

Data Redundancy Reduction in IoT Weather Station

Praveen K^{1*}, Rajalakshmi K², M. Malathi², Dhanagopal R⁴

^{1,3,4}Department of ECE, Chennai Institute of Technology, Kandrathur, Chennai,

²Department of CSE, Chennai Institute of Technology, Kandrathur, Chennai.

¹praveenjere@gmail.com, ²rajalakshimik2citchennai.net, ³malathiyoga08@gmail.com, ⁴dhanagopal.phd@gmail.com

Abstract

Portable weather station allows us to track fluctuations in temperature, humidity, wind speed, wind direction, rainfall and barometric pressure to create a prediction of forthcoming weather conditions. Weather forecast helps farmers/gardeners plan for crop irrigation & protection, helps power generation industry to plan & optimize power production & utilization, and people to plan their outdoor activities. The scope of this work is to retrieve the above said environmental parameters at a place of interest and then remove the redundancy in the retrieved parameters before uploading them in to the IoT Cloud. Technical experts have predicted that 44 zettabytes of data generation expected in 2020, data storage optimization is one of the biggest challenges that are ahead of us, this work aims at reducing the redundant data and to use the storage space efficiently, by doing so the traffic and congestion at the cloud layer can be averted and also the memory requirement can be reduced where unprecedented data are stored.

Keywords: Internet of Things (IoT), Embedded Computing System, Cloud Services

1. Introduction

Intelligent systems capable of assimilating environmental data and man-made utilities, to improve efficient use of resources are required for sustainable development. The ability of the weather station is to create ambiguous future a little more certain. It is frequently used to measure the environment conditions with the support of sensors and computing system. Typical measurements include temperature, atmospheric pressure, humidity, wind speed, wind direction, and drizzle amounts. As regional weather may vary at the micro level, it is convenient and profitable for numerous businesses and industries to have on-site, reliable weather data. Some areas that require such data are power production and distribution industry, agriculture, mining, etc. The agricultural operations are achieved by measuring local weather conditions through weather stations. Real-time, local weather monitoring mechanisms may provide appreciated information for process control, employee security, incident response, conservational temperature, wind loads, and rainfall volume. Environmental factors like heat stress, heavy rains, winds, and lightning must be closely monitored in the mining industry to guarantee personnel protection and comfort of processes. Being able to have ambient weather circumstances readily accessible can help make quality decisions in everyday responsibilities in the field of forestry, assisting with choices nearby replanting, tree care/thinning, firefighting, erosion control, habitat improvement, and recreation maintenance. India's search for green energy sources has intensified, due to the ever increasing demand for energy, adverse effects of fossil fuels on climate & health and growing awareness on the ill effects of pollution. Renewable energy sources contribute to only 2.2% of the total energy distribution of India. Optimization of the generation of power from renewable energy & its distribution would contribute to the economy, reduction of pollution and mitigation of adverse effects of climatic change. The

initiatives of the Government of India like the Electricity Act, 2003 makes Renewable Purchase Specifications (RPS) mandatory for all states in order to increase the generation of electricity from renewable energy sources and The Jawaharlal Nehru National Solar Mission (JNNSM) seeks to take advantage of the enormous possible of solar power as a future energy source in the country. Power generation through wind turbines and wind mills can be made more effective and efficient, if the parameters like wind velocity and directions are known. Efficiency of power generation from solar energy and hydro energy can be improved by proper and accurate prediction of the rainfall. Portable weather station can be used to accomplish these.

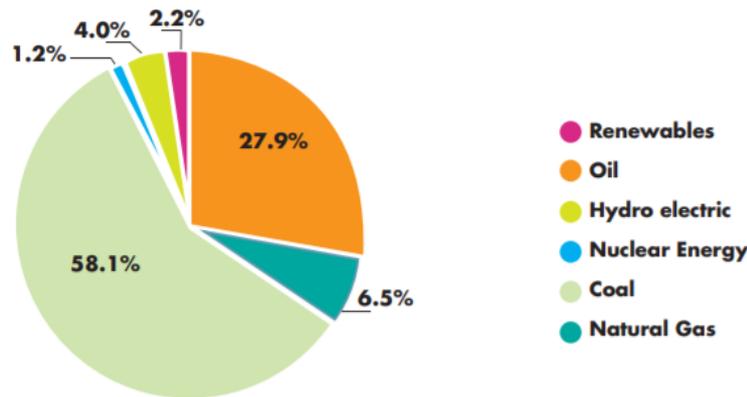


Fig. 1. The Energy Distribution of India

Another area of concern is the adverse effects of air pollution caused. The foremost cause for anthropogenic activities on climate and health is adverse effects of air pollution. A contemporary report from the World Health Organization (2016), based on quantities from air monitoring stations in 4,300 cities across the world begins undoubtedly that air pollution is a global problem. A whopping nine in 10 people on Earth breathe highly polluted air, and additional 80 percent of urban dwellers have to bear outdoor pollution that surpasses health standards, conferring to the WHO's World Global Ambient Air Quality Database. The Adulteration level in the air is reduced via Constant pollution monitoring. Effective utilization of environmental forecasted variables help a lot to improve the productivity and reduce the losses in agricultural applications [1,2,3] knowing the forecasted data on behalf of applications such as photovoltaic plants and micro-grids, helps improving the amount of solar harvest [4,5,6], also predicted data play a vital role to enhance the effectiveness of greenhouse environmental control and irrigation management [7,8]. Predictions of solar radiation, air temperature and relative humidity diminishes the redundancy and in turn increases the energy savings, forecasts of solar radiation decreases the threat of temporary depletion and improves their efficacy [9]. Weather monitoring station usually deploys sensor units, data acquisition & processing units, storage unit and a display. The measured parameters are periodically logged into local storage units for data analysis. Restrictions on the number of nodes, restrictions in coverage area, high data redundancy and limited storage are some of the main challenges in such systems. In IoT based weather monitoring, measured parameters are stored in IoT clouds and retrieved for processing without any constraints on distance, place or time.

2. LITERATURE SURVEY

[10] In this paper, fog computing has been recommended as a resource which makes it possible or enables the internet of things domain. It is said that this fog computing has a very low delay period in the end of the networking. In this they have suggested the use of

this fog computing to make it possible for the consumers of the internet of things in a hierarchical model of computing. They have designed an offloading game. This offloading game in turn helps to effectively allocate the required resources of fog computing to the users of IoT. It is also said that through this method, the equilibrium can also be attained in a very short period of time. By this methodology that is proved, it is also seen that it is more beneficial than the other existing methodologies. This offloading game is based on the Q- maximization problem. In this, it is also seen that the computation speed from the IoT devices has been increased and also the delay has been significantly decreased. They have also mentioned that they are on their way to build an online mechanism for the same purpose as their future work. [11] In another reference, they have proposed a double-matching approach. This strategy is suggested for overcoming the problem of the resource sharing in the field of the fog computing networks. This also oversees that the cost is managed effectively and efficiently. This strategy is said to be the next step of the acceptance algorithm. By this strategy which is mentioned above, it is seen that a minimum cost result can be obtained which is at the same time effective too. In this strategy it is improved from two participants to three participants. This is seen that it is effectively used amongst the CDCs. As a result, it is also seen that the most minimum cost involving great results can be obtained by this kind of DADMS strategy. As a future work, it is also seen that this can also be extended to aerial vehicles also. [12] In a similar reference, which also uses fog computing, it has also combined cloud computing with the fog computing for the internet of things. Generally data which is generated from the IoT devices is found to be extensively large. This data cannot be stored in the devices which are present at the end or it can be said that it is not wise to store in the devices because of the limitation of the storage space. These difficulties or the challenges are addressed in this paper by using different techniques. For this they have combined or in other words they have integrated the fog computing technique and the cloud computing for overcoming these difficulties and challenges. Over here, it is designed such that the data that are generated from the IoT devices are directly redirected to the cloud server instead of the local storage. By this, the data generated can be seen to be stored and secured effectively relieving us from the problems of data insecurity and unavailability of the storage space. [13] In a paper, there is a proposal which consists of the technique of fog computing which is used in a multi-tier format for the data analytics process of the IOT devices that are used in large scale in the cases of the emerging smart cities. It is seen that the data that are generated from the sensors of the IoT devices is said to be an important criteria for the smart cities. In this method of fog computing, it is seen that the resources are seen to be very dedicated. Therefore this fog computing model is seen to be very effective and resourceful for the building of the smart cities. In this, experiments were conducted on ad-hoc as well as the fogs to care the QoS aware research management schemes. These experiments have shown that this method is very efficient for the smart city scheme. [14] In this paper, they have come up with a solution for checking the health of the patients who live in the smart home. For this, they have used the technology of the internet of things. The data that is generated by this is found to be redirected to the cloud. But because of this, there still lies a latency in the data transfer and the other processes. As the previous cases, here also they have made the usage of fog which has been used as an assistant for the IoT. For the reason of increasing the speed they have used a gateway. By this, it is seen that this method is more effective for remote data sensing without any time delay. [15] In a supportive response, they have initiated a method or a technique called indie fog which can be considered as a successor of the fog. It is because the limitations of the fog technique are overcome by this method. That is the disadvantages like the network availability, etc., are taken over by this method. This poses a serious advantage of this method for working with internet of things technology. By this this proves to be an efficient and effective infrastructure. [16] In another paper, the challenges and solutions like the security problems, etc., that are faced by this methodology are

addressed in this paper. So by this it has proposed a method for securing the fog. There are many challenges that arise in this method like the latency, increase in data are proving to be a considerable disadvantage to this effective system. This discussion performed in this paper gives us a clear idea of the challenges and its solutions prevailing in this area. By this a clear work can be performed by knowing the above mentioned in the near future.[17] A reference on the Internet of things clearly shows the analysis of the performance as well as shows how a weather forecasting system works in real-time which is also adaptive at the same time. In this, they have used the technique of machine learning. In the domain of machine learning the concept of KNN that is the k's nearest neighbor is put into use. Along with this many other algorithms have also been tested. By this only an error of around seven percentage is seen which is much lower than the error which is generated out of the non-adaptive ones which have a percentage error around fourteen percent.

3. EXISTING SYSTEM MODEL

Sensors measures the numerous parameters from the environment and it can be linked to data logger. The data loggers have a precise a lesser amount of storage capacity and are powered by power supplies. Conventional data loggers do not have provision of modularity reprogramming of the devices are not possible, to include control actions and prediction. The acquired parameters can be programmable in some data loggers. Prior to the usage of wireless communication, data loggers shown in fig (2) would be detached from the unit and connected to processor. With the advancement in technology, a few weather stations have a integrated interface wirelessly or as an optional part. IEEE 802.11 or the IEEE 802.15.4 standards are used for the wireless part, Programmable wireless communication interfaces such as the NPort Z3150 (Moxa, Brea, CA, USA) [18] or the Waspnote (Libelium, Zaragoza, Spain) [19] can be used if a on board interface is not available.

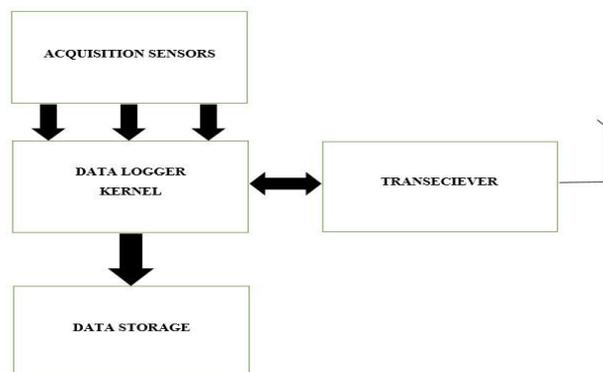


Fig. 2. Data Logger

Intelligent weather station measures the Solar radiation, temperature and relative humidity is described in [20]. These measurements are completed at a definite time interval defined by the user; it also forecasts the development of each variable in a prediction horizon up to 48 steps-ahead. The measurement sampling time must agree with prediction step or it can be a multiple of it. Further for every variables average value is calculated, over the prediction step, finally the prediction step is 5 min and sampling time interval is 1 min sampling time interval is 1 min and the main objective of the system is, it may be a independent, regarding electrical energy. In addition, to enable its arrangement, it integrates wireless communication based on the wireless IEEE 802.15.4 [21] standard. The functional

diagram for ZigBee based wireless network[22] is shown in the fig 3

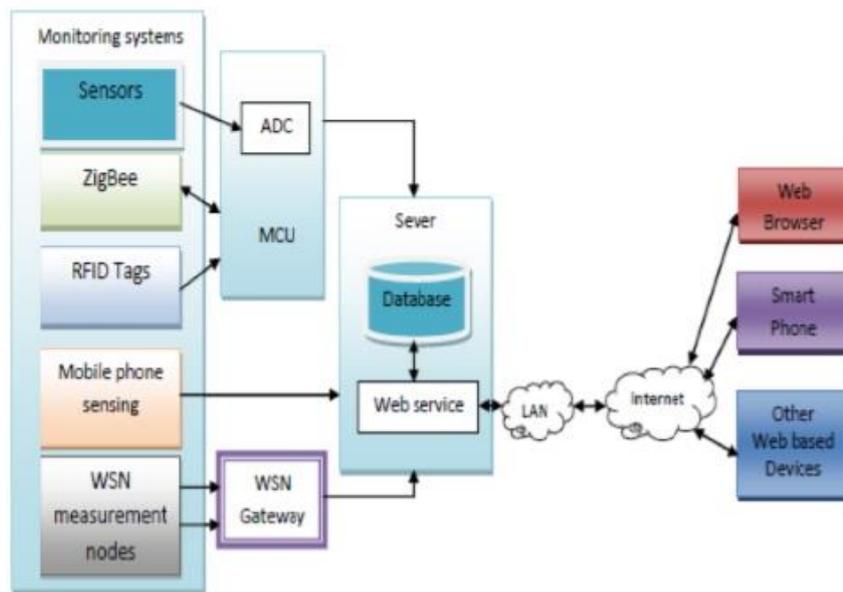


Fig. 3. Zig bee based wireless Sensor Network

RFID stores and retrieves data to and from an RF compatible IC through electromagnetic transmission. Tracking and labelling the items in shops and factories is done through RFID. It comprises of tags and readers. A tag has a unique identification (ID) number and a memory that accumulate data like environmental factors such as temperature, humidity, etc. Through reader writing data to and reading data from tags through wireless transmissions is possible. Usually, tags are devoted into objects that are needed to be identified or tracked. Established on the power source tags are categorized into: active tags, passive tags, and semi-active tags. Mobile phone sensors that are used in urban areas are categorized into two types: participatory sensing and opportunistic sensing in the prior user is directly involved, further it id]s not in use since the constraints on static information and mobility restrictions. Cheap cost sensors capable of retrieving, storing and processing environmental information are used in Wireless Sensor Networks They are also capable of communicating with neighboring nodes [22].

It is more convenient to receive the data via WSN gateway at any instant and place. The gateway takes care of node authentication, message buffering. End device router, gate way node and monitoring center makes Wireless sensor network management model. The end node has the ability to accumulate the wireless sensor network data, further it send them to parent node. Next the data transmitted to gateway node from parent node by way of router. The gateway node eliminates data after scrutinizing and wrapping them into Ethernet format data forwarded to the server.

4. PROPOSED MODEL

The foremost scope of this work is to diminish the redundancy present in the acquired data, Redundancy removal is done at the stage just before uploading the data to the cloud and hence overloading at gateway is averted and also processing

time is minimized, The proposed model consists of five different units, Main Processing unit, Sensory Units, IoT Gateway Communication Protocol, IoT Cloud Interaction

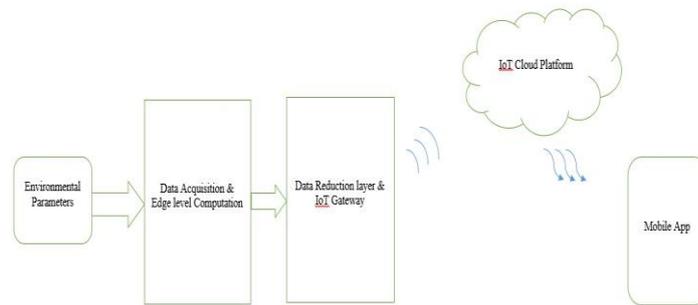


Fig. 4. Functional Blocks of Smart Weather Station with Redundancy Removal Technique

i) Sensory Units:

The environmental parameters are retrieved by high accuracy and reliable sensors , It includes Temperature, Humidity , Absolute Pressure, Relative Pressure, Air Quality , and wind velocity. Capacitive humidity sensor helps to measure the moisture content in the air and equivalent digital value send to data pin. The absolute pressure of the surrounding air can be dignified using Barometric pressure sensors

ii) Data Acquisition and Edge Level Computational Unit

It regulates the usefulness and exactness of the entire system, it should have enough ports (serial, parallel and analog ports)to interface all the required sensors. Arduino is preferred in this work because of its ready to use structure, and compliance with gateway nodes, moreover many libraries are available to incorporate sensors with it.

iii) Data Reduction Layer &IoT Gateway

he connection point among the cloud and controllers, sensors are established via an Internet of Things (IoT) gateway. It may be implemented with the help of physical device or software program. Intelligent gateway or a control tier is the alternative name of Internet of Things (IoT).

The communication stack overhead time can be reduced through aESP32. It is a complete standalone system as a device to host MCU. It provide Wi-Fi and Bluetooth functionality to other system through its SPI / SDIO or I²C / UART interfaces

iv) Communication protocol:

In this work, IEEE 802.11b/g/n protocol based Arduino WiFi shield has been provided since of its simplicity of installation and compatibility. This is a 2.4 GHz Ultra High Frequency (UHF) connectivity it's a appropriate communicating protocol which accommodates the connectivity to a radius around 100m WiFi access point or router to get permitted to internet, Physical Arduino WiFi shield is used for his purpose.

V) IoT Cloud Platform

Cloud platforms preferred for this investigates is Thing Speak, Application

Programming Interface (API) based interconnectivity established with recommended system. API is a set of routines, protocols, and tools for building software applications exclusively in cloud platforms.

This supports the developer to correlate the cloud services with the hardware for data visualization, data storage, data analytics, and triggering purposes. Plotly has made the study of graphical plotting very easy, whereas Thing Speak offers the plug-in based capability to screen the current value of MISSENARD index.

5. REDUNDANCY REDUCTION ALGORITHM (OVERVIEW)

Step1: Initialize the Sensors and Ports

Step2:

```
if(sensors_status == not working) // check whether the sensors
```

```
{
```

```
    Notify the status through Terminal
```

```
    While(sensors_status== not working)// wait until the sensors work
```

```
}
```

Step3: Acquire the environmental parameters

Step4: Processing and normalizing the data

Step5: Check for the Redundancy by comparing with the previously stored value

```
    if ( Present value != Previous value)
```

```
    {
```

```
        Open serial port &
```

```
        Upload the data to IoT Cloud
```

```
    }
```

```
    Else
```

```
    {
```

```
        Do not upload the value to the cloud
```

```
    }
```

Step6: wait for a while// Delay before acquiring next set of parameters

Step7: Repeat from Step no 2 to Step no 7

V. Computational Analysis on Environmental Parameters

Parameters recorded by the Smart Weather Station are given in Table II, along with typical range of values. These raw text files were generated every week.

Table 1. Typical environmental parameter ranges

	Temperature	Humidity	Absolute Pressure	Relative Pressure
Resolution	1°C	1%	0.01 hpa	0.01 hpa
Range	-40°C to 65°C	1 to 100%	300-1100hpa	300-1100hpa
Accuracy	± 2°C	3%	0.12 hpa	0.12 hpa
Update Interval	1min	1min	1min	1min

Redundancy Removal Temperature

The plot was taken for the duration of 30 minutes, It is evident from the graph that most of the time the temperature is constant and there is no drastic change, Temperature lies in the range between 28 degree Celsius and 29 degree Celsius throughout the period of observation, Fig 4. Shows Instead of uploading the redundant data to the IoT Cloud, Temperature should be uploaded only when there is significant or moderate change.

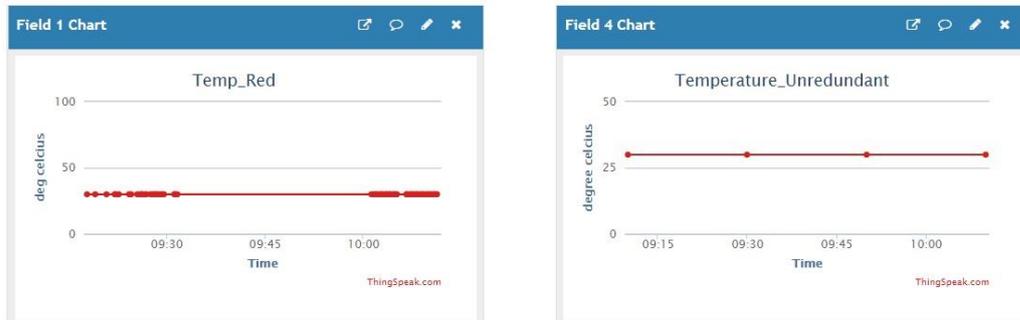


Fig. 5. Redundancy chart for temperature

Redundancy Removal Humidity

Water content in the air is one of the important parameter, that can be used to predict the rainfall, in industrial environment also it plays a major role, typically it ranges from 40-60 % during sunny days and 75-100% during rainy days, graph depicts that this parameter is also slowly changing throughout the observing test period, if humidity had to be uploaded every minute , then 30 entries are needed for the observing duration of 30 minutes, if the discernment humidity percentage is set to 3% , considerable reduction is possible in the entry, from the Fig. 5 it is seen that only few entries are sufficient to represent all data during the entire time range.



Fig. 6. Redundancy chart for Humidity

Redundancy Removal Pressure

Even though there are abrupt changes seen in the Fig. 7 for both absolute and relative pressure, they (most of the variations) are within the tolerable range (10hpa), These data are considered as redundant data, the data type used for pressure is “double” which means for single entry 8 bytes are needed ,for 30

entries 240 bytes are needed to be uploaded, if the redundancies are removed only 10 to 12 entries are needed ,it means 80 to 92 bytes,



Fig. 7. Redundancy chart for Pressure

The following table 2 shows the redundancy table for the various environmental parameters

Table 2.Redundancy table for various environmental parameters

	Temperature	Humidity	Absolute Pressure	Relative Pressure
No of Entry for sample Period (30mins)	60	60	60	60
No of Bytes/Entry	2	2	8	8
No of Bytes for the Sample period with out Redundancy	2	8	304	256
No of Bytes /Day with Redundancy	5760	5760	23040	23040
No of Bytes /Day without Redundancy	96	192	14784	12288
Compression Ratio	60:1	30:1	1.55	1.875
Space Saving	98.3%	96%	35.8%	46%

The following Fig. 8 shows the compression ratio and amount of space saving for various environmental parameters

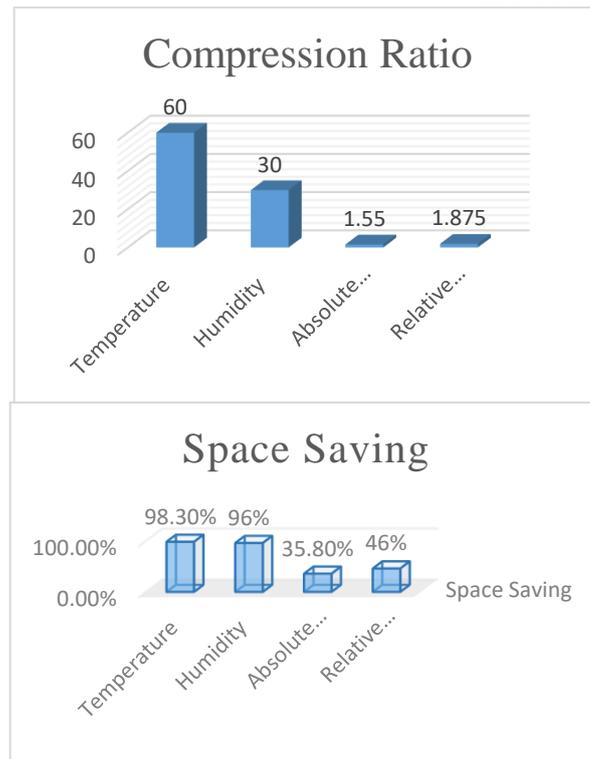


Fig. 8. Compression Ratio and Amount of Space Saving for Various Environmental Parameters

It is evident from the graph in Fig. 8 that there is a significant amount of memory have saved by reducing the redundant data at the edge level itself, Notice that Sine Temperature and Humidity changes very slowly the space saving is very high for these Parameters

6. CONCLUSION

In the suggested architecture functions of different modules were discussed. The weather monitoring system with Internet of Things (IoT) concept experimentally tested for observing three parameters. It also directed the sensor parameters to the cloud (Google Spread Sheets). This data will be useful for upcoming analysis and it can be easily shared to other end users. The thoroughreport of the environment and to make robust decision a number of observations can be composed and joined from several nodes. During this collection of data only some are useful the other are reduced. The performance of the system may be affected due to this data redundancy. In applications like weather station the frequency of uploading the data to cloud is very often, and most of the time the data are redundant, The redundant data removal could be done at cloud layer but that might overload the cloud and may lead to congestion and performance degradation, so in this work redundant data are removed at the edge level instead of cloud, this increases the overall performance in terms of reduced latency, minimized data of the IoT Application. In future depending on the application and with the help of degree of impact the size of reduction can be increased further. Hence the implemented weather station with data redundancy the proposed system avoids the redundancy to develop the overall performance of the system. In future this kind of weather station helps the former in the agricultural

areas for irrigation system. This model offers an efficient and low cost solution for continuous monitoring of environment.

REFERENCES

1. Das M, Ghosh SK 2014, A probabilistic approach for weather forecast using spatio-temporal inter-relationships among climate variables Gwalior, India *Proceedings of the (2014), 9th International Conference on Industrial and Information Systems (ICIIS) 15–17 December*,
2. C. D. Scott and R. E. Smalley, “Diagnostic Ultrasound: Principles and Instruments”, *Journal of Nanosci. Nanotechnology*, vol. 3, no. 2, (2003), pp. 75-80.
3. Ferreira PM , Ruano AE (2008), *Proceedings of the 17th IFAC World Congress 11 July*, Discrete modelbased greenhouse environmental control using the branch & bound algorithm Seoul, Korea.
4. Luo Y, Chang X , Peng S, Khan S, Wang W, Zheng Q (2014) ‘Short-term forecasting of daily reference evapotranspiration using the hargreaves-samani model and temperature forecasts’, *Agricultural water management*, vol. 136, pp.42-51
5. Chen SX, Gooi HB, Wang MQ (2013) ‘Solar radiation forecast based on fuzzy logic and neural networks’, *Renewable energy*, vol.60, pp.195–201.
6. Inman RH, PedroH, Coimbra C (2013) ‘Solar forecasting methods for renewable energy integration’, *Progress in Energy Combustion Science*, vol.39, pp.535–576.
7. Sharma N, Sharma P, Irwin D., Shenoy P. Predicting solar generation from weather forecasts using machine learning; *Proceedings of the (2011) IEEE International Conference on Smart Grid Communications (SmartGridComm); Brussels, Belgium. 17–20 October 2011; pp. 528–533.*
8. Ferreira PM, Ruano AE, Silva S, Conceicao (2012), ‘Neural networks based predictive control for thermal comfort and energy savings in public buildings’, *Energy and Buildings*, vol.55 pp.238–25.
9. Collotta M, Messineo A, Nicolosi G, Pau GA (2014), ‘Dynamic fuzzy controller to meet thermal comfort by using neural network forecasted parameters as the input. *Energies*’, vol. 7, pp. 4727–4756.
10. Renner C, Nguyen P (2015), *Proceedings of the Sustainable Internet and ICT for Sustainability (SustainIT) 14–15 April*, ‘A system for efficient dissemination of weather forecasts for sustainable solar-powered sensors’, Madrid, Spain.
11. Hamed Shah Mansouri, Vincent Wong (2018) , ‘Hierarchical Fog-Cloud Computing for IoT Systems: A Computation Offloading Game’, *IEEE Internet of things journal*, vol.9, pp. 1-12
12. Boqi Jia, Honglin Hu, Yu Zeng, Tianheng Xu, Yang Yang (2018), ‘Double-Matching Resource Allocation Strategy in Fog Computing Networks Based on Cost Efficiency’, *Journal of Communications and Networks*, vol. 20, no. 3, pp. 237-246
13. Jun Song Fu, Yun Liu, Han Chieh Chao, Bharat K, Zhen Jiang Zhang (2018), ‘Secure Data Storage and Searching for Industrial IoT by Integrating Fog Computing and Cloud Computing’, *IEEE Transactions on Industrial Informatics*, vol.14, no.10, pp. 4519 – 4528
14. Jianhua He, Jian Wei, Kai Chen, Zuoyin Tang, Yi Zhou & Yan Zhang (2018),

- ‘Multi-tier Fog Computing with Large-scale IoT Data Analytics for Smart Cities’, *IEEE Internet of Things Journal*, vol. 5, no. 2, pp.1-10
15. Prabal Verma, Sandeep K Sood (2018), ‘Fog Assisted- IoT Enabled Patient Health Monitoring in Smart Homes’, *IEEE Internet of Things Journal*, vol.5, no.2, pp 1-10.
 16. Chii Chang, Satish Narayana Srirama, Rajkumar Buyya (2017), ‘Indie Fog: An Efficient Fog-Computing Infrastructure for the Internet of Thing’, *The IEEE Computer Society*, pp. 92-98
 17. Jianbing Ni, Kuan Zhang, Xiaodong Lin, Xuemin Shen(2018), ‘Securing Fog Computing for Internet of Things Applications: Challenges and Solutions’, *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 601-628
 18. Fowdur a TP, Beeharrya Y, Hurbungs b , Bassooa V, Ramnarain-Seetohul c & Chan Moo Lun E (2018), ‘ Performance analysis and implementation of an adaptive real-time weather forecasting system, , *Internet of Things*, vol. 3, no. 4, pp. 12-33
 19. Moxa Inc. Nport z3150 Series User’s Manual. [(Accessed on 6 December 2015)]. Available online: http://www.moxa.com/doc/man/NPort_Z3150_Series_Users_Manual_v2.pdf.
 20. Libelium Wasmote Overview. [(accessed on 6 December 2015)]. Available online:<http://www.libelium.com/products/wasmote/overview/>
 21. Ferreira PM, Gomes JM, Martins IAC, Ruano AE (2013), ‘A neural network based intelligent predictive sensor for cloudiness, solar radiation, and air temperature’, *Sensors*, vol. 13 no. 7 pp. 9547-9548.
 22. IEEE IEEE Standard for Local and Metropolitan Area Networks. [(accessed on 6 December 2015)]. Available online: <http://standards.ieee.org/getieee802/download/802.15.4k-2013.pdf>.
 23. Shizhuang Lin , Jingyu Liu & Yanjun Fang(2007), *IEEE International Conference on Automation and Logistics18-21 Aug*, ZigBee Based Wireless Sensor Networks and Its Applications in Industrial, Jinan, China