ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Reducing Energy Consumption in WBSNs using Statistic Test to detect Duplicated Data

Saleh Massoud[1], Youssef Sobh Ahmad[2], Mothanna Alkubeily[3], Ahmad S. Ahmad[4]

Third Year Student, Faculty of Medical Engineering, Al-Andalus University, AlQadmous - Tartus, Syria[1]

Fourth Year Student, Facultyof Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria[2]

Assistant Professor, Facultyof Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria [3]

Assistant Professor, Facultyof Medical Engineering, Al-Andalus University, AlQadmous - Tartus, Syria[4]

ABSTRACT: Duplication and redundancy in WBSNs refer to collect the same vital signs, using sensor attached on the patient body, over a period of time. Transmitting duplicated and similar data well consume the sensor battery faster, and increase the traffic in the network. Therefore, in this paper, we introduce a lightweight technique to detect duplicates data, using statistic test to identify the similarity between the data. Then preventing the sensor from sending this duplicated data,which leads to increase the lifetime of the sensor, and reduce energy consumption.

KEYWORDS: Wireless Body Sensor Networks; data duplication; Redundancy; Lightweight technique.

I. INTRODUCTION

The increasing use of wireless network and the constant miniaturization of electrical devices has empowered the development of Wireless Body Sensor Networks (WBSNs), which is defined of an autonomous system that is used to monitor the daily life activities of patient [1]. It consists of intelligent sensor nodes which do not hamper the daily life activities of the patient. The sensors are attached on clothing, on the body, or even implemented under the skin. they allow the individual to closely monitor the changes in their vital signs and provide feedback to help maintain an optimal health states [2].

WBSNs are useful in detecting chronic diseases such as heart attack, asthma, diabetes, EEG …etc, and to warn the patient in case of emergency conditions. Fig. 1 shows the sensors on human body and their utilization.

Using WBSNs, the patient experiences a greater physical mobility and is no longer compelled to stay in the hospital. Like any innovations, WBSNs seek to reduce risks and give more services to the patient. However, any innovation introduces new challenges. The challenges will be tacked as the following [3]:

- *Reliability:* the main challenge is to ensure that the information reliably gets to its destination.
- *Privacy and security:* there are several security issues to be considered, such as eavesdropping, identity spoofing. Security can be improved using data encryption [4].
- *Energy consumption:* it is very important to increase the lifetime of the sensor by reducing energy consumption [5]. The battery in the sensor is designed in small size and should has a long-life time, so the patient doesn't have to replace it every short amount of time.
- *Delivery time:* since medical data that are collected from the patient might be, in some cases, sensitive and critical, which may threaten his life. The data should be delivered immediately to the hospital or the doctor, so the required transmission time should be shortened.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

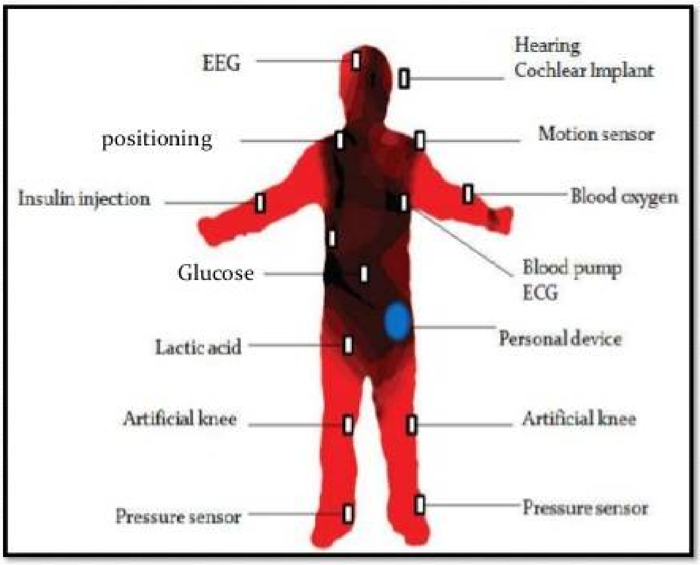

Fig.1.Sensors on Body and their utilization

As we know, vital signs can vary from an emergency value to normal values (duplicated) and vice versa. The changing rate of the signs can be quick or slow regarding the patient's situation and the vital signs. Since the sensor has low capabilities for far transmission, it sends all his collected data to another sensor, which forward it to another sensor and so on, until the data are delivered to the Access Point and then to the hospital, like multi-Hop communication [6]as shown in fig. 2.By doing this, the sensor's energy will be consumed in the network, and we are not considering one vital sign in a day. We are interested in continuously generated data and that drains the sensors power. consequently, leading to several collisions in the network.

The significant amounts of the data sent from the sensor are either redundant or almost similar, it initiates the requirement of deduplication of the sensor data [7]. Instead of sending all the normal data collected from the sensor (which are often duplicated), the sensor should send only some specific normal measures, and of course the emergency values, leading to reduce the rate of unnecessarily transmission, thus improving power efficiency.

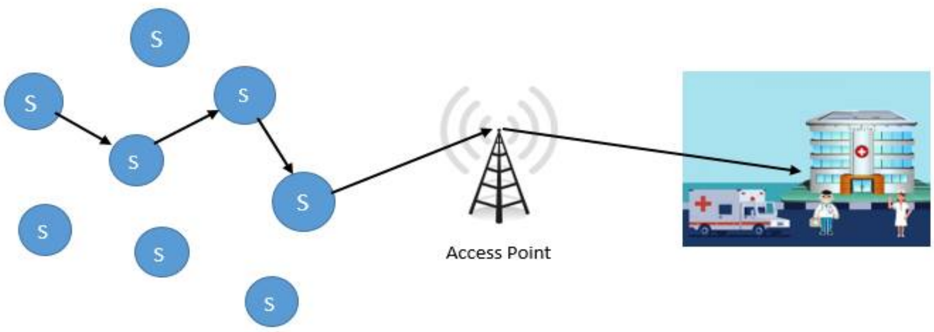

Fig.2. Multi-hop communication in WBSN

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

II. Related work

In this section, we describe popular researches about deduplication techniques and energy consumption.

In [6] the authors find an energy efficient MAC protocol design for swallowable BSNs. They present a new protocol where they proved that multi-hop communication is more energy efficient than single hop communication within the human body when the circuitry is low.

Authors in [7] introduced detection of duplicated data from sensors. They aim to de-duplicates of similar and redundancy data, in the processing server, by extracting the features from the data, using Local Binary Pattern (LBP), and then matching the patterns with each other to find if they are similar or duplicated. The major drawback when using LBP that it sometime creates a zero values, and if so; when to match the patterns using Chi-square formula, they may divide by zero, and that make their work have some problems.

A data reduction with low overhead was introduced in DARE [8]. The idea behind DARE is to employ a scheme, call Duplicate-Adjacency based Resemblance Detection (DupAdj), by considering any two data chunks to be similar (i.e. candidates for delta compression) if their respective adjacent data chunks are duplicate in a deduplication system. Then furtherenhance the resemblance detection efficiency by an improved super-feature approach. Here the data fragmentation in data reduction systems left over a serious problem.

In [10], authors worked on adaptive sampling approach, based on fisher test theory, that estimates and adapts the sensing frequency based on previous readings and the patient criticality, to optimize the energy consumption.

Another approach is introduced in [11], which proposed a scheme to deduplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption. It integrates cloud data deduplication with access control.

A learning-based deduplication system is presented in [12], it uses a novel method of interactively discovering challenging training pairs using active learning. Deduplication integrates data from multiple sources.

III. Proposed algorithm

A. Overview:

For the pre_mentiond reasons, we present our proposed work, in order to reduce energy consumption, which leads to increase the lifetime of the sensor, and considering minimum transmission time.

The generated data are analyzed and matched with patterns for removing similar and duplicated data. The proposed work shows that we can detect duplicated data and prevent the sensor from sending them, with lightweight software program, and minimum transmission time, in both cases: if there is an emergency vital signs or not. Fig. 3 shows the flowchart for our proposed work.

Our idea concentrates on using statistic test in order to prevent the sensor from sending duplicated data. However, if the data appeared to be not duplicated, we should ensure minimum transmission time because we are dealing with patient information which needs to be delivered immediately to the hospital or the doctor, to take the fastest procedure.

For this reason, we use Chi-Square test [13] to identify the degree of similarity between the new out coming and the saved data, to decide whether it should be sent or not.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

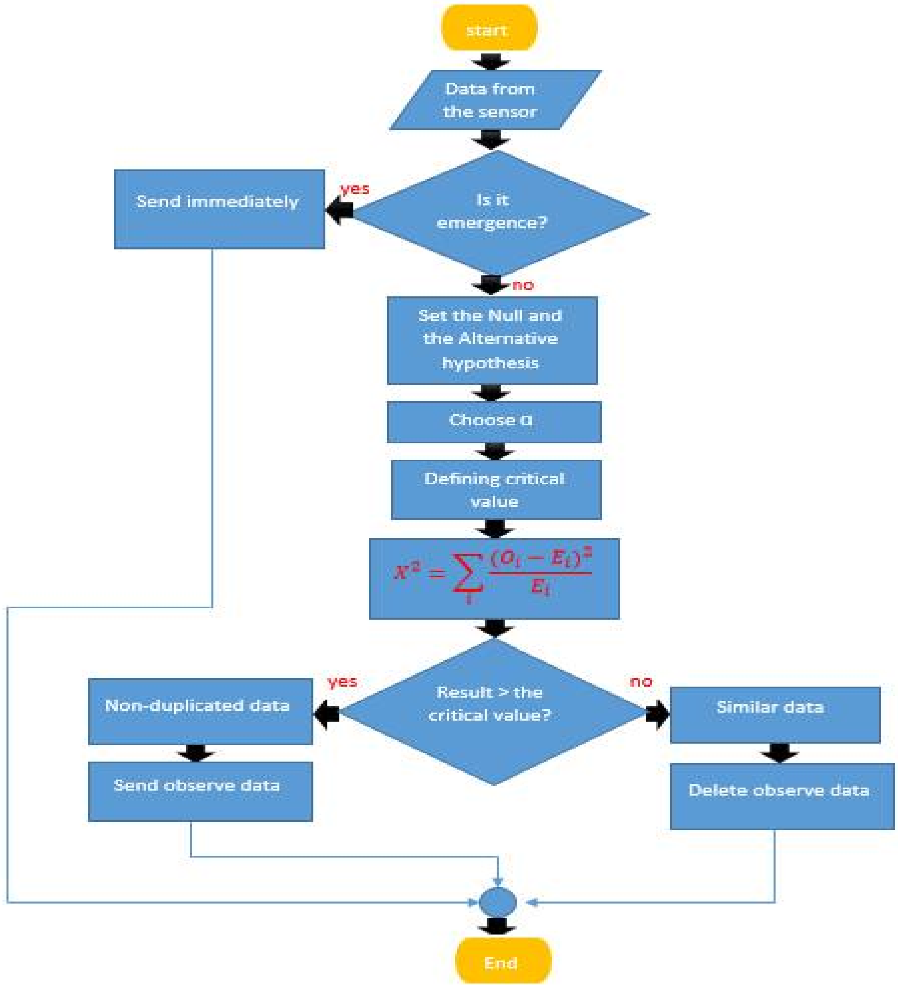

Fig.3.Flowchart of our proposed Algorithm

B. Description of our Proposed Algorithm:

There is the so-called "the Expected data" in the sensor, which contains the vital signs that the sensor has already collected and saved, to be the original pattern that are going to be compared to the other set of data in order to find the degree of similarity. The other set of data, on the other hand, that are collected, are called the "observed data", which are the new sensed data.

In the next section, Hypothesis Testing Steps [14], are going to be applied to find if the observed data are close to the expected data.

<u>First step:</u> we set the null and the alternative hypothesis:

- The null hypotheses(or H_0): the data are duplicated and similar, indicates that the observed data are close to the expected data; the new sensed data are similar to the saved ones (no change or no difference from the past (Mathematical definition)). This kind of data shouldn't be sent, *but rather deleted.*
- The alternative hypotheses (or H_1): the data are not duplicated, this shows that the observed data are not close to the expected data; the new sensed data are different from the saved data, this kind of data *should be sent.*

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Second step:choose level of Significance:α

The level of significance is just the area in the tail in chi-square distribution, fig. 4 shows this area. The smaller the value the more significant the results. Since we are dealing with medical data, we choose the smallest valuefor (α) which is $(\alpha = 0.001)$.

The previous mentioned area is called the Rejection Region; it is with highly important as it allows for a final conclusion at the end of the statistic test. If the result of the statistic test (Chi-square test) falls into the rejection region, that means we can reject the null hypothesis, and accept the alternative one. In other word, we can detect that the observer data are not close to the expected data, leading to, sending the observed data. On the other contrary, if the result of the test falls out of the rejection region, that means we can reject the alternative hypothesis and accept the null hypothesis, and detect that the observed data are close to the expected one, and delete the observed data, and not doing the transmission party, leading to reducing the energy.

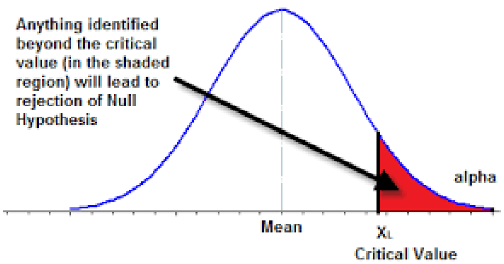

Fig.4. Area in Chi-square distribution

Third step:defining the critical value:

The critical value is the point that separates the tail from the rest of the curve. To find the critical value we need two parameters: the level of significance, and the degrees of freedom.The degree of freedom is equal to the (possible outcomes -1). Since we have 25 reading from the sensor, the degree of freedom equal to 24, and the critical value in this case, from Chi-square table, fig. 5, is going to be (51.1786) [which we are going to use in the simulation].

Fourth step:Apply the statistic test using chi-square test:

The formula that is used for this test is: [which we are going to use in the simulation].

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

After finding the result of chi-square test, the comparison can be made between this result and the critical value. If the result is greater than the critical value, which in return means the result falls into the rejection region, we reject the null hypothesis, that leads to send the observed data. On the other hand, if the result is less than the critical value; which means the result is out of the rejection region, we accept the null hypothesis, in this case we detect the duplicated and similar data, then we prevent the sensor from sending the observed data.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Table1:Chi-square table

df	α					
	0.100	0.050	0.025	0.010	0.005	0.001
20	28.4120	31.4104	34.1696	37. 5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031

IV. SIMULATION RESULTS

In order to execute our simulation, critical value is needed, as well as finding the statistic test "using chi-square test", and compare the result with the critical value to determine if the data are duplicated or not. Note that the result for the equal patterns is zero.The simulation tool is MATLAB 13.
We take three scenarios for diabetes patient, with real-values [18].

A. First scenario:the data are duplicated

the Expected data are:
(120,120,125,130,125,110,119,111,119,121,125,130,127,119,121,126,127,121,120,113,135,120,125,125,130)
The observed data are:
(120,120,120,130,125,110,119,111,120,120,125,130,139,130,125,126,127,125,129,120,125,130,129,120,130)
In the first scenario we can see that the observed and the expected data are close to each other. The simulation result for chi-square test is (5.3856) which is less than the critical value, leading to the undertaken procedure is, then, automatically deleting.
The required time to do this is (1.63 * 10^{-3} s).

B. Second scenario:The data are not similar

The observed data are:
(90,100,110,90,110,120,120,125,110,120,130,135,130,135,100,120,135,140,110,120,90,120,110,110,110)
We can see that the observed data are not close to the pre_mentiond expected data. The simulation result for chi-square is (71.785) which is greater than the critical value, so the undertaken procedure is to send the observed data.
The required time for the process to do this is (1.6 * 10^{-3} s)

C. Third scenario: There is an Emergence value

The observed data are:
(90,100,110,50,110,120,120,125,110,120,130,135,130,135,100,120,135,140,110,120,90,120,110,110,110)
We can see here that we have an emergence value "50".
In this case, we don't do any kind of processing on the observed data. It is, then, sent immediately.The simulation time for this is (7.5 * 10^{-4} s). Fig. 6 shows the required time for each scenario

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

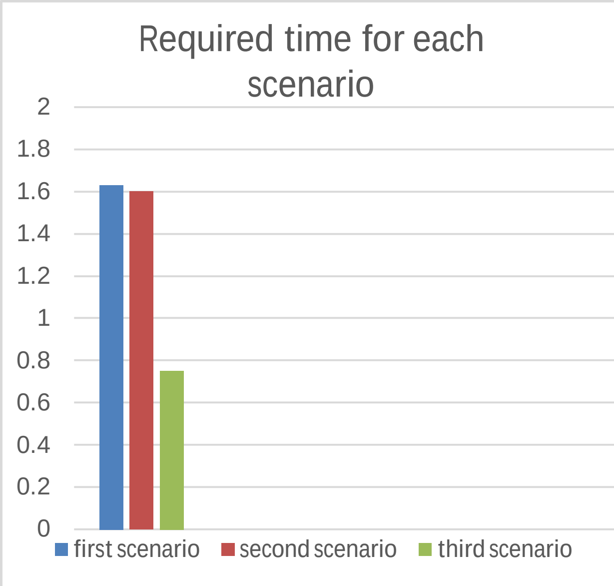

Fig.5. Required time in each scenario in ms

V. Conclusion and Future Work

This research aims to develop a lightweight deduplication technique. Thistechnique would help to detect similar, redundant and duplicated data generated from the sensor; which drain its energy. Our proposed algorithm considerably reduced energy consumption by dropping redundant and similar data, with lightweight software program, and minimum transmission time.

As a future work, we seek to apply our algorithm on many sensors that are sensing the same vital sign, in order to increase their lifetime, and to reduce the huge amount of data that are received in the hospital or the doctor, to make the processing of this data smoother.

REFERENCES

1. A. Sangwan and P.P. Bhattacharya, "Wireless Body Sensor Networks: A Review". International Journal of Hybrid Information Technology, 8(9): 105-120, (2015).
2. C. Otto, A. Milenkovic, C. Sanders and E. Jovanov, "System Architecture of a Wireless Body Area Sensor Network for Ubiquitous Health Monitoring", Journal of Mobile Multimedia, 1(4): 307-326, (2006).
3. Y. Hao and R. Foster, "Wireless body sensor networks for health-monitoring applications", Physiological Measurement,29(11), (2008).
4. D.H. Gawali and V.M. Wadhai, "RC5 Algorithm: Potential Cipher Solution for Security in Wireless Body Sensor Networks (WBSN)". International Journal of Advanced Smart Sensor Network Systems (IJASSN), 2(3):1-7, (2012).
5. S. K. Makki, M. Stangl and N. Pissinou, " Energy Consumption in Wireless Sensor Networksusing Data Fusion Assurance", International Journal of Wireless & Mobile Networks (IJWMN), 5(2), (2013).
6. L. Lin, C. Yang, K.J. Wong, H. Yan, J. Shen and S.J. Phee, " An energy efficient mac protocol for multi-hop swallowable body sensor networks", Sensor, 14(10): 19457-19476, (2014).
7. S. Beulah and F.R. Dhanaseelan, " Detection of duplicated data with minimum overhead and secure data transmission for sensor big data", Springer, (2017).
8. W. Xia, H. Jiang, D. Feng and L. Tian, "DARE: A deduplication-aware resemblance detection and elimination scheme for data reduction with low overheads", IEEE Trans. Compute. 65(6), 1692–1705 (2016).
9. T. Guo, L. Zhang, W. Liu and Z. Zhou, "A novel solution to power problems in implanted biosensor networks", 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2006, pp. 5952-5955, Aug 2006.
10. C. Salim, A. Makhoul, R. Darazi and R. Couturier, " Adaptive sampling algorithms with local emergency detection for energy saving in Wireless Body Sensor Networks". IEEE/IFIP Network Operations and Management Symposium (NOMS), 2016.
11. Z. Yan, W. Ding, X. Yu, H. Zhu and R.H. Deng, "Deduplication on encrypted big data in cloud". IEEE Trans. Big Data 2(2), 138–150 (2016).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

12. S. Sarawagi and A. Bhamidipaty, "Interactive deduplication using active learning", proceedings in 8th ACM SIGKDD international conference in knowledge discovery data mining, 269-278, (2002).
13. https//helthiack.com, last visit 20/11/2017.
14. Hogg, McKean, Craig, "Introduction to Mathematical Statistics", [Book] Pearson, Seven Edition,2005
15. I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani and S.U. Khan, "The rise of "big data" on cloud computing: Review and open research issues ". Inf. Syst. 47, 98–115 (2015)
16. K. Akhila, A. Ganesh and C. Sunitha, "A Study on deduplication techniques over encrypted data". Proc. Comput. Sci. 87, 38–43 (2016)
17. J. Stanek, A. Sorniotti, E. Androulaki and L. Kencl, "A secure data deduplication scheme for cloud storage". In: International Conference on Financial Cryptography and Data Security. Springer, Heidelberg (2014)
18. Albert Satorra, Peter M. Bentler: "A scaled difference chi-square test statistic for moment structure analysis". Psychometrika, 2000.