



Redundant Data Filtering in WSN and RFID Network Integration

Ashwini W. Nagpurkar¹, Siddhant K. Jaiswal²

M. E. Student, Dept. of CSE, WCC, GHRCE, Nagpur, India¹

Assistant Professor, Dept. of CSE, CSE, GHRCE, Nagpur, India²

ABSTRACT: The integration of Wireless Sensor Network (WSN) and Radio Frequency Identification (RFID) is an emerging technology which utilizes identified advantages of both systems making it more efficient and reliable. The Hybrid network formed by integration of WSN and RFID systems provide excellent infrastructure to process, acquire and distribute data in decentralized dynamic environments. The integrated network involves various challenges among which redundant data is critical as it is coupled with time delay and energy consumption which results in waste of various network resources. In this paper redundant data issue is taken into consideration and it is eliminated further to improve the performance of the Hybrid network. Redundant data cleaning is discussed in detail along with its effects on system. An algorithm is proposed to overcome redundant data problem which performs efficient redundant data filtration. Proposed algorithm is compared with previous available technique. Simulation results show that the proposed algorithm reduces data redundancy maintaining the integrity of original data.

KEYWORDS: Redundant data, Data filtering, Integration, Radio Frequency Identification (RFID), Wireless Sensor Network (WSN).

I. INTRODUCTION

Wireless Sensor Network (WSN) and Radio Frequency Identification (RFID) has received great attention due to their phenomenal advances in very large integrated systems, micro-electro-mechanical systems and highly integrated low power digital electronics. They are emerging as the most ubiquitous computing technology due their broad applicability and various noteworthy advantages [2]. The evolution of WSNs and RFID has followed separate development and research path which has led to distinct technologies.

WSN is a network consisting of a Sink Node also known as base station and a number of small, light weight and wireless nodes called sensor nodes. These sensor nodes sense the environmental conditions like temperature, humidity, pressure, light, sound, and vibration and accordingly collect the information [3]. The sensor nodes has computational capabilities allowing processing of collected information. This collaboratively collected and processed information is further transmitted to the base station. WSNs provide cost effective monitoring of critical applications including border monitoring, industrial control, military, environmental monitoring and healthcare application [4]. In contrast RFID technology enables detection and identification of an object. An RFID network consists of readers and tags. A tag consisting of a chip and an antenna is incorporated on a target object. A reader obtains information by scanning these tags and accordingly transmits information to the server. General applications of an RFID systems are supply chain management, highway toll collection, controlling building access, public transportation, developing smart home appliances, animal tracking etc [4].

RFID technology is been extensively adopted in industrial applications and on the other hand, sensor networks found important applications in harsh environment conditions. Nevertheless, there are many applications where information retrieved by sensing environmental conditions is not sufficient and additional information like identity or the location of an object is important [3]. Though sensor networks can be used in these environments as well, but the location and identity of an object still remain critical. In these cases, integration of both WSN and RFID technologies is the optimal solution as they complement each other. The integration of RFID and WSNs maximizes their effectiveness and give innovative perspectives to a broad range of applications [4].



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

The Hybrid network formed by integration of WSN and RFID networks has various identified challenges viz. energy conservation, data cleaning and filtering, real-time performance, anti-collision, localization, and authentication [5]. Among all above challenges, Data cleaning and filtering is important to circumvent inefficient utilization of network resources [6]. In an RFID network, a reader interrogates tags multiple times to increase the observed read rate due to which several copies are attached to single object. Also the sensor nodes and RFID tags are densely deployed in an area of interest in order to cover each and every section of area under observation. This results in data redundancy as some sections of area are covered by more than one device node. Redundant data is nothing but repeated readings or duplicate copies of reading. The filtration of this redundant data is essential as it does not provide any useful information but unnecessarily consumes valuable resources. The process of removing redundant information from gathered information is known as data cleaning. More precisely, redundant data cleaning is a process of modifying, replacing or deleting the inaccurate, incorrect or irrelevant part of data [7].

The remainder of this paper is as follows. Section II consists of related work. Section III provides a brief idea about the architecture used for evaluating the performance of proposed scheme. Section IV deals with a detailed description of proposed technique. Simulation result performance analysis is provided in Section V and finally Section VI concludes the paper.

II. RELATED WORK

Shawn R. Jeffery in [8] addressed various shortcomings of static smoothing filters to correct missed readings and proposed SMURF algorithm which employs adaptive smoothing scheme for RFID data filtration. It is the first declarative algorithm and act as adaptive smoothing filter for data cleaning. SMURF aims at two cleaning mechanism viz; Per tag cleaning and Multi tag cleaning. As SMURF is an adaptive smoothing scheme, there is no need to set the window size as it adjusts its size automatically considering the characteristic of data streams. In order to adapt window size automatically, the algorithm should be able to differentiate between tag movement and dropped reading. A statistical sampling-based approach is used by SMURF to achieve above goal and thus produces accurate data streams with appropriate balance between tag motion and missed reading.

Extensible receptor Stream Processing (ESP) proposed by Shawn R. Jeffery [9] is a declarative query based framework. It is designed to clean data streams produced in sensor network. ESP is a pipeline framework for data processing which cleans data of receptor data streams online. ESP organizes the receptor data stream into five stages: Point - Smooth - Merge - Arbitrate - Virtualize. It exploits spatial and temporal aspects of receptor data streams by processing multiple receptor streams and produces more enhanced and accurate data stream. ESP provides considerable development over raw sensor data and accurately reflects the physical world.

In [6] EIFS i.e. Energy Efficient In-Network RFID Data Filtering Scheme proposed by Dong-Hyun Lee, redundant data is divided into two types: inter-cluster and intra-cluster cases. The type of RFID data packet received by the cluster head is decided from its f field. The sender is an intra cluster node if the value of f is 1 and cluster head should carry out data filtering. After data cleaning the field f is set to 0. Thus packets with field f 0 will not be filtered again by any cluster head which reduces computation cost. After intra-cluster filtering, CHs send their data towards sink. EIFS algorithm first detects the inter-cluster redundant data and accordingly sends a feedback messages to intermediate nodes which results in reduction in unnecessary transmissions. If a redundancy is detected by a cluster head then through feedback it informs to its intermediate CHs. Then according to the feedback, updation of tag list is carried out.

Efficient In-Network Redundancy Filtering EIRF proposed in [10] by Ali Kashif Bashir, make use of hash tables to determine the redundancy in data streams. It also utilizes height-balanced tree to revise the expired chronological data which is faster than previous methods. EIRF has less network delay and processing latency. Barjesh Kochar have proposed a novel data cleaning technique along with data transformation and data loading techniques in [6]. The cleaning is been performed on the basis of the probability of response of the tags and the window size. The window size is enlarged if the strength of the dirty data is low and decreased if it is high in the past window of interrogation.

In Improved Cross Redundant Cleaning Algorithm (ICRDC) proposed by Li Wang in [5], a structure called cross tags list (CTL) which consists of tuple queue ensuring inserting and deleting operation needs to be maintained in the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

memory. Then a verification operation can be carried out to determine whether the tag of waiting arbitration is nearby the reference tags in CTL. If there are no such arbitration tags then Euclidean distance of waiting arbitration tag can be calculated and with minimum relative position method the cross redundant data can be eliminated. If there are some waiting arbitration then based on sliding window mechanism the tuple cache queue in tuple groups can be detected and check for redundancy. This CTL tuple containing redundant data is filtered and thus a reasonable memory can be maintained.

III. ARCHITECTURE

Figure1. shows the architecture of RFID and WSN integrated network. The architecture consists of four types of nodes [11]:

- Conventional sensor node (CSN)
- Cluster head sensor node (CH)
- Sensor tag (ST)
- Hybrid Sensor node (HSN)

Conventional sensor nodes are the most fundamental and simple component of integrated network. CSN are normal sensor nodes which sense physical conditions of the environment and accordingly transfer information to the cluster head (CH). The main role of CSN is to act as relays while providing additional information about the environment [11]. CH nodes are identical to CSN but perform certain additional functions like data aggregation and processing of information. Each sensor cluster contains one CH.

A sensor tag (ST) is the node formed by the integration of RFID tag and Sensor. Thus it has sole identity due to RFID tag and sensing capability due to Sensor incorporated in it. There three types of ST, viz. Active, Passive and Semi-active. A Passive sensor tags are used in above architecture which do not utilize battery power for sensing and communication purpose. Another key component is Hybrid Sensor Node (HSN) which is a combination of a sensor node and an RFID reader in one single node. The resultant device can perform:

- Sensing of environmental conditions.
- Wireless communication with each other
- Reading of identification numbers from tagged objects or persons
- Transmission of this information to the sink or next HSN node.

As the integrated network consists of four different types of node, successful communication between them is important. Entire network is divided into clusters. This clustering is achieved by Data Density Correlation Degree Clustering Method. Thus the formation of cluster takes place considering the density of data in particular region. The clusters of Sensor tags and wireless sensor nodes are formed, where a CH is elected. A cluster may or may not contain an HSN. CSN transfer data to its respective CH. The CH aggregates the entire received data along with its own sensed data and forward it to Base Station (BS). The CH can forward data to BS directly or via another CH whichever is more efficient. Sensor tags are located within HSN reading range. The integrated sensor tags form their own clusters and HSN acts as its CH. The role of HSN is similar to CH. The CH and HSN can communicate with each other and they can also act as a relay for each other to transfer information to BS in an efficient manner.

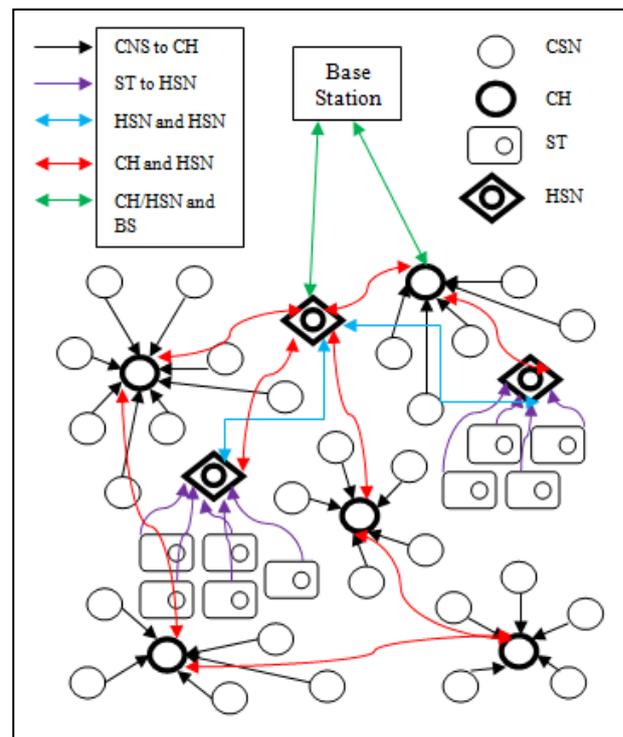


Figure 1. Architecture

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

IV. PROPOSED ALGORITHM

A. Assumption

There are:

1. m Hybrid sensor node (HSN)/cluster head (CH)
2. $m \times n$ fixed reference tags/Conventional Censor Nodes
3. l tags/CSNs for arbitration.

The attributes of the reference tags/CSNs are known and each HSN/CH can detect and locate reference tags/CSNs in its working area [5].

B. Definitions

Definition 1: $D_{i,j}$ is the Euclidean distance between the signal intensity of tag i /CSNs i waiting for arbitration on HSN j /CH j and the signal intensity of m reference tags/CSNs of HSN j /CH j . If the value of $D_{i,j}$ is small to HSN j /CH j , it more likely will belong to HSN/CH of the tag i / CSNs i :

$$D_{i,j} = \sqrt{\sum_{k=1}^m (W_{i,j} - G_{j,k})^2}, \quad i \in (1 \dots l), j \in (1 \dots m)$$

Where,

$G_{j,k}$ - Signal intensity of reference tag k /CSN k belonging to HSN j /CH j such that $j \in (1 \dots m), k \in (1, \dots n)$.

$W_{i,j}$ - Signal intensity of reference tag i /CSNs i waiting for arbitration on HSN j /CH j such that $i \in (1, \dots l), j \in (1, \dots m)$ [5].

Definition 2: A triple $d(R_{id}, T_{id}, \tau)$ is a set of original data. R_{id} is denoted as unique identification of the HSN/CH. T_{id} is denoted as unique identification of the tag/CSN. τ is timestamp which the tag/CSN is detected. Through a triple d , the mapping relationship of the RFID system between physical device and logic identity can be built [5].

Definition 3: Affiliation is possible as tag i /CSNs i respond to HSN j /CH j within the induction scope of HSN j /CH j at the moment t . Further, we denote it as follows:

$$F(i, j, t) = \frac{1}{3} \sum_{k=0}^2 p(i, j, t - k) \quad i \in (1 \dots l), j \in (1 \dots m)$$

Where,

$p(i, j, t)$ is the frequency that HSN j /CH j links to tag i /CSNs i at the moment t [5].

C. Conditions

Cross redundant data is produced by some nodes which are located in the same working area. Following conditions are satisfied by cross redundant data:

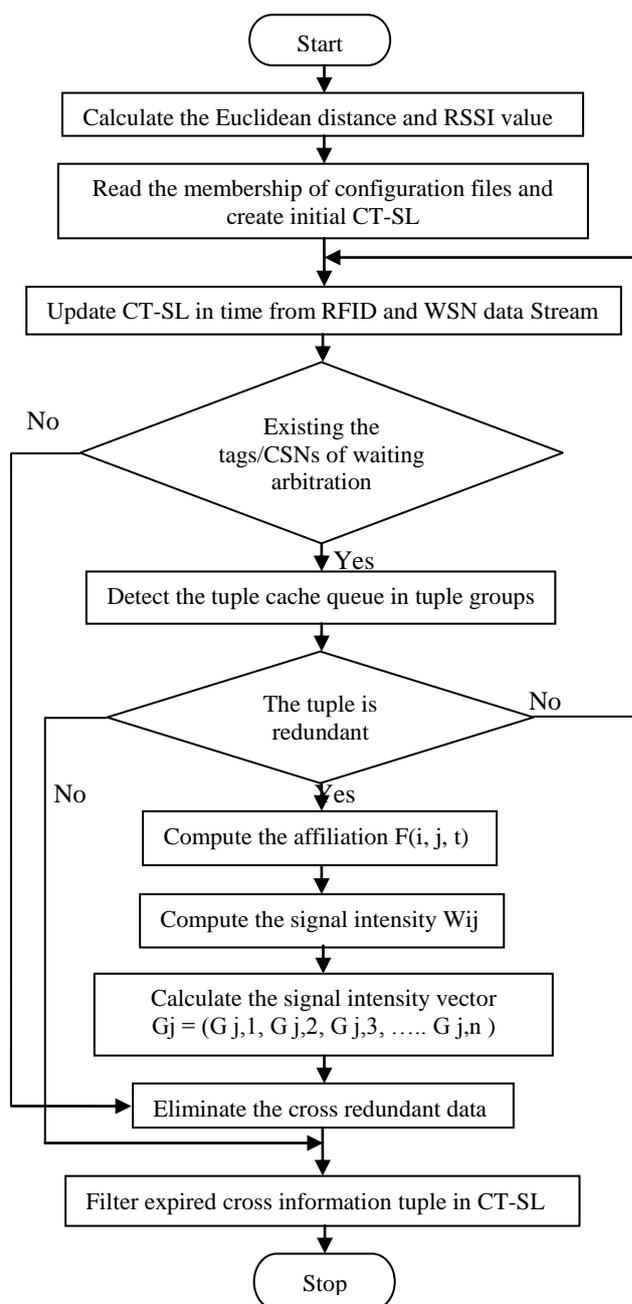


Figure 2. Flow Chart of Data Cleaning Algorithm



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

1. For two arbitrary triples $d_i (R_i d_i, T_i d_i, \tau_i)$ and $d_j (R_j d_j, T_j d_j, \tau_j)$ from the same cross space, the condition of $(R_i d_i \neq R_j d_j) \cap (T_i d_i = T_j d_j)$ is satisfied.
2. At the same time, the relation of $(\tau_i = \tau_j) \cap (|\tau_i - \tau_j| \leq \sigma)$ is also satisfied, where σ is the time threshold [5].

D. Description of the Proposed Algorithm

The redundancy in data causes unnecessary utilization of network resources. To avoid this situation following algorithm is employed. A structure called Cross Tag-Sensor List (CT-SL) is maintained incrementally in the memory. CT-SL consists of a tuple queue which ensures inserting and deleting operation. The time sliding window mechanism used can complete the arbitration using the units of time window on many orderly tuple simultaneously [5]. The algorithm works as follows:

1. Initially, Every HSN and CH calculates the Euclidean distance of each and every node present within its cluster. Also the RSSI value of every incoming data packet is calculated. Data packets having RSSI value above threshold value are forwarded for further process.
2. According to the Euclidean distance, Cluster head further creates an initial CT-SL. This is achieved by reading the membership of configuration files. The CT-SL contains all the data packets which are sorted out in above step considering RSSI value. Thus initial CT-SL is created taking Euclidean distance and RSSI value into consideration.
3. The primarily created CT-SL is updated from time to time in accordance with the data streams received from RFID tags and Conventional sensor nodes (CSNs) present in the network.
4. Further, it is tested whether the existing tags/CSNs of waiting arbitration are close to reference tags/CSNs in CT-SL. If the answer is NO, go to step 10) and if YES go to step 5).
5. Then based upon the sliding window mechanism, the tuple cache queue can be detected in tuple groups.
6. Further the algorithm checks whether the tuple is redundant or not using the specification standards in definition 2. If it does not satisfy the condition, return to step 3; otherwise, go to step 7).
7. The affiliation $F(i, j, t)$ is computed which indicates that tag i / CSNs i responds to HSN j / CH j at the moment t .
8. Further, the signal intensity W_{ij} is computed to see whether the tag i / CSNs i of waiting arbitration is relative to the HSN j / CH j that already detected tag i / CSNs i in sliding window.
9. Calculate the signal intensity vector $G_j = (G_{j,1}, G_{j,2}, \dots G_{j,n})$ which denotes whether the reference tags/CSNs belongs to HSN/CH or not.
10. The method of minimum relative position and $F(i, j, t)$ are applied to arbitrate the HSN/CH of cross redundant data. The cross redundant data can be eliminated through an exclusive process.
11. The expired cross information tuple in CT-SL is filtered to maintain a reasonable memory and then return to step 3).

V. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

Modified ICRDC algorithm is similar to ICRDC with minor modifications as follows:

1. First, the HSN and CH calculates the Euclidean distance of each and every node present within its cluster.
2. The RSSI value of every incoming data packet is calculated and only data packets having RSSI value above threshold value are forwarded for further process.
3. The initial CT-SL is formed in accordance to the Euclidean distance of node and it contains data packets having appropriate RSSI value.

With these modifications the Modified ICRDC scheme is implemented in Integrated Network and its simulation is carried out in Network Simulator-2 (NS-2). The integrated network is similar to the network developed by researchers in [11] which is described in section III. It consists of total 75 nodes among which some are sensor nodes and some are integrated nodes (refer section III). The performance of the Modified ICRDC algorithm is tested in terms of Compressibility and Accuracy.

Figure 3 shows the Compressibility graph plotted against simulation time. The compressibility parameter here denotes the amount of compression achieved in gathered information due to elimination of redundant data. In graph, red line represents the behaviour of proposed scheme and green line indicates ICRDC behaviour in Integrated Network. The generated graph shows that initially, the performance of ICRDC is better than the proposed technique but as the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

time progresses modified ICRDC continuously overcome ICRDC. This shows that though the ICRDC performs well in the primary stage, the overall performance of Modified ICRDC is superior.



Figure 3. Comparison of Compressibility between ICRDC and Modified ICRDC

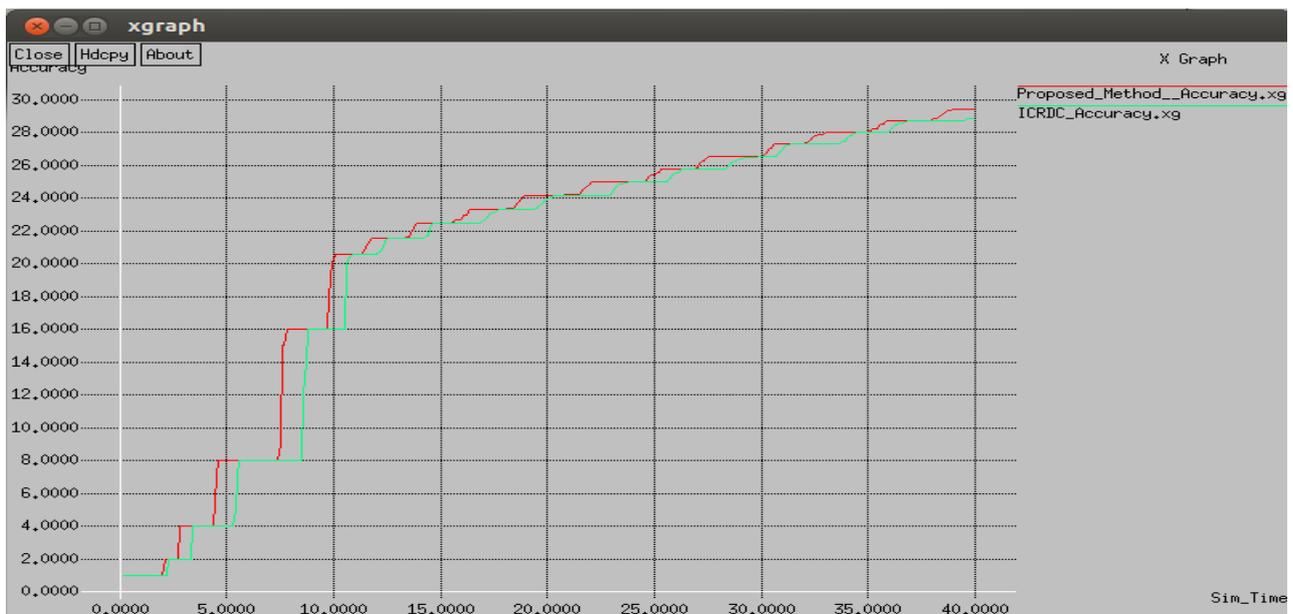


Figure 4. Comparison of Accuracy between ICRDC and Modified ICRDC

Figure 4 shows the comparison between ICRDC and proposed technique in terms of accuracy. Redundant data cleaning is a critical task in a sense since data integrity must be maintained simultaneously with redundancy elimination. For this purpose, the accuracy of redundant data cleaning algorithm is important. The accuracy parameter



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

over here denotes how accurately the data filtering algorithm performs redundant data elimination. In graph, red line indicates the behaviour of proposed scheme and green line shows ICRDC behaviour in Integrated Network. A little improvement is achieved in terms of accuracy as comparison to ICRDC.

VI. CONCLUSION

In this paper, the issue of redundant data in WSN and RFID integrated network is taken into consideration. Redundant Data transmission causes waste of time, energy and other network resources. To overcome this problem a Modified ICRDC technique is proposed which is build upon ICRDC algorithm. Also a brief information is provided about the architecture where the proposed scheme is implemented. Simulation result obtained validates the efficiency of the proposed algorithm. The proposed scheme successfully performs redundant data filtration making integrated network a resource efficient network.

REFERENCES

- [1] Ashwini W. Nagpurkar, Siddhant K. Jaiswal, "An Overview of WSN and RFID Network Integration", *IEEE Sponsored Second International Conference On Electronics And Communication Systems(Icecs '2015)*.
- [2] J. Cho, Y. Shim, T. Kwon, and Y. Choi, "SARIF: A novel framework for integrating wireless sensor and RFID networks", *IEEE Wireless Communications*, 14(6), 50–56, December 2007.
- [3] Fadi M. Al-Turjman, Ashraf E. Al-Fagih1 and Hossam S. Hassanein, "A Novel Cost-Effective Architecture and Deployment Strategy for Integrated RFID and WSN Systems", *International Conference on Computing, Networking and Communications, Network Architecture & P2P Protocol Symposium*, 2012.
- [4] Hai Liu, Miodrag Bolic, Amiya Nayak and Ivan Stojmenovic, "Taxonomy and Challenges of the Integration of RFID and Wireless Sensor Networks", *Network, IEEE*, Vol.22, No.6, pp.26-35, November-December 2008.
- [5] Li Wang, Li Da Xu, "Data Cleaning for RFID and WSN Integration", *IEEE Transactions On Industrial Informatics*, Vol. 10, No. 1, February 2014
- [6] Ali Kashif Bashir, Se-Jung Lim, Chauhdary Sajjad Hussain and Myong-Soon Park, "Energy Efficient In-network RFID Data Filtering Scheme in Wireless Sensor Networks", *Sensors 2011*, 11, 7004-7021; doi:10.3390/s110707004.
- [7] Barjesh Kochar and Rajender Chhillar, "An Effective Data Warehousing System for RFID Using Novel Data Cleaning, Data Transformation and Loading Techniques", *The International Arab Journal of Information Technology*, Vol. 9, No. 3, May 2012.
- [8] Shawn R. Jeffery, Minos Garofalakis and Michael J. Franklin, "Adaptive Cleaning for RFID Data Streams", in *Proc VLDB*, Seoul, Korea, 2006, pp. 163–174.
- [9] S. R. Jeffrey, G. Alonso, M. Franklin, W. Hong, and J. Widom, "A pipelined framework for on line cleaning of sensor data streams," in *Proc. IEEE Comput. Soc. Atlanta*, 2006, pp. 140–142.
- [10] Dong-Hyun Lee, Eun-Mook Lee, Ali Kashif Bashir and Myong-Soon Park, "Efficient In-Network Redundancy Filtering in RFID System Integrated with Wireless Sensor Networks", *6th International Conference on Networked Computing (INC)*, 2010.
- [11] D. M. Boneke Nokonoko, G. A. Lusilao-Zodi, A. B. Bagula, M. E. Dlodlo, "An Energy-Efficient Routing Protocol for Hybrid-RFID Sensor Network", *IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia*, 13 - 15 September 2011.
- [12] Meng Ma, Ping Wang, Chao-Hsien Chu, "Data Management for Internet of Things: Challenges, Approaches and Opportunities", *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*.
- [13] Ferrer-Vidal, A. Rida, S. Basat, L. Yang, and M.M. Tenzeris, "Integration of sensors and RFID's on ultra-low-cost paper-based substrates for wireless sensor networks applications", in *Proceedings of the 2nd IEEE Workshop on Wireless Mesh Networks (Wi Mesh 2006)*, Reston, VA, September 25, pp. 126–128, 2006.
- [14] JN. Cho et al., "A 5.1-uW UHF RFID Tag Chip Integrated with Sensors for Wireless Environmental Monitoring", *European Solid-State Circuits Conf. (ESSCIRC)*, Grenoble, France, Sept. 2005.
- [15] A.G. Ruzzelli, R. Jurdak, and G.M.P. O'Hare, "On the RFID wake-up impulse for multi-hop sensor networks", In *Proceedings of 1st ACM Workshop on Convergence of RFID and Wireless Sensor Networks and their Applications (Sense ID) at the Fifth ACM Conference on Embedded Networked Sensor Systems (ACM SenSys 2007)*, Sydney, Australia, November, 2007.
- [16] L. Zhang and Z. Wang, "Integration of RFID into wireless sensor networks: Architectures, Opportunities", In *Proceedings of the Fifth International Conference on Grid and Cooperative Computing Workshops (GCCW'06)*, Changsha, China, pp. 463–469 October 2006.
- [17] G. Yang, M. Xiao, and C. Chen, "A simple energy-balancing method in RFID sensor networks", In *Proceedings of 2007 IEEE International Workshop on Anti-Counterfeiting, Security, Identification*, Xiamen China, pp. 306–310, April 16–18, 2007.
- [18] Myo Min Aung, Yoon Seok Chang and Jong-Un Won, "Emerging RFID/USN Applications and Challenges", In *International Journal of RFID Security and Cryptography (IJRFIDSC)*, Volume 1, Issues 1-4, Mar-Dec 2012.



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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

BIOGRAPHY

Ashwini W. Nagpurkar is pursuing Masters of Engineering in Wireless Communication and Computing (WCC) Department of Computer Science and Engineering (CSE) from G.H. Raisoni College of Engineering, Nagpur, MS, India. She received Bachelor of Engineering (BE) degree in Electronics and Telecommunication Department from Bhausaheb Mulik College of Engineering, Nagpur, MS, India. Her research interests include Wireless Sensor Network, Radio Frequency Identification etc.

Sidhant K. Jaiswal received his Bachelor of Engineering (BE) degree in Computer Engineering from S.R.P.C.E., Nagpur, MS, India. He received his Master of Technology degree in Computer Science and Engineering (CSE) from SRCOEM, Nagpur, MS, India. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering. He has totalled 3 National and International papers published. His research interest includes Networking.