, pp.1461-1468, (2014)
[3] Jun Liu, Zheng Yan, Laurence T. Yang, *Fusion - An aide to data mining in Internet of Things,* Information Fusion, (2014)
[4] Giancarlo Fortino, Stefano Galzarano, Raffaele Gravina, Wenfeng Li, *A framework for collaborative computing and multi-sensor data fusion in body sensor networks*, Information Fusion, Vol.22, pp.50-70 (2015)
[5] Tifenn Rault, Abdelmadjid Bouabdallah, Yacine Challal, *Energy efficiency in wireless sensor networks: A top-down survey*, Computer Networks, Vol. 67, pp. 104-122, (2014)
[6] Junaid Ahmed Khan, Hassaan Khaliq Qureshi, Adnan Iqbal, *Energy management in Wireless Sensor Networks: A survey*, Computers & Electrical Engineering, (2014)
[7] Xiaoli Xu; Yunbo Zuo; Guoxin Wu, *Design of Intelligent Internet of Things for Equipment Maintenance*, Proc. of 2011 International Conference on Intelligent Computation Technology and Automation (ICICTA), vol.2, no., pp.509,511, 28-29 March 2011
[8] I. Stojmenovic, *Machine-to-Machine Communications With In-Network Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical Systems*, Internet of Things Journal, IEEE, vol.1, no.2, pp.122-128, (2014)
[9] Xiaoying Zhang; Hong Chen; Ke Wang; et al. *Rotation-based privacy-preserving data aggregation in wireless sensor networks*, Proc. of 2014 IEEE International Conference on Communications (ICC), pp.4184-4189, (2014)