ISSN: 2229-6948(ONLINE) ICTACT JOURNAL ON COMMUNICATION TECHNOLOGY, DECEMBER 2020, VOLUME: 11, ISSUE: 04
DOI: 10.21917/ijct.2020.0342

A HETEROGAMOUS HIERARCHICAL RELATIVE DISTANCE CLUSTERING BASED CONGESTION CONTROL IN WIRELESS SENSOR NETWORKS

Meena S. Doibale[1] and G.D. Kurundkar[2]
[1]*Department of Computer Science, Bharatiya Mahavidyalaya, India*
[2]*Department of Computer Science, Shri Guru Buddhiswami Mahavidyalaya, India*

Abstract
Congestion control is an important aspect in Wireless sensor networks when the nodes produce more traffic than the network can handle. In the current era where Internet of things (IoT) is being used extensively in every filed possible, congestion control algorithms play a vital role in maintaining the communication between nodes and the server. The congestion control algorithms take into account the resources available in the network and the nodes trying to communicate to resolve the congestion, achieving the maximum throughput. This paper presents a hybrid heterogeneous hierarchical relative difference clustering technique to avoid congestion in a given WSN. The clustering process is performed on the nodes taking into account their different characteristics. The distance between the nodes, the amount of data transmitted from the node, the energy required for transmission etc are used to cluster the nodes. Once the clustering of the nodes on individual parameters is performed, a hybrid combination of the nodes is formed in such a way that the congestion in the clusters in minimum. The network architecture considered for the experiments consists of a base station, gateways and the nodes that want to communicate the data. The proposed method reduced the congestion in the network to a great extent.

Keywords:
Congestion Control, Wireless Sensor Networks, Hierarchical Relative Distance, Clustering

1. INTRODUCTION

Wireless networks use radio interfaces as a communication medium. Several protocols exist for broadcasting data on this type of media. The most common wireless protocols are those used in 802.11b/g type WiFi networks [1]. This choice has a significant impact on the use communication channels. Therefore, a station can only receive one data packet all at once, otherwise a collision results. This phenomenon can be illustrated by a group of three characters. Two people address the same individual at the same time. This will receive two pieces of information and it is inaudible to him. To solve this problem in everyday life, rules of politeness are used which stipulate that each person speaks in turn. In the 802.11b/g networks, the principle is the same, and the polite rule used is the protocol access to the CSMA/CA network [2]. This protocol aims to avoid collisions. But this protocol has its limits: if a station has a lot of interlocutors, it is extremely difficult to avoid collisions without maintaining a satisfactory flow. If for example there are large number of people speaking to the same individual, and if to avoid collisions, each person is allowed to pronounce a word in turn, it will take a significant time for each nobody can say his sentence. So there is a traffic congestion phenomenon that occurs.

Congestion phenomena in Ad-Hoc networks is a broad research topic studied and it is still relevant today. Even recently, an article [3] offers us a congestion control mechanism for Ad-Hoc networks. Such mechanisms can act on different levels: access protocols, routing, etc Article [4] presents some of these protocols. Certain publications [5] propose the multiplication of radio interfaces to limit the collisions. The modification of the access protocols to the radio layer also remains an alternative interesting [6]. Nowadays, multilayer protocols (access and routing levels) provide even more efficiency [7]. Theoretical developments have come to support this type of study [8]. Bandwidth is also a criterion that has been widely studied [9].

Congestion is the phenomenon observed when there are too many packets present in the network. Accumulation of packets awaiting transmission in the nodes of the network needs retransmission of packets wrongly or wrongly considered as lost [10]. This leads to the increased traffic and degraded performance. In the literature two types of solutions are presented: prevention (or anticipation) or regulation (a posteriori). The techniques used can be divided into the following categories:

1.1 PRE-ALLOCATION OF RESOURCES

This category of congestion control algorithms use reservation of resources namely storage space, portion of bandwidth. This can only be used as part of a connection and is efficient but may cause unnecessary monopolization of resources. Mingyan et al. [11] presented a model based on the pre allocation model.

1.2 ARRHYTHMIC CONGESTION CONTROL

This category of congestion control algorithms use keeps a constant number of packets circulated in the network by using tokens. The drawback in the network is that the distribution of tokens according to the load is poor. Khalel et al. [12] modelled the concept of arrhythmic congestion control.

1.3 RETRO-CONTROL

This category of congestion control algorithms uses transmitter flow control. Here each node monitors the utilization rate of its links explicitly by sending congestion packet to nodes sending packets to be routed on loaded links. Implicitly the absence of acknowledgment indicates congestion in the reaction time. Carlucci et al. [13] and Srivastava et al. [14] presented models using retro control.

1.4 PRIORITY/DESTRUCTION OF PACKETS

This category of congestion control algorithms is not used unless massive pre-reservation is present and therefore inefficient and congestion is inevitable. Here one needs to define criteria to choose the packets to destroy or forward as a priority. It is inevitable but as a last resort. Argoubi et al. [15] presented a concept on priority model.

The WSN architecture used in this paper is presented in the Fig.1. The base station receives the data of the nodes through the gateways placed in tactical position in the grid. The Fig.1 illustrates the working mechanism.

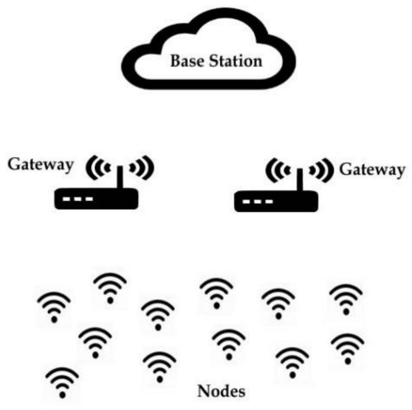

Fig.1. Proposed WSN architecture

The nodes in the WSN connect to the gateways to transmit the data. The gateway intern collects the data form the nodes and sends it to the base station. For example, if the base station can receive the data at a speed of 100kb/s, this speed is divided uniformly among the four gateways. Each gateway gets a transmission speed of 25kb/s which is used to collected the data form the nodes.

2. UNSUPERVISED MACHINE LEARNING

Unsupervised learning consists of teaching an artificial intelligence algorithm information that is neither labelled, nor classified to allow this algorithm to react to this information without human intervention. In addition, the algorithm processes the data without any prior training, it "trains itself" with the data it receives [16]. An unsupervised learning algorithm uses categories associated with the data submitted to it, but it must make them emerge itself, in order, for example, to recognize that a cat is a cat, or that an article of the IA journal is an article of the IA journal. In supervised learning, thousands of images of dogs are provided to the algorithm with the label "dog". In this way, if it is given any other image it can determine whether it represents a cat or not. In unsupervised learning, no label is provided, by processing thousands of images, the algorithm must be able to create a "dog" category on its own, even if it doesn't know what it is, and it notice the similarities between the images. The algorithm only grouped all the images of dogs together because they all had a certain number of points in common. size, four legs, shape of the face, shape of the muzzle etc. Nieva et al. [17] presented an unsupervised machine learning model.

The most common unsupervised learning problem is clustering. Clustering is the stage where the data is separated into categories. It is the cornerstone of unsupervised learning. This is what makes the parallel between unsupervised learning and the human way of reasoning since artificial intelligence is then autonomous. There is no need for prior human intervention to create the categories, which is exactly the case with humans. The categorization phase of supervised learning (done by humans therefore) is a very resource intensive process. Letting artificial intelligence take care of it therefore has an undeniable advantage [18]. Unsupervised learning algorithms can perform more complex processing tasks than supervised learning systems, but they can also be more unpredictable. When asked to sort through different data, an artificial intelligence that relies on unsupervised learning can also add categories in unexpected and unwanted ways, which can create confusion instead of tidying up.

At the moment, unsupervised learning is in its infancy, but there is a lot of scope for it, especially for data classification [19]. Unsupervised learning is still too little exploited in a concrete way, this is largely due to the fact that it can have black box phenomena. In the sense that one cannot always explain perfectly why the algorithm gives such or such result. This makes it unusable (for the moment) in particular for the field of health or armaments, for which the requirements in terms of data analysis are greater.

3. CLUSTERING

Before the artificial intelligence of computers became capable of detecting similarities between individuals, it was human intelligences that implemented clustering algorithms. There are several dozen, of which here are several major categories. For each method, it is necessary to choose how to measure the similarity between two individuals, which we can imagine as two points of the real space in dimension *p*. Hierarchical type clustering methods are different. They form connections between individuals step by step, and use a distance matrix between individuals to find the cluster closest to another. In practice, the algorithm is initiated from *n* sets which are individuals as singletons. The first connection is therefore made between the two closest individuals. Yau et al. [20] presented hierarchical type clustering.

To start the second step, the distance matrix is updated by removing a box, because of the grouping of two individuals. But how to calculate the distance from one set to another if it is not a singleton. This is precisely one of the choices to be made at the start. the aggregation strategy. There are many, the simplest being to choose the minimum distance between individuals of the two groups (single linkage), maximum (complete linkage) or average (average linkage). At the end of this second step, the two closest groups are connected. And so on for the following steps, until you connect the last two groups which cover all the individuals.

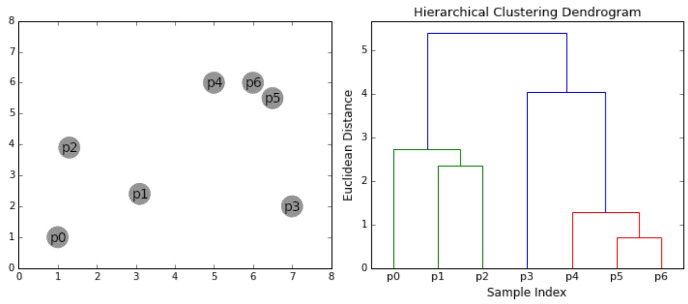

Fig.2. Clustering by hierarchical method

On the left, the individuals represented in R_2. On the right, the associated dendrogram and the stages of its creation. If we want two classes, we will choose (p_0, p_1, p_2) and (p_3, p_4, p_5, p_6). If we want classes that are more than one unit of measurement apart, we will choose (p_0, p_1, p_2), (p_3) and (p_4, p_5, p_6).

The successive connections are represented on a dendrogram shown in Fig.1. The distance associated with each connection is on the y axis of it. The algorithm ends of course with the choice of our clusters. Here again, the criterion is not unique. One can want a precise number, or fix a criterion of inter-class distance. The dendrogram to highlight the clusters is shown in Fig.1.

The option calculates a hierarchy from a matrix of distances. It offers three procedures. A hierarchy over a finite set A is a set A of parts of A (the classes), i.e. a subset of the set of parts of A such that.

Step 1: *A* is an element of *A* (the largest class contains all the elements);

Step 2: If *a* is an element of *A*, {*a*} is an element of *A* (the smallest classes contain only one element);

Step 3: If *H* and *K* are two classes of A, of three things one or they are without common elements, either *H* contains *K*, or *K* contains *H* (two classes are either disjoint or nested) 1.

The classification is hierarchical with a genes and a family either do not of species in common or the genus belongs to the family. A hierarchical classification algorithm is a method of constructing hierarchy. The ascending hierarchical classifications (CAH) start from the set of classes with only one element and at each step bring together two most similar classes until the class containing all the elements are obtained. The notion of resemblance or difference of two classes uses a notion of proximity between classes which derives from the distance between starting elements 2. Choosing an initial distance, a principle of construction, an index of proximity between classes generate a multitude of possibilities which exclude that a hierarchy is true or false and it can only be useful.

3.1.1 Agglomerative Hierarchical Clustering Technique:

In this technique, initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or *K* clusters are formed.

The basic algorithm of Agglomerative is straight forward.

Step 1: Compute the proximity matrix

Step 2: Let each data point be a cluster

Step 3: Repeat - merge the two closest clusters and update the proximity matrix

Step 4: Until only a single cluster remains

Key operation is the computation of the proximity of two clusters. A sample proximity model is shows in Fig.3.

	A	B	C	D	E
B	16				
C	47	37			
D	72	57	40		
E	77	65	30	31	
F	79	66	35	23	10

Fig.3. Proximity Matrix

A pictorial representation of the Agglomerative Hierarchical clustering Technique is shown in Fig.4. Consider six data points {*A*,*B*,*C*,*D*,*E*,*F*}.

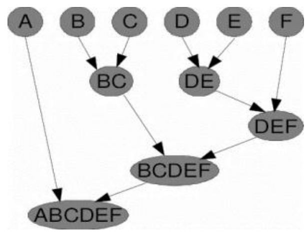

Fig.4. Hierarchical Clustering Technique

Step 1: In the initial step, the proximity of individual points is calculated and consider all the six data points as individual clusters as shown in Fig.4.

Step 2: In step two, similar clusters are merged together and formed as a single cluster. Let's consider B, C, and D, E are similar clusters that are merged in step two. Now, four clusters remain which are A, BC, DE, F.

Step 3: We again calculate the proximity of new clusters and merge the similar clusters to form new clusters A, BC, DEF.

Step 4: Calculate the proximity of the new clusters. The clusters DEF and BC are similar and merged together to form a new cluster. We're now left with two clusters A, BCDEF.

Step 5: Finally, all the clusters are merged together and form a single cluster.

The Hierarchical clustering Technique can be visualized using a Dendrogram. A Dendrogram is a tree-like diagram that records the sequences of merges or splits as shown in Fig.5.

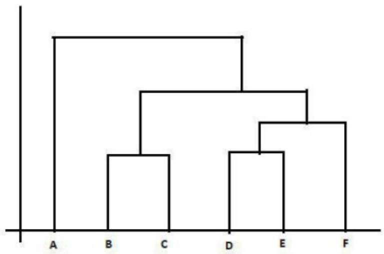

Fig.5. Dendrogram

4. EXPERIMENTAL RESULTS

The experimental setup consisted of a 100×100 grid with 25 nodes placed in random positions. The nodes are plotted in Fig.6.

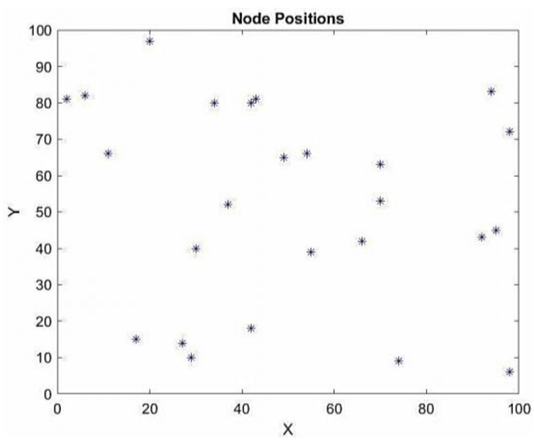

Fig.6. Node Positions

The parameters extracted from the nodes in the experiment are:

- Distance between the nodes
- Energy of the nodes
- Packet length
- Frequency of Transmission

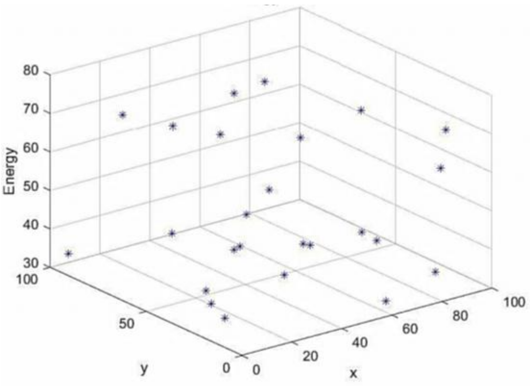

Fig.7. Energy of the nodes

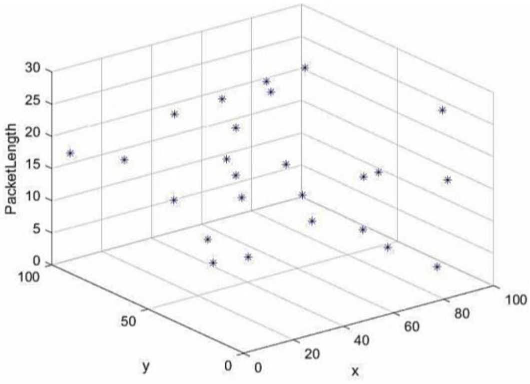

Fig.8. Packet length of nodes

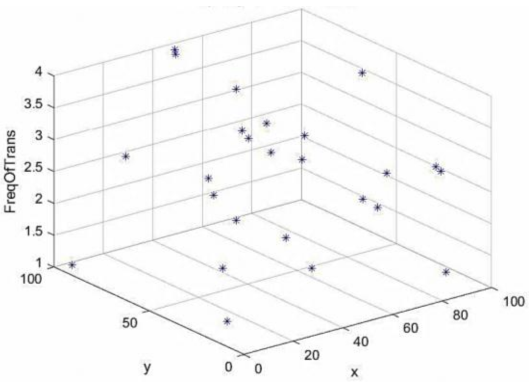

Fig.9. Frequency of Transmission of the nodes

The Fig.7-Fig.9 show the energy, packet length and frequency of transmission respectively. The nodes are clustered using hierarchical relative distance based clustering on the above parameters and the clustering results are shown in Fig.10-Fig.12.

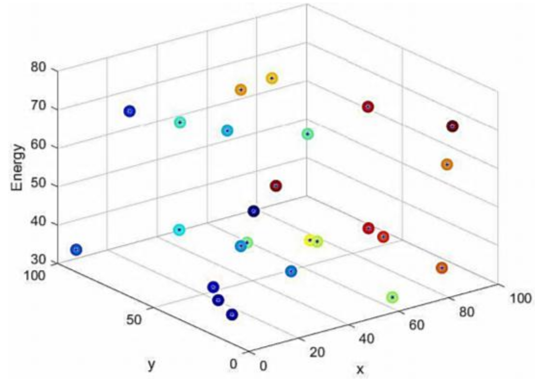

Fig.10. Clustering results of Energy

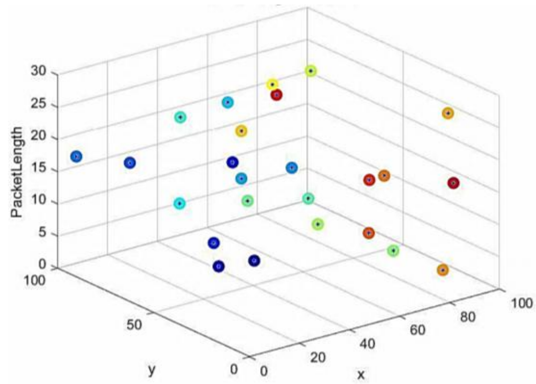

Fig.10. Clustering results of Packet length

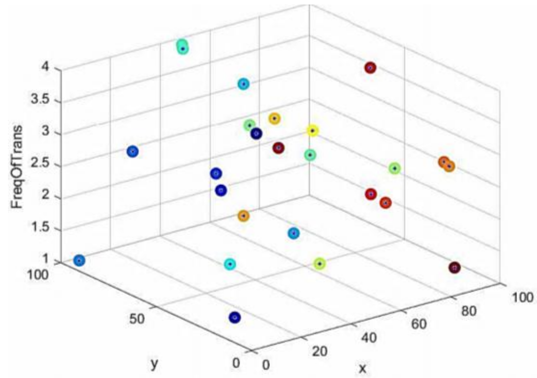

Fig.10. Clustering results of Frequency of Transmission

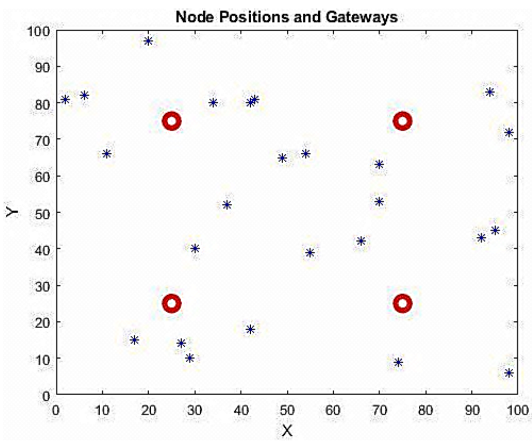

Fig.11. Gateway locations

The Fig.11 shows the gateway positions in red circles. For instance, if distance based clustering is employed, all the nodes close to gateway 1 will transmit through it. But because all the nodes have high energy, congestion occurs at gateway 1. The clustering is shown in Fig.12.

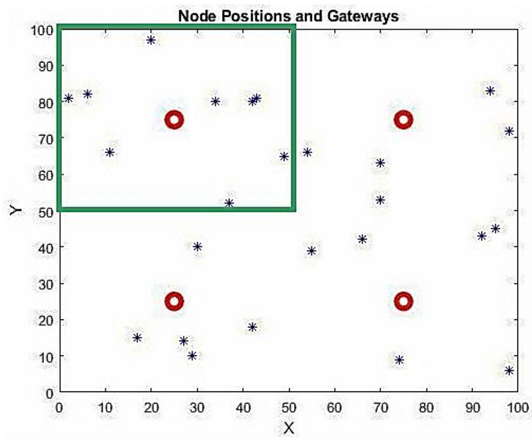

Fig.12. Clustering result of distance

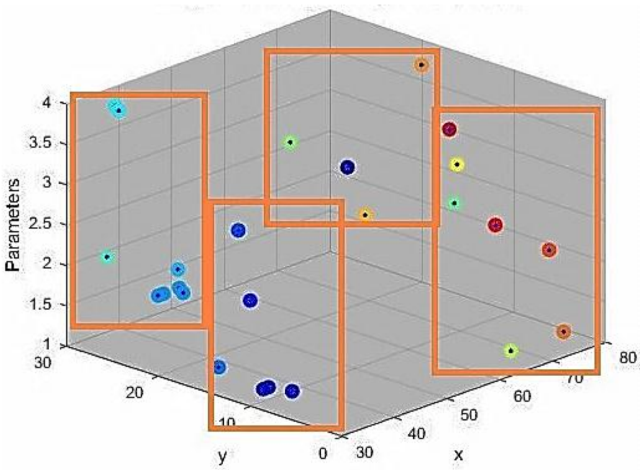

Fig.13. Final Clusters formed after proposed clustering model

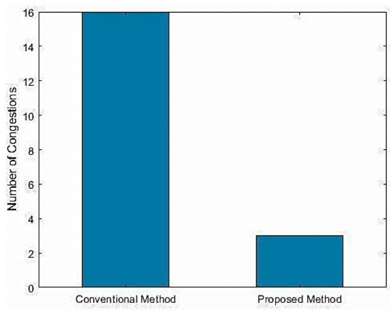

Fig.14. Number of congestions

In the proposed model, four final hybrid clusters are formed from the results of the individual clusters. The nodes are chosen in such a way that each gateway handles minimum congestion. The nodes with high energy requirement are scattered between all the gateways. The same approach is followed for packet length and frequency of transmission in such a way that the resources of the gateway are optimally utilized. For example, the gateway 1 get a node with high energy, long transmission time and less transmission frequency. The proposed method thus reduced the congestion at the gateways.

The Fig.14 shows the congestion with normal distance based clustering and the proposed method. The number of congestions in the proposed method is very less when compared to the conventional method.

5. CONCLUSION

Congestion control is an important task to be performed in a WSN to increase the performance and prevent loss of data. Conventional distance and energy based clustering techniques could not achieve good congestion control. The method proposed in this paper takes into account four parameters of the nodes namely inter node distance, energy, transmission packet length and frequency. The hybrid clustering performed in the end makes sure that the nodes are equally distributed among the gateways in the best way possible to avoid congestion.

REFERENCES

[1] O. Semiari, S. Walid, B. Mehdi and D. Merouane, “Integrated Millimeter Wave and Sub-6 GHz Wireless Networks: A Roadmap for Joint Mobile Broadband and Ultra-Reliable Low-Latency Communications”, *IEEE Wireless Communications*, Vol. 26, No. 2, pp. 109-115, 2019.

[2] Trang Tien and Hoon Oh. “SCSMA: A Smart CSMA/CA using Blind Learning for Wireless Sensor Networks”, *IEEE Transactions on Industrial Electronics*, Vol. 67, No. 12, pp. 10981-10988, 2019.

[3] Anam Javaid, Tariq Umer and Mubashir Husain Rehmani, “A Comprehensive Survey of Network Coding in Vehicular Ad-Hoc Networks”, *Wireless Networks*, Vol. 23, No. 8, pp. 2395-2414, 2017.

[4] V. Lakshman and C.R. Bharathi, “Multi-Mode Routing Mechanism with Cryptographic Techniques and Reduction of Packet Drop using 2ACK Scheme MANETs”, *Smart Intelligent Computing and Applications*, pp. 649-658, 2019

[5] Marco Brocanelli and Xiaorui Wang, “Smartphone Radio Interface Management for Longer Battery Lifetime”, *Proceedings of IEEE International Conference on Autonomic Computing*, pp. 93-102, 2017.

[6] Marco Centenaro, Lorenzo Vangelista, Stephan Saur, Andreas Weber and Volker Braun. “Comparison of Collision-Free and Contention-Based Radio Access Protocols for the Internet of Things”, *IEEE Transactions on Communications*, Vol. 65, No. 9, pp. 3832-3846, 2017.

[7] Y.E. Sagduyu, Yi Shi, Tugba Erpek, Sohraab Soltani, Sharon J. Mackey, Derya H. Cansever, Mitesh P. Patel, Bart F. Panettieri, Boleslaw K. Szymanski and Guohong Cao, “Multilayer MANET routing with social Cognitive Learning”, *Proceedings of International Conference on Military Communications*, pp. 103-108, 2017.

[8] R. Ferrus, Oriol Sallent, Jordi Perez-Romero and Ramon Agusti, “On 5G Radio Access Network Slicing: Radio Interface Protocol Features and Configuration”, *IEEE Communications Magazine*, Vol. 56, No. 5 pp. 184-192, 2018

[9] Jeffrey G. Andrews, Francois Baccelli, Thomas D. Novlan and Charlie Jianzhong Zhang, "Design and Analysis of Initial Access in Millimeter Wave Cellular Networks", *IEEE Transactions on Wireless Communications*, Vol. 16, No. 10, pp. 6409-6425, 2017.

[10] Raquel Caballero Aguila, Aurora Hermoso Carazo and Josefa Linares Perez, "Distributed Fusion Filters from Uncertain Measured Outputs in Sensor Networks with Random Packet Losses", *Information Fusion*, Vol. 34, pp. 70-79, 2017.

[11] Xinping Guan, Cunqing Hua, Cailian Chen and Ling Lyu, "Predictive Pre-Allocation for Low-Latency Uplink Access in Industrial Wireless Networks", *Proceedings of International Conference on Computer Communications*, pp. 306-314, 2018.

[12] Tarik Ismail Hassan, "An Experimental Study of Congestion Control in Wireless Sensor Networks using a Combination of Resource and Traffic Control", PhD Dissertation, Department of Computer and Electrical Engineering, University of Leicester, pp. 1-213, 2018.

[13] Gaetano Carlucci, Luca De Cicco, Stefan Holmer, and Saverio Mascolo, "Congestion Control for Web Real-Time Communication", *IEEE/ACM Transactions on Networking*, Vol. 25, No. 5, pp. 2629-2642, 2017.

[14] Sachin Tripathi and Karan Singh, "Energy Efficient Optimized Rate-Based Congestion Control Routing in Wireless Sensor Network", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 3, pp. 1325-1338, 2020.

[15] Karima Maalaoui, Mohamed Houcine Elhdhili and Leila Azouz Saidane, "Priority-MAC: A Priority based Medium Access Control Solution with QoS for WSN", *Proceedings of International Conference on Computer Systems and Applications*, pp. 1-6, 2016.

[16] Junaid Qadir, Aunn Raza, Hunain Arif, Kok-Lim Alvin Yau, Yehia Elkhatib, Amir Hussain and Ala Al-Fuqaha, "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges", *IEEE Access*, Vol. 7. pp. 65579-65615, 2019.

[17] Joaquin F. Rodriguez Nieva and Mathias S. Scheurer, "Identifying Topological Order through Unsupervised Machine Learning", *Nature Physics*, Vol. 15, No. 8, pp. 790-795, 2019.

[18] Jianwei Yang, Devi Parikh and Dhruv Batra, "Joint Unsupervised Learning of Deep Representations and Image Clusters", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5147-5156. 2016.

[19] Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M. Poskitt and Jun Sun, "Anomaly Detection for a Water Treatment System using Unsupervised Machine Learning", *Proceedings of IEEE Conference on Data Mining Workshops*, pp. 1058-1065, 2017.

[20] Christopher Yau, "PCA Reduce: Hierarchical Clustering of Single Cell Transcriptional Profiles", *BMC Bioinformatics*, Vol. 17, No. 1, pp. 140-165, 2016.