IOP Conf. Series: Journal of Physics: Conf. Series **1065** (2018) 072016 doi:10.1088/1742-6596/1065/7/072016

Interval data fusion with preference aggregation in wireless sensor network: energy-accuracy trade-off in presence of outliers

L I Khudonogova, S V Muravyov
Division of Automation & Robotics
National Research Tomsk Polytechnic University
Pr. Lenina, 30, Tomsk, 634050, Russian Federation

E-mail: muravyov@tpu.ru

Abstract. For balancing measurement accuracy and energy consumption in a wireless sensor network in presence of outliers it is proposed sensor accuracy enhancement algorithm SensAcc and active node selection algorithm ActiveNode based on the interval data fusion method IF&PA. The results of numerical experimental investigation of the developed algorithms are presented. It is shown that the SensAcc provides the reduction of the uncertainty of measurement result at least tenfold comparing with the uncertainty of multisensor readings under possible existence of failed nodes. Simulation results have shown the ActiveNode allows to reduce the cluster nodes energy consumption approximately threefold.

1. Introduction

In wireless sensor network (WSN), important problems of measurement accuracy enhancement and energy conservation are deemed to be closely connected. Due to environmental impacts, battery discharge, noise and communication overload, multisensor measurement data, received at the base station (BS), are tend to be unreliable (inaccurate and/or incomplete). The typical solution in this case is the detection of outliers, i.e. sensor observations significantly deviated from expected behavior [1]. However, this process is attended with extra energy consumption because of growing number of transmissions that leads to faster battery exhaustion, thus shortening the network lifetime.

One of extensively applied techniques for the outlier detection is data fusion (aggregation) [2-6]: recursive principal component analysis based algorithm for solving data fault and redundancy problems [2], fault tolerant data aggregation technique for eliminating false data using locality sensitive hashing scheme [6], etc. Despite most of the above algorithms have high communication and computational cost they were designed without systematic accounting energy consumption.

In this paper, we propose (1) sensor accuracy enhancement algorithm and (2) active node selection algorithm to provide measurement accuracy and energy consumption trade-off. Both algorithms are based on our IF&PA method (interval fusion with preference aggregation) presented in [7].

By *interval data fusion* we understand a procedure of determination of a resulting interval $[x^* \pm \varepsilon^*]$ which is consistent with the maximal number of initial intervals $\{I_k\}$, $k = 1, \ldots, m$, and with maximum likelihood contains a point x^* (*fusion result*) that can be considered as a representative of all I_k with an uncertainty ε^* (*fusion uncertainty*). The IF&PA is implemented by means of representation of intervals on the real line by rankings over a set of discrete values belonging to these intervals. A *range of actual values* is a set $A = \{a_1, a_2, \ldots, a_n\}$ of fully ordered discrete values a_i, $i = 1, \ldots, n$, used to represent the intervals as rankings. Values a_i are obtained by partition of the union of initial intervals into $n - 1$ equal subintervals. Ranking of n elements of the set A in the form $\lambda = (a_1 \succ a_2 \succ \ldots \sim a_s \succ \ldots \sim a_n)$ specifies a binary relation of weak order on the set A. Ranking λ_k induced by interval I_k (called *inranking* by us) is composed of elements of the set A and satisfies the following conditions for $i, j = 1, \ldots, n$: (1) $a_i \in A_k \wedge a_j \notin A_k \Rightarrow a_i \succ a_j$; (2) $a_i, a_j \in A_k \vee a_i, a_j \notin A_k \Rightarrow a_i \sim a_j$; (3) $a_i \notin A_k \wedge a_j \in A_k \Rightarrow a_i \prec a_j$; (4) $a_i, a_j \in A_k$ are neighboring natural numbers $\Rightarrow j \equiv i + 1$. Set of m inrankings $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ of n elements is called a *preference profile*. For the preference profile one can determine a *consensus ranking*, which is such a single ranking (linear order of alternatives) that a distance (defined in terms of a number of pair-wise disagreements between rankings) from it to the initial inrankings is minimal. In [7] it is shown that the best value in consensus ranking, accepted as a fusion result, guarantees improved accuracy and robustness of the fusion procedure.

Let the WSN have a cluster tree topology, where every cluster includes m sensor nodes $S = \{s_1, s_2, \ldots, s_m\}$ located in some area under investigation. Nodes in one cluster can directly communicate with each other. We take into account spatio-temporal correlation among sensor data, i.e. assume that all nodes in one cluster determine the same quantity value. Each node contains a multisensor for measuring p parameters [8].

Being aware of uncertainty ε_k^i of i-th sensor, k-th node provides measurement data of i-th parameter as an interval $d_k^i = [x_k^i \pm \varepsilon_k^i]$, where x_k^i is a sensed value. A node generates a set of data $D_k = \{d_k^1, d_k^2, \ldots, d_k^p\}$ concerning all measured parameters and transmits it to a cluster head (CH) along with information about its own residual energy. The CH of j-th cluster, in its turn, gathers a collection of data sets $D^j = \{D_1, D_2, \ldots, D_m\}$ from all m cluster members and then sends them to the BS, which stores node location.

2. Increasing measurement accuracy

The problem is to determine the maximum feasible accurate value of the measurand in a cluster using incomplete, inaccurate and/or inconsistent data D^j. To solve this problem we propose sensor accuracy enhancement algorithm SensAcc applicable for every i-th measurement quantity in j-th cluster. The SensAcc algorithm is based on our IF&PA method and consists of the stages illustrated in figure 1.

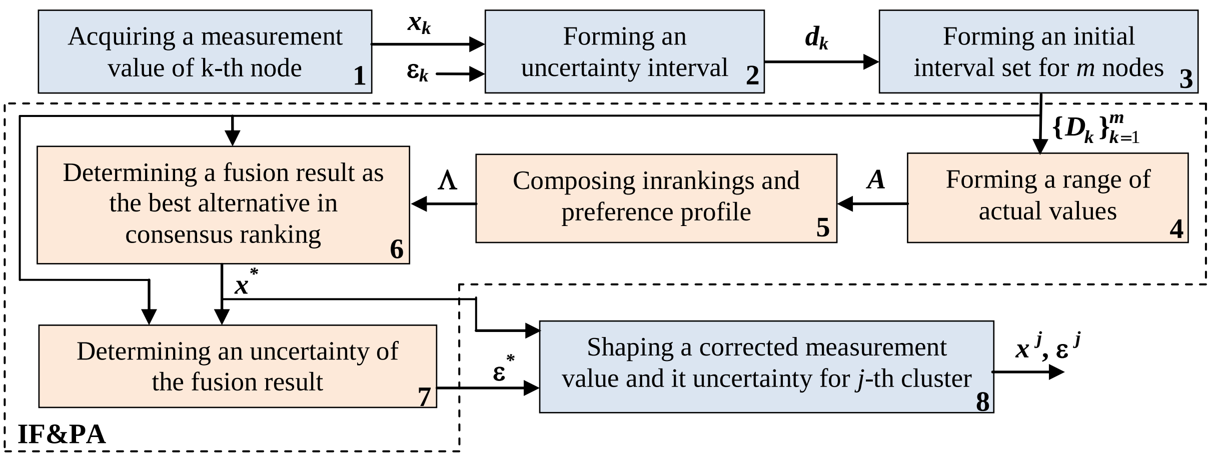

Figure 1. Main stages of the SensAcc algorithm.

3. Decreasing energy consumption

Since data transmission requires much more energy than measurement and computing operations, energy consumption in WSN is usually reduced by decreasing a number of transmissions in duty-cycling work mode which allows sensor nodes to alternate between active and dormant states. We propose an active node selection algorithm ActiveNode (preliminarily introduced in [9]) intended for selecting a set of active nodes which consume minimum possible energy in cluster and at the same time provide data sufficient for keeping required accuracy level. ActiveNode is launched on the BS at the beginning of every round and contains four stages as follows.

1. ***Cluster head selection.*** Nodes are ranked by residual energy E_{res}, distance to the BS, and previous activity. The best alternative in a consensus ranking is selected to be the CH.
2. ***Active node selection.*** Nodes are ranked by E_{res}, distance to the CH, and node accuracy Δ_q^k. The first $g - 1$ nodes in the consensus ranking, together with the CH, are included into a set of active nodes $S_a = \{s_1, s_2, ..., s_g\}$. The number g of active nodes in each round may vary from 5 to any bigger odd number in dependence on the required accuracy.
3. ***Nodes activation.*** The BS sends activation message to all selected CHs, and every CH activates all members of S_a in its cluster. Measurement data D^j with E_{res} are transmitted to the BS.
4. ***Node accuracy estimation.*** Differences Δ_q^k are calculated and transmitted to the next round: $\Delta_q^k = | x_q^* - x_q^k |$, where $q = 1, \ldots, p$; $k = 1, \ldots, m$; x_q^k is a value of a quantity q, obtained by k-th node; x_q^* is a fusion result of x_q^k values obtained by the IF&PA.

4. Numerical experimental investigations

To verify the proposed algorithm SensAcc authors developed special software IntFusion in the NI LabVIEW environment. The software generates pseudo-random interval data (simulating measurement outcomes of sensor nodes) and implements the IF&PA method for determination of the data fusion result x^* and its uncertainty ε^*.

The modelled WSN was intended for ecological monitoring of the three soil parameters: temperature t, moisture h, and electrical conductivity G; their nominal values are given in second column of table 1. Uncertainties ε of the sensors of temperature, moisture and conductivity were set to 1 °C, 3 % and 3 % respectively. In the generated data set there were included outliers, i.e. erroneous data from faulty nodes or/and missing data from dead nodes. The WSN was consisted of 151 nodes divided into 10 clusters with $m = 15$ nodes (multisensors) in each cluster.

The corrected measurement values x^j, their uncertainties ε^j, and deviations ξ of x^j from its nominal value, obtained by the SensAcc algorithm for arbitrary chosen cluster are shown in third-fifth columns of table 1. The values of uncertainty $\overline{\varepsilon}_k$ and deviation $\overline{\xi}_k$, averaged over 15 cluster nodes, are in sixth and seventh columns of table 1.

Table 1. Values of x^j, ε^j, ξ and averaged over cluster nodes values of $\overline{\varepsilon}_k$ and $\overline{\xi}_k$ for one cluster.

Parameter	x_{nom}	x^j	ε^j	ξ	$\overline{\varepsilon}_k$	$\overline{\xi}_k$
t, °C	15	14.989	0.033	0.011	1.092	1.031
h, %	22	21.975	0.062	0.025	0.710	0.588
G, mS/m	3.5	3.592	0.025	0.092	0.454	0.573

One can see from table 1 that deviations ξ of the corrected results x^j improved by one-two orders of magnitude in comparison with averaged deviations $\overline{\xi}_k$ of the measured values x_k. Furthermore, the uncertainties ε^j are 10-30 times smaller than averaged uncertainties $\overline{\varepsilon}_k$ of the measured values.

The software IntFusion was also used to investigate experimentally the proposed algorithm ActiveNode in mode of energy consumption monitoring in clusters. There were done 10 experiment runs where the modelled WSN contained 10 clusters with 15 nodes each. Every node was equipped by the sensors of temperature, moisture, electrical conductivity, illuminance and pH. Number g was accepted to be equal to $0.45m$, rounded up to the nearest odd integer. The polling period was 6 s.

In every round, all active nodes in a cluster transmitted synthetic random measurement values to the BS, and energy consumption was proportional to the transmission distance. In modelled clusters, the following three scenarios were investigated: (1) node selection by ActiveNode algorithm; (2) random node selection by RandSel algorithm; (3) without active node selection. In each experiment run we evaluated the energy consumption of individual nodes, total consumption in cluster and total cluster lifetime. By the lifetime of the cluster we understood the time from the start of WSN functioning until the battery discharge of more than 50 % of cluster nodes [10].

Figure 2 shows the results of comparison in terms of number of dead nodes round by round for the three scenarios. As evident from figure 2, the ActiveNode algorithm allowed to increase cluster lifetime 2-3 times as against Scenario 3, and approximately 1.5 times as compared with Scenario 2. Figure 3 demonstrates the total energy consumption of the cluster for three modeled scenarios over 166 rounds. It can be observed that the proposed algorithm shows the best performance in terms of energy consumption during all operation period. The ActiveNode provided the least energy consumption 97.09 mJ and the longest cluster lifetime 321 rounds in comparison with Scenario 2 (114.37 mJ at 218 rounds) and Scenario 3 (116.36 mJ at 166 rounds).

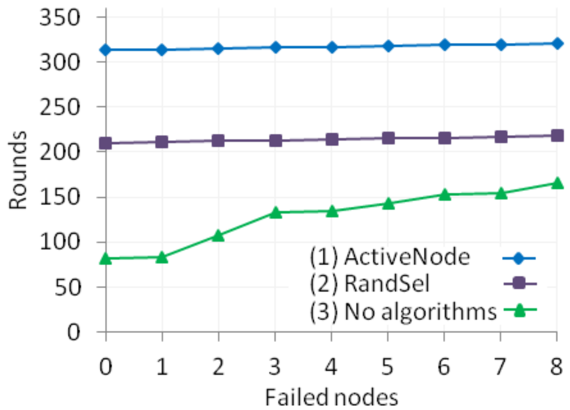

Figure 2. Number of dead nodes vs number of rounds for three scenarios.

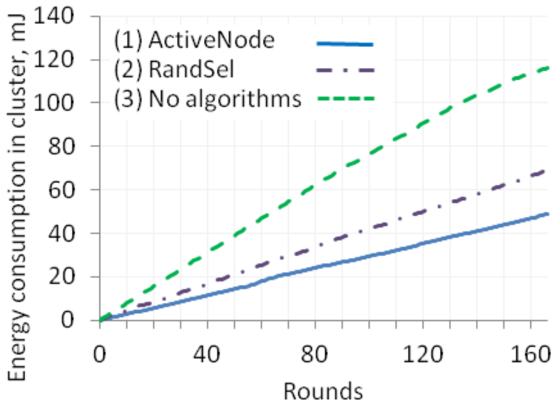

Figure 3. Total energy consumption in a cluster for three scenarios.

5. Conclusion

It was proposed two algorithms for WSN based on the interval data fusion method IF&PA: sensor accuracy enhancement algorithm SensAcc and active node selection algorithm ActiveNode. Numerical experimental results showed that cooperative use of the proposed algorithms provides essential decrease of energy consumption while obtaining the measurement result with the required accuracy.

Acknowledgment

This work was supported by the Russian Science Foundation, project # 18-19-00203, in part of the interval data fusion by preference aggregation approach development, and by the Russian Ministry of Education and Science, basic part of the state task "Science", project 2.5760.2017/8.9, in part of numerical experimental investigations.

References

[1] Han J and Kamber M 2006 *Data Mining: Concepts and Techniques* (San Francisco: Morgan Kaufmann).
[2] Yu T, Wang X, and Shami A 2016 *Proc. Int. Conf. on Communications* (Kuala Lumpur, Malaysia) vol 1 (IEEE) p 7511419
[3] Guo X, Wang D, and Chen F 2015 *Int. J. Distrib. Sens. Netw.* **11** 943532
[4] Al-Doghman F, Chaczko Z, Jiang J 2017 *Proc. 25th Int. Conf. on Systems Engineering* (Las Vegas, Nevada, USA) vol 1 (IEEE) pp 480-487
[5] Pourghebleh B, Navimipour N J 2017 *J. Netw. Comput. Appl.* **97** 23-34
[6] Ozdemir S, Xiao Y *Proc. 5th Int. Conf. on Application of Information and Communication Technologies* (Baku, Azerbaijan) 2011 vol 1 (IEEE) p 6110983
[7] Muravyov S V, Khudonogova L I, and Emelyanova E Y 2018 *Measurement* **116** 621-630
[8] Muravyov S V, Khudonogova L I 2015 *Proc. 7th Int. Cong. on Ultra Modern Telecommunications and Control Systems* (Brno) vol 1 (IEEE) 293-297
[9] Khudonogova L I, Muravyov S V 2016 *Proc. Int. Siberian Conf. on Control and Communications* (Moscow) vol 1 (IEEE) 7491835
[10] Dietrich I, Dressler F *ACM Trans. Sens. Netw.* 2009 **5** 1-38