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. 6, Issue 2, February 2018

An Efficient Approach to Aggregate Big Data in Wireless Sensor Network

Raju B N

Dept. of Computer Science, Akkamahadevi Women's University, Vijayapura, India

ABSTRACT*:* In this paper the proposed system work for the data aggregation in Wireless Sensor Network in Big Data, a lot of people with small devices and mobile devices such as actuators, sensors and robots generates tremendous amount of data. Big data where it powerfully demands a network infrastructure having the ability to efficiently process, collect, share, cache and deliver the data instead of simple transmissions, such network designs show the requirements of availability, energy efficiency, data aware intelligence and high performance. The big data is characterized by volume, variety; value, velocity and its applications face challenges in processing, transmitting, sharing, acquiring, visualizing, storing and analyzing data. The proposed system concentrates on network designs for big data sharing. To reach these requirements, we adopt Information-Centric Networking (ICN) approach where in-network caching is utilized and data are retrieved from names.

KEYWORDS: Big data, Information centric network, CCN, Energy-efficiency

I. INTRODUCTION

Sensor networks are collection of sensor nodes which co-operatively send sensed data to base station. As sensor nodes are battery driven an efficient utilization of power is essential in order to use networks for long duration hence it is needed to reduce data traffic inside sensor networks, reduce amount of data that need to send to base station. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offers that sensor nodes need less power for processing as compared to transmitting data. It is preferable to do in network processing inside network and reduce packet size. One such approach is data aggregation which is an attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. Wireless sensor networks have limited computational power and limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols. In this paper, a data aggregation framework on wireless sensor networks is presented. The framework works as a middleware for aggregating data measured by a number of nodes within a network.

Big data are generally generated by and collected from geographically distributed devices and stored in data warehouses for processing in powerful data centers with massive interconnected servers. Its applications face challenges in acquiring, storing, processing, sharing, transmitting, analyzing and visualizing data with very large quantities. This works focus on the network designs for big data sharing. Video sharing applications allow users to upload multimedia contents to data centers and share them with their friends in real time. For IoT services the data generated from vast amount of sensors are collected, stored, processed, visualized and delivered to the users. On the other hand, the Internet is originally designed for End-to-end communications where the networks serve as the data transmission pipes that connect data sources, data centers, and users. Big data will overwhelm the current communication networks because huge amounts of data sharing applications produce redundant and duplicate traffic if networks simply act as transmission pipes. This huge volume of traffic hinders efficient data flows and gives internet the difficulties in providing the highly available services for these applications.

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. 6, Issue 2, February 2018

II. LITERATURE SURVEY

Sanjeev SETIA et al [1] described many sensor applications; the data collected from individual nodes is aggregated at a base station or host computer. To reduce energy consumption many systems perform in-network aggregation of sensor data at intermediate nodes and enroute to the base station. Most existing aggregation algorithms and systems do not include any provisions for security and consequently these systems are vulnerable to a wide variety of attacks. In particular, compromised nodes can be used to inject false data that leads to incorrect aggregates being computed at the base station. The user discuss the security vulnerabilities of data aggregation systems, and present a survey of robust and secure aggregation protocols that are resilient to false data injection attacks.

Tsung-Yi Tsai et al [2] proposed a method for wireless sensor networks which are used extensively in environment and habitat monitoring; the large volume of data transmission can increase the workload of the sensor nodes and reduce their useful lifetime. The compressive sampling techniques have been proposed to reduce the volume of data transmission when the data is sparse in certain domain, while finding the optimal routing path that minimizes data traffic is an NP-complete problem, a near-optimal routing protocol in the literature requires omniscient knowledge of the entire network and thus incurs extensive message exchanges in real applications. This work proposes for a distributed algorithm that uses local minimization to dynamically construct a routing path to reduce the data traffic for compressive sampling based aggregation. This algorithm does not require the omniscient knowledge of the global network topology and incurs much lower overhead than the near optimal solution and therefore is more suitable for practical applications.

Mohammad Abu et al [3] proposed a method for wireless sensor networks to monitor dynamic environments that change rapidly over time. This dynamic behavior is either caused by external factors or initiated by the system designers themselves. To adapt to such conditions, sensor networks often adopt machine learning techniques to eliminate the need for unnecessary redesign. Machine learning also inspires many practical solutions that maximize resource utilization and prolong the lifespan of the network. In this paper, the user present an extensive literature review over the period 2002-2013 of machine learning methods that were used to address common issues in wireless sensor networks (WSNs). The advantages and disadvantages of each proposed algorithm are evaluated against the corresponding problem. It also provides a comparative guide to aid WSN designers in developing suitable machine learning solutions for their specific application challenges.

Jyoti Rajput et al [4] described that in wireless Sensor network data aggregation is an important technique to achieve power efficiency in the sensor network. In some applications such as: wireless sensor network, data mining, cloud computing data aggregation is widely used. Because sensor node has limited battery power so data aggregation techniques have been proposed for wireless sensor networks. A challenge to data aggregation is how to secure aggregated data from disclosing during aggregating process as well as obtain accurate aggregated results. Also described various protocols for securing aggregated data in wireless sensor networks.

III. PROPOSED SYSTEM

The architecture of proposed system is shown in figure:1 initiated with network establishment which is followed by clustering of communication nodes using K-means And collected data is aggregated to the cluster head memory. Aggregately name-based routing (ANBR) [5] method is used for data retrieval and information retrieved is sent to data center. ANBR comprises two level registration and retrieval. Wireless transfer of data is achieved using Wi-Fi module ESP8266.

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. 6, Issue 2, February 2018

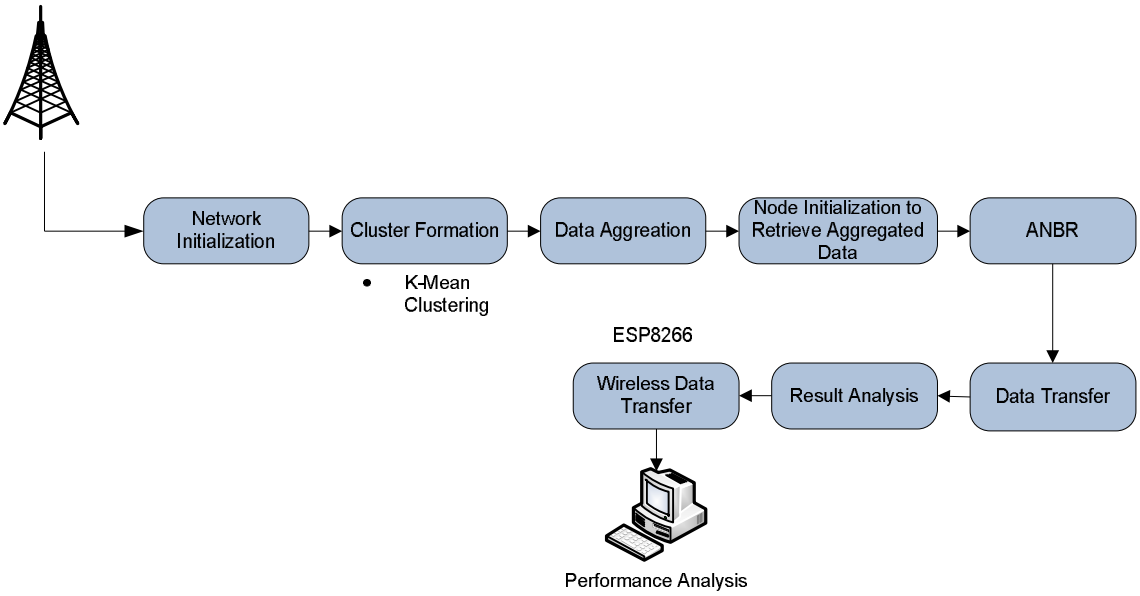

Figure 1: Architecture of Proposed System

a. Network Initialization

Initialization is the process of locating and using the defined values for variable data that is used by a computer program. For example, an operating system or application program is installed with default or user-specified values that determine certain aspects of how the system or program is to function. Typically, these values are stored in initialization files (in Windows, these can be identified as files with an INI suffix). When the operating system or an application program is first loaded into memory, a part of the program performs initialization - that is, it looks in the initialization files, finds definite values to substitute for variable values and acts accordingly. For example, the desktop appearance and application programs that are to be started along with the operating system are identified and loaded.

Network initialization starts with grouping sensor nodes under single cluster head [6] called parent node. Cluster head collects aggregated data from all nodes in cluster and sends to sink or base node. There will be many such clusters in network under single sink node. Every node in cluster is identified by node ID.

b. Cluster using K-means Algorithm

In the K-means algorithm Cluster Head (CH) is selected based on the residual energy in the nodes and Euclidian distance, hence the node present at the centre collects all the information about the node id, position of the nodes and remaining energy of all the nodes. The collected information is stored in the list of centre node, once the information is collected than the clustering process starts [12]

1. 'k' defines the number of clusters, in the network 'k' number of centroids are initialized at random places.
2. The Euclidian distance is calculated from each node to all centroids and nearest nodes is assigned to it, by this 'k' number of clusters is done.

   Suppose there are n nodes are given such that each one of them belongs to Rd. The problem of finding the minimum variance clustering of this node into k clusters is that of finding the k centroids $\{m_j\}_{j=1}^{k}$ in R_d such that

$$\left(\frac{1}{n}\right) X \sum \left(\frac{min}{j} \, d^2 \left(x_i, m_j\right) \right), for\ i = 1\ to\ n \qquad (1)$$

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. 6, Issue 2, February 2018

Where $d(x_i, m_j)$ represents the Euclidean distance between x_i and m_j, are called clusters centriod or as cluster means.

3. In each cluster the position of the centroids is calculated again and check the change in position with the previous one
4. If any changes in the position of the centriod then go to STEP 2, or else the clusters are finalized and the clustering process stops.

The 'k' number of clusters is done in the network and the CHs are chosen

c. Data Aggregation

Data aggregation is a process of aggregating the sensor data using aggregation approach. The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approaches K-means, TAG (Tiny Aggregation) etc. This aggregated data [9] is transferred to sink node by selecting the efficient path. In these data aggregation algorithms we gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor network offers, that a sensor nodes need less power for processing [8] as compared to transmitting data. It is preferable to do in network processing inside network and reduce packet size. One such approach is data aggregation with attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. Wireless sensor networks have limited computational power, limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols.

d. ESP 8266 Module

The ESP8266 module is a TTL "Serial to Wireless Internet" device. Providing your microcontroller has the ability to talk to a TTL serial device (most do) you'll be in business! The original instructions have been translated from Chinese into cryptic data sheets. We'll try to change that with this Instruct able. The ESP8266 module is a 3v device, but it's no wimp. It draws quite a bit of power. In fact, you'll probably need to make sure that your circuit's power supply can handle at least 1 amp of power. In order to provide security over collected data from sensors a Blowfish algorithm [12] is used for encrypting and decrypting data sent from cluster head to sink node.

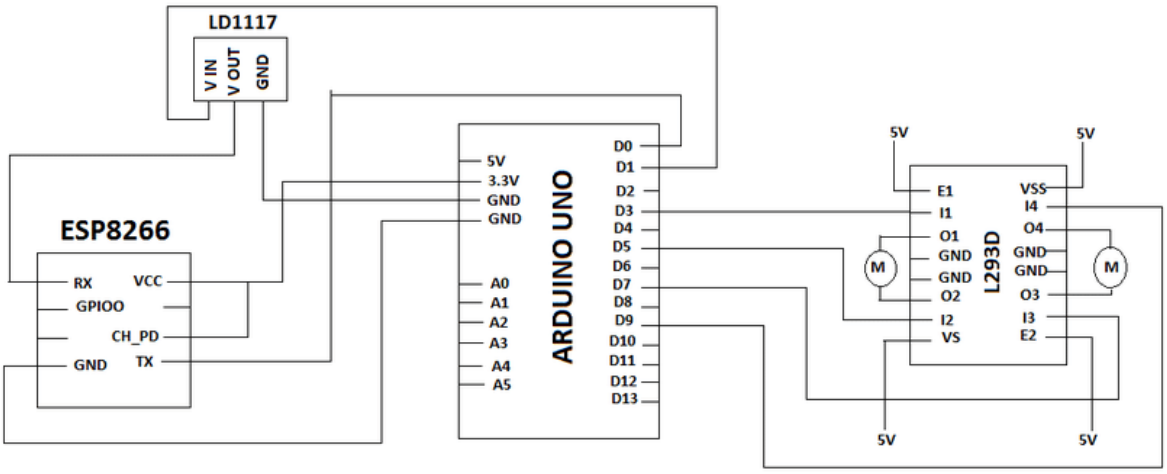

Figure 3: Overall Block Diagram for Arduino UNO Connection

A symmetric key distributed by sink node to all nodes in cluster. Data collected at sink node is sent through ESP8266 Wi-Fi module to the server.

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. 6, Issue 2, February 2018

e. Data Security

Data security helps keeps to keep data private. Secure data transmissions prevent contact lists and personal e-mail from being read by someone other than the intended recipient, keep firmware upgrades out of devices they don't belong in, and verify that the sender of a piece of information is who he says he is. The sensibility of data security is even mandated by law in certain applications: in the U.S. electronic devices cannot exchange personal medical data without encrypting it first, and electronic engine controllers must not permit tampering with the data tables used to control engine emissions and performance. Data security techniques have a reputation for being computationally intensive, mysterious, and fraught with intellectual property concerns. While some of this is true, straightforward public domain techniques that are both robust and lightweight do exist. One such technique, an algorithm called Blowfish, is perfect for use in embedded systems which encrypt data for end- end services with the help of symmetric key distribution.

IV. EXPERIMENTAL RESULTS

Results from the proposed system can be inferred that, as number of nodes increase in a network there is a decrease in loss of system energy when compared to existing system. Below graph shows a result for proposed system. Initially a set of network nodes are initialized, the respective pictorial representation depicted in Figure 4(a). Results from the below graph i.e. Figure 4 (d), (e) can be inferred that, when number of nodes increase in a cluster packet drop from respective nodes also increase.

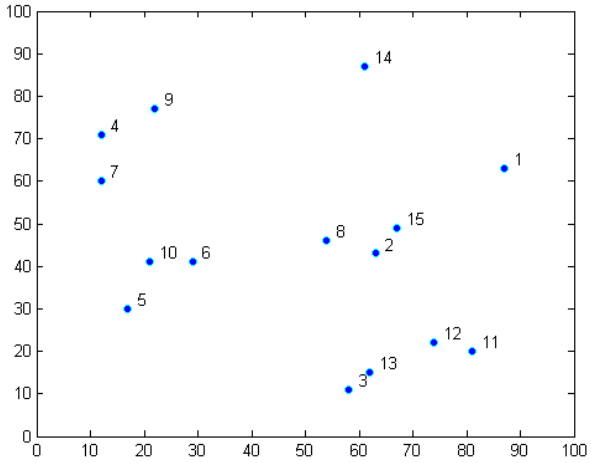

Figure 4(a): Network Initialization

In proposed system from stimulation results in MATLAB it is easy to find a particular node that drops the packet and that particular node are avoided in further rounds. Therefore, with implementation of K-means protocol [10] proposed system improves system energy and authentication. The respective cluster formation and cluster head selection (i.e. CH) is depicted in Figure 4(b) and 4(c).

The nodes are randomly created distributed around with a single base station. The idea is to form cluster of sensor nodes based on signal strength and use the cluster-head as a router to forward data of other nodes in cluster to the base station.

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. 6, Issue 2, February 2018

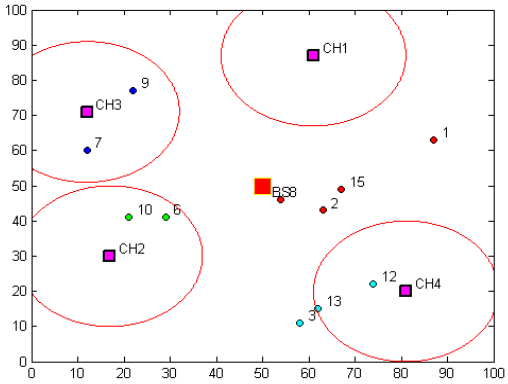

Figure 4(b): Cluster formation

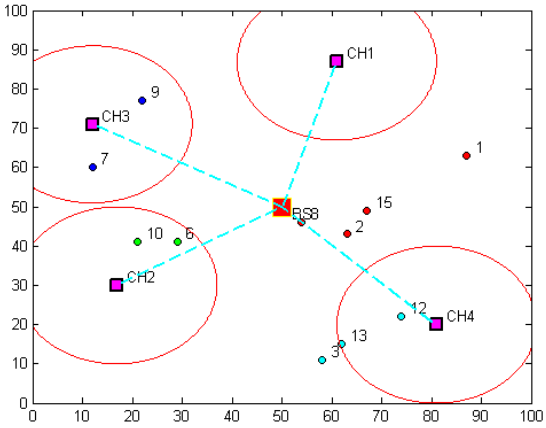

Figure 4(c): Cluster Head Selection and Key Distribution

The data processing is performed at cluster-heads. K-means is a dynamic clustering protocol where symmetric key is distributed by the base station to all cluster head in order to increase the security of the system. The respective pictorial output is depicted in Figure 4(c).Blow fish algorithm is used to generate the key for encrypting and decrypting data among sensor nodes. And also represents data passing through the Arduino module because it stores the information and then passes it through Wi-Fi module through the network connection. The respective AT commands and connections are shown in below Figure 4(f), (g) and (i).

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. 6, Issue 2, February 2018

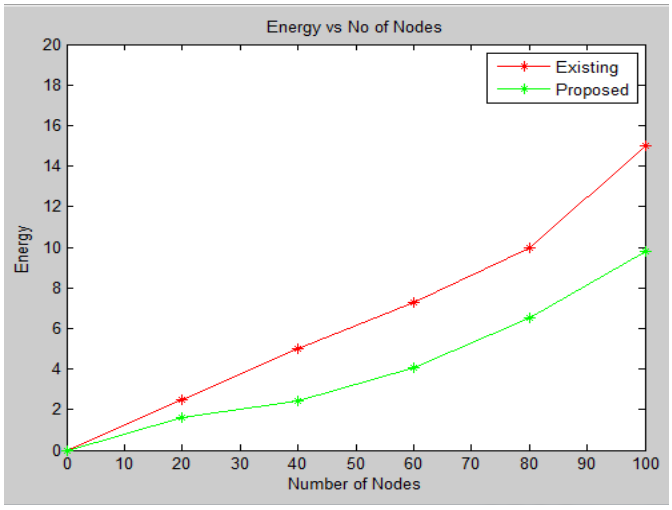

Figure 4(d): Number of Nodes Vs System Energy

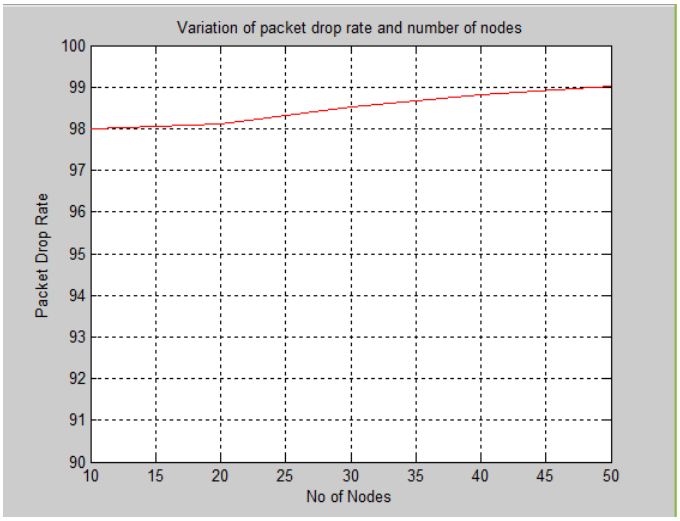

Figure 4(e): Number of nodes vs packet drop

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. 6, Issue 2, February 2018

Function	AT Command	Response
Working	AT	OK
Restart	AT+RST	OK [System Ready, Vendor:www.ai-thinker.com]
Firmware version	AT+GMR	AT+GMR 0018000902 OK
List Access Points	AT+CWLAP	AT+CWLAP +CWLAP:(4,"RochefortSurLac",-38,"70:62:b8:6f:6d:58",1) +CWLAP:(4,"LiliPad2.4",-83,"f8:7b:8c:1e:7c:6d",1) OK
Join Access Point	AT+CWJAP? AT+CWJAP="SSID","Password"	Query AT+CWJAP? +CWJAP:"RochefortSurLac" OK
Quit Access Point	AT+CWQAP=? AT+CWQAP	Query OK
Get IP Address	AT+CIFSR	AT+CIFSR 192.168.0.105 OK
Set Parameters of Access Point	AT+ CWSAP? AT+ CWSAP= <ssid>,<pwd>,<chl>, <ecn>	Query ssid, pwd

Figure 4(f): AT Commands

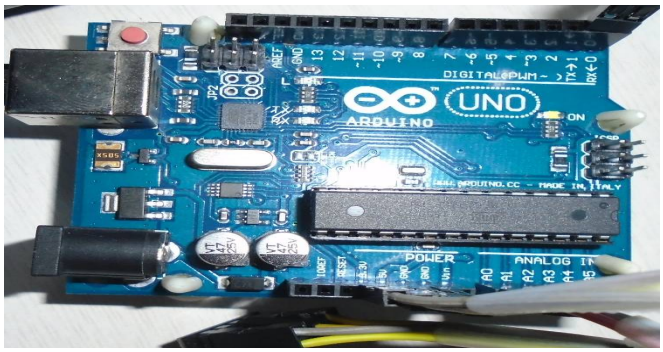

Figure 4(g): Arduino Board

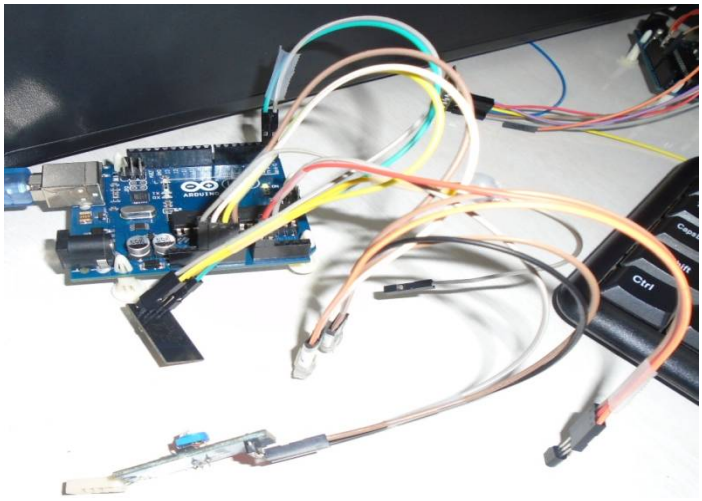

Figure 4(i): Overall System Connection using ESP8266 Wi-Fi Module

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. 6, Issue 2, February 2018

V.CONCLUSION

Wireless Sensor Networks are more helpful in different applications such as health, office military surveillance, home monitoring and in many smart and intelligence systems. There are some problems to the security of the network in wireless sensor networks and safe data aggregation is also one of the big problems. The proposed system presents brief discussion of the wireless sensor network, security needs to data aggregation, data aggregation, overview of different security protocols and their comparison, different techniques of data aggregation in wireless sensor network [11].

REFERENCES

[1] Sanjeev SETIA a, Sankardas ROYb and Sushil JAJODIA b, "Secure Data Aggregation in Wireless Sensor Networks".
[2] Tsung-Yi Tsai, Wei-Chi Lan, Chunlei Liu, and Min-Te Sun, "Distributed Compressive Data Aggregation in Large-Scale Wireless Sensor Networks", Vol. 1, 2013.
[3] Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato and Hwee-Pink Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications".
[4] Jyoti Rajput, Naveen Garg, "A Survey on Secure Data Aggregation in Wireless Sensor Network", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 4, Issue 5, 2014.
[5] Nandini. S. Patil, Prof. P. R. Patil, "Data Aggregation in Wireless Sensor Network".
[6] Wenbo He, Xue Liu, Hoang Nguyen, Klara Nahrstedt, "PDA: Privacy-Preserving Data Aggregation in Wireless Sensor Networks".
[7] Kiran Maraiya, Kamal Kant, Nitin Gupta, "Efficient Cluster Head Selection Scheme for Data Aggregation in Wireless Sensor Network". International Journal of Computer Applications,Vol 23, 2011.
[8] Md. Zair Hussain, M. P. Singh and R. K. Singh, " Analysis of Lifetime of Wireless Sensor Network", International Journal of Advanced Science and Technology, Vol. 53, 2013
[9] Anindita Ray, Debashis De, "Data Aggregation Techniques in Wireless Sensor Network: A Survey",International Journal of Engineering Innovation & Research, Vol 1, Issue 2.
[10] Chunyao FU, Zhifang JIANG, Wei WEI2and Ang WEI, "An Energy Balanced Algorithm of LEACH Protocol in WSN", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, 2013.
[11] Joseph Polastre, Robert Szewczyk, Alan Mainwaring, David Culler and John Anderson, "Analysis of Wireless Sensor networks For Habitat monitoring".
[12] Sasikumar P and Khara S, "K-means Clustering in Wireless Sensor Networks", In Computational intelligence and communication networks (CICN), fourth international conference. IEEE, pp. 140-144, 2012.