Acta Technica 62 No. 2B/2017, 541–552

Application of hybrid compression and converging technology in wireless sensor network routing protocol in China[1]

Yang Chen[2]

Abstract. In order to give full play to the advantages of wireless sensor and make use of the restriction of it, it is very important to study and design the proper routing protocol and data fusion. In this paper, the basic structure, characteristics and classification of wireless sensor networks were discussed systematically, and several typical routing protocols were introduced and compared. Based on the analysis of existing network layer routing protocols and key technologies, the network layer routing protocols and their key technologies in wireless sensor networks were studied. Finally, the conclusions that the data compression rate is basically a fixed value, and the influence of the depth of the ternary tree is very small, which can greatly increase the scalability of the network were drawn.

Key words. Wireless sensor networks, hybrid routing model, data fusion.

1. Introduction

In the past ten years, the rapid development of network information technology has made the connection between people in every corner of the world more and more close, and people can talk and communicate conveniently anytime and anywhere. Because of the passive nature of things, the communication between people and things will not be very smooth. Therefore, in order to make the physical world useful to us, other tools and techniques must be used to explore everything in the world, such as various sensors. Based on the development of communication technology, sensing technology and micro electromechanical systems, sensors are so small that they can be encapsulated into a millimeter scale chip, with low cost, but high

[1]The study was supported by the Key natural science research project supported by Anhui province (KJ2016A303) Research on data fusion technology of wireless sensor network for privacy protection; Quality engineering project supported by Anhui province (2016sjjd041) Anhui Xinhua University & Hefei City Cloud Data Center Co., Ltd, practice education base.

[2]College of Information Engineering, Anhui Xinhua University, Hefei, Anhui, China, 230088

robustness and more powerful functions. The "intelligence" of sensors can help people complete the collection, arrangement and communication of data, thus greatly expanding the application field of sensors. Wireless sensor networks will fundamentally change the interaction between human and the natural world, and make the virtual network world and the physical world and the various fields communicate with people. People can get rich and reliable information from the physical world by means of tools, and realize the desire of "ubiquitous computing". It is no exaggeration to say that the widespread application of wireless sensor networks is the trend of the development of science and technology, which will always lead to changes in the wireless sensor and related research fields to the human world.

2. State of the art

Wireless sensor networks are at the forefront of new technologies, and there are still a lot of hot issues to be discussed, such as data fusion. In recent years, a large number of routing protocols have been developed in the process of data collection, most of which can support data fusion functions, so as to reduce energy consumption and communication costs [1]. In the process of building a cluster, considering the influence of the residual energy and the distance from the sink node to the routing protocol design, a routing protocol supporting data fusion was designed. The node structure was upgraded to heterogeneous, and the establishment of cluster and data transmission path under heterogeneous condition were analyzed [2]. Considering the coverage of nodes, the most comprehensive information was transmitted with the least nodes. Linear regression model was used to fuse data and reduce data transmission. At present, most data fusion research focuses on the shortest path tree junction, but the focus of different application scenarios is certainly different. Much of the work of data fusion is focused on intra-data fusion in wireless sensor networks [3]. For fusion queries, a certain amount of computation within the network is acceptable. The sensor sensing data is processed into individual parts and then constantly updated during the delivery to the sink node or base station. Each sensor node only delivers one or relay a small amount of data to other nodes, thus forming an energy efficient method [4]. Current data fusion protocols are mainly divided into three categories: cluster based data fusion, tree based data fusion, and multipath based data fusion [5].

2.1. Methodology

A large number of wireless sensor nodes are deployed in a certain range of monitoring target areas, and a network is generated among nodes through self-organizing and adaptive methods, and the structure of the wireless sensor network is shown in Fig. 1.

The whole network is usually composed of sensor nodes, sink nodes sink, Internet and satellite console in the monitoring area. The working principle is as follows: a large number of sensor nodes are randomly distributed within a certain range of target regions, and the data parameters in the region are collected according to

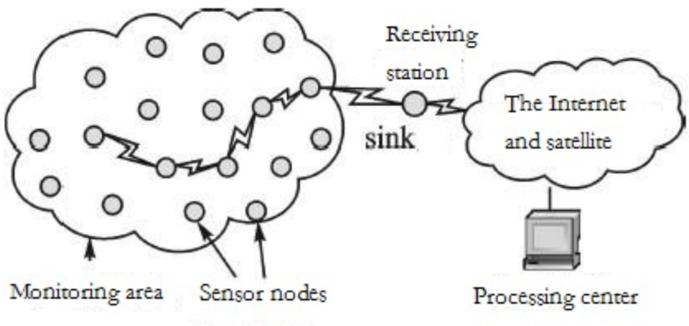

Fig. 1. Structure of wireless sensor network

specific requirements. And a single hop or multi-hop communication mode is used to send the collected data to the sink node. Then, after the connection with the console and the satellite, the user can find data at the end of the console or send instructions to each node in the network [6].

In general, sensor nodes can be considered as a micro embedded system powered by batteries, and the storage capacity and communication processing capability are weak, which limits the direct communication between the sensor nodes and the sink node. From the point of view of network function, in addition to local data acquisition and fusion, sensor nodes also relay and forward data from other nodes, which acts as double roles of terminals and repeaters [7]. The sink node can be a powerful node with enough energy support and more memory and processing capabilities, and can also be used as a gateway device to connect the wireless sensor network and realize the conversion between the two protocol stacks.

Wireless sensor network protocol stack provides the necessary software support for communication between sensor nodes. According to the identification of the majority of scholars, wireless sensor network protocol stack is divided into five layers, namely application layer, transport layer, network layer, data link layer and physical layer. And there are three supporting parts, energy management platform, mobile management platform and task management platform stack [8].

Data fusion technology includes the collection, transmission, synthesis, filtering, correlation and synthesis of useful information given by various information sources, so as to assist people in the determination, planning, detection, verification and diagnosis of situation environment [9]. Sensor data fusion technology needs to get the internal relations and rules of various information, and combine with the spatio-temporal correlation of sensor nodes, thus to eliminate redundant and erroneous information, retain the key and correct information components, and ultimately simplify and optimize the information. A single sensor may only obtain partial information segments of the environment or the object under test, while the multi-sensor information can reflect the characteristics of the environment perfectly and accurately after fusion. Data fusion technology plays an irreplaceable role in the research of intelligent information processing technology [10].

According to the level classification, data fusion can be divided into pixel level fusion, feature layer fusion and decision level fusion. Pixel level fusion is the fusion directly on the original data layer, which is synthesized and analyzed before the raw data of various sensors are preprocessed, belonging to a low level of convergence. Feature layer fusion extracts features from the original data of the sensor, and analyzes and processes the features comprehensively, belonging to the intermediate levels of convergence [11]. For each sensor, the decision level fusion firstly completes the basic processing locally such as preprocessing, feature extraction, recognition or decision, and establishes the preliminary conclusion, and then makes decision fusion by association processing, and finally obtains the joint inference result, which belongs to high-level fusion.

When the deployment density of nodes is large, the sensor sensed data will be redundant, that is, multiple neighboring sensor nodes will acquire the same information. In order to save the energy of sensor nodes, it is necessary to remove the redundant information and reduce data transmission [12]. In a network with thousands of sensor nodes, the energy consumed to deliver a message is far greater than the energy consumed by the computation, and data fusion can remove the redundant information very well, which makes the transfer of information small.

Space means that different kinds of data come from different places and spaces, while time means that the database can adapt to changes in the objective environment over time. If the nodes in a wireless sensor network are similar in the near or continuous time period, then the sensed data or attribute values are similar or have a regular change. Even if there are still a few nodes that fail to be sensed at a certain point in time, a set of data can also be obtained by using the similarity of data to have very little error or a slight change to real data. Then, these nodes have time correlation [13]. In a wireless sensor network, because of the random deployment of nodes, especially in some remote locations, the use of aircraft to broadcast nodes will result in high density of some nodes. At the same time, the data sensed by some geographically similar nodes have a high degree of similarity. And then the data induced by these nodes have spatial correlation.

In order to store and find the convenience, the tree structure is used to store and transmit data, and a complete trinomial tree can be constructed. A child node transmits only a set of coefficients to its parent node, which is used for data fusion and data query, greatly reducing energy consumption and effectively prolonging the life cycle.

In a network where a large number of sensors are deployed, the minimum spanning tree is used to solve the problem of how the sensor nodes pass to the base station via the optimal path. Greedy algorithm is a commonly used algorithm. But the greedy algorithm is a local optimal algorithm, which just makes what seems the best choice at the moment, and if it is not considered from the overall optimum, it is only a locally optimal solution in some sense [14]. In this paper, a complete trinomial tree is constructed for propagating data. The role of each node of a fully ternary tree is the cluster head of a hierarchical network that performs only the functions of receiving data, transmitting data, and merging data. And the structure of the CTETREG data fusion tree is shown in Fig. 2.

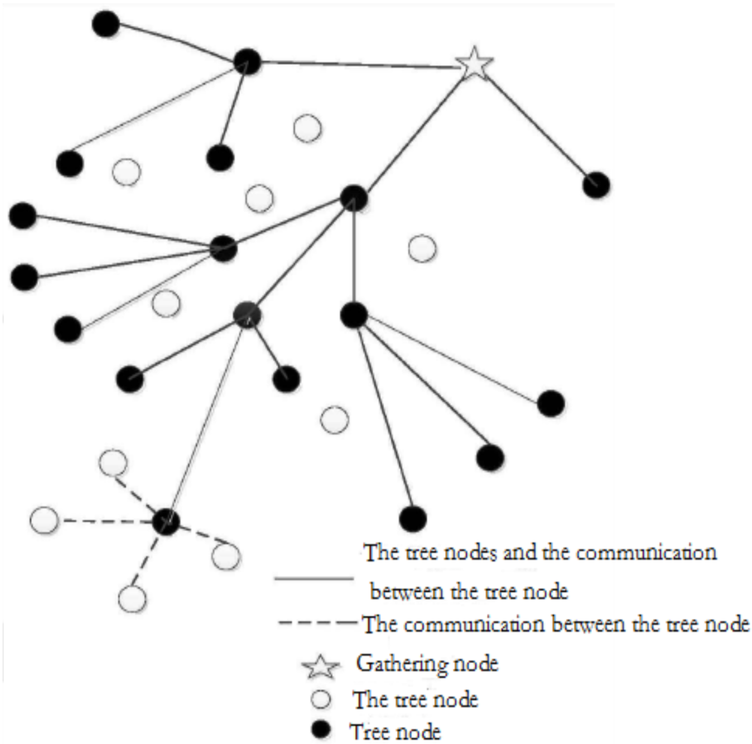

Fig. 2. The structure of the CTETREG data fusion tree

For the choice of roots, in order to facilitate the coordinate representation, the wireless sensor network area is set to a square area, and the length is L. If the length of each subregion is L_{s}, then, the whole network can be divided into L/L_{s} subregions. The roots of adjacent areas must be as close as possible to meet the needs of querying multiple regions while minimizing the propagation path. So, the root of each query subtree is best deployed in the corner of the square sub region. For odd sub regions, the root is selected in the upper right corner of the subregion, while the even region is selected in the upper left corner, so that the roots of the neighboring four sub regions are clustered together [15].

When the roots of each tree are determined, tree analysis and selection are necessary. Using the topological structure of the complete trinomial tree, each parent node has a 3 child node, which ensures the full coverage of the tree and the accuracy of the data processing. The entire sensor network area is S, and the total number of nodes deployed is N, and then the node density N/S is ρ, each subregion is S_s, and the average number N_{s} of nodes in a subregion is ρS_{s}.

In a full trinomial tree, the total number of nodes is N_{at} as shown in equation (1):

$$N_{\mathrm{at}} = \frac{3^{p+1} - 1}{2}\,. \tag{1}$$

Here, p is the depth of the tree of the complete trigeminal tree.

Tree depth can be obtained by type 1, as shown in type 2:

$$p = \log_3 \left(\frac{2N}{n_s + 1} \right) - 1 \,. \tag{2}$$

Here, n_s is the number of non-tree nodes corresponding to a tree node.

Using formula (2), the following formulae (3) and (4) are obtained:

$$N = n_s N_{at} + N_{at} \,, \tag{3}$$

$$N_{at} = \frac{N}{n_s + 1} \,. \tag{4}$$

In this paper, it is assumed that there are $N = 3000$ nodes in sensor networks, area$S = 800 \times 800$ is a square area, and the subregion is $S_s = 400 \times 400$. The average number of nodes in each subregion is

$$N_s = \rho \, S_s = \frac{N}{S} \, S_s = \frac{3000}{800 \times 800} \times 400 \times 400 = 750 \,.$$

The depth of the tree is $p = 3$, and the number of tree nodes in each subregion is

$$T_c = N_{at} = \frac{3^{p+1} - 1}{2} = 40 \,.$$

Further, $n_s = 15$, and, therefore, the number of nodes in the subregion is

$$n_s t + t = 15 \times 40 + 40 = 640 < 750 \,.$$

From the above calculation, it can be seen that the maximum tree depth $p = 3$ is appropriate.

In a wireless sensor network, a large number of distributed nodes are constantly sensing the data of the surrounding environment and transmitting it to other nodes. If the environment changes suddenly or the sensor node receives interference, an abnormal signal value is generated. In order to keep the accuracy of the data, it is necessary to identify abnormal data in time and analyze them or discard them directly. which was very important in some application fields. For example, in forest fire monitoring, a large number of sensor nodes are deployed, and each sensor node will continue to transmit information to the base station in an active or passive manner. In general, the temperature of sensor induction is within a fixed range. Once a fire occurs, the sensor sends out an abnormal data value that touches the warning signal, which requires timely identification to control the spread of the fire. But sometimes because of human factors, for example, when outdoor enthusiasts cook in the forest, just the furnace is located near a sensor, the abnormal data value will also be generated. If the node is restored to normal in a short time, and the value of the neighboring sensor is normal, it can be done without any processing.

There are three performance indexes of CTETREG data fusion tree, percentage error, compression ratio and fusion data output size.

Percentage error is the representation of the error between the induced data and the real data in the form of a percentage.

The compression ratio is the ratio between the number of output data of the neutron nodes in the query tree and the total data received by the other query tree nodes and the non-query tree nodes, which is the ratio between output and input. The greater the compression ratio, the smaller the ratio of the reduction will be, and the more redundancy of the data will be. And the smaller the compression ratio, the less redundancy of the data will be. Generally, the smaller the compression ratio, the better the effect can be. However, considering the completeness and accuracy of data, the compression ratio is not as small as possible.

For the fusion data output size, it is expected that when the tree depth is different and the node density is changed, the output of the data is a fixed value, which is beneficial for network expansion and energy savings.

3. Result analysis and discussion

Based on the conclusions obtained by the simulation platform MATLAB, the advantages of the method were compared. The specific simulation reference data is shown in Table 1.

Table 1. Simulation parameters

Notation	Value	Unit
Z	800×800	m^2
R	40	m
S	3000	
ρ	0.0047	
Z_s	400×400	m^2
p^n	0.33	
p	3	
n_s	15	

The simulation results are shown in Fig. 3. Symbol Z is measured by the width of the node coverage in the monitoring area, and R is measured in terms of the accuracy of the area being covered. Node Z_s is randomly distributed in the 400 m×400 m squared plane region.

As shown in the figure, when the tree depth was 1, there were only 4 nodes in each subregion for processing the whole sub region parameter value. In the distribution of dense nodes, the received data was limited, and sometimes the real data value can't be accurately reflected, so that the percentage error was 120 %. When the tree depth was 2, the node of the tree reached 13. Compared with the 4 nodes, the number of nodes involved in the data fusion increased, and the accuracy of the data fusion value was improved. However, the simulation can also see up to 65 %, but when the tree depth was 3, the fusion effect was greatly improved, and the percentage error rate was only about 5 %. All of the above results were better than two fork tree, even if the two fork tree was deep; its percentage error rate was 5 %.

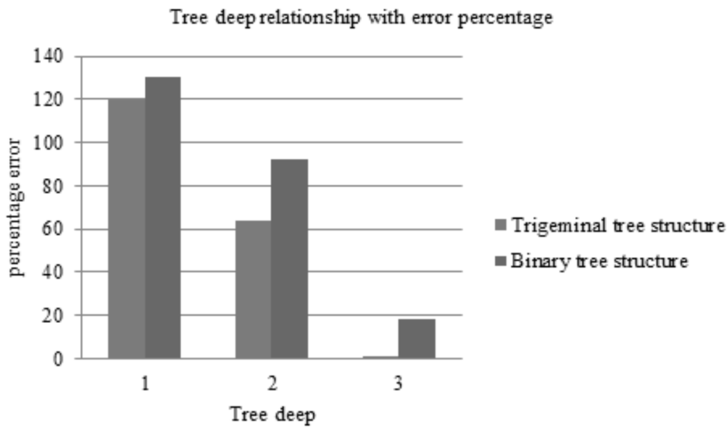

Fig. 3. Comparison of the error percentage and tree depth of the complete two fork tree and the complete trinomial tree

Figure 4 shows the comparison of the compression ratio and tree depth of the full two-fork tree and the complete trinomial tree.

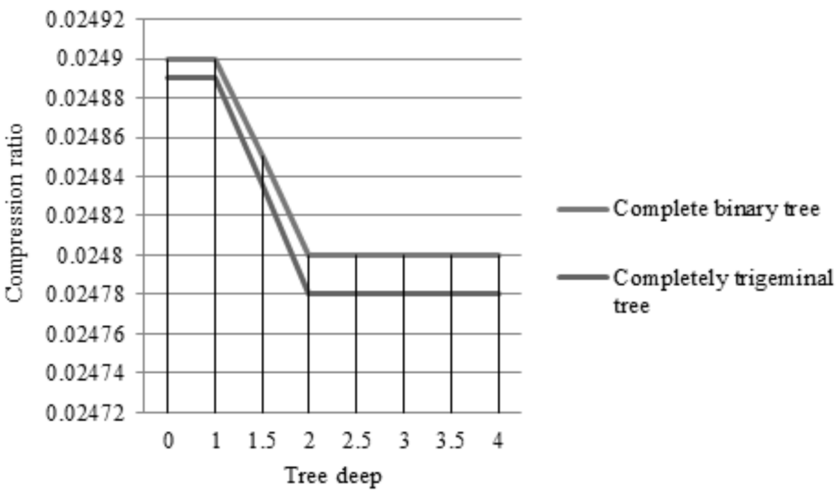

Fig. 4. Comparison of the compression ratio and tree depth of the full two-fork tree and the complete trinomial tree

A low compression ratio can reduce data redundancy, reduce broadband duty cycle and save transmission energy. As can be seen from the simulation Fig. 2, the compression ratio of a fully two fork tree was between 0.0248 and 0.0249, while the compression ratio of the complete trinomial tree was almost 0.02478 to 0.02489. The full trinomial tree node was more than the full two fork tree, and the data was much more. The base of the denominator was large, but the compression ratio was smaller. Therefore, the amount of data transmission was less, but the difference between the two was not large, thus having little impact on the accuracy of the data. From the change of the compression ratio, it can be seen that the range of change was very small, basically a constant, which was beneficial to the expansion of wireless sensor network.

From the simulation Fig. 5, it can be seen that the output size of the data fusion was $(S_x + S_y + S_c)$ bit value based on the regression approximation of the trinomial tree model. When a node was passed to another node and base station, the final output data value was a constant, which also confirmed that the data values included the coefficients and coordinate ranges obtained from data fusion. In theory, the coefficient is a fixed number, and the coordinate range is a value, and the length of the data added by it is considered to be a constant value. From the simulation results, it also can be seen that the output of the data was independent of the number of nodes and the tree depth, which greatly increased the scalability of the network, and facilitated the application of this method in a wider range.

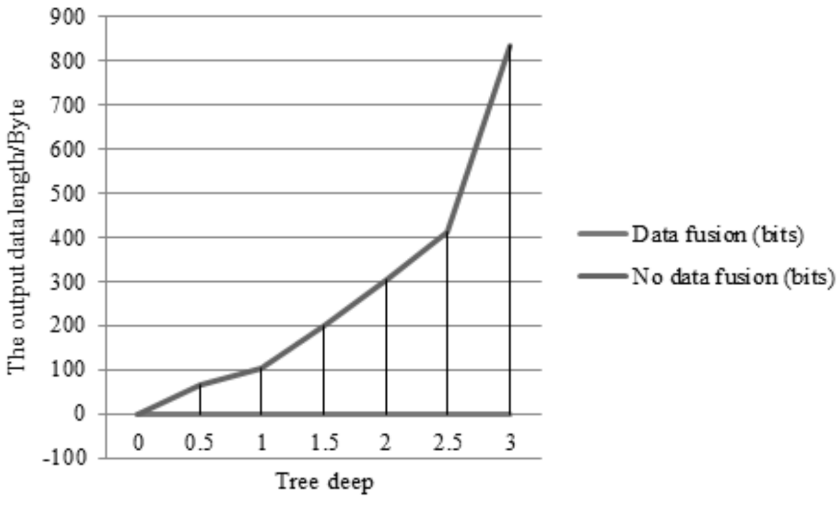

Fig. 5. The effect of a full Trident tree with and without data fusion

4. Conclusion

In order to solve the limitations of wireless sensor and make full use of its advantages, in this paper, on the basis of studying the systematicness of routing protocols, the application scope of the protocol was extended and simulated, and the energy saving and security of data fusion in data transmission were studied. Some conclusions were drawn as follows: a regression analysis data fusion algorithm based on full trinomial tree structure was proposed to reduce the data transfer and the energy consumption. Moreover, the data packets transmitted by each tree node were always of the same size, and even if the number of nodes and the tree depth increased, the size of packet transmission would not be affected, so that there was a good scalability in the network. In addition, every parent node of a full trinomial tree had more child nodes than that of a full two fork tree. In the same case, the complete trinomial tree required less tree depth, which makes its performance better than that of the fully two fork tree. Simulation results show that, compared with the algorithm of fully two tree structure, the data fusion method based on the complete trinomial tree structure can greatly prolong the network life cycle. Of course, there are some problems that deserve further study, such as how to apply fuzzy control and intelligent control to data fusion.

References

[1] C. Karlof, D. Wagner: *Secure routing in wireless sensor networks: Attacks and countermeasures.* IEEE International Workshop on Sensor Network Protocols and Applications, 11–11 May 2003, Anchorage, AK, USA, IEEE Conference Publications (2003), 113–117.

[2] C. Sha, R. C. Wang, H. P. Huang, L. J. Sun: *Energy efficient clustering algorithm for data aggregation in wireless sensor networks.* Journal of China Universities of Posts and Telecommunications *17* (2010), Suppl. No. 2, 104–109, 122.

[3] D. Kumar, T. C. Aseri, R. B. Patel: *EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks.* Computer Communications *32* (2009), No. 4, 662–667.

[4] H. F. Chen, H. Mineno, T. Mizuno: *Adaptive data aggregation scheme in clustered wireless sensor networks.* Computer Communications *31* (2008), No. 15, 3579–3585.

[5] M. Mostafizur, R. Mozumdar, G. F. Nan: *An efficient data aggregation algorithm for cluster-based sensor network.* Journal of Networks *4* (2009), No. 7, 598–606.

[6] W. M. Lee, V. M. S. Wong: *E-Span and LPT for data aggregation in wireless sensor networks.* Computer Communications *29* (2006), Nos. 13–14, 2506–2520.

[7] W. H. LiaoH, Y. C. Kao, C. M. Fan: *Data aggregation in wireless sensor networks using ant colony algorithm.* Journal of Network and Computer Applications *31* (2008), No. 4, 387–401.

[8] Y. Y. Mao, F. R. Kschischang, B. Li, S. Pasupathy: *A factor graph approach to link loss monitoring in wireless sensor network.* IEEE Journal on Selected Areas in Communications *23* (2005), No. 4, 820–829.

[9] Y. Yu, R. Govindan, D. Estrin: *Geographical and Energy Aware Routing: A recursive data dissemination protocol for wireless sensor networks.* Marine Pollution Bulletin *20* (2011), No. 1, paper 48.

[10] J. Kulik, W. Heinzelman, H. Balakrishnan: *Negotiation-based protocols for disseminating information in Wireless Sensor Networks.* Wireless Networks *8*, (2002), Nos. 2–3, 169–185.

[11] G. Anastasi, M. Conti, M. Di Francesco, A. Passarella: *Energy conservation in wireless sensor networks: A survey.* Ad Hoc Networks *7* (2009), No. 3, 537–568.

[12] B. H. Liu, W. C. Ke, C. H. Tsai, M. J. Tsai: *Constructing a message-pruning tree with minimum cost for tracking moving objects in Wireless Sensor Networks is NP-complete and an enhanced data aggregation structure.* IEEE Transactions on Computers *57* (2008), No. 6, 849–863.

[13] J. Lian, K. Naik, G. Agnew: *Data capacity improvement of wireless sensor networks using Non-uniform sensor distribution* . International Journal of Distributed Sensor Networks *2*, (2006), No. 2, 121–145.

[14] V. Rajendran, K. Obraczka, J. J. Garcia-Luna-Aceves: *Energy-efficient collision-free medium access control for wireless sensor networks.* Journal Wireless Networks *12* (2006), No. 1, 63–78.

[15] B. Khaleghi, A. Khamis, F. O. Karray, S. N. Razavi: *Multisensor data fusion: A review of the state-of-the-art.* Information Fusion *14* (2013), No. 1, 28–44.

Received September 19, 2017