

Data acquisition system for football based on ZigBee

QING NING¹

Abstract. More and more attention is paid to football player's scientific training and ways to collect their movements' data, thus team skills and tactics are improved. Ways to achieve a better result is now a popular research topic among many professional football players. Wireless network is built based on ZigBee technology in order to collect movement data of football players in the field. Centroid locating algorithm and trilateration algorithm are studied from the aspect of location, thus improved location algorithm-centroid multilateral algorithm is put forwarded. Based on this algorithm, football players' movements' data are collected and analyzed by k-means clustering. Therefore, players' running ability are analyzed. Movement data of players in 5 groups are taken as sample data and above method is used to analyze the relevant data. Experiment verifies that this method is good for analyzing football players' physical power and the corresponding strategic cooperation.

Key words. ZigBee, location algorithm, k-means, movement data, application.

1. Introduction

Wireless data-built network is made up of nodes with signal acquisition ability, transmittability and data-handling capacity. Those node are used for monitoring observing zone as much as possible, thus smart task can be completed. All nodes are made up of data acquisition module, data receiving and storage module, data transmitting module and battery module. ZigBee is widely applied because of its low-power dissipation, networking and good expansibility. One of its significant advantages is that it can be used to build large network freely, whose coverage area is enormous.

Comparing to other network construction mode, ZigBee-based network has more practical functions than others. However, its construction is also more complex. Manufacturer who thinks highly of ZigBee technology, chip maker and soft giant establishes a technical alliance in order to standardize ZigBee technology.

Advantages of ZigBee can be summarized as (1) low power consumption because

¹School of Physical Education, Henan University of Science and Technology, Luoyang, 471000, China; E-mail: qnqingning@126.com

ZigBee node is of short working time and low transmission rate; (2) reliable signal data transmission because mechanism of TCP signal transmission is reliable; (3) 8 bits kernel-needed because requirement of hardware storage is relatively low; (4) relatively short corresponding time for different states; (5) large network capacity and many sub-ZigBee network can be added in one area; (6) improved network security because encryption algorithm of 128 bits can be used to set security level at will.

2. Literature review

After the introduction, practical ZigBee technology was applied to many fields of production in a short period of time [1]. Comparing to other countries in the world, ZigBee was introduced to China very late. For ZigBee network, data collected by sensor were analyzed and transmitted by ZigBee controller [2]. ZigBee is a wireless communication technique which is often compared to WIFI technology. WIFI technology is of great transmittability [3] in a limited time because of large power dissipation. Corresponding standards for physical layer and network layer in network structure construction were established by ZigBee technology. In physical layer, wireless transmission function was achieved by enable control [4] and network layer was used to construct and allocate resource of ZigBee network [5].

Appropriate network topology was needed when constructing network. Generally, star topology featuring simple construction, easy networking and convenient adding or deleting node [6] was used as long as site area was not too large. Though it is of low practical security, practical security of star topology is higher than that of mesh topology structure. Network constructed with star topology was coordinately working with nodes. Nodes were equal, thus, data can transmit among different nodes [7]. When designing hardware of ZigBee network based on node, embedded single chip microcomputer (SCM) of CC2530 [8] or DHT11 type [9] can be used as control chip. As for software, IAR compiler [10] was adopted, which supported different types of single chip computer.

Wireless location technology can be used to collect data once ZigBee network was built. Location technologies based on ZigBee can be classified into ranging and range-free [11] according to whether it was related to measuring distance or not. Trilateration of ranging algorithms and centroid locating of range-free algorithms were selected as the research objects. Considering that trilateration was demanding for hardware and centroid locating was demanding for signal quantity, centroid multilateral algorithm was put forwarded which integrated the advantages and disadvantages of these two algorithms.

3. Research method

3.1. Network covering football court built based on ZigBee

Star topology (see Fig. 1) was adopted to construct network before acquiring data when the running range of football players was relatively small. This network was made up of a main coordinator and several terminal nodes. Coordinator (hereinafter referred to FFD) was core of the network, which was used to receive data transmitted from different terminal nodes. Generally, FFD was located near to terminal node for the sake of convenient networking and maintaining. Terminal nodes (hereinafter referred to RFD) were used to assist coordinator to collect data acquired from football players and sent signals to players' sensor in order to interact with them.

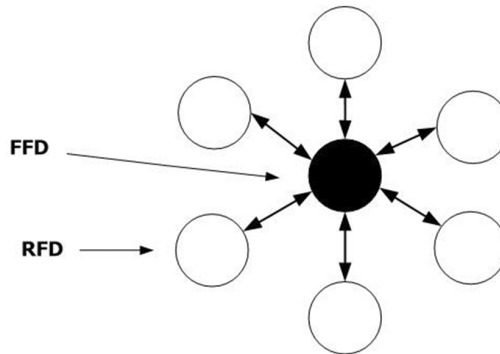


Fig. 1. Star network

One of the biggest advantages of star topology structure was easy management because its structure was simple and total path direction was relatively less. In order to cover the whole football court, several star ZigBee networks were set (see Fig. 2). Coverage area of star network was relatively more concentrated than the others'. Every football players in the field was carrying a small signal tracker, thus RFD can track their movement trail, collect their movement data and send those data back to coordinator. Then, all data received will be sent to PC by coordinator.

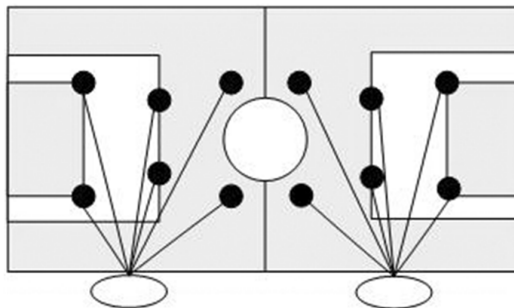


Fig. 2. Layout of ZigBee nodes

Hardware can be classified into coordinators, terminal nodes and PC processing and display terminal. For coordinator, CC2430 single chip microcomputer (SCM) was used as core control unit to receive data from network built by ZigBee and those data were transmitted to PC centralized controller through corresponding communication port (see Fig. 3), thus network was maintained in real time. These data collected was sent to CC2430 processor. ZigBee technology was used to communicate with surrounding terminal nodes (see Fig. 4) and DHT11 chip was used by terminal nodes to collect movement data of players those send by their sensors. Generally, this sensing device was used together with CC2430 single chip microcomputer (SCM).

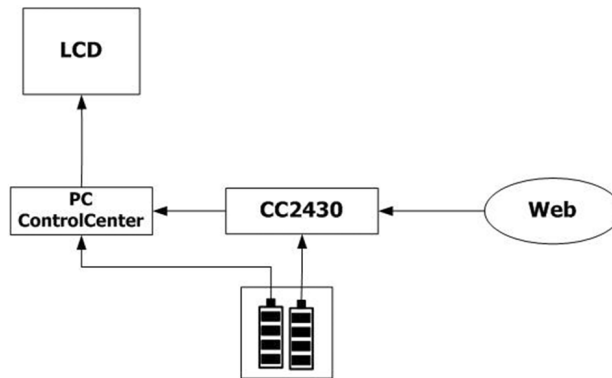


Fig. 3. Master node structure

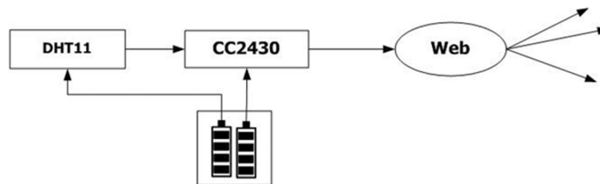


Fig. 4. Slave node structure

CC2430 was used as control chip for coordinator because it had many advantages of ZigBee technology, such as communication technology module of standard IEEE802.15.4 is included.

CC2430 single chip microcomputer (SCM) was also featured: (1) high signal receiving sensitivities; (2) low power dissipation in non-working state and RTC was used to start it; (3) CSMA/CA communication standard was supported; (4) lowered energy consumption because supply voltage relocatability is good; (5) easy to analog signal receiving because port of 14 bits was integrated in CC2430 chip to change modulus.

Terminal node was used to collect players' movement data, which was embedded with DHT11 chip. Because serial bus communication was used in DHT11 chip, number of IO port was reduced and transport protocols was simple. When transmitting

signals through DHT11 sensing device, coordinator was used to send request signal. After around 40 microseconds, Response signal at terminal node was read. If signal status bit was 0, it means Response signal has been sent by sensing device; when it was sending, the status bit will be changed into 1. Thus, data can be transmitted between terminal induction system and coordinator. When transmitting data and status bit of data package was start from 0, high level signal was sent, which means DHT11 equipment was abnormal and examination was needed.

3.2. Location algorithms

Given X , Y and Z as fixed nodes, corresponding coordinates are (t_X, f_X) , (t_Y, f_Y) and (t_Z, f_Z) . Distance between those three nodes and destination node D were Length_X , Length_Y and Length_Z , respectively. Given center of a circle with coordinate of $D(t, f)$ and radius was distance between node and node, the below equations can be obtained

$$\text{Length}_X = \sqrt{(t - t_X)^2 + (f - f_X)^2}, \quad (1)$$

$$\text{Length}_Y = \sqrt{(t - t_Y)^2 + (f - f_Y)^2}, \quad (2)$$

$$\text{Length}_Z = \sqrt{(t - t_Z)^2 + (f - f_Z)^2}. \quad (3)$$

According to equations (1-3), the coordinate of the destination node D can be given as:

$$\begin{aligned} & \begin{bmatrix} t \\ y \end{bmatrix} = \\ & = \begin{bmatrix} 2(t_X - t_Y) & 2(f_X - f_Z) \\ 2(t_Y - t_Z) & 2(f_Y - f_Z) \end{bmatrix}^{-1} \begin{bmatrix} t_X^2 - t_Z^2 + f_X^2 - f_Z^2 + \text{Length}_Z^2 - \text{Length}_X^2 \\ t_Y^2 - t_Z^2 + f_Y^2 - f_Z^2 + \text{Length}_Z^2 - \text{Length}_Y^2 \end{bmatrix}. \end{aligned} \quad (4)$$

Range-based localization algorithm was of high accuracy and hardware requirement is demanding. From the point of cost and power dissipation, some algorithms irrelevant to ranging were studied by many scholars. This Range-based localization algorithm, different to ranging-based algorithm, does not need ranging information and localizes node based on network connectivity.

Centroid locating method, one of the typical algorithms, was put forward by scholar of University of Southern California. This algorithm was easy to apply because it was to solve mean value of corresponding polygon vertices of polygon geometry area. This algorithm scene was facing a similar network structure. In this structure, a signal was sending signals around continuously. When the corresponding destination node received this signal up to a certain amount or a certain period of time, this destination node was regarded as within the communication range. Meanwhile, when destination node received a certain amount of signal that distributed in several geometry areas, corresponding centroids can be calculated. And those cen-

troids can be used to locating destination node. There are obvious drawbacks in this algorithm. For example, though layout of signal range was reasonable, insufficient signal number would cause sparse area and decreasing number of centroid. Thus, locating accuracy would be lowered.

Two above algorithms were typical for ranging algorithm and range-free algorithm, respectively. In complex engineering, locating algorithm was based on trilateration, so trilateration was widely applied. However, this algorithm was limited by hardware of node and odds power dissipation. Generally, expected locating result cannot be achieved when measuring distances between destination node and three or more fixed node. Centroid locating algorithm based on centroid has a relatively low hardware requirement and communication overhead but requires a certain number of signals. Based on those two algorithms, centroid multilateral algorithm was put forwarded.

This ranging-based algorithm was similar to trilateration in many aspects. However, it can be seen from Fig. 5 that this algorithm was different to trilateration because regions were used in this algorithm.

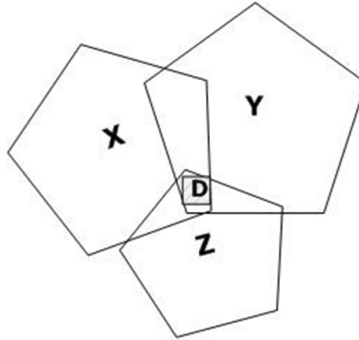


Fig. 5. Multilateral algorithm

In this figure, X , Y and Z are fixed nodes and D is the destination node. Distances between the destination node and these fixed nodes XD , YD and ZD can be measured by trilateration and these were expressed as L_x , L_y and L_z . Under ideal condition, X , Y and Z are the centers of three polygons that mutually intersect. The intersection area was simplified into a small rectangular area (see shadow area in Fig. 5). Centroid of the rectangular area are regarded as destination node D , and the coordinates of destination node D are identified.

The algorithm is based on finding intersections of several polygons, so this can be simplified by solving rectangle intersection of two polygons. Given two polygons and rectangles T1 and T2 correspond to these two polygon intersections. Four vertex coordinates of rectangles T1 and T2 were marked. The reason for simplifying these two polygons as rectangle was that calculation can be much easier than used to because less vertex coordinates mean less calculation. According to the vertex coordinates, the larger value of the X -axis was obtained by comparing the left margins of T1 and T2; the less value of X -axis was got by comparing right margins of T1 and T2; the less value of Y -axis was got by comparing upper boundaries of T1

and T2 and the larger value of Y -axis was got by comparing lower boundaries of T1 and T2. Thus, location of the rectangular region is defined using these four values.

Given four vertexes of T1 and T2 distributed according to 2×2 and described by two-dimensional array, the vertex coordinates of the rectangle can be described as

$$R_{T1} = \{(X_{T11}, Y_{T11}) (X_{T12}, Y_{T11}) (X_{T11}, Y_{T12}) (X_{T12}, Y_{T12})\}$$

$$R_{T2} = \{(X_{T21}, Y_{T21}) (X_{T22}, Y_{T21}) (X_{T21}, Y_{T22}) (X_{T22}, Y_{T22})\}$$

Based on solving method for solving the above question, the larger or lesser value can be calculated by comparing the corresponding sides of the two rectangles. Thus, location area of R_{T1} and R_{T2} is shown as $\{(X_{Z11}, Y_{Z11}) (X_{Z12}, Y_{Z11}) (X_{Z11}, Y_{Z12}) (X_{Z12}, Y_{Z12})\}$

$$\begin{cases} X_{Z11} = \max(X_{T11}, X_{T21}), \\ X_{Z12} = \min(X_{T12}, X_{T22}), \\ Y_{Z11} = \min(Y_{T11}, Y_{T21}), \\ Y_{Z12} = \max(Y_{T12}, Y_{T22}). \end{cases} \quad (5)$$

Coordinates of these three fixed nodes are $X(a_1, b_1)$, $Y(a_2, b_2)$ and $Z(a_3, b_3)$ The distances between these three fixed nodes and destination node were Lx , Ly and Lz , which were calculated by the same method as before. Thus, the corresponding vertex coordinates of rectangles Rx , Ry and Rz corresponding to the three polygons can be expressed as

$$Rx = \{(a_1 - Lx, b_1 + Lx), (a_1 + Lx, b_1 + Lx), (a_1 - Lx, b_1 - Lx), (a_1 + Lx, b_1 - Lx)\},$$

$$Ry = \{(a_2 - Ly, b_2 + Ly), (a_2 + Ly, b_2 + Ly), (a_2 - Ly, b_2 - Ly), (a_2 + Ly, b_2 - Ly)\},$$

$$Rz = \{(a_3 - Lz, b_3 + Lz), (a_3 + Lz, b_3 + Lz), (a_3 - Lz, b_3 - Lz), (a_3 + Lz, b_3 - Lz)\}.$$

Intersection rectangles of two rectangle areas were solved and marked as $Z_{Ri, Rj} = (a_{11}, b_{11}), (a_{12}, b_{11}), (a_{11}, b_{12}), (a_{12}, b_{12})$. For rectangle $Z_{Ri, Rj}$, left margin was expressed as a_{11} , right margin was expressed as a_{12} , upper boundary was expressed as b_{11} and lower boundary was expressed as b_{12} . The intersection rectangle of Rx and Ry was derived according to the above solution method and equation (5).

$$\begin{cases} a_{11} = \max(a_1 - Lx, a_2 - Ly) \\ a_{12} = \min(a_1 + Lx, a_2 + Ly) \\ b_{11} = \min(b_1 + Lx, b_2 + Ly) \\ b_{12} = \max(b_1 - Lx, b_2 - Ly) \end{cases} \quad (6)$$

The area of the location rectangle can be obtained as long as the same calculation was made in the other polygon areas based on equation (5).

3.3. Analysis to movements data based on K-Means

Motion trail of football players in court was obtained by the designed location algorithm. Those data including position coordinates based on which data mining can be used to obtain speed and running range. There were many algorithm cat-

egories and clustering algorithm was one of the typical algorithms. In clustering algorithm, every record was seen as an integral whole and several records formed as a set. Thus, clustering algorithm was about a process for dividing this set according to a certain standards.

One of the widely used clustering algorithms (see Fig. 6) was partial clustering. Under partial clustering, all records were divided into clusters, thus set of center points were obtained. Every record was then put into appropriate set of center points. Center points would be re-calculated by subset and all records would be divided by new set of center points. This was a process of repeated iteration which stops when the center points would not change when iterating. Take *K*-Means algorithm as an example, and Mahout algorithm for data mining (DM) was used to further analyze data.

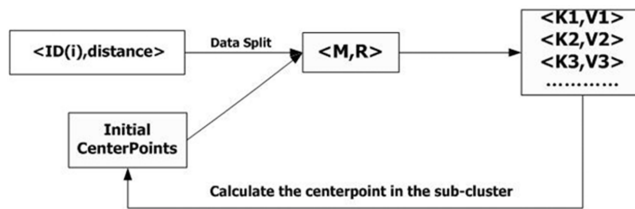


Fig. 6. Processing of *K*-Means with Mahout

Mahout is a machine learning library (MLL) which included many algorithms such as data mining algorithm and machine learning algorithm. Those algorithms were MapReduced to make large-scale data analysis on distributed cluster and Hadoop.

```

[root@newHad1 ~]# mahout kmeans \
> -i /mahout/kmeans-ex/reuters-sparse/tfidf-vectors \
> -c /mahout/kmeans-ex/reuters-clusters \
> -o /mahout/kmeans-ex/reuters-kmeans \
> -k 10 \
> -dm org.apache.mahout.common.distance.CosineDistanceMeasure \
> -x 10 -clustering -ow
  
```

Here, the operating parameters (-i) and (-o) denote input, parameter (-c) denotes information of initial center point, parameter (-k) denotes cluster numbers after dividing and parameter (-x) denotes the maximum number of iterations.

With *K*-Means cluster, running range and physical power of players in court can be obtained by analysis. According to previous location algorithm, players' data can be obtained, which is shown as follows.

Wang Ming: (15.063, 17.042, 5.64, 10.85, 11.75, 16.95, 30.45, 42.75.....)

Li Liang: (4.953, 7.102, 6.75, 5.47533, 6.12, 8.745, 7.11, 8.175.....)

These are the movement data of players at a certain position. All center data of players and distance between players and center can be obtained through *K*-Means, thus:

Wang Ming: (c: 18.811875, 1:3.748875, 2:1.769875, 3:13.171875, 4:7.96175.....)

Li Liang: (c: 6.8037912, 1:1.850791, 2:0.208208, 3:0.0537812, 4:1.32846.....)

Parameter c denotes the mean distance of movement distances and after being analyzed with K-Means, it can be know that Wang Ming's running range was larger and mean movement distance was longer while Li Liang's mean movement distance was much shorter. Thus, Li Liang should have more physical training.

4. Experimental results and their analysis

4.1. Transmission performance of ZigBee

Reliability of data transmission and acceptance of wireless network built based on ZigBee was tested in different speed and different numbers of data packages-sent.

Given the sent data package was fixed, transmission distance was progressively increasing, data units- sent were initially set as 2000 and sensor carried by players sent 5 times per minute. Area of football court was $50\text{ m} \times 95\text{ m}$, thus two subnets of star topology were set to collect data. The principal data concerning the wireless transmission is in Table 1.

Table 1. Wireless transmission 1

Transmission distance	Data units sent	Data units received	S-R efficiency
15	10000	10000	100 %
25	10000	9989	99.8 %
30	10000	9965	99.6 %
40	10000	9932	99.32 %
50	10000	9875	98.75 %

It can be seen from Table 1 that data acceptance efficiency was not lower than 90%. In the longest transmission distance, such as 50 meters, successful rate of data transmission can be further tested by changing sending rate of sensor.

Under the premise of same transmission distance (see Table 2), acceptance efficiency was reduced with the increasing number of data units- sent. Sending times of sensor was set less than 25, thus collection of players' movement data would not be influenced and acceptance efficiency was ensured no lower than 80 %.

4.2. Analysis to location algorithm

Central positions of different rectangles were obtained through simulation. Based on the improved centroid multilateration algorithm, the experiment verified that the central position of rectangle was at the centre of intersection area. However, there are errors in practice. Provided the maximum rectangle included in intersection area is $A'B'C'D'$, the test position is the midpoint of corresponding crosswire which is marked as O .

Practical center point is around point O and point P is defined as the maximum

error. Rectangle $A'B'C'D'$ and centroids X, Y, Z of three different polygons were put in a plain domain $F1F2YD$. Point Y is centroid of the larger polygon. $F1F2$ parallel to point O is the shortest side of rectangle $A'B'C'D'$. Line L is constructed vertically to $F1F2$. Make a line across centroid Y and vertical to line L and their point of the intersection is marked as point D . Thus, trapezoid $F1F2YD$ is built (see Fig. 7).

Table 2. Wireless transmission 2

Transmission distance	Data units sent	Data units received	S-R efficiency
	20000	19654	98.27 %
	30000	26864	89.54 %
	40000	35345	88.36 %
	50000	41752	83.5 %
	60000	47531	79.21 %

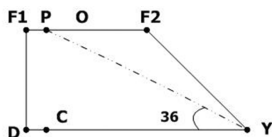


Fig. 7. Scope of $F1F2YD$

In practical calculation, it is hard to perform calculations based on rectangle $A'B'C'D'$ because of these errors and calculation time is quite long. After a simple logical derivation, absolute error point is located around point P on $F1F2$.

$$\text{Deviation : } e = PO = F1O - F1P. \tag{7}$$

Make a line crossing point P and vertical to line DY , which intersects line DY at point C . Thus, the below equation can be obtained

$$F1P = CD = YD - YC. \tag{8}$$

Given $YD = 2F1D = 2d$, the below equation can be obtained

$$F1P = 2d - 2d \times \cos(\overrightarrow{PYD}). \tag{9}$$

Lines $F1D$ and $YF2$ intersect at point T . Thus, there is an imaginary triangle DYT . According to theories relevant to centroid, the following equation was deduced.

$$F1F2 = YD \times (1/2). \tag{10}$$

According to equations (9) and (10), we get the relation

$$\text{Deviation : } e = YD * (1/2) \times (1/2) - F1P \approx 0.12d. \quad (11)$$

After deduction, it can be seen that the error value is in direct proportion to distance between the fixed nodes. In other words, if positions of the fixed nodes are identified, there would be an error.

After analyzing the error value, the maximum location range of players is set as $2\text{ m} \times 2\text{ m}$. The average error and standard deviation of three location algorithm in different coordinate range are compared. In Table 3, TM denotes trilateration, LC denotes locating centroid algorithm and TC denotes centroid multilateration algorithm.

Experiment was made in the maximum range of $2\text{ m} \times 2\text{ m}$ to collect movement coordinates of players. Tape was used to record actual position of players and several points were selected in each area. Signal was received at each point and these three algorithms were used to locate. Location times were not less than 50 times. Average error and standard deviation (see Table 3) were obtained according to errors between coordinates and practical results. Thus, it was known that centroid multilateration algorithm was more accurate than that of trilateration and locating centroid algorithms.

Table 3. Test results of $2\text{ m} \times 2\text{ m}$

Coordinates (m)	Algorithms	Average coordinates (m)	Error	St. deviation
(0.46,0.18)	TC	(0.67,0.157)	0.280	0.09
	LC	(0.69,0.07)	0.521	0.360
	TM	(0.73,0.13)	0.476	0.22
(1.78,1.05)	TC	(1.50,0.88)	0.241	0.129
	LC	(1.76,0.82)	0.359	0.322
	TM	(1.65,0.82)	0.257	0.307
(1.32,1.89)	TC	(1.54,1.52)	0.380	0.036
	LC	(1.79,1.58)	0.560	0.128
	TM	(1.70,1.70)	0.442	0.136

5. Conclusion

ZigBee technology was used to collect data of players in the field and two subnets were built under new topology to cover the whole court. Area of Football court was $50\text{ m} \times 95\text{ m}$, and the maximum movement range of players were set as $2\text{ m} \times 2\text{ m}$. Each player was carrying a sensor. More accurate centroid multilateration algorithm that covered wider range was put forwarded based on improved trilateration and locating centroid algorithm. Running distance of players and their physical abilities were

analyzed based on collected data. Thus, reasonable tactics were identified.

References

- [1] F. CUOMO, A. ABBAGNALE, E. CIPOLLONE: *Cross-layer network formation for energy-efficient IEEE 802.15.4/ZigBee wireless sensor networks*. *Ad Hoc Networks 11* (2013), No. 2, 672–686.
- [2] P. S. PURNIMA: *ZigBee and GSM based patient health monitoring system*. Proc. IEEE International Conference on Electronics and Communication Systems (ICECS), 13–14 February 2014, Coimbatore, India, IEEE Conference Publications (2014), No. 2, 1–5.
- [3] J. WANG: *ZigBee light link and its applications*. *IEEE Wireless Communications 20* (2013), No. 4, 6–7.
- [4] N. BAKER: *ZigBee and Bluetooth strengths and weaknesses for industrial applications*. *Computing & Control Engineering Journal 16* (2005), No. 2, 20–25.
- [5] H. C. TUNG, K. F. TSANG, K. L. LAM, H. Y. TUNG, B. Y. S. LI, L. F. YEUNG, K. T. KO, W. H. LAU, V. RAKOCEVIC: *A mobility enabled inpatient monitoring system using a ZigBee medical sensor network*. *Sensors (Basel) 14* (2014), No. 2, 2397–2416.
- [6] M. DAI, Y. H. WANG, Q. Y. PAN: *A design for smart home system based on ZigBee*. *Computer Measurement & Control 21* (2013), No. 03, 706–708.
- [7] M. J. HUANG, T. WANG: *Self-healing research of ZigBee network based on coordinator node isolated*. *Applied Mechanics and Materials 347–350* (2013), 2089–2094.
- [8] W. ZHENG, Z. M. LI, D. H. LUO: *Design and implementation of wireless network in smart home*. *Video Engineering 37* (2013), No. 21, 56–59.
- [9] L. X. WU, J. ZHAN, W. SHI: *The low power study based on ZigBee network*. Proc. IEEE International Conference on Wireless Communications, Networking and Mobile Computing WiCOM '08, 12–14 October 2008, Dalian, China, IEEE Conference Publications (2008), 1–4.
- [10] D. S. YUN, S. H. CHO: *A data transmission method in ZigBee networks using power efficient device*. Proc. IEEE International Conference on Advanced Technologies for Communications, 6–9 October 2008, Hanoi, Vietnam, Conference Publications (2008), 162–165.
- [11] P. KINNEY: *ZigBee technology: Wireless control that simply works*. Proc. Communications Design Conference, 29 September–2 October 2003, San Jose, CA, *USD 2* (2003), 1–20.

Received May 7, 2017