Vol. 35, No. 02, pp. 1078-1090/Diciembre 2022

ISSN-E 1995-9516
Universidad Nacional de Ingeniería

http://revistas.uni.edu.ni/index.php/Nexo
https://doi.org/10.5377/nexo.v35i04.15550

A Novel Algorithm for Redundant Data Filtering in WSN and RFID Integrated Networks

Un algoritmo novedoso para el filtrado de datos redundantes en redes integradas WSN y RFID

Mahia Samani[1], Ahmad Khademzadeh[2*] , Kambiz Badie[2]

[1] Department of Computer Engineering, Qom Branch, Islamic Azad University, Qom, Iran.
[2] Iran Telecom. Research center, Iran.
[*]Corresponding author E-mail: Zadeh@itrc.ac.ir

(recibido/received: 18-septiembre-2022; aceptado/accepted: 15-noviembre-2022)

ABSTRACT

Wireless sensor networks (WSN) and radio frequency identification (RFID) are base technologies employed in decentralised dynamic environments. In the hybrid network formed by integrating RFID and WSN, RFID data can be used applying WSN protocols for multi-hop communications. However, RFID data contain excessive duplication, which increases time delay and energy consumption, resulting in wastage of different resources. There exist four popular RFID–WSN integration architectures: hierarchical RFID-sensor topology, network RFID-sensor topology, reader-sensor nodes topology, and mixed topology. In this paper, we propose a new plan for a WSN–RFID integrated network. The entire network is divided into clusters, and the clustering hierarchical routing algorithm is employed to send data from the head nodes to the base station. Further, we propose two algorithms to overcome redundant data on the hybrid network. Our simulation results demonstrate that the proposed method reduces redundant data and processing time compared with existing methods.

Keywords: redundant data; data filtering; integration; radio frequency identification; wireless sensor network; algorithm

RESUMEN

Las redes de sensores inalámbricos (WSN) y la identificación por radiofrecuencia (RFID) son tecnologías básicas empleadas en entornos dinámicos descentralizados. En la red híbrida formada por la integración de RFID y WSN, los datos RFID se pueden utilizar aplicando protocolos WSN para comunicaciones multisalto. Sin embargo, los datos RFID contienen una duplicación excesiva, lo que aumenta el tiempo de demora y el consumo de energía, lo que resulta en el desperdicio de diferentes recursos. Existen cuatro arquitecturas populares de integración RFID-WSN: topología de sensor RFID jerárquica, topología de sensor RFID de red, topología de nodos de lector-sensor y topología mixta. En este documento, proponemos un nuevo plan para una red integrada WSN-RFID. Toda la red se divide en grupos, y el algoritmo de enrutamiento jerárquico de agrupación se emplea para enviar datos desde los nodos principales a la estación base. Además, proponemos dos algoritmos para superar los datos redundantes en la red híbrida. Nuestros resultados de simulación demuestran que el método propuesto reduce los datos redundantes y el tiempo de procesamiento en comparación con los métodos existentes.

Palabras clave: datos redundantes; filtrado de datos; integración; identificación de frecuencia de radio; red de sensores inalámbricos; algoritmo

1. INTRODUCTION

Wireless sensor networks (WSNs) comprise a sink node (or base station) and several small, light, wireless, and cost-effective devices called sensor nodes (Yarinezhad, 2019; Yetgin et al., 2017). Sensor nodes are capable of data processing and sensing; they sense environmental conditions (such as temperature, pressure, light, sound, and vibrations) and capture data based on the parameters (Randhawa & Jain, 2019). Such captured and processed data are delivered to base stations through head nodes. These sensor networks are used in numerous applications, such as environment monitoring, particularly border control, industrial control, army, and healthcare (Das & Harrop, 2014; Yuan et al., 2014). Nevertheless, such sensor networks are not capable of identifying objects around them. On the other hand, networks that integrate WSN and radio frequency identification (RFID) facilitate the simultaneous sensing of environmental conditions and identification of objects (Nagpurkar & Jaiswal, 2015). Such networks are employed by a number of facilities such as chain store management (Alfian et al., 2017) and healthcare (Zhao, 2014).

RFID technology implementation comprises a data reader, tag, and software. The data reader reads the tags attached to objects and saves data in its memory (Jin et al., 2015). RFID does not support a multi-hop from one data reader to another. However, the integration of RFID and WSN can enable the transfer of data from one data reader to another by the sensor network protocol, so that it can finally be delivered to the base station. (Piramuthu & Zhou, 2016; Zhang & Qi, 2014).

The remainder of the paper is organized as follows. Section 2 explains the research methodology. Section3 contains the related works while the architecture of the hybrid network is explained in section 4. The proposed algorithm is described in section 5. Simulation result is presented in section 6 and finally, section 7 concludes the paper.

2. METHODOLOGY

The steps of the research methodology used in this work are as follows:

- Review and synthesis of existing knowledge on WSN and RFID network integration.
- Investigation of existing scenarios and problems
- Formulation of a hypothesis that solves existing problems
- Hypothesis testing
- Evaluation and conclusion

2-1) Review and synthesis of existing knowledge on WSN and RFID network integration

Four popular RFID–WSN integration architectures exist:

- Hierarchical RFID-sensor topology,
- Network RFID-sensor topology,
- RFID reader-sensor nodes topology
- Mixed topology

<u>Hierarchical RFID-sensor topology</u>: Adding a sensor to an RFID tag is an important way to integrate the two technologies (Figure 1). Sensor tags with limited communication capacity communicate with the reader. Each tag is assigned to an object, person or animal with a unique ID, while the sensors on the tag sense the information needed (Sundaresan et al., 2014). According to the tag's power, this integration can be divided into three groups: passive tag with a sensor, semi-passive tag with a sensor and active tag with a sensor (Sundaresan et al., 2014, 2015, Wang et al., 2013).

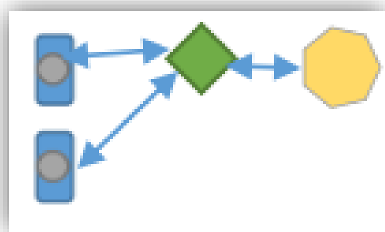


Figure 1. Hierarchical RFID-sensor topology

Network RFID-sensor topology: RFID tags can be added to the sensor nodes (Figure 2). Consequently, these merged sensor tags can communicate with each other as well as with other wireless devices. Therefore, these integrated sensor tags have increased their communication capacity and can form a collaborative ad hoc network.

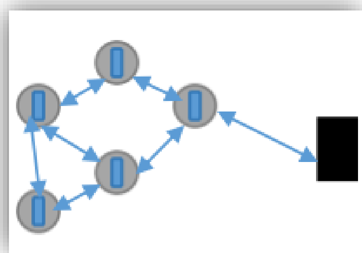


Figure 2. Network RFID-sensor topology

RFID reader-sensor node topology: Another possible strategy for integrating RFID and sensor networks is to add a sensor to the RFID reader (Figure 3). These nodes can be considered RFID data readers with enhanced sensor capacity and the ability to transmit information (Aung et al., 2015). They can also act as a router and deliver messages to the destination. These nodes are responsible for collecting information from simple RFID tags within their control range. They also communicate with each other to transmit this information to the sink (Sundaresan et al., 2015, Wang et al., 2014).

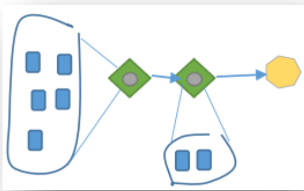


Figure 3. RFID reader-sensor nodes topology

Mixed topology: In mixed architecture, RFID tags and sensor nodes are separate devices, but they physically exist in an integrated network and work independently (Zhang et al., 2009).

2-2) Investigation of existing scenarios and problems

RFID and WSN integrated networks have their own challenges, such as energy consumption, data filtering, data cleaning, real-time performance, and authentication, among which, data filtering and cleaning are more important. Redundant data cleaning is a type of processing that modifies, replaces, or deletes incorrect or irrelevant data (Kochar & Chhillar, 2012; Kapoor & Piramuthu, 2010). Redundant data may generally be classified into three types (Bashir et al., 2011; Sarma, 2001):

1) A tag may be read by the data reader several times without any change in its data (reading tags in close time intervals).

2) Some tags may be placed within the range of several readers, and the data in the overlapped area are read by several readers(figure 4: The data of two tags-sensor(st1,st2)are read by two data reader(SR2,SR1))

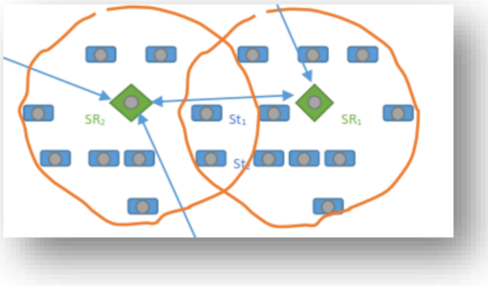


Figure 4. Redundant Readers

3) Data level duplication occurs when in order to increase reliability and decrease the missing data rate, a few tags with one EPC may be attached to the same object (Figure 5) (Bashir et al., 2011; Sarma, 2001)

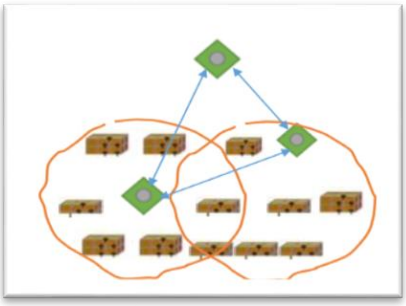


Figure 5. Data level duplication

2-3) Formulation of a hypothesis that solves existing problems

Many theories have been proposed to solve the data redundancy problem in RFID-based systems. This paper discusses INPFM, CLIF and EIFS and solutions for the data redundancy problem. Two algorithms are presented and explained in detail in Section 5.

2-4) Hypothesis testing

In this work, simulation is used for hypothesis testing. This simulation is implemented and performed using the C++ language. A more detailed explanation of hypothesis testing is presented in section 6.

2-5) Evaluation and conclusion

The evaluation and conclusion are given in sections 6 and 7, respectively.

3. RELATED WORK

Redundant data filtering is one of the key challenges in RFID applications. Most organisations want access to only one copy of data (i.e., data without redundancy). Many studies have been conducted with the aim of filtering redundant data in servers. The transfer of redundant data from nodes to the server may affect the energy of the nodes, resulting in network overhead and shortening network life. In order to reduce such unnecessary transmissions, redundant data must be filtered within the network. INPFM[1] (Choi & Park, 2007)

[1] In-network phased filtering mechanism

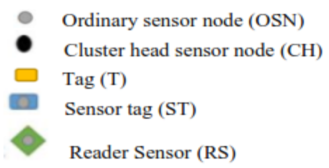
and CLIF[2] (Kim et al., 2008) filter redundant data inside the network, but these approaches involve high calculation expenses and do not remove all overheads. The INPFM algorithm employs the tree structure and filters redundant data in each K-hop reader. The problem is that if INPFM checks redundant data in each hop, calculation expenses increase; on the other hand, if such data are checked after a few steps, some redundant data will enter the network and increase network load. CLIF applies clustering topology and includes two phases: (a) identification of redundant data, and (b) filtering of redundant data. This scheme classifies redundant data into two classes: 1) redundant data within a cluster, and 2) redundant data between clusters. If a node receiving the data is the cluster head, it checks whether the received data belongs to that cluster or the adjacent cluster and filters the data; otherwise, the data will be transferred to the next cluster. The problem with CLIF is that only the redundant data of each cluster and its adjacent cluster will be filtered, while some redundant data may remain unfiltered in overlap zones and may be transferred to the sink. This problem may be more prominent when there is more redundant data in the overlap zones. The EIFS[3] algorithm uses a clustering structure (Bashir et al., 2011). The problem in CLIF can be eliminated by transferring the redundant data to the sink without filtering, and if this process is repeated several times, the data delivery route will change (for this purpose, a threshold is taken into consideration). However, in some cases, such a route change may result in the extension of the data delivery route to the sink. A summary of the limitations of all three above-mentioned approaches is given in Table 1.

TABLE 1. Weak points of evaluated approaches

APPROACHES	WEAKNESSES
INPFM (Choi & Park, 2007)	• some duplicate data might not meet at a filtering point • computation cost is high
CLIF (Kim et al., 2008)	• computation cost relatively high and inducing delays • computation Overhead • inter-cluster duplication may be or not filtered at an inter mediate CH
EIFS (Bashir et al., 2011)	• inducing delays • amount overhead • sometimes changing of route causes that route becomes longer

4. ARCHITECTURE

Figure 6 reveals the architecture of hybrid network from WSN and RFID in proposed plan that includes the five types of nodes as follows:



[2] cluster-Based in network phase Filtering scheme
[3] Energy Efficient RFID data filtering scheme

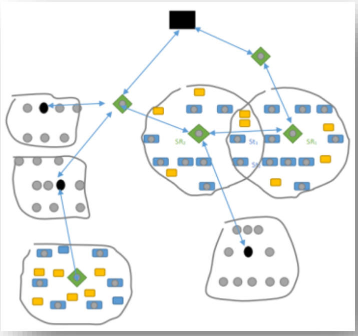


Figure 6. the architecture of hybrid network from WSN and RFID in proposed plan.

Ordinary Sensor node is usual sensor node that sense environmental conditions and relay data to the CH. The cluster head node is a type of OSN, moreover it can perform additional functions, for example: integrating and processing data. Each cluster has only one CH.

In this architecture, a ST node is composed of a RFID tag and a sensor, and is capable of sensing the environmental conditions in addition to identification of the objects.

There are three type of ST: Active, passive and semi-active. In this architecture, passive ST do not use batteries, for Sensing or Communication. Another component, Reader sensor (RS) is composition of an RFID reader and a sensor node which acts as a node. The RS nodes can perform following functions (Piramuthu & Zhou, 2016):

- Sensing of environmental conditions.

RS can be considered an RFID data reader with sensing capabilities thus, it can senses environmental conditions (such as temperature, pressure, sound, and …)

- Reading ID number from tag.

These nodes are responsible for gathering information from simple RFID tags within their range. They also communicate with each other to transmit this information to the base station where the data is collected and processed by a human.

- Wireless connection with each other and creating ad_hoc communication network.
- They are able to transfer of data to another SR node or sink node as a router

In this architecture, RSs act as head node. RSs and CHs may communicate with each other may act as relays and may transfer data from one CH or RS to another, or to base station.

5. PROPOSED ALGORITHM

Entire network is divided into clusters. Clustering hierarchical routing algorithm (CHRA) (Samani, 2011) is used for sending data from head nodes to the base station.

When an RFID data reader reads a tag, the following information is read by RFID data reader from a tag:

- ➢ EPC-Code
- ➢ Time Stamp

When the reader reads two item of data (for example: A, B), duplication happens If the Following conditions are satisfied:

1) EPC -code A == EPC -code B.

2) $|T_A - T_B| < T$ (Constant T is a threshold for difference between T_A and T_B.)

The duplication data can be divided into the following two types:

1) *The redundant data inside the cluster*: All the nodes of a cluster transfer data to their own cluster head, and after eliminating redundant data, the cluster head filters information along the routing path towards the base station using CHRA routing algorithm.

2) *The redundant data between two clusters*: Proposed algorithm filters these duplication data at every k hop reader.

Two algorithms are presented. These algorithms are used for detecting duplicate of data and data filtering. As before said, in this architecture there are five type of nodes (RS, ST, T, CH, OSN) that is used in these algorithms. Pseudocodes of algorithms are explained in the following. (Algorithm 1 and Algorithm 2).

```
Filtering decision (params : in coming data )
Begin
 Updateflg =0;  k=0;
 Loop for incoming data
       If  I am a OSN        then  send sensored data to CH
       If  I am a ST or T    then  send data to RS.
       If  I am a CH         then  send data through the route.
       If  I am a RS
           Then  CL-id ← cluster id of incoming data
                 If  Cl-id == my clustered
                    Then   call   INTERNAL FILTERING STEP.
                    ELSE   if   k <  3
                           Then  {send data through the route; exit}
                            ELSE  if  k ==  3  then  Do  Updateflg
=1;
                                                                     call
EXTERNAL FILTERING STEP;  End.
                           End– if
                 End – if
       End – if
   If   Updateflg==1       then k=0;
                           Else k++;
 End loop
End Begin
```

Algorithm 1. Filtering decision

```
 Tag {Epc ,
      t .  / * time – stamp * /
      Sensored – data }
EXTERNAL – FILTERING ( params : Tag )
{       Pos = hash ( tag.epc.)
 If arry1( pos ).epc = = tag.epc
   Then If  | arry1 [pos].t  –  tag.t |   <   T
            Then drop it / * dublicate data * /
             Else update tag to arry1 .
    Else if arry2[ pos] . epc = = tag . epc
            Then if  | arry2[ pos].t – tag.t  |   <   T
                   Then drop it .
                   Else update tag to arry2 .
            Else if  arry1 = = 0
                 Then arry1[ pos] = tag
                 Else if  arry 2 = = 0
                        Then arry2[ pos] = tag
                         Else Wrong Position.
}
```

Algorithm 2. External-Filtering step

Algorithm2 is explained using the read data shown in table 2. Figure 7 clarifies the sequences of read data mappings. In this example, duplication data happens when it is read only for one Tag ID and the time difference is under 3 seconds.

Table 2. a sample of read data

time	Tag-EPC
0	9
1	15
3	17
4	17
7	15
10	23
14	27

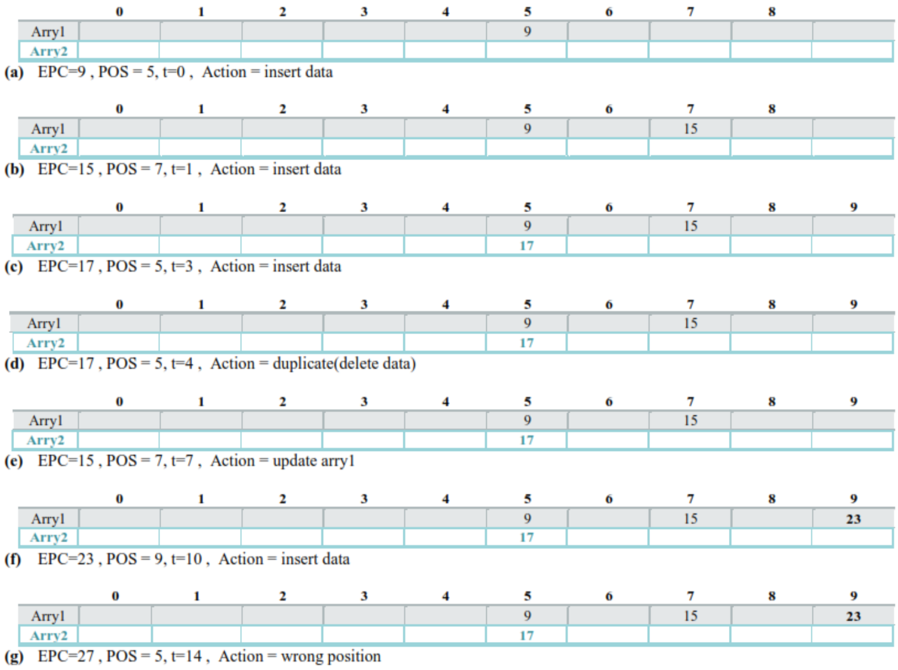


Figure 7. the states of arrays before and after insertion of data.

6. SIMULATION RESULTS

In this section, simulation is used to analysis the performance of the proposed algorithms. This simulation is implemented and performed under C++ language, and as we mentioned before, the nodes are classified the shape of cluster. The detailed simulation environment is given in table 3.

Table 3. simulation environment

parameters	values
Field area	100 × 100 m
Number of cluster	35
Member of cluster	10 - 25
Number of node	800
Reading range	5m
Transmission range	10m
Reading interval	3 s
Duplication ratio	10% and 35%

6-1) Wrong position rate (WPR)

In this paper, WPR is unsuccessful rate of elements in the array for each certain number of readings. Following is equation for WPR, where 'N' is number of reading and 'WP' is number of incorrect positions.

WPR = $\left[\frac{WP}{N}\right]\%$

In order to find the best size of array for a certain number of "N" reading, some tests have been conducted.

Figure 8 illustrate the WPR for different number of array size i = 10000, i = 15000, i = 20000. The number of readings is changed from 1000 to 10000 with increment of 1000 for each step. For all 'I' values, WPR hit the lowest point when the number of reading is 1000. The results show to get lowest WPR, the array size "I" must be five time as much as the number of reading.

6-2) Comparative analysis of filtering performance

In this experiment, the array size is 50,000, which is five times the maximum read (10,000). Consequently, the WPR value is very low and almost zero. Because of the very small amount of WPR, almost all duplicate data are filtered after every three steps.

The filtering performance of four approaches are compared in Figure 9 (with 10% and 35% duplication). As can be seen, the proposed algorithm exhibits better filtering performance than the other approaches. This is especially evident when the number of readings increases.

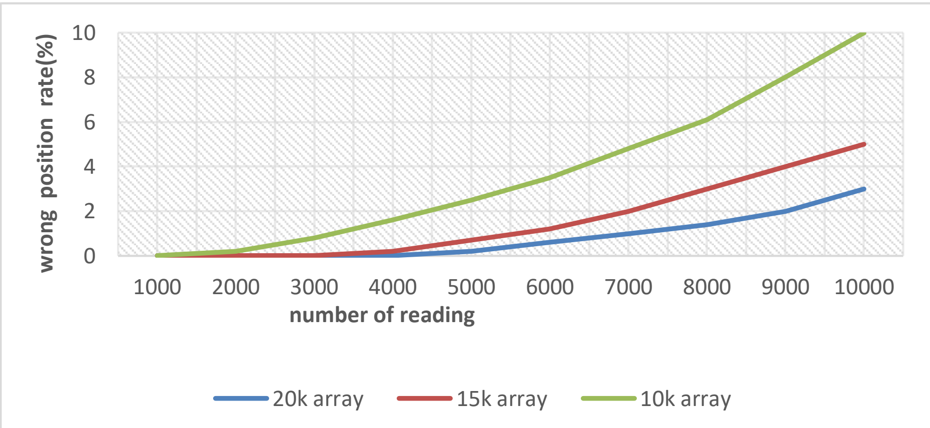


Figure 8. Wrong Position Rate. (WPR)

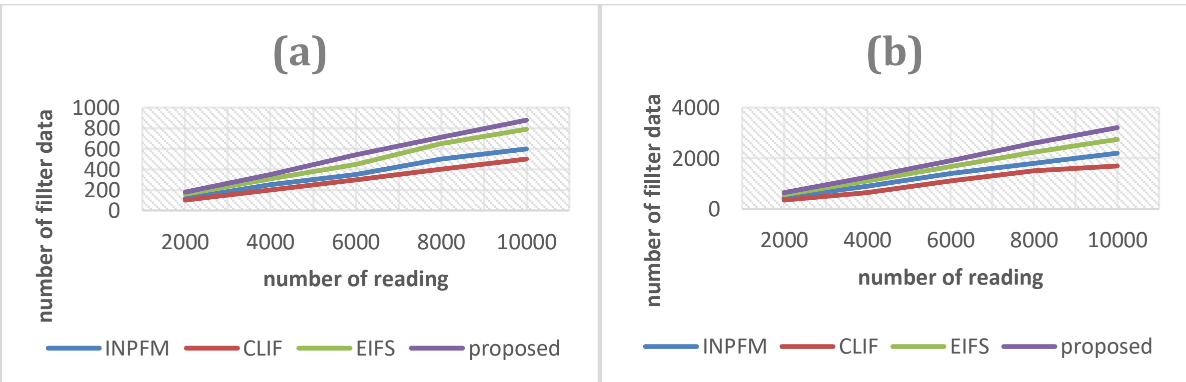


Figure 9. Amount of data filtered (a) 10% replication. (b) 35% replication.

6-3) Comparative analysis of processing time

In the proposed method, the external-filtering function is the only function that is repeated after every three steps. This function takes only O(1) time. Therefore, the processing time of the proposed method will be very short.

Figure 10 compares the processing times of the schemes for filtering duplicate readings. The number of readings is increased from 20,000 to 100,000. In this experiment, the array size is set to the maximum number of readings as in the previous experiment.

According to this graph, INPFM, CLIF and EIFS consumed more time for filtering than the proposed scheme.

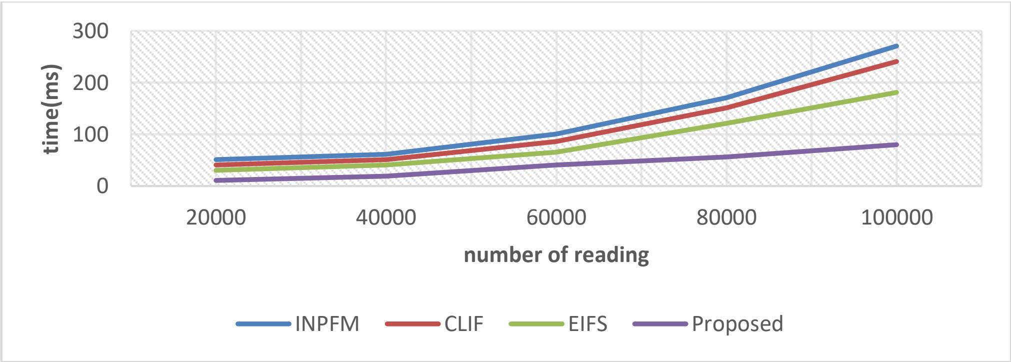


Figure 10. Comparison of processing time for filtering duplicate readings.

7. CONCLUSION

In this study, WSN and RFID were integrated into a hybrid system that employed the advantages of both technologies. WSN and RFID can be integrated into different forms. In this paper, first, the challenges in the RFID–WSN integrated network are presented. Three schemes of INPFM, CLIF and EIFS are investigated, and the advantages and disadvantages of each are expressed. Furthermore, a new architecture for the RFID–WSN integrated network is proposed. Subsequently, the problem of redundant data in the proposed technology is considered. We propose two algorithms to solve this problem.

In this experiment, the array size was selected five times as much as the number of reading, therefore wrong position rate value is very low and almost zero. As a result, almost all duplicate data are filtered. Moreover, these algorithms have a reduced runtime since they take only O(1) time.

Finally, our simulation results validate the efficiency of the proposed algorithms .In the clustering method, owing to traffic congestion, head nodes near the base station lose their energy early (hot zone problem). In the future, the proposed scheme can be improved by providing a way to reduce the hot zone problem.

REFERENCES

Alfian, G., Rhee, J., Ahn, H., Lee, J., Farooq, U., Ijaz, M. F., & Syaekhoni, M. A. (2017). Integration of RFID, wireless sensor networks, and data mining in an e-pedigree food traceability system. *Journal of Food Engineering*, *212*, 65-75.

Aung, M. M., Chang, Y. S., & Won, J. U. (2012). Emerging RFID/USN applications and challenges. *International Journal of RFID Security and Cryptography (IJRFIDSC)*, *1*(1–4).

Bashir, A. K., Lim, S. J., Hussain, C. S., & Park, M. S. (2011). Energy efficient in-network RFID data filtering scheme in wireless sensor networks. *Sensors*, *11*(7), 7004-7021.

Choi, W., & Park, M. S. (2007). In-network phased filtering mechanism for a large-scale RFID inventory application. In *4th International Conference on Information Technology and Applications, ICITA 2007* (pp. 401-405). Charles Sturt University.

Das, R. & Harrop, P., (2014). Printed, Organic & Flexible Electronics Forecasts, Players & Opportunities 2014-2024: Market data; market and technology appraisal; case studies: the complete picture for printed, flexible and organic electronics. IDTechEx. Available from: http://www.idtechex.com/research/reports/printed-organic-and-flexible-electronics-forecasts-players-and-opportunities-2014-2024-000404.asp.

Jin, C., Xu, C., Zhang, X., & Zhao, J. (2015). A secure RFID mutual authentication protocol for healthcare environments using elliptic curve cryptography. *Journal of medical systems*, *39*(3), 1-8.

Kapoor, G., & Piramuthu, S. (2010). Single RFID tag ownership transfer protocols. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(2), 164-173.

Kim, D. S., Kashif, A., Ming, X., Kim, J. H., & Park, M. S. (2008, June). Energy efficient in-network phase RFID data filtering scheme. In *International Conference on Ubiquitous Intelligence and Computing* (pp. 311-322). Springer, Berlin, Heidelberg.

Kochar, B., & Chhillar, R. S. (2012). An effective data warehousing system for RFID using novel data cleaning, data transformation and loading techniques. *Int. Arab J. Inf. Technol.*, *9*(3), 208-216.

Nagpurkar, A. W., & Jaiswal, S. K. (2015, February). An overview of WSN and RFID network integration. In *2015 2nd International Conference on Electronics and Communication Systems (ICECS)* (pp. 497-502). IEEE.

Piramuthu, S., & Zhou, W. (2016). *RFID and sensor network automation in the food industry: Ensuring quality and safety through supply chain visibility*. John Wiley & Sons.

Randhawa, S., & Jain, S. (2019). MLBC: Multi-objective load balancing clustering technique in wireless sensor networks. *Applied Soft Computing*, *74*, 66-89.

Samani, M. (2011, July). A new routing algorithm to reduce consumed energy in wireless sensor. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks* (pp. 95-98). IEEE.

Sarma, S. E. (2001). *Towards the five-cent tag*. Technical Report MIT-AUTOID-WH-006, MIT Auto ID Center, 2001. Available from http://www. autoidcenter. org.

Sundaresan, S., Doss, R., Piramuthu, S., & Zhou, W. (2014). A robust grouping proof protocol for RFID EPC C1G2 tags. *IEEE Transactions on Information Forensics and Security*, *9*(6), 961-975.

Sundaresan, S., Doss, R., Zhou, W., & Piramuthu, S. (2015). Secure ownership transfer for multi-tag multi-owner passive RFID environment with individual-owner-privacy. *Computer Communications*, *55*, 112-124.

Wang, L., Da Xu, L., Bi, Z., & Xu, Y. (2013). Data cleaning for RFID and WSN integration. *IEEE transactions on industrial informatics*, *10*(1), 408-418.

Wang, X., Ji, Y., & Zhao, B. (2014, September). An approximate duplicate-elimination in RFID data streams based on d-left time bloom filter. In *Asia-Pacific Web Conference* (pp. 413-424). Springer, Cham.

Yarinezhad, R. (2019). Reducing delay and prolonging the lifetime of wireless sensor network using efficient routing protocol based on mobile sink and virtual infrastructure. *Ad Hoc Networks*, *84*, 42-55.

Yetgin, H., Cheung, K. T. K., El-Hajjar, M., & Hanzo, L. H. (2017). A survey of network lifetime maximization techniques in wireless sensor networks. *IEEE Communications Surveys & Tutorials*, *19*(2), 828-854.

Yuan, F., Zhan, Y., & Wang, Y. (2013). Data density correlation degree clustering method for data aggregation in WSN. *IEEE Sensors Journal*, *14*(4), 1089-1098.

Zhao, Z. (2014). A secure RFID authentication protocol for healthcare environments using elliptic curve cryptosystem. *Journal of medical systems*, *38*(5), 1-7.

Zhang, Z., & Qi, Q. (2014). An efficient RFID authentication protocol to enhance patient medication safety using elliptic curve cryptography. *Journal of medical systems*, *38*(5), 1-7.

Zhang, Y., Yang, L. T., & Chen, J. (Eds.). (2009). *RFID and sensor networks: architectures, protocols, security, and integrations*. CRC Press.