ISSN: 2366-1313

INTERNET-OF-THINGS DATA REDUNDANCY SCHEME FOR REDUCTION OVER WIRELESS SENSOR NETWORK

[1]Betam Suresh, [2]Dr. S. TamilSelvan
[1]Research Scholar, Department of CSE, Annamalai University, Annamalainagar 608002, Tamil Nadu, India
[2]Assistant Professor, Department of Information Technology, Annamalai University, Annamalainagar 608002, Tamil Nadu, India.

Abstract:

In Wireless Sensor Network, sensor nodes are randomly deployed where the sensor nodes are not situated faraway from each other. Thus, an overlapping area is generated due to intersection of their sensing ranges. If an event occurs within the overlapping area, all the sharing nodes sense the same event and produce redundant and correlated data. Data redundancy exhaust network resources and increase network overhead. Data aggregation and numerous data redundancy reduction algorithms are employed to solve this problem. This paper reviews modern data redundancy reduction used sleep schedule model to solve the redundancy. All proposed algorithms are classified on the basis of network coverage and similarity among sensory data which can be used in reducing redundancy in WSN effectively.

Keywords - Coverage Control Protocol; Efficient Data Redundancy; Intelligent Algorithms; Locality Sensitive Hashing; Sleep Schedule.

1 INTRODUCTION:

Smart technologies play a crucial role in sustainable economic growth. They transform houses, offices, factories, and even cities into autonomic, self-controlled systems without human intervention [1]. This modern automation trend and ever-increasing use of cuttingedge technologies are boosting the world's economy [2]. The Internet of Things (IoT) and Wireless Sensor Networks (WSN) both play vital roles in this modernization [3]. IoT is a branch of engineering primarily concerned with offering thousands of miniature, physical connected objects, which may collaborate to achieve a shared goal. IoT has gained much importance due to the abundant usage of these tiny networked devices. These are smart, yet basic things that can sense and communicate wirelessly [4]. WSN is a collection of sensor and routing nodes, as shown in Figure 1, which may be put together in the environment to predict physical conditions, such as wind, temperature, and many others. These networks collect and process data from tiny nodes and than transfer it to the operators. WSN also plays a significant role in aquaculture and the oil industry, including data collection, offshore exploration, disaster prevention, tactical surveillance, and pollution monitoring.

WSN can be held to denote finite set of sensor devices geographically distributed in each indoor or outdoor environment. In some applications like commercial applications, monitoring system, food safety, military surveillance, and the prediction of natural disasters [5], there is presence of WSN. To ensure reliability, WSN deployment is densely and randomly based on the kind of application being used and it is of benefit to inaccessible area [6]. WSN has limited resources of the sensor nodes, since the main key of resource in WSN are energy and communication. Approximately, 80% of energy consumed in individually sensor node is used for data transmission. Rechargeability or replacement of the battery, source of energy, is impossible. Hence, energy efficiency is a main design issue that needs to be considered for enhancing WSN lifetime. The above WSN features facilitate the ability of sensor nodes to detect same event concurrently and produce redundant and correlated data, due to overlap their sensing ranges within monitoring area [7].

Moreover, to improve data accuracy, system lifetime, sensing reliability and security issues, redundancy must be exploited [8]. This will be of benefit to the redundancy in WSN. However, generating redundant and correlated data, which can give rise to network overhead and drain resources which is considered a disadvantage for the redundancy in WSN. WSN suffers from redundancy, especially when it comes to data transmission, because sensor nodes consume huge energy while transmitting redundant data [9]. Thus, data aggregation remains essential in WSN, because it removes redundant data which in turn conserve energy by reducing energy consumed by sensor nodes [10].

Data mining is performed on sensors' data in two ways both temporal and spatial. In a spatial manner typically, WSN applications need details or information from spatially dense sensor nodes to provide a sufficient coverage for content gathering. In the end,

ISSN: 2366-1313

many sensors can record data about one event in their sensor fields. Because of the high density of the topology of the network the spatially proximal sensor data are closely linked to decreasing inter-node distance [11].

To analyze performance to determine the performance with the suggested NDM strategy is done using the weighted low energy adaptive Clustering Hierarchy (W-leach) and Chronological Tree (CT), Novel Stream Mining (NSM), K-means Data Relay (K-MDR), Data-centric routing (DCR) as well as the Energy Efficient as well as Balanced Cluster Based Aggregation (EEBCDA).

2 RELATED WORK

Because of the rapid growth and rapid expansion of WSN The number of nodes has been growing. This leads to a huge cost of cost of operation and maintenance because of the huge volume of data. Recently, numerous research studies have been conducted on various technologies for smart homes, including IoT smart control, intelligent and energy management, home automation and wearable devices utilizing WSN. In [8] the Markov random field-based max is employed to take an energy-efficient decision with precision, employing a posteriori estimate method for an IoT-focused WSN. The method addresses the mechanism for synchronization between nodes to reduce energy consumption. However, it is afflicted by the delay parameter which is one of the most important performance indicators for IoT-based WSN.

In [11] a K-MDR clustering algorithm has been suggested to create the sensor nodes group. It is more difficult that the sensors network recognize and gather a huge amount of data that is continuous across time. The collected data must be transmitted to the sink to aid in subsequent decision-making processes. The clustering of sensory data functions as a core job for data mining. In addition it is using the Conserve as well as the Observe Mode (COM) algorithm decreases the amount of data in the cluster without diminishing the coverage. There are a number of challenges to the COM algorithm, like the limited bandwidth for communication and power supply constraints. supply, and storage resource.

in [12], a brand new framework known as bespoke WSN is being proposed to allow the integration of compressed streams of information with to mine for information streams of events that occur in a short time and stream. The proposed framework was

ISSN: 2366-1313

assessed using synthetic data from a home simulator as well as with actual data generated by a self-organizing WSN. It can analyze the patterns that are heterogeneous from various sensors that are connected to a variety of software and applications. Based on the results of simulations the author claims that the framework was identified as self-adaptable, energy efficient, less redundant and more scalable.

With the use of low-cost WSN devices, RFID and web of technologies, IoT system comes into account that enables communication between machine-to-machine and machine-to-human. In this regard, the physical objects and devices of the digital world can be integrated and utilized over the virtual environment by producing various types of applications and services. However, there are certain major challenges have to be consider in the deployment of the IoT system. The sensors used in the devices generate enormous amount of data rapidly. These generated data can be noisy, continuous and raw in nature

In order to understand the raw data in [12] Compressed Sensing (CS) tory is utilized by creation of the sampling point. It directly corresponds to the generated large volume of data by the sensor nodes. It uses both the CS-based signal and information acquisition/compression paradigm in the form of non-linear cluster aparse reconstruction algorithm with the feature of random sampling. Motivated by this, a Service Oriented Architecture (SOA) is proposed by the author because heterogeneous devices generates different pattern of data. Hence, first architecture of corresponding IoT should be learned and afterwards based on data management framework, an essential web service has to be designed

3 PROPOSED WORK:

3.1 Data packet classification in IoT-oriented WSN

The majority of research focused on the use of data mining as a way of reducing intrusions or malfunctioning nodes on the network. However , in this research the emphasis is on eliminating redundancy in data packets detected by nodes prior to sending them to the base station or the cluster head to be processed. Sensor nodes in IoT is operated in two ways. The first is that the data is collected and sensed by its neighbors in a spatial manner. Additionally, the data is analyzed and monitored over a defined interval of time, in a temporal manner. The strategy proposed works best in

ISSN: 2366-1313

both situations. To eliminate redundant data, specific attributes and features of the data that are analyzed are necessary. In order to filter the data and make sure that it is a proper mining of packets, the packet classification method is used in the proposed method. The proposed NDM approach allows load balance, traffic accounting, and monitoring of data sensed through the ability to provide Quality-of-Service (QoS) for various software and applications.

There are two kinds of data that are generated and analyzed by sensors both raw and intermediate. The raw data is created by leaf nodes (spatial). But, the raw data that is in the forwarding status between the sensor nodes is referred to as intermediate data (temporal). In this regard, two decision rules were employed in the classifier used by TCAM sensors' hardware chips. nodes. First rule R1 is used for raw data, and the the second one R2 can be employed to determine the redundancy removal for intermediate data transmitted by adjacent nodes. In this research the proposed NDM strategy is applied to get rid of redundant data packets detected by sensor nodes through the aid of rules for classifiers.

The notions of classifiers, attributes, and redundant rules are defined in detail. An attribute field iis an undefined variable length. The field's domain is i of V bits is represented with D(A i) (2v 1). A packet that has two fields (A 1, A 2, . . . An K) of properties is an D-tuple (P 1, P 2 . . . (P (d) and is part of D(A i). The relation between the matching set of rule M(R i) and the non-matching set NM(R i) of the classifier ID.

To reduce further the amount of redundancy of data between the packets detected by sensors nodes, an general probabilistic estimation model is developed known as Prob. A heuristic method of solving the problem is suggested to meet the balance between the characteristics associated with the information packets. Suppose C1 . . . CN represents the N classes of the classifier. In addition, R1 . . . RN is used to represent the set of N rules for the NDM strategy. The general form can thus be defined as:

$$Pre = \{R_1 Prob(C_1|R_1) . . . Prob(C_N|R_1)\} . . . \{R_N Prob(C_1|R_N) . . . Prob(C_N|R_N)\}$$

Table 1 : Nodes data distribution and the accuracy

ISSN: 2366-1313

No of nodes	NSM	CT	K-NDM	W-Leach	NDM	DCR	EEBCDA
50	52.28	68.67	51.86	87.90	95.12	55.23	49.00
100	60.90	58.90	64.89	83.34	98.23	62.34	49.85
150	34.90	52.19	67.23	82.45	92.87	61.89	39.10
200	32.89	66.20	72.11	86.39	91.90	59.09	37.28
250	26.89	59.20	78.23	79.89	89.23	57.12	35.97
300	23.12	54.56	70.12	75.34	84.24	52.78	32.99
350	18.20	43.90	57.67	60.12	87.56	49.34	29.87
400	16.91	56.89	50.23	62.45	86.12	48.45	23.67
450	14.78	67.13	54.34	63.13	80.13	47.90	25.90

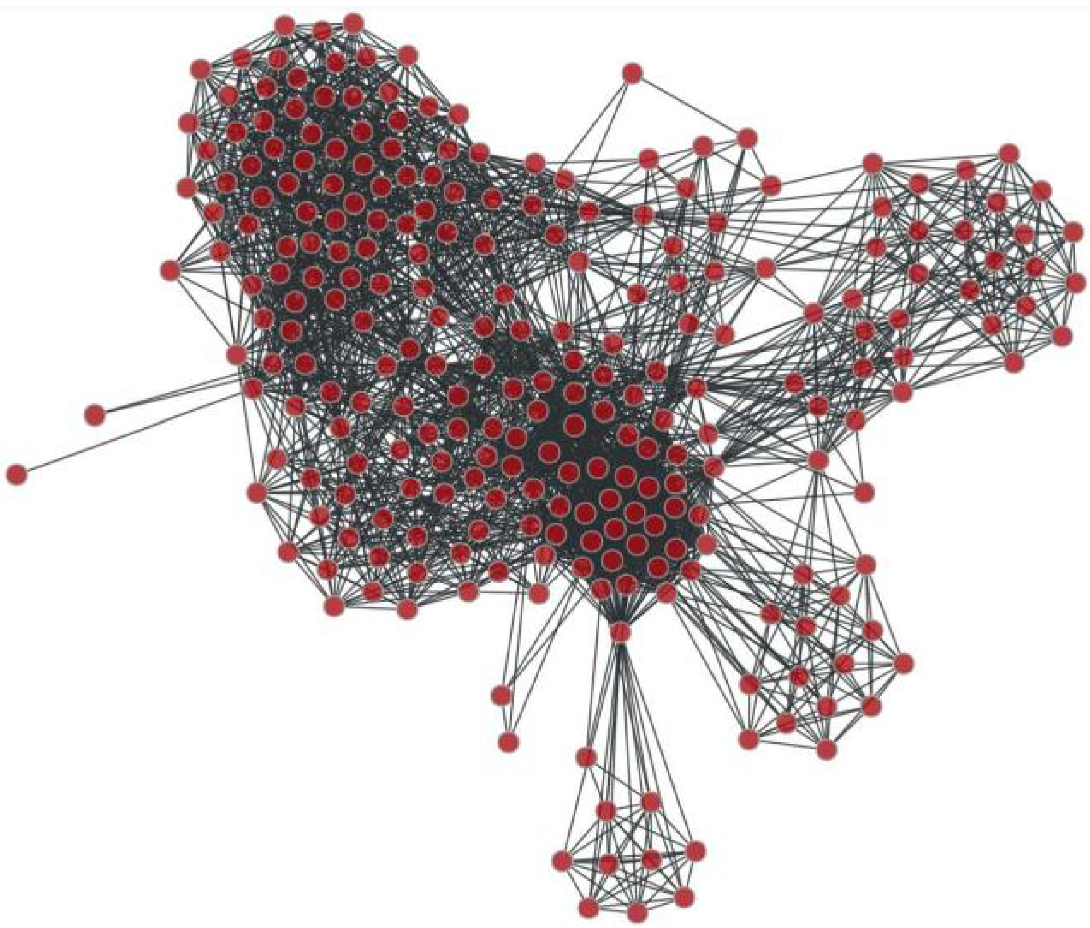

Fig 1: A random IoT oriented WSN using 500 nodes

Table 2: Values of simulation parameters for NDM strategy

Parameters	Values

ISSN: 2366-1313

Communication range	30 m
Total Buffer Space (TBS)	10KB
E_{mp}	0.0013pJ/bit/m^4
E_{elec}	50nJ
E_{fs}	10pJ/bit/m^2

4 SIMULATION RESULTS

Implementation scenarios of proposed scheme described in the above sections by using SimpleIoTSimulator simulator. By using this simulator, WSN test environments are created using a large number of nodes. It uses 64-bit RedHat Enterprise Linux for work. The total simulation time is set to be 800 seconds. We ran the simulation in 1500 rounds for WSN test environments and presented an average performance. The size of the packet is fixed to be 5000 bytes, in which 100 bytes are reserved for the header. The packet size can be upto 65535 bytes. However, it is depend on different sizes of payload. That is the reason, the used packet size in the simulation is 5000 bytes. The number of clusters is 10 % of the total number of nodes. The initial energy of the network is set to be 13000 Joules. W-leach clustering method is utilized by NDM strategy.

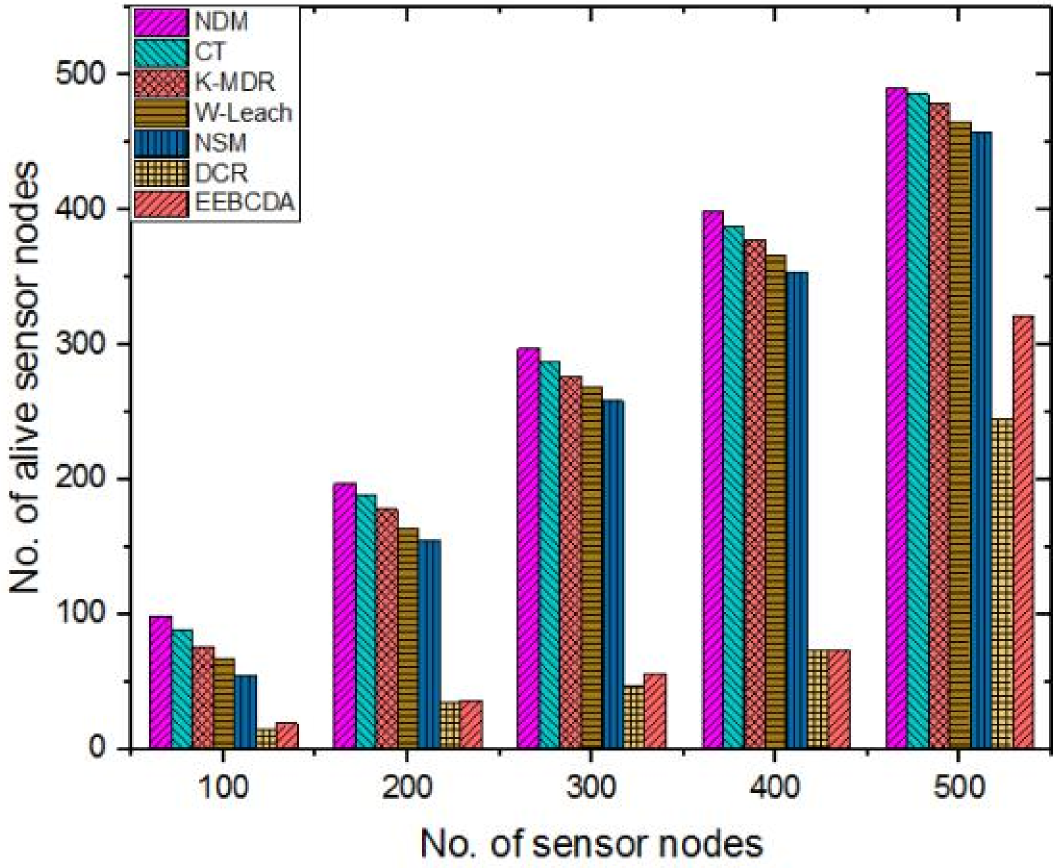

Fig 2: Performance of NDM scheme using no. of alive sensor nodes

4.1 Number of alive sensor nodes

The performance of the proposed NDM scheme is evaluated with a comparison to another scheme in terms of the number of alive sensor nodes. As the name indicates, the performance measure states the number of working sensors after the simulation. The number of sensors included in the simulation is 500 nodes . The the proposed NDM scheme is superior to W-leach by having the highest amount of active sensors. In the graph that resulted you will see that 490 of the nodes alive out of the 500 nodes of NDM scheme, when compared with W-leach. The reason for this is that every sensor node in the W-Leach scheme used data mining to analyze the packet that was received. From the graph, performance can be seen for K-MDR scheme that is approximately similar to the W-Leach scheme. The reason is, K-MDR scheme utilized the multi-sensor model for data mining by discovering the run-time of sensor nodes.

ISSN: 2366-1313

4.2 End-to-end delay

The performance metric known as end-to-end delay is analyzed in the new NDM scheme and compared to other schemes for data mining that are included in the survey of literature. The data mining method proposed has achieved 0.7861 milliseconds of end-to end delay when the number of nodes exceeds 100. The NDM strategy ensures that the minimal delay around the nodes' ranges from 1 to 300. When the simulation has reached the maximum amount of nodes i.e. 500, the NDM method achieves 46.8973 milliseconds of delay between the end of the simulation and. Other data mining strategies such as W-Leach, K-MDR, CT, DCR, EEBCDA and NSM are able to achieve 52.8731, 63.0986, 65.742, 42.307, 76.99 and 80.125 milliseconds, respectively. This was more than the time attained by the NDM data mining algorithm.

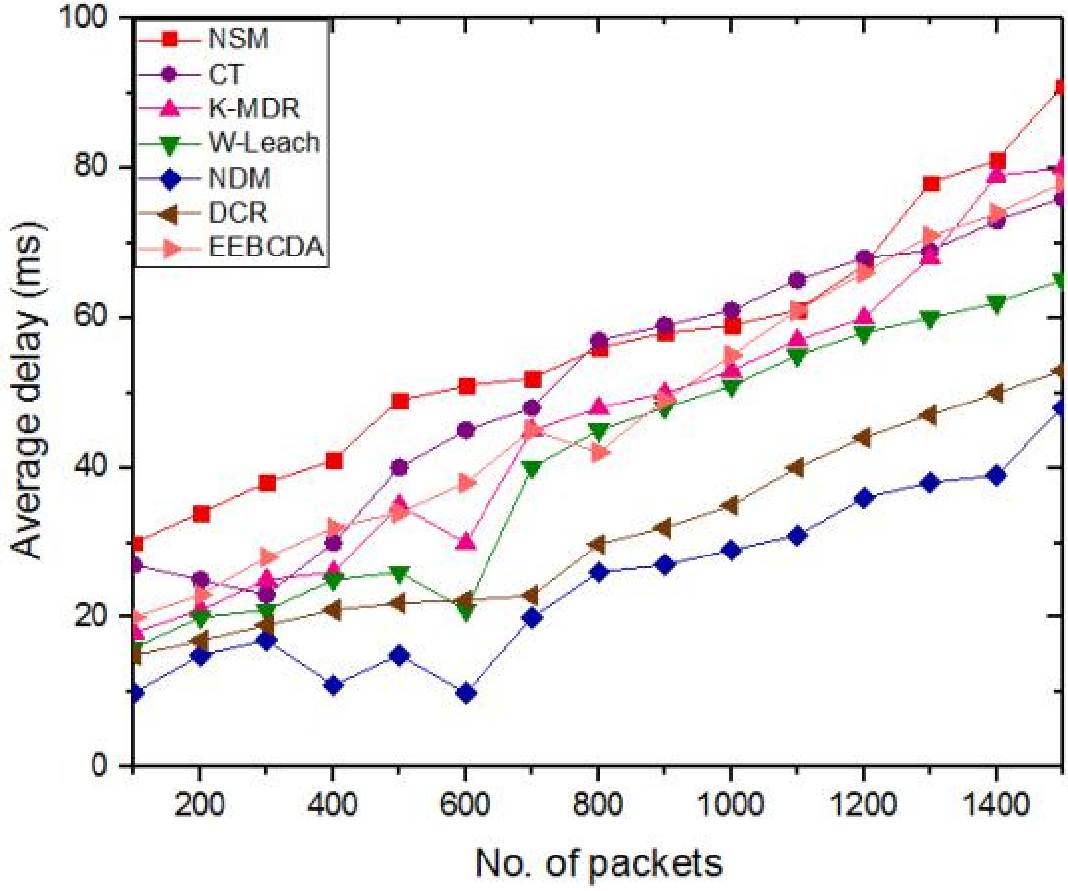

Fig 3: Performance in terms of end-to-end delay

4.3 Energy in the network

ISSN: 2366-1313

The performance of the proposed NDM algorithm for data mining is in contrast to another scheme with regard to energy. In the graph, we is evident that the performance of the NDM scheme is superior in comparison to other schemes. The energy that the network produces is about 97.89 J when the simulation used an unlimited amount of sensors i.e. 500. A different scheme called W-Leach was able to achieve just 69.87 Jjoules over this period. The reason isbecause each sensor node is outfitted with a controller for data mining for data aggregate. This way, the energy consumption of the network is diminished when compared to the NDM strategy.

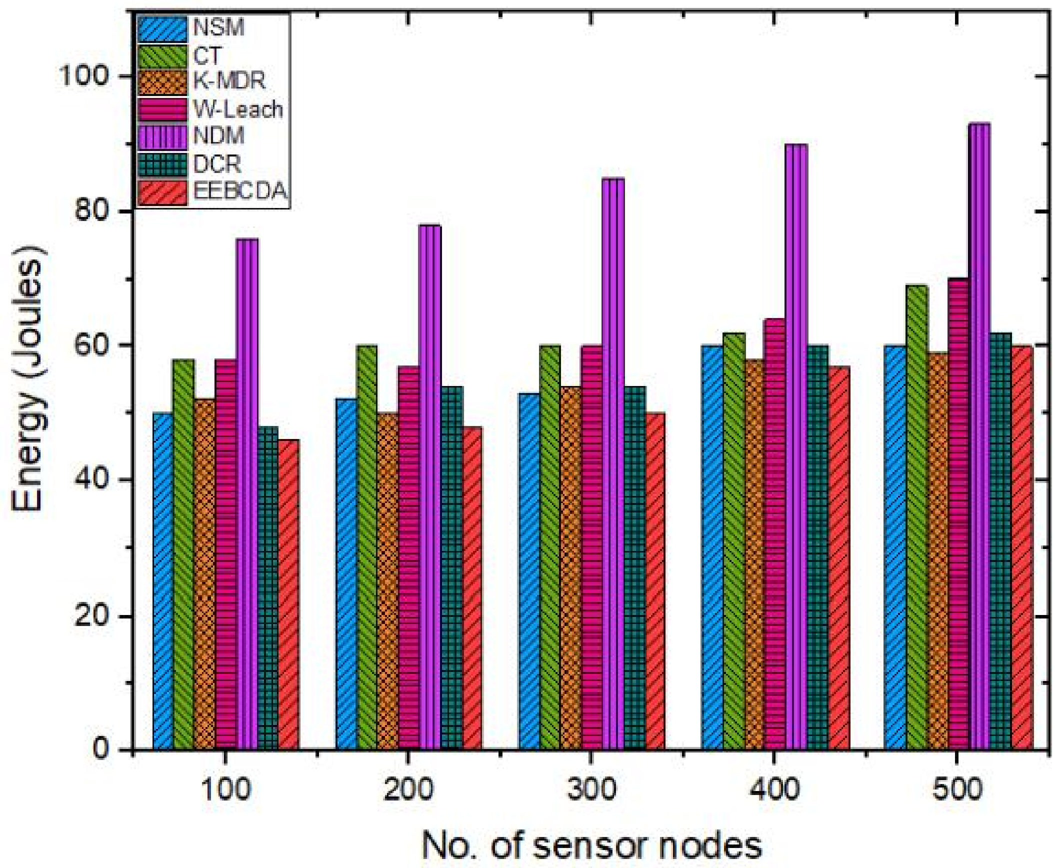

Figure 4: Performance in terms of energy

4.4 Lifetime of the network

The life span and energy are closely related and directly affect how well the system performs. If the energy of sensors is high, this leads to an increase in the lifetime of the entire network. Figure 3.6 shows the life span of the network that is proposed by the NDM Data Mining scheme. Figure 3.5 already show that the energy saved through the NDM scheme amounts to 97.89 Joules. This means that the lifetime of the network is quite long for the proposed strategy when contrasted to other schemes.

ISSN: 2366-1313

Thus, this proposed NDM data mining method exceeds its competitors and provides a longevity of 95 percent.

4.5 Throughput

The average throughput in terms kbps is displayed for all of the methods of data mining and the proposed strategy. To increase the effectiveness of the NDM in the long run, the average throughput must be greater than that due to an algorithm for data mining. From the chart we can see that the average speed of throughput attained from NDM NDM is 94.45 milliseconds in the event that the total number of sensors is 100. However, it can achieve 76.76 per second of throughput when the sensors is just 500. The reason is

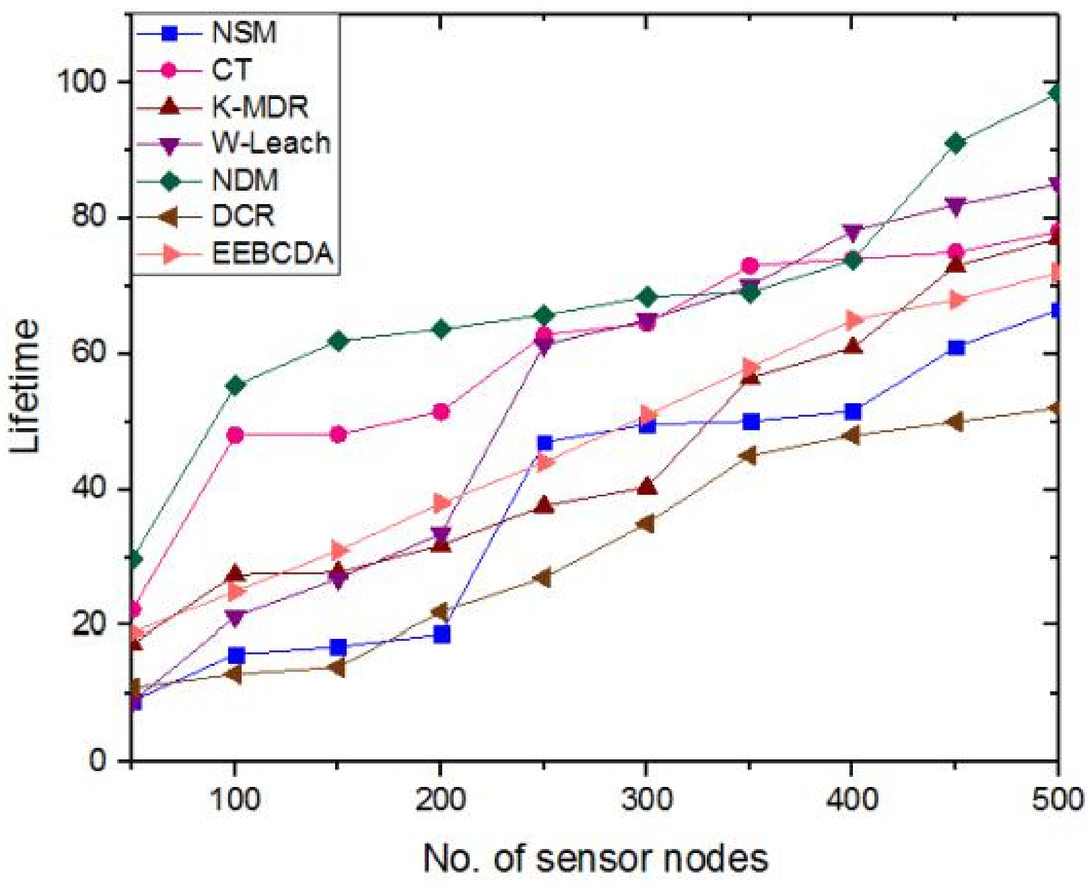

Figure 5: Performance in terms of network lifetime

ISSN: 2366-1313

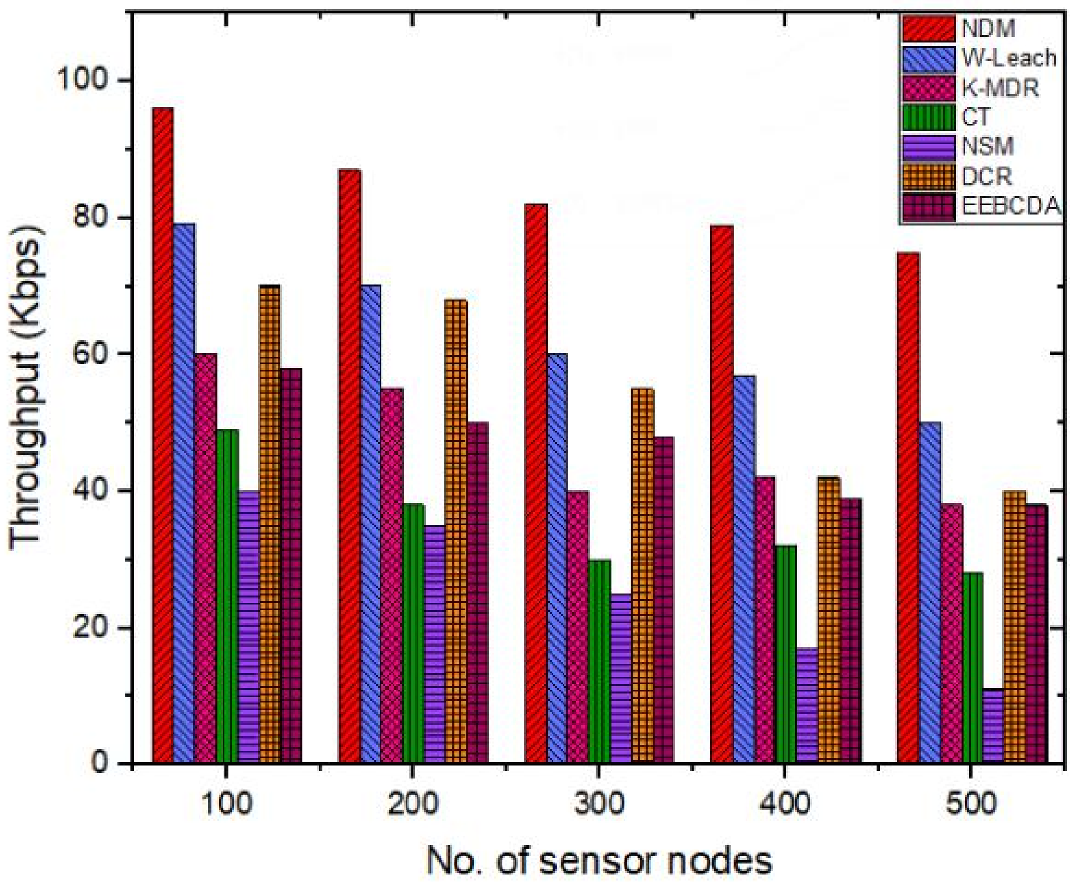

Figure 6: Performance in terms of throughput

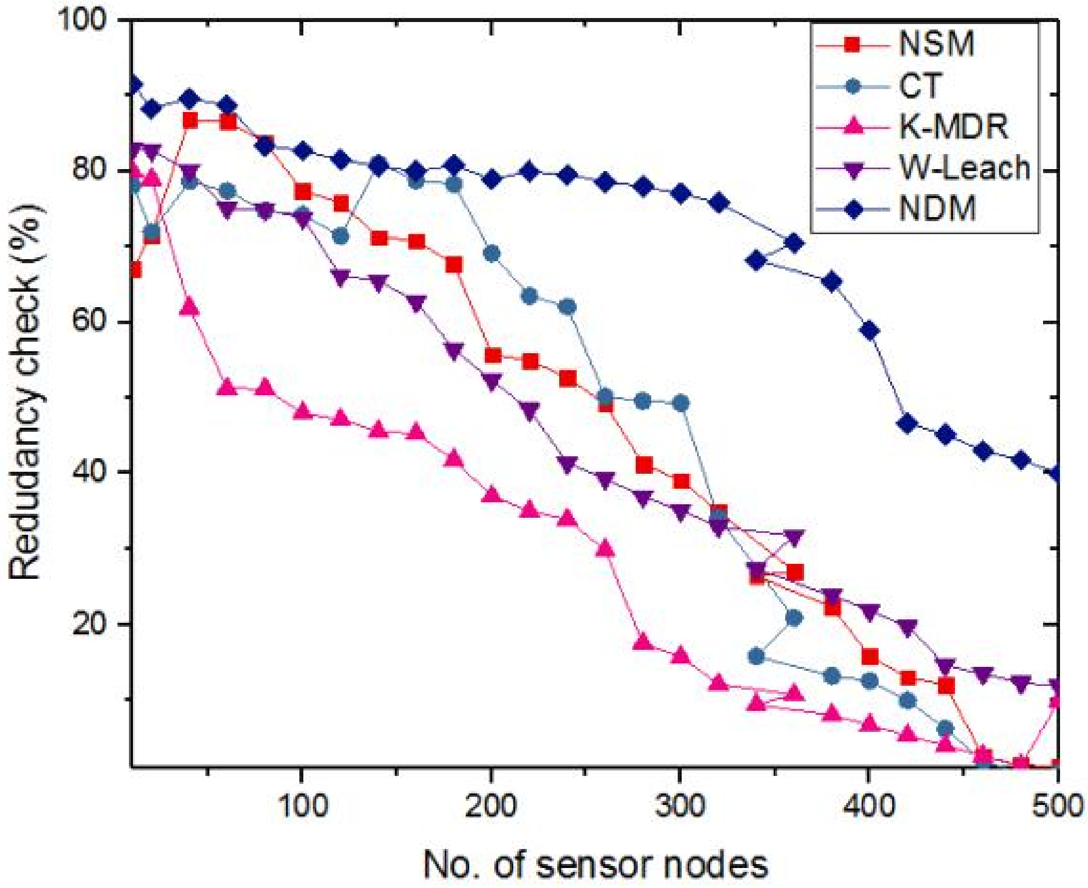

Figure 7: Performance in terms of packet redundancy check

is, the The proposed NDM scheme has higher performance because of the choice of a suitable decision regarding forwarding the data packets.

4.6 Percentage of Packet Redundancy Check

The performance metric known in the name of packet Redundancy Check (PRC) is demonstrated using the simulation set up. The table shows redundancy represented by the letters Rd. The PRC numbers for every one of schemes are normalized to (0 100, 100) and converted to percent. In Figure 3 in Figure 4, the PRC values are presented for all schemes. In the graph, we is evident that the amount of PRC is extremely high for the scheme proposed; i.e. 91 %. The paper [13] discusses different problems have been identified by the authors in order to complete the mining task on data aggregation process in the IoT-driven WSN. The authors demonstrate that mining strategies and ingratiation must make use of energy efficiently. The efficiency must take into account most important aspects, such as the ability to scale, less consumption of bandwidth, and fewer of packets, and delays.

ISSN: 2366-1313

A study has been published that is built on QoS data aggregation using WSN. The study shows how the mining technique can be utilized to decrease the amount of redundant data in order to conserve energy, and can also cause delays to the network. Thus, a balanced trade-off is required to be reached. In this manner, one is clear that the NDM data mining approach is more efficient in comparison to other methods that have been proposed in this area. The NDM method has high throughput, PRC energy savings, long-term energy savings and the amount of live sensors with less of an end-to-end delay. Due to the wireless nature of IoT-focused WSN the proposed method makes use of data mining on parameters of the network to choose the appropriate packet to forward. It offers the highest probability of success in transmitting the packet to the destination using the highest redundancy.

5 CONCLUSION:

The IoT-related WSN is an incredibly active research field due to the variety of applications and services that are available in diverse areas. In this respect there is a significant quantity of data gathered by sensor nodes and some of this data is redundant. The redundancy affects the performance of the network and causes certain disadvantages. To combat this problem an approach based on data mining known as an NDM strategy is suggested based on the parameters of network-sensed packets in order to choose the most suitable data to forward on to the subsequent node. Its performance as an IoT-based WSN that employs the NDM strategy is assessed using a variety of performance indicators. The simulation results revealed that the proposed method beat other strategies in this area. In the near future an application relating to a real-time challenge will be created with the framework proposed for dynamic scenarios.

6 REFERENCES:

[1] M. Carlos-Mancilla, E. López-Mellado and M. Siller, "Wireless Sensor Networks Formation: Approaches and Techniques", Journal of Sensors, vol. 2016, pp. 1-18, 2016.

[2] Y. Wang, X. Wang, B. Xie, D. Wang and P. D & Agrawal, “Intrusion detection in homogeneous and heterogeneous wireless sensor networks”, IEEE transactions on mobile computing, vol. 7, no. 6, pp. 698-711, 2008.

ISSN: 2366-1313

[3] N. Rani and M. Gupta, "Review on Key Pre-Distribution Schemes in Wireless Sensor Networks", International Journal of Advanced Smart Sensor Network Systems, vol. 6, no. 1, pp. 01-15, 2016.

[4] K. Sohraby, D. Minoli, and T. Znati, Wireless Sensor Networks: Technology, Protocols, and Applications. Hoboken, New Jersey: John Wiley and Sons, 2006.

[5] A. Chatterjee and D. Mukherjee, "Variety event detection in Wireless Sensor Networks through single hop cluster topology," IFIP Int. Conf. Wirel. Opt. Commun. Networks, WOCN, pp. 1–5, 2013.

[6] Y. Zhuang, J. Pan and L. Cai, "Minimizing energy consumption with probabilistic distance models in wireless sensor networks", In INFOCOM, 2010 Proceedings IEEE , pp. 1-9, 2010.

[7] A. Pal, "Localization algorithms in wireless sensor networks: Current approaches and future challenges", Network Protocols and Algorithms, vol. 2, no. 1, pp. 45-73, 2010.

[8] T. S, M. R. Tiwari and S. Shringi, "Event Driven Routing Protocols for Wireless Sensor Network- A Survey", International Journal on Computational Science & Applications, vol. 3, no. 2, pp. 1-16, 2013.

[9] M. M. Bhuiyan, L. Gondal and J. Kamruzzaman, "CODAR: Congestion and delay aware routing to detect time critical events in WSNs", In Information Networking (ICOIN), 2011 International Conference on IEEE, pp. 357-362, , 2011.

[10] M. Ndiaye, G. Hancke and A. Abu-Mahfouz, "Software Defined Networking for Improved Wireless Sensor Network Management: A Survey", Sensors, vol. 17, no. 5, p. 1031, 2017.

[11] D. Sharma, S. Verma and K. Sharma, "Network topologies in wireless sensor networks: a review 1", International Journal of Electronics & Communication Technology, vol. 4, no. 3, pp. 93-97, 2013.

[12] N. Zaman, L. Tang Jung and M. M. Yasin, " Enhancing Energy Efficiency of Wireless Sensor Network through the Design of Energy Efficient Routing Protocol". Journal of Sensors, vol. 2016, pp. 1-16, 2016.

[13] N. Kimura and S. Latifi, "A survey on data compression in wireless sensor networks," Inf. Technol. Coding Comput. 2005. ITCC 2005. Int. Conf., vol. 2, pp. 8–13, 2005.

ISSN: 2366-1313

[14] H. Chen, H. Mineno, and T. Mizuno, “A Meta-Data-Based Data Aggregation Scheme in Clustering Wireless Sensor Networks,” 7th Int. Conf. Mob. Data Manag., vol. 2006, pp. 1–7, 2006.

[15] I. Sharma, R. Singh and M. Khurana, “Comparative study of LEACH, LEACH-C and PEGASIS routing protocols for wireless sensor network”, In Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in IEEE, pp. 842-846, 2015.

[16] R. Rajagopalan and P. K. Varshney, “Data-aggregation techniques in sensor networks: A survey,” IEEE Commun. Surv. Tutorials, vol. 8, no. 4, pp. 48–63, 2006.

[17] H. Zhang and J. C. Hou, “Maintaining Sensing Coverage and Connectivity in Large Sensor Networks,” Ad Hoc Sens. Wirel. Networks, vol. 1, pp. 89–124, 2005.

[18] D. Curiac, C. Volosencu, D. Pescaru, L. Jurca And A. Doboli, “Redundancy and Its Applications in Wireless Sensor Networks: A Survey”, Wesas Journal, WSEAS Transactions on Computers, vol. 4, no.8, pp. 705-714, 2009.

[19] D. E., Boubiche, S., Bilami, A., and H. Toral-Cruz, “Toward Adaptive Data Aggregation Protocols in Wireless Sensor Networks”, In Proceedings of the International Conference on Internet of things and Cloud Computing (p. 73). ACM, 2016.

[20] W.S. Jung, K.W. Lim, Y.B. Ko, and S.J. Park, “A Hybrid Approach for Clustering-Based Data Aggregation in Wireless Sensor Networks,” 2009 Third Int. Conf. Digit. Soc., pp. 112–117, 2009.

[21] J. Kiepert and S. M. Loo, “A wireless sensor data fusion framework for contaminant detection,” in 2009 IEEE Conference on Technologies for Homeland Security, HST 2009, 2009, pp. 214–220.

[22] R. Daman, S.Dhiman and K. Rakesh, “A Review on Various Approaches of Data Aggregation in WSN,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 6, no. 6, pp. 861–864, 2016.

[23] S. Sirsikar and S. Anavatti, “Issues of data aggregation methods in Wireless Sensor Network: A survey,” Procedia Comput. Sci., vol. 49, no. 1, pp. 194–201, 2015