Two Layer data Prediction and secured data transmission in WSN

Shashikumar R[1], Dr. Anupama A Deshponde[2], Dr. B. Mohankumar Naik[3]

Research scholar, JJT University, Vidhyanagari, Jhunjunu, Rajasthan India

Abstract *— This Wireless Sensor Network (WSN) is one of the advance technologies for transmitting and receiving the sensor related information. The development of cyber-physical system (CPSs) can reduce the gap between physical world and the cyber world. The sensors sense the environmental data periodically. In continuous data sensing there are more redundant data, which leads to transmission of unnecessary bits and consumes energy. The prediction based approach has been implemented to reduce the data redundancy in the network and save the energy in the sensor nodes. It is a challenge to design a system which supports efficient method to sense and predict the data from different WSNs. Least Mean Square (LMS) and Kalman filters are used to predicate the data based on the actual value. Before transmitting the data has to be encoded. Blowfish Algorithm (BA) is implemented to encode the data. By encoding, data can be transmitted across wide range. The proposed system with LMS-Kalman filter can achieve better prediction accuracy, increase in network lifetime and privacy.*

Keywords *— Wireless Sensor Network (WSN), Cyber-Physical System (CPS), Least Means Square (LMS), Kalman Filter and Blowfish Algorithm (BA).*

I. INTRODUCTION

Internet is the technology grown very fast and acquired the space. Internet provides the easy way to interact between the people and exchanging the useful information around world within a fraction of second. Internet changes the way how we conduct the studies, research, entertainment, business and services. However, there is a space between the cyber world, where data is transformed and exchanged and the physical world in which we live. The cyber-physical systems allow an advanced idea for societal-level services that is a break point at scale never possible before. There is space between the physical and virtual world. Cyber-physical world provided a promising way to progress the contact between the virtual and the physical world. The aims of CPS are to monitor the activities of physical processes and also to change actions and its behaviours to build the physical world work in the efficient and better way. Usually, a CPS includes two major parts, a cyber system and a physical process. In general, the physical method is controlled or monitored by the cyber system, which forms a networked system of number of miniature devices with computing, communication and sensing capabilities.

The physical procedure includes may be a man-made physical system, ordinary phenomenon or a combination of more complex systems. As the advancement in the communication between the cyber system and physical process, the physical system security issues increasingly more risk at cyber system.

A WSN is a scattered network and it consists of an enormous number of self directed, distributed and small low powered devices called sensor nodes. WSN mainly used to sense the physical data, and send to destination. WSN usually consists of more number of small, spatially spread, embedded and battery-operated devices that form a network to process, collect and convey the information to users and it also has restricted processing and computing capabilities. Sensor nodes are the multi-functional and energy efficient wireless devices. Fig. 1 shows the WSN architecture.

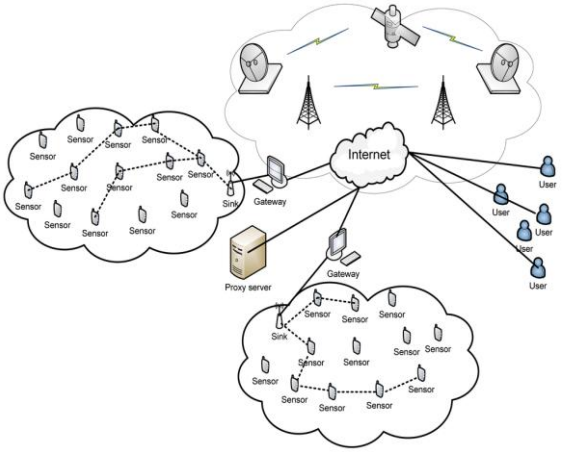


Fig 1: WSN architecture

The set of nodes collects the data from the environment to achieve the particular function. There are number of network topology are available to link the nodes. Transceivers devices are used to provide the communication bridge between nodes. There are number of sensors to monitor the environmental or physical data, like humidity, noise, light, pressure, temperature and many others.

The WSNs can monitor the real-time environment data like temperature, humidity, light

and voltage. In number of application once the nodes are initialized with some amount of energy, it is not possible to recharge them again. Generally the nodes require minimum energy to do any operation. Data generated by the sensor nodes periodically with high temporal redundancy. The redundant data is unnecessary to transmit and wasting energy. Prediction based data sensing and transmission reduce the redundant or unnecessary transmission of data.

To achieve the increase in network lifetime, eliminate the redundant bit in the data and to save the energy, prediction based system is proposed. The proposed system predicts the data depends on the actual data in the sensor node. The predicted value is very close to the actual value, hence it reduce the data transmission. The privacy of the data is achieved through encoding the data before transmission.

II. RELATED WORK

Naveed Ilyasa et.al [01] has presented a novel AUV-aided Efficient Data Gathering Routing Protocol (AEDG) to deliver the data efficiently. These protocols can extend network lifetime, AEDG provides an Autonomous Underwater Vehicle (AUV) to assemble the data from the different gateways and create a criterion to control the number of nodes. The gateway is rotated to balance the consumption of energy.

Ravinesh C. Deo et.al [02] has proposed a system to authenticate a computationally high-speed, simple and efficient non-linear algorithm called Extreme Learning Machine (ELM). In eastern Australia by use of the data from 1957 – 2008, Effective Drought Index (EDI) data is predicated and also considering the monthly data. Mean, highest and least air temperature, rainfall is predicted by the large scale climate data.

Samer Samarah et.al [03] has proposed a system to predicate the data efficiently. The prediction system that builds within the nodes of the sensor and cloud are applied to produce the data. The main aim the proposed system is to free the sensor nodes from transmitting a vast amount of data and thus minimize the consumption of energy of the sensor's battery. It assumes that the data is distributed linearly and the reading is done by n-dimensional space.

Mariam Alnuaimi et.al [04] has proposed a ferry based approach; the ferry based system reduces the need of multi-hop forwarding within the sensing nodes, Ferries tour across the entire sensing field to gather all the sensed data. By eliminating the multi-hop the energy consumption by nodes is reduces or energy is saved particularly to the nodes which are near to base station they are utilized by neighbour nodes to transfer information to base station..

Lahouari Ghouti et.al [05] has a proposed an Extreme Learning Machine (ELMs), it is well-known for universal approximation to model and calculate mobility of random nodes in a Mobile Ad-hoc Network (MANET). The proposed system is compared with the existing algorithm to evaluate the performance, accuracy is increased.

Huang Lu et.al [06] has proposed system to secure the data during transmission. Secure and Efficient Transmission Identity-Based digital Signature (SET-IBS) and Secure and Efficient Transmission Identity-Based Offline/Online digital Signature (SET-IBOOS) are the two algorithms to secure the data for cluster-based wireless sensor network. The SET-IBOOS and SET-IBS, by using the IOOBS and IBS scheme are used.

Yanjun Yao et.al [07] has presented research on Energy-efficient Delay Aware and Lifetime balancing (EDAL) algorithm to increase the network lifetime. An energy efficient delay aware lifetime balancing protocol is proposed to collect the data without any loss in the data packets in WSN. The protocol is based the Open Vehicle Routing Problems with Time Deadline (OVRP-TD). The aim the algorithm is to discover the shortest paths that cover all the source nodes with the minimal cost, with consideration of some constrains of load balancing and the delay required of packet delivery.

Number of research work done on the sensing and prediction of data in WSN. It is clear that the system need to be improved to provide better successful prediction ratio. The proposed system in the section below can provide a better predication ratio with confidential data. The below section describes about methodology of proposed system.

III. METHODOLGY

WSN is one of the key mechanisms to collect the sensor data from the physical environmental. The neighbour nodes with similar characteristics form clusters. The unequal clustering is used to avoid the overload problem in the CHs. CHs are responsible for data collection and transmission in the network. Hence the Hybrid Unequal Clustering Layer protocol is used for the forming the unequal clusters. Based on the load at the sensor node and distance the clustering is done. To gather the data from sensor, Energy Efficient Data Gathering algorithm is implemented. The length constrain is considered to route the data from the CHs to BS.

By the prediction method data redundancy is reduced, both at the source and the sink node. Least Means Square (LMS) and Kalman Filter (KF) are applied for data prediction. LMS algorithm is used for the initial prediction of the data from the actual value. To adjust the coefficient of the predicted value is done through KF. The predicted output values of the KF is compared with the threshold value, if the prediction error is greater than the threshold, the sensor node encodes the data by Blowfish Algorithm (BA) and depends on the Shortest Path Tree (STP) routing protocol the actual

data is send to BS. Fig 2 shows the proposed system architecture.

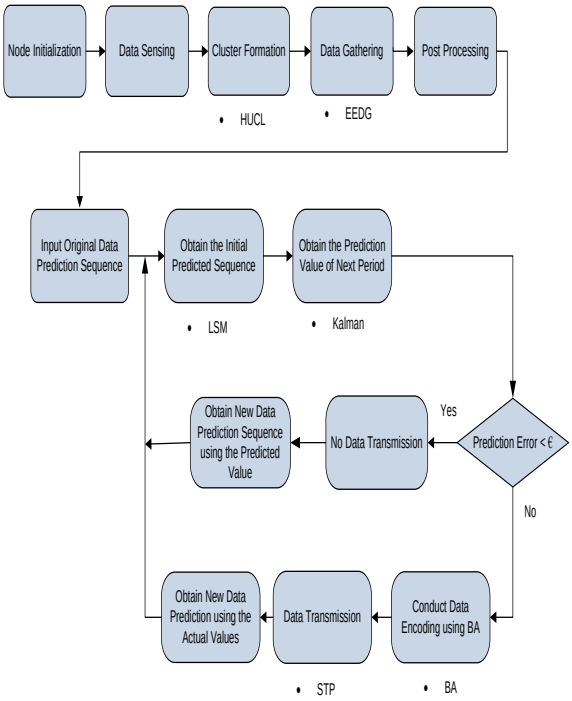


Fig 2: Block diagram of proposed system

A. *Least Mean Square*

The LMS algorithm first introduced by Widrow and Hoff in 1960 is one of the steepest descent methods. The LMS approach is adopted only based on the data error at the current time in order to reduce the mean square error. The Fig 3 shows the LMS prediction algorithm. The LMS algorithm provides low overhead computation and compared with other filters, LMS gives more flexible to use [10].

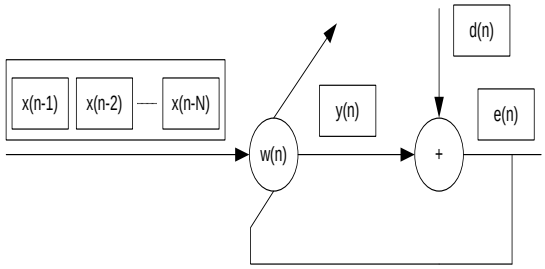


Fig. 3: Least Means Square Prediction Algorithm.

$$X(n) = [X(n-1), X(n-2)...X(n-N)]^T \quad (1)$$

Which is composed of the previous N values from the direct n-1, this gives an input signal to the filter. Y(n) is output which predicts the true value x(n) at n time, can be obtained by

$$Y(n) = W^T(n)\, X(n) \quad (2)$$

Where weight coefficient of the filter is

W (n) = [W1 (n), W2 (n)... WN (n)] T, e(n) is the error between the desired signal d(n) and the output is given by

$$e(n) = d(n) - y(n) \quad (3)$$

N+1 can be updated by next instant by,

$$W(n+1) = W(n) + \mu e(n) X(n) \quad (4)$$

The step size parameter is μ. The Eq. (1) to (4) describes the predication based least means square algorithm [08].

B. *Kalman Filter for data prediction*

The LMS technique is implemented to initial prediction of data with the reference of actual data in the node. The coefficient of the initial prediction has to be adjusted with the help of Kalman Filter. The best recursive filter to estimate the position of a linear dynamic structure from order of noisy calculation is a Kalman filter. It gives high accuracy in prediction depends on the obtainable quantity of data. It has been used to propose adaptive direction mechanisms in WSN. Much research done on the Kalman filter, olfati-Saber has proposed a system with peer-peer continuous-time distributed Kalman filter by make use of the local aggregation of the sensor data but tries to attain a consensus on detecting with other sensor nodes in the network. Yu et done a research on Kalman filter, in that all the sensor nodes can interact with all the neighbouring nodes and the process of filtering can be spared among all the sensor nodes. By utilizing a pinning sensor control scheme, only a fraction of nodes need to measure the destination data. In the system, the Kalman filter is used to approximate the information sequence for all the nodes rather than to select nodes [09].

1) *Kalman Filter Based Prediction Model*

In a sensor node, sequence information from a non continuous time data series, which can be modelled by the following linear stochastic difference Eq. (5)

$$X(K) = A(K)\, X(K-1) + B(K)\, U(K) + W(K) \quad (5)$$

X (K) denotes the data predicted at the time K. A(K) denotes the status of transmission model which is helpful for the information of the last data (K-1). B(K) denotes the control-input model applied to the control vector U(K). Noise at the prediction time is given by W(K) , which is believed to follow a zero mean multivariate normal distribution with the covariance Q(k).

In the Eq. (6) Z(k) represents the true sensed information at the time K,

$$Z(k) = H(K)X(k) + V(k) \quad (6)$$

The study model H(k) which maps the information predicted space into the true data sensed.th noise V(k) assumed to be Zero Means Gaussian white noise with covariance V(k).

2) *Kalman Filter Based Prediction Algorithm*

Kalman filter designed with two phase: Update and prediction model. In the prediction phase the information predicted from the previously sensed data and to generate the data for the current period.

The covariance model and prediction model are given by the Eq. (7) and Eq. (8). Update phase, measurement data at the present period is utilized to refine this prediction to achieve a new, extra accurate data prediction and continued for the current period. The updated prediction model and its covariance model are given by the Eq. (9) and Eq. (10). Eq. (11) is used to compute the gain of the Kalman filter.

$$\hat{X}(K+1|K) = A(K)X(K|K) + B(K)\,U(K) \quad (7)$$

$$P(K+1|K) = A(K)\,P(K|K)A(K)^T + Q(K) \quad (8)$$

$$\hat{X}(K+1|K+1) = \hat{X}(K+1|K) + Kg(K+1)\big(Y(K)\big) - H(K+1)\hat{X}(K+1|K) \quad (9)$$

$$P(K+1|K+1) = \big(1 - Kg(K+1)H(K+1)\big)P(K+1|K) \quad (10)$$

$$Kg(K+1) = P(K+1|K)H(K)^T[H(K+1)P(K+1)K)H[K+1)^T + R(K)]^{-1} \quad (11)$$

$\hat{X}(n|m)$ gives the perdition of X at time n, given sense information series of recent m times. P(n|m) denotes the error covariance matrix according to $\hat{X}(n|m)$. To predicate the temperature sensor data Kalman filter is considered as a single measurement of a single model. A(k)=1, (k)=Q(k)=0, R(k)=H(k)=1,andP(0j0)=1.Let Y(k) ={y(1, Y(2),----Y(k)} denotes the past sensed information sequence.

Algorithm: Kalman Filter for Data Prediction
Input: Previous Data Value
Output: Next Sequence of Data

Step 1: Start
Step 2: Initialize state estimation from first observation
Step 3: Prediction for state vector and Covariance.

$$\hat{X}(K+1|K) = A(K)X(K|K) + B(K)\,U(K)$$
$$P(K+1|K) = A(K)\,P(K|K)A(K)^T + Q(K)$$

Step 4: Compute Kalman gain factor

$$Kg(K+1) = P(K+1|K)H(K)^T[H(K+1)P(K+1)K) * H[K+1)^T + R(K)]^{-1}$$

Step 5: Correction abased on observation.
Step 6: Stop.

3) Implementation of the Blowfish Algorithm

The LMS and Kalman Filter are two best methods implemented to predicate the next sequence of data in the WSN. LMS algorithm predicts the data with the input of actual data of the sensor. The Kalman Filter takes the input data from the LMS to predict the second sequence of the data. The predicted data is validated by comparing with the threshold value$\in$. The rate of error is less than the predict data rate, the data is encrypted by Blowfish Algorithm.

All the nodes in the wireless sensor network have a unique secret key. At the beginning session key is generated by the sink node and transmitted to all the nodes in the network. The sensor nodes will compute Needham–Schroeder Symmetric Key (NSSK) with the unique secrete key and also with the session key for encrypting the data and to decrypting the sensor nodes. Because of the sink node knows all the unique secrete keys of the sensor nodes, it might compute the NSSKs to decipher the information. Data transmission between the sensor nodes and the sink nodes use Blowfish algorithm for encryption. The functional flow of the data prediction is as shown in the Fig.4

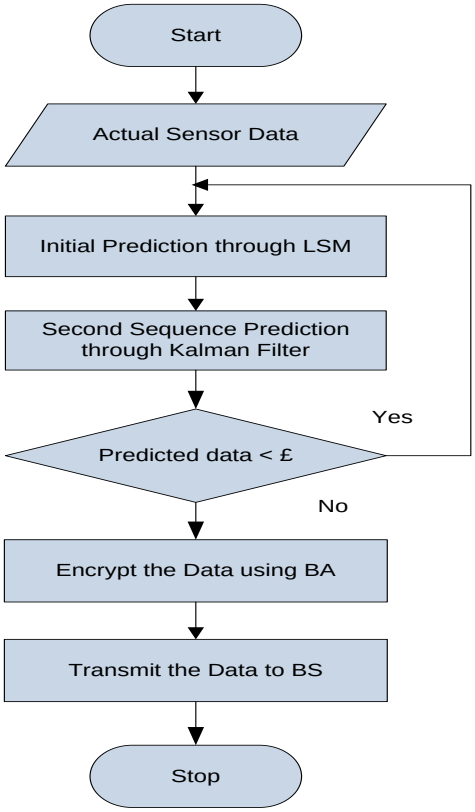


Fig.4: Functional Flow of Sensor Data Prediction in WSN.

BA is designed by Bruce Schneier which can be available in the open domain. BA is introduced in 1993 for the first time, it has not yet split yet, Blowfish, a 64bit block cipher is an excellent choice for data encryption of sensor data, because it is a lightweight, open domain and secured even after wide analysis. Blowfish algorithm can be implemented by two ways via hardware or software implementation. While compare both hardware and software implementations, hardware implementation of BA has many advantages.

The basic block diagram of the blowfish algorithm is as shown in the Fig. 5

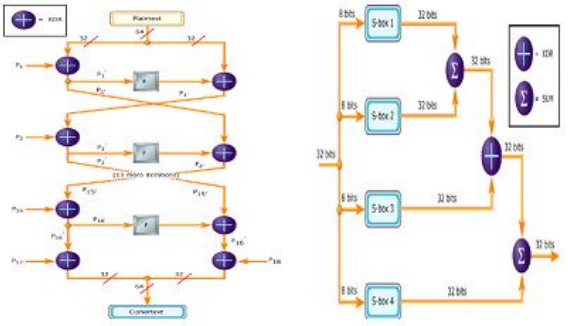
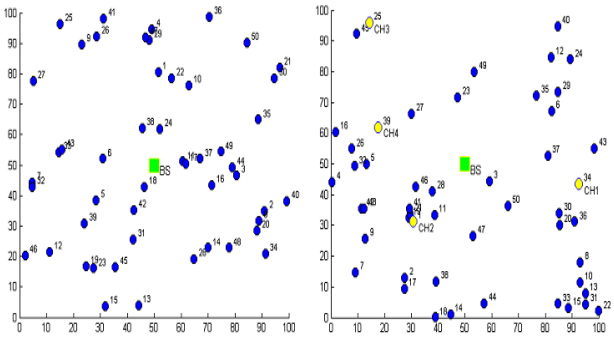


Fig .5: Block Diagram of Blowfish Algorithm

There are two phase in blowfish algorithm data encryption and key Expansion as shown in the Figure 5. P-array and S-box starts key expansion by deployment of many number of sub-key dependent permutation and a key-dependent and key dependent data-dependent replacement are conducted in the entire round. Each round includes XOR and added extras on 32-bit words. The more compound part of this algorithm is F-function because it is the lone part which make use of the S-box. From the recent research, to simplify the complexity of processing, a novel f-function was calculated to create dynamic S-box and XOR operator and also many advanced methods are introduced to create S-box and P-array.

IV. EXPERIMENT RESULT

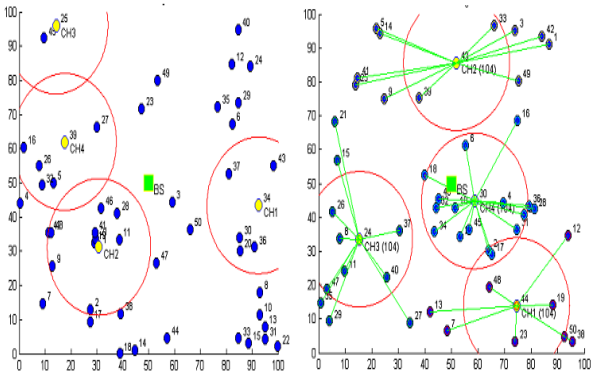
The proposed system simulation results are explained in this section. To simulate the proposed system, MATLAB 2012a tool is used. The network consists of 50 nodes with the initial energy of E_o. At the centre of the network BS is initialized shown in the Figure 6 (a). The neighbour nodes which exhibit more similar characteristics are combined into form a group called clustering. The clusters with the uneven number of nodes called unequal clusters. The CHs are selected as shown in the Figure 6 (b).



(a) Node Initialization **(b) CHs Selection**

Fig 6: Sensor Nodes Initialization, BS and CHs Selection

The neighbour nodes are selected by considering the distance parameter. The neighbour nodes are selected to form a cluster as shown in the Figure 7 (a). Once the clustering is done, load at each node is collected to determine the mean of the load as shown in the Figure 7 (b). If the load in the CHs are not balanced, Re clustering of the node is done to balance the load at each CHs. Number of iteration are conducted to balance the load at the CHs.



(a) Cluster Formation **(b) Load Calculation**

Fig 7: Cluster Formation and Load Balancing

While gathering the data, the node which is having low energy transfers the data to its neighbour node before node die as shown in the Fig 8 (red node). The Intel Berkeley research lab dataset is considered as a load. Different physical environment sensing sensor data is considered (Temperature, Humidity, Light and Voltage). The actual data of the sensor nodes are considered for the prediction of next sequence.

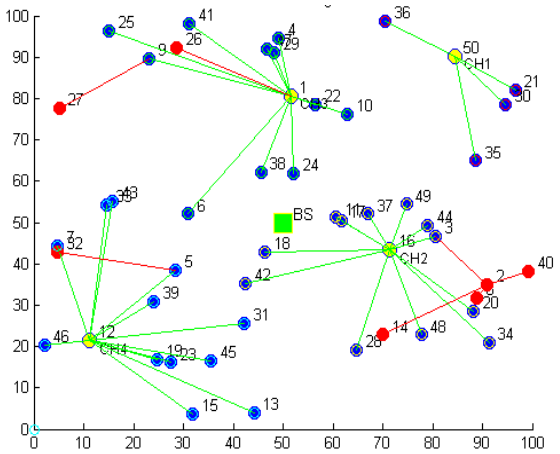


Fig .8: Data Gathering by CH

After the data gathering form all the member nodes by the cluster head, thereafter forward it to the base station. The data transmission form all the CHs to the base station is as shown in the Fig 9.

The actual data is encoded before transmitting to base station. If the predicted data is almost equal to actual data or the prediction error acceptable than there is no need to transmit the data to base station, this avoids the data redundancy in the network. The predicted data from the actual data are evaluated by the performance metric successful prediction rate and the packet delivery ratio.

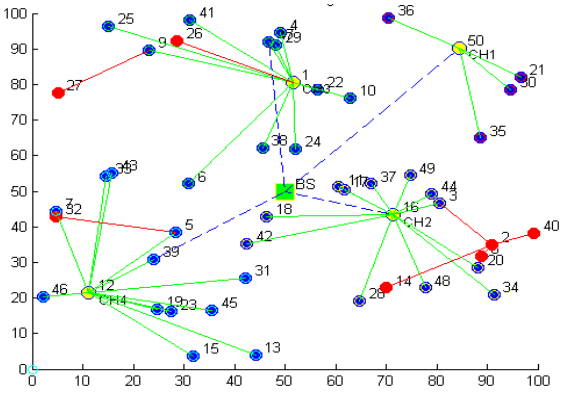


Fig 9: Data Transmission to BS

The transmission from all the CHs to the Base station is shown by blue dotted lines.

The actual data is encoded before transmitting to base station. If the predicted data is almost equal to actual data or the prediction error acceptable than there is no need to transmit the data to base station, this avoids the data redundancy in the network. The predicted data from the actual data are evaluated by the performance metric successful prediction rate and the packet delivery ratio. Fig 10 shows the data prediction sequence with the actual data.

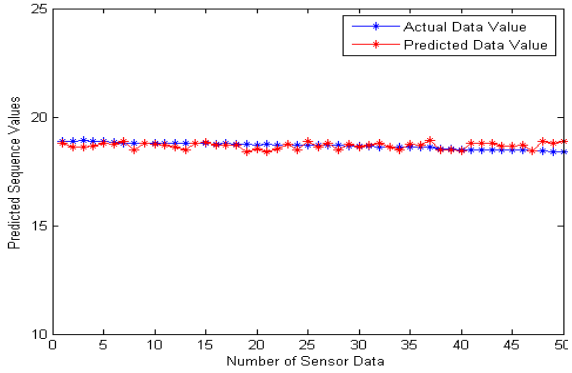


Fig 10: Data Prediction

The energy consumed by the nodes to perform data transmission is as shown in the Fig 11.

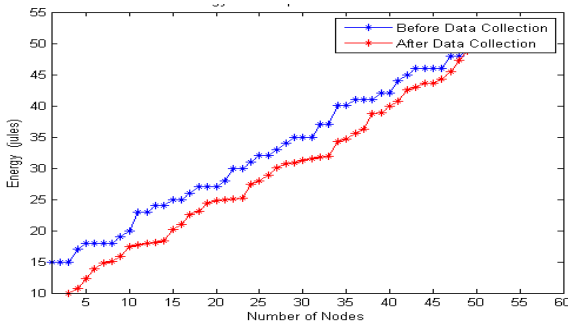


Fig 11: Energy consumption

The proposed two layer prediction system using LMS-Kalman filter has been compared with the existing GM-KRLS [11]. The successful prediction ratio is computed by taking the difference of the predicted and actual value.

SPR = Predicted value – actual value (13)

The SPR is computed for all the four sensor nodes which we have been selected for the test purpose. The SPR computed for of them is as shown in the Fig 11. The comparison Table 1 shows the comparison of existing method and the proposed method

Table 1: Existing and Proposed Model Performance Comparison Table

S.No	Author	Methods	SPR
1	Luo X [11]	GM-KRLS	43 %
2	Proposed model	LMS-Kalman Filter	90 %

The comparison Table 1 shows the comparison of existing method and the proposed method, the SPR is computed for all the four sensor data with different threshold values as shown in the Fig 12. The existing system SPR of light sensor data is considered with threshold value of 0.5 for comparing the proposed system as shown in the Fig 12.

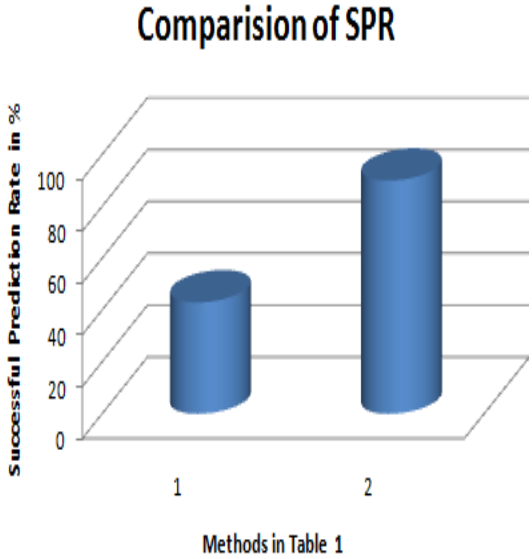


Fig 12: Comparison Graph for Successful Prediction Rate.

The Packet Deliver Ratio (PDR): It is the ratio of original data received to the data transmitted and is given by the Eq. (12) and shown in the Fig 13.

$$PDR = \frac{\text{actual packets received}}{\text{total number of packets transmitted}} * 100 \quad (12)$$

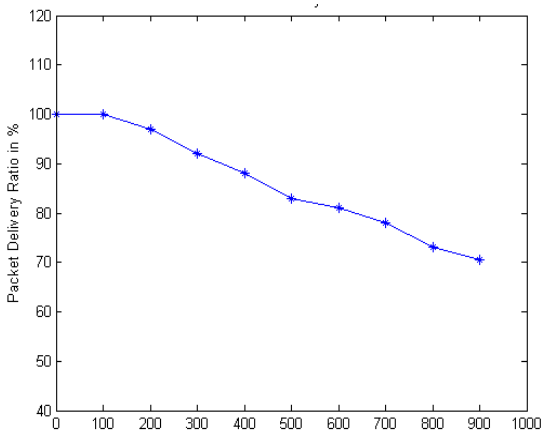


Fig 13: Packet delivery ratio

V. CONCLUSIONS

In WSNs load based clustering and prediction based data transmission are the effective way to reduce the redundant data. The paper presents the Hybrid Unequal Clustering Layer (HUCL) to balance the load through unequal clustering. The residual energy is considered to select node as a CHs. All the CH in the network consists of equal amount of load with uneven number of member nodes. To route the data to sink node, Energy Efficient Data Aware (EEDA) protocol is approached. The actual data at the sink node is considered to predicate the next sequence. The initial value is predicted by Least Square Mean (LSM) technique with the actual value. The predicted value coefficients are adjusted with Kaman filter. The predicted data is encoded with blowfish algorithm (BA) to transmit the data confidentially. To evaluate the performance of the proposed system, publicly available data at Intel Berkeley lab is considered. Packet delivery ratio and successful prediction rate is increased with the reduction of data redundancy. Thus reduces the energy consumption in the node and increase the lifetime of the network.

REFERENCES

[1] [1] Naveed Ilyasa, Turki Ali Alghamdib, Muhammad Nauman Farooqa, Bilal Mehbooba, Abdul Hannan Sadiqa, Umar Qasimc, Zahoor Ali Khand and Nadeem Javaida, "AEDG AUV-Aided Efficient Data Gathering Routing Protocol for Underwater Wireless Sensor Networks", Procedia Computer Science. Elsevier, Vol. 52, pp. 568-575, 2015.

[2] Ravinesh C. Deo and Mehmet Şahin, "Application of the Extreme Learning Machine Algorithm for the Prediction of Monthly Effective Drought Index in Eastern Australia", Atmospheric Research. Elsevier, Vol. 152, pp. 512 – 525, 2015.

[3] Samer Samarah, "A Data Predication Model for Integrating Wireless Sensor Networks and Cloud Computing", Procedia Computer Science. Elsevier, Vol. 52, pp. 1141 – 1146, 2015.

[4] Mariam Alnuaimia, Khaled Shuaiba, Klaithem Alnuaimia and Mohammed Abdel-Hafez, "Ferry-Based Data Gathering in Wireless Sensor Networks with Path Selection", Procedia Computer Science. Elsevier, Vol. 52, pp. 286 – 293, 2015.

[5] Lahouari Ghouti, Tarek R. Sheltami and Khaled S. Alutaibi, "Mobility Prediction in Mobile Ad Hoc Networks Using Extreme Learning Machines", Procedia Computer Science. Elsevier, Vol. 19, pp. 305 – 312, 2013.

[6] Huang Lu, Jie Li, and Mohsen Guizani, "Secure and Efficient Data Transmission for Cluster-Based Wireless Sensor Networks", IEEE, Vol. 25, No. 3, pp. 750 – 761, 2014.

[7] Yanjun Yao1 and Qing Cao, "EDAL: an Energy-efficient, Delay-aware and Lifetime-balancing Data Collection Protocol for Wireless Sensor Networks", Mobile ad-hoc and sensor systems (MASS), 2013 IEEE 10th international conference. IEEE, Vol. 23, issue 3, pp. 810 – 823, 2015.

[8] Mou Wua, Liansheng Tan and Naixue Xiong, "Data Prediction, Compression, and Recovery In Clustered Wireless Sensor Networks For Environmental Monitoring Applications", Elsevier, Vol. 329, pp. 800-818, 2016.

[9] Guiyi Wei, Yun Ling, Binfeng Guo, Bin Xiao and Athanasios V. Vasilakos, "Prediction-Based Data Aggregation in Wireless Sensor Networks: Combining Grey Model and Kalman Filter", Elsevier, Vol. 34, pp. 793 – 802, 2011.

[10] Liansheng Tan and Mou Wu, "Data Reduction in Wireless Sensor Networks A Hierarchical LMS Prediction Approach", IEEE, Vol. 16, No. 6, pp. 1708 – 1715, 2016.

[11] Luo X, Zhang D, Yang L T, Liu J, Chang X and Ning H, "A Kernel Machine-Based Secure Data Sensing and Fusion Scheme in Wireless Sensor Networks for the Cyber-Physical Systems", Future Generation Computer Systems. Elsevier, Vol. 61, pp. 85-96, 2016.