# Divide-And-Conquer Approach In Machine-To-Machine Communication

**P.Nivetha,**
Asst.Prof, Dept. of BCA
Bon Secours College for Women
Thanjavur, India
nivyabi@gmail.com

**V.UmaMaheswari,**
Asst.Prof, Dept. of IT
Bon Secours College for Women
Thanjavur, India
gumatnj76@gmail.com

*Abstract*— **The interconnection takes place between two machines with the help of physical objects like sensors, satellites, cars, buildings, smart phones etc.,. The purpose of objects is to sense, communicate, store, manage and process the data. Mesh network helps to get the large amount of information about our living and non-living environment. Big data applications took part in different fields of knowledge such as machine-to-machine (M2M) communication, wireless sensor network (WSN) and internet of things (IoT). Big data in M2M communication does not handle large data sets effectively in the existing big data architecture. Some other drawbacks are cost, data management, data quality, data control, computational efficiency, power consumption, time duration etc. To overcome this problem, an efficient divide-and-conquer (D&C) algorithm is used to recursively break down the original problem into two or more sub-problems of the same or related type. The solution to the sub-problems is combined to give a solution to the original problem. This work mainly focuses on the earthquake measurement to predict the destruction before occurrence since it involves huge data sets of the region which are increasing from Gigabytes to Petabytes. Few more analysis could be done with the help of Hadoop MapReduce framework. This framework generates key-value pair as an output that provides security and accuracy of the original data set. Many real-time applications may use this algorithm to get better result and also helps to take decision making**.

***Keywords— Big data analytics, M2M, Divide-and-conquer (D&C) approach, Real time data processing, Data collection.***

## I. INTRODUCTION

Big data is a term for data sets consisting of the large amount of data in which traditional data processing are inadequate to process those data. Big data analytics is growing faster because of processing a massive volume of data in our day-to-day life. The input for big data analytics are social media, online transactions, business processing, bank transactions etc. Big data mainly welcomes real-time and offline data processing. Offline processing is performed by "batch" type programs or systems and is not bounded by "response time". Real-time data processing is the short time execution of data and provides instantaneous output. Some challenges may involve data analysis, capturing the data, data sharing, updating the query, privacy and security.

M2M communication needs to satisfy the characteristics of Big Data such as Volume, Variety and Velocity. Big data generally involves an enormous amount of data that are generated by various machines, networks and humans. The basic definition of volume is to talk about the quantity of generated and stored data. Variety involves type and nature of the dataset. Velocity deals with speed of the data that are generated by various objects. Earthquake measurement needs to be implemented in Hadoop map reduce framework by using divide-and-conquer algorithm whereas Hadoop is an open source, Java–based programming framework for storing the data and also process high level data sets in the distributed environment.

Map reduce framework largely helps to makes the process faster with the help of job tracker and task tracker. Also it has the capacity to manage the storage part with the help of name node and data node.

Some requirements are not specific to divide-and-conquer approach in M2M and it can be applied to M2M / IoT platforms.

1. Heterogeneity: It involves various machines like satellites, sensors, various data formats, protocols, vehicles, etc.
2. Reconfiguration: It can be deployed in any broad range environment. It has the capability to access remotely.
3. Scalability: It is the capacity of the system to manage the growing data. It has two dimensions such as vertical and horizontal scalability. Vertical scalability deals with storage whereas horizontal scalability deals with processing purpose.

4. Services: It handles the method which supports the large number of users at the right level of abstraction.

## II. Related Work

In [1] mainly focuses on the internet of things. The main factor is to integrate the several technologies with communication solutions by the use of smart objects. The purpose for the smart objects are identifying and tracking the things that are used in various field of knowledge like telecommunications, social media and bio informatics. Some others are used in both wired and wireless sensor networks for sensing the objects which is used to manage several things. Smart objects are used to enhance the communication protocols which are shared in the Next Generation Internet along with the distributed intelligence.

The basic concept is to make use of RFID tags, sensors, actuators and mobile phones to identify the particular things and helps to communicate between various machines. RFID's are used in many applications such as e-health, smart card systems that are used in schools, colleges and universities. RFID are used in industrial fields to manufacture small things, logistics and business or process management. Some others are used in transportation to detect particular objects. Many issues are addressed in both technological and social fields since it does not provide trust, privacy and security.

The major problem in RFID is to make full interoperability of interconnected devices that provides higher degree of smartness. Since the communications are wireless so that eavesdropping is simple in it. RFID tags are not able to handle too many messages at a time for communicating two things.

In [2] the authors developed a big data mechanism which involves the numerous data to innovate new big things. One of the best examples for big data is IoT. It focus on the software architecture that involves numerous sensor-based data is used to transfer the data to the data storage. It supports the user for experimental purpose and are used on a project called"SMARTCAMPUS" to build the innovative applications. It mainly focused on data collection, scalability problem and data virtualization. Mediation layer acts as an interface between sensed data and developers.

The SMARTCAPMPUS project focused on data storage and collection. It involves large datasets to handle the critical path in the big data platform. Example for such large datasets is temperature evolution, social media and parking details etc. This survey talks about the software application on the data sets for creating their own services. It faces the challenging issues in the data storage and also security.

In [3] there are about billions of devices are used for sensing, communicating, computing and validation purpose. This survey discusses about Internet of Things, federating sensor networks and cloud computing. It mainly involves for solving the problem of the data management, data storage, data archiving and data processing. Traditional approaches are not used for managing those data streams. Many applications like environmental monitoring, business and process management involves billions of devices that are used for sense, capture and processing the real time data. Vehicles are the major resource that produces the large amount of data about our surrounding environment. Some other areas are smart homes, finance, traffic control system, search quality and also in trade analysis. Nike+ipod/ iPhone application is an application used to collects and tracks information without the details of the jogger.

Its main purpose is to sense the jogger during jogging. Another application is iSmooth Run used to upload the data in the fitness social networks. It involves millions of user to use that application for their purpose. A greenhouse gas sensor networks gathers large amount of real-time information about their gases and its behaviour throughout the state of California. The following statistics are the examples of big data. Big data consists of high-valued data with raw data. ICT researchers and practitioners handle the challenges in the big data analytics and information implosion.

The authors Ramaswamy, Lawson and Gogineni [4] developed a framework called quality-centric big data architecture for federated sensor services. The large number of independently-managed sensors is shared throughout the platform. An eco-system will have many stakeholders like sensor data providers, cloud infrastructure providers and data consumers. Challenging tasks involved in the cloud-based platforms are still emerged. This survey points on data quality (DQ) centric big data infrastructure for the federated sensor service clouds. It proposes a big data architecture in which DQ is pervasive throughout the platform. It involves a mark-up language called SDQ-ML for sensor services mainly for sensor feed requirements. The paper analyses the advantages and limitations of big data technologies. Since the DQ is pervasive in nature throughout the environment. The data quality issues are occurred in the cloud based framework. Data Quality for sensor feeds is the broad concept that has quality attributes involves accuracy, availability and latency. In general, sensor feeds are continuous in nature that evolves data streams and DQ is dynamic.

The authors Lindsey and Raghavendra [5] developed a Power-efficient gathering in sensor information systems. In Aerospace conference proceedings tells about the sensor webs that are deployed in the working environment to gather the useful information about the nodes. Sensor webs consist of nodes with the help of wireless communication and limited power supply. The main objective is to collect the useful information in an energy efficient manner for a longer duration. One of the famous clustering algorithm techniques with the static node is LEACH. Each sensor node is responsible for data transfer to the ground base station.

The LEACH protocol is the perfect solution to form the clusters for fusing the data into the base station before transmission. It takes place the equal distribution of the energy for re-clustering in the network. To gather the sensed information using LEACH takes much time for data transformation to overcome the time spent on each sensor nodes, this paper talks about PEGASIS. It largely helps to achieve uniform energy consumption which is based on the position of cluster nodes and also limits the communication range. PEGASIS is used to communicate the neighbouring nodes in the efficient manner and also decreases the energy consumption spent on each round. The disadvantage is to consider the communication range so the efficiency of the sensor nodes can be easily reduced.

The authors Mayilvaganan and Sabitha [6] explained about smart grid with a detailed cloud-based architecture. A smart grid is an electrical grid has numerous operations which are used to measure energy in the various fields. An electric grid consists of generating stations to deliver the electricity from suppliers to consumers. It carries high voltage transmission lines for power supply. The electrical utilities are basically helps for the improvement of economy and reliability. In the concept of distributed generation, there is a lot of grid system usage in the economy leads to grid defection.

In order to overcome this problem, smart grid is introduced in the market. Smart grid outperforms an electric grid to satisfy the customer's needs. In this survey mainly addresses the demand and power supply in order to create the cloud-based environment for it. The smart grid deals with big data consist of data regarding the customer's power usage patterns. It should maintain both the demand and power supply in the balanced state. Smart grid formation deals with big data mainly focuses on historic weather data monitoring, current power supply details that are used in industrial purpose. The required data set could be taken from the cloud storage with the help of renewable energy resources.

The purpose of smart grid is to usage of power production effectively. Other important aspect need to consider as a power distribution which involves big data analytics and cloud architecture. It helps to make the smart grid much better to effectively handle the demand and power supply in the equal state.

The combined works of Mai and Choo [7] developed a connectivity–based clustering scheme for mobile adhoc networks. This paper deals with the effectiveness, stability and reliability of a multi-hop clustering scheme. When the cluster size becomes larger, we move on to the multi-hop clustering scheme than one-hop clustering. To create the new cluster selection we can use intra-connection degree between the sensor nodes. It simply increases the lifetime of the mobile nodes within the cluster that are used in the wide area networks. Some challenging issues in the mobile ad hoc networks are maintenance, scalability issue, reliability and effectiveness. To improve the overall efficiency, multi-hop clustering needs to make the cluster heads apart from the neighbourhood clusters. Two main issues need to focused here is to form the cluster effectively and avoids the small clusters that took part in the maintenance phase. In the wide –scale networks, high-level applications uses the cluster backbone in the larger level.

The combined work of Nakayarna and Ansari [8] proposed methods which are based on KAT. Many research processes is going on the wireless sensor networks (WSN) and VLSI with the help of tremendous use of tiny sensors. The purpose of the mobile sink is to collect the data from the tiny sensors for data dissemination. The new challenging issues take part in the routing part and the distribution of sensors. The KAT (K-means And TSP) mobility classifies the nodes into small clusters that are based on the node location.

The KAT algorithm is structured that does not depend on the communication range. It produces the optimal cluster with minimum energy consumption. The drawback in the mobile sink is the failure probability of sensor nodes. Due to malfunctions of some sensors, any intruders can attack the nodes easily leads to fault resilience. Better throughput can be attained with the help of K-means clustering algorithm along with TSP-derived migration in the mobile nodes. It found the optimal route for the damaged sensor to recover from the intruders. Since it assigns the appropriate number of clusters and mobile sink helps to improve their overall performance. For security purpose it uses the self- regulating technique for each level that gives the best solution.

The combined work of Baurngartner, Ferrari and Rao [9] developed a method which is the most general and effective approach is carried out in the dynamic systems is the optimal control system. It helps to concentrate more on the quality of the sensor network service. This survey tells about the trajectory optimization of control system. The mechanism which follows the integral objective function and system models for providing the better quality service.

It plays a vital role in the track detection system by using the geometric traversal approach. This method should be applied well in subsystems such as on-board sensors, oceanographic fields and underwater vehicles. It simply formulates the problem of optimal sensors trajectory into OC problem. The numerical results from the mechanism gives the better performance compared to area-coverage and path-planning methods. It can give the best quality of service than others. It also helps to sensing part with less energy constraints and produce pareto-optimal trajectories.

The authors Juan, Tzu-Chuan, Shih-En Wei, and Hung-Yun Hsieh[10] addresses the one key technology that is used for machine-to-machine (M2M) wireless networks is the

clustered communication. The existing clustering techniques are involved with the numerous service quality that are needed for the individual machine. The best example for that is to "machine-centric" clustering technique by M2M applications. It does not handle the big data problem more effectively. Data-centric clustering is much more helpful for data gathering and also for data correlation by various machines.

The aim is to reach the target and solve those problem effectively. Since it involves cluster formation and effective power control for the regular power supply to all nodes. The best solution for problem solving is to formulate data-centric cluster formation and uplink power control for the data gathering process in it. Data–centric clustering attains the significant performance compared to conventional machine-centric clustering. It produces the better result in terms of considerable amount of energy resources.

The combined study of Rahul C, Roy, Jain, and Brunette [11] describes the details about three-tier architecture in the spare sensor networks. The aim to collect the sensor data for the exploitation of mobile entity called it as MULEs. The step-by-step process is to pick up data, store it in a particular database and finally submitted into the wired access points. MULE helps to reduce the power consumption that are used in the short-range communication. It focused the analytical model for the entire system performance and the scalability of the system.

The analytical model focuses on the two-dimensional random walk and key system variables. Data success rate, latency, buffer capacity needs to be observed for the sensors and MULEs. The obtained result is used for analysis and also gives the guidelines for the deployment of the system. The main idea is to better utilization of the mobile entity present in the environment that provides the transportation medium for the sensor data with less energy It addresses to develop a stochastic model for simplification purpose like infinite bandwidth, random walk mobility model and error-free communication. MULE–to-MULE communication and reliability are the major issue in it. The limitation is to continuously monitor the sensor for the identification of the MULE's presence. The sleep time for the sensors degrades the system performance.

The combined study of [12] Lin, Chiu, Hsiao addresses the dementia disease. The combination of technology and advanced knowledge of manpower helps to reduce the elderly person suffering from dementia. It provides the stay protection system without interfering with the daily activities of the dementia patient. The help of GPS integration with of various technology helps to find out the exact place of the missing person where he/she is. The system gives four monitoring schemes which include indoor residence monitoring, outdoor activity area monitoring, emergency rescue and remote monitoring nodes. In order to implement those scheme, the system which needs service platform. The platform needs four major server such as web service server, database server, message controller server and health-GIS server. The necessary parameters are volunteer workers to find out the position of the person by using mobile phones through the service platform. It also helps to find the system performance and reliability. The major issues are aging of the population and it's very tough in the real health care environment.

## III. Proposed Methodology

The proposed divide-and-conquer (D&C) data analytical architecture for big data in M2M has the ability to handle enormous amount of data set and also provides accuracy of data generated by various machines. Few challenges need to consider such as capturing the data, data formation, storing the data within the memory, handling the data management, sharing the data resources, analysing the data and visualizing the big data.

The basic process starts with the sensor which is largely helps to collect the real time data set of the region. The real time input data needs to be stored in HDFS for large storage. Then, the data to be sent into MapReduce for simply dividing the input data into smaller one that take part in mapper function. It produces key-value pair as an input to the next process. Grouping process takes place with respect to their values and goes into the reducer function that simply combines the result and produces the better output and brings it into the server. An authorized application can first copy the input data from the local host into HDFS and brings into next step. Mapper and reducer function process the data in parallel manner after that the resultant output to be moved from MapReduce into local host.

### A. *Proposed System Architecture:*

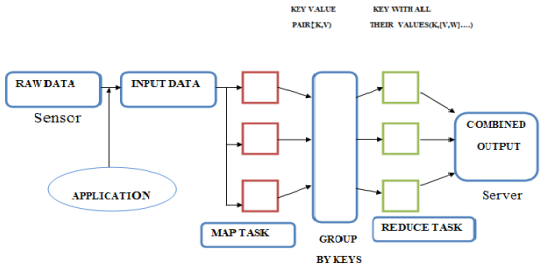


Fig 3.1 Proposed System Architecture

## IV. Experiments and Results

### A. *Data Set:*

The data set consists of serial number, source name, equivalent identification, version, date and time, latitude, longitude, magnitude, depth, NST and full form of the region.

Since it involves numerous data set which are ranging from gigabytes to petabytes.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Src | Eqid | Version | Datetime | Lat | Lon | Magnitud | Depth | NST | Region | |
| 2 | nc | 71920701 | 1 | Saturday, | 38.7865 | -122.763 | 1.5 | 1.1 | 27 | Northern California | |
| 3 | nc | 71920696 | 0 | Saturday, | 40.3742 | -122.112 | 1.5 | 10.6 | 9 | Northern California | |
| 4 | nc | 71920686 | 1 | Saturday, | 38.8083 | -122.823 | 1.2 | 2.3 | 24 | Northern California | |
| 5 | ci | 15272537 | 0 | Saturday, | 33.3733 | -117.17 | 1.3 | 14.4 | 17 | Southern California | |
| 6 | ak | 10637537 | 1 | Saturday, | 60.6897 | -141.662 | 1.9 | 32.4 | 12 | Southern Alaska | |
| 7 | ak | 10637523 | 2 | Saturday, | 61.6978 | -147.005 | 3 | 20.1 | 63 | Southern Alaska | |
| 8 | ci | 15272529 | 0 | Saturday, | 33.5647 | -117.451 | 1.9 | 0 | 13 | Southern California | |
| 9 | nc | 71920666 | 1 | Saturday, | 38.8215 | -122.794 | 1.4 | 3.1 | 34 | Northern California | |
| 10 | ak | 10637498 | 1 | Saturday, | 60.9345 | -152.023 | 2.1 | 77.8 | 12 | Southern Alaska | |
| 11 | us | b000emzu | 6 | Saturday, | -2.9911 | 129.8079 | 5.3 | 7.3 | 185 | Seram, Indonesia | |
| 12 | ci | 15272505 | 0 | Saturday, | 33.4792 | -116.532 | 1.5 | 15.6 | 84 | Southern California | |
| 13 | ak | 10637473 | 1 | Saturday, | 61.3581 | -146.551 | 1.6 | 26.3 | 5 | Southern Alaska | |
| 14 | nc | 71920611 | 0 | Saturday, | 38.7593 | -122.875 | 1 | 4.2 | 12 | Northern California | |
| 15 | ci | 15272489 | 0 | Saturday, | 34.3395 | -116.88 | 1.5 | 2.1 | 47 | Southern California | |
| 16 | uw | 60491097 | 2 | Saturday, | 47.6203 | -120.269 | 2.2 | 1.3 | 11 | Washington | |
| 17 | nn | 399504 | 9 | Saturday, | 40.3547 | -119.844 | 1.2 | 9.7 | 13 | Nevada | |

Fig 4.1 Data Set About various Regions

### *B. Loading the Input Data Set into HDFS:*

The linux command is used to get the input from host system into hadoop HDFS. Hadoop Distributed File System (HDFS) is the primary storage by Haadoop applications.It gives high-performance aaccess to data across hadoop clusters.It becomes a key tool for managing pools of big data and its applications.There are about numerous data set to upload the region into hadoop environment.

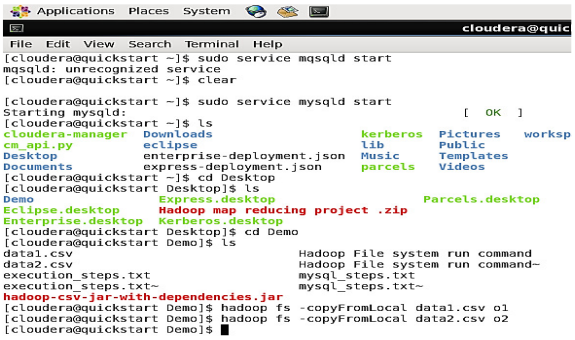


Fig 4.2 Copy Region dataset from Host to Guest

### *C. Uploading the Files:*

The necessary command is used to upload the jar files for displaying the output. HDFS is used to upload large data set and also generates the result within time limit.

### *D. Login Page:*

The user is accessing the welcome page by registering the new user name and password with the help of LINUX command. The next step is to enter into the login page, if it is a new user, it should be registered first by giving username and password or else directly enter and run the program.

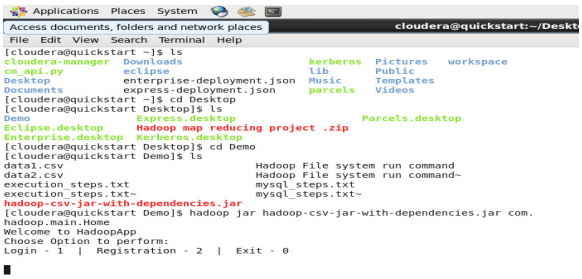


Fig 4.3 Registration for the new user

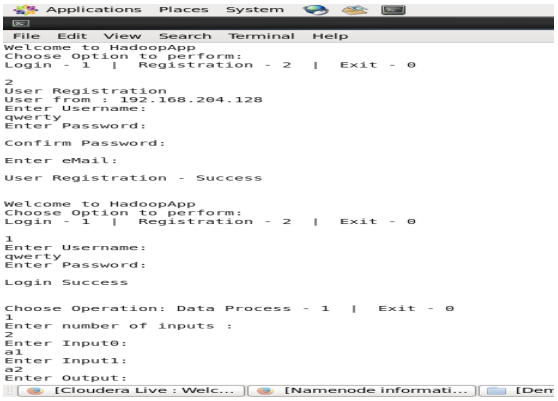


Fig 4.4 Login into the System with UserName and Password

### *E. Map Reduce Freamework:*

It simply takes the numerous data set to do map reduce and produces file system counters, job counters, map-reduce framework, shuffle errors, file input format counters, file output format counters.

*Fig 4.5 Map Reduce Freamework*

### *F. Output Screen:*

The output result is displayed in the form of id, $, region name and maximum magnitude. Likewise it can generate huge volume of data set within the short durable time period.

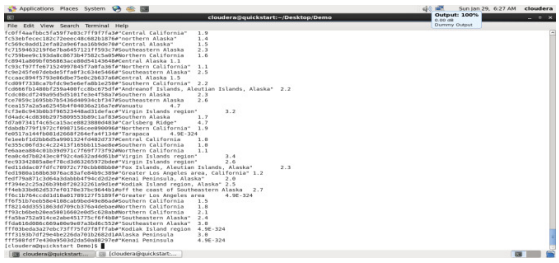

Fig 4.6 Output Screen

*G. Performance Graph about various Regions :*

The output depicts the maximum magnitude of the region on the horizontal bar. It shows the various earthquake regions in which it results in highest magnitude of one particular region.

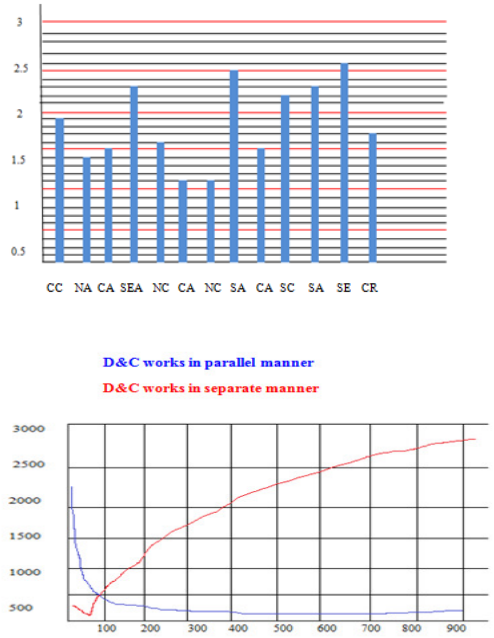


Fig 4.7 Various earthquake Regions of higher magnitude regions and performance of D&C in serial and parallel manner

## Conclusion

Big data in M2M communication that uses divide-and-conquer approach is a generic model for the big data analysis. In the earlier days divide-and-conquer approach takes place separately with the help of their respective objects. The earthquake data set takes huge volume of data to tell about the detail of the region. Hadoop is the best platform to handle those data sets effectively. Map reduce framework works well in case of enormous amount of database in which data storage needs to be maintained by HDFS and data processing takes place with the help of map reduce. The main advantage of using divide-and-conquer (D&C) algorithm in earthquake measurement is to enhance the overall performance because the working process should be carried out by parallel manner such that we can easily attain the better output. Also it considers more data quality, data management and time consumption. It could provide more accurate data to do the next process. Simultaneous work leads to cost reduction for the basic requirements and also leads the working flow to be more secure with the help of MD5 algorithm. The final result is considered for measuring the region details to detect the safety of the region.

## References


[1] Atzori, Luigi, Antonio Iera, and Giacomo Morabito. "The internet of things: A survey." Computer networks 54, no. 15: 2787-2805, 2010.

[2] Cecchinel, Cyril, Matthieu Jimenez, Sébastien Mosser, and Michel Riveill. "An architecture to support the collection of big data in the internet of things." In Services (SERVICES), 2014 IEEE World Congress on, pp. 442-449. IEEE, 2014.

[3] Zhang, Jia, Bob Iannucci, Mark Hennessy, Kaushik Gopal, Sean Xiao, Sumeet Kumar, David Pfeffer et al. "Sensor data as a service--a federated platform for mobile data-centric service development and sharing." In Services Computing (SCC), 2013 IEEE International Conference on, pp. 446-453. IEEE, 2013.

[4] Ramaswamy, Lakshmish, Victor Lawson, and Siva Venkat Gogineni. "Towards a quality-centric big data architecture for federated sensor services." In Big Data (Big Data Congress), 2013 IEEE International Congress on, pp. 86-93. IEEE, 2013.

[5] Lindsey, Stephanie, and Cauligi S. Raghavendra. "PEGASIS: Power-efficient gathering in sensor information systems." In Aerospace conference proceedings, 2002. IEEE, vol. 3, pp. 3-3. IEEE, 2002.

[6] Mayilvaganan, M., and M. Sabitha. "A cloud-based architecture for Big-Data analytics in smart grid: A proposal." In Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on, pp. 1-4. IEEE, 2013.

[7] Mai, Khac Tiep, and Hyunseung Choo. "Connectivity-based clustering scheme for mobile ad hoc networks." In Research, Innovation and Vision for the Future, 2008. RIVF 2008. IEEE International Conference on, pp. 191-197. IEEE, 2008.

[8] Nakayama, Hidehisa, Nirwan Ansari, Abbas Jamalipour, and Nei Kato. "Fault-resilient sensing in wireless sensor networks." Computer Communications 30, no. 11: 2375-2384,2007.

[9] Baumgartner, Kelli AC, Silvia Ferrari, and Anil V. Rao. "Optimal control of an underwater sensor network for cooperative target tracking." IEEE Journal of Oceanic Engineering 34, no. 4: 678-697,2009.

[10] Juan, Tzu-Chuan, Shih-En Wei, and Hung-Yun Hsieh. "Data-centric clustering for data gathering in machine-to-machine wireless networks." In Communications Workshops (ICC), IEEE International Conference on, pp. 89-94. IEEE, 2013.

[11] Shah, Rahul C., Sumit Roy, Sushant Jain, and Waylon Brunette. "Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks." Ad Hoc Networks 1, no. 2: 215-233, 2003.

[12] Lin, C-C., M-J. Chiu, C-C. Hsiao, R-G. Lee, and Y-S. Tsai. "Wireless health care service system for elderly with dementia." IEEE Transactions on Information Technology in Biomedicine 10, no. 4: 696-704, 2006.