ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Overlay Network Construction Based On Node Location in Neighbour Selection to Manage Data

Namrata B. Jadhav[1], Prof. S. N. Shelke [2]

M. E Student, Dept. of Computer Engineering, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India

Assistant Professor, Dept. of Computer Engineering, Sinhgad Academy of Engineering, Savitribai Phule Pune University, Pune, India

ABSTRACT: Currently there is the rapid growth of data and communication techniques in educational, industrial, government applications, and business. To manage such big amount of data, mining provides techniques such as clustering, parallel and distributed architecture. In the past the overlay based data mining architecture used parallel and distributed hadoop framework to construct network. The proposed work of this paper uses parallel and distributed architecture and clustering to construct an overlay network. An overlay network construction based on node location removes failed super node and then reconstructs the network again by taking new super node from same cluster. This proposed work used to keep all connection active with service availability. The comparison of existing overlay and proposed overlay network shows that available nodes are remain more in proposed work than existing.

KEYWORDS: Big data, Data Mining, Overlay Network, Node Location, Neighbour selection, Failed Node Removal, Network Reconstruction.

I. INTRODUCTION

Big data mining is the ability of extracting large amount of data. In addition, big data requires result to be efficient data as response to demands of real time services [2]. Today there is rapid increment of information or data not only in academic but also in business, government applications, and industrial data. To handle and manage such big data is becoming critical issue. To manage such big data, mining provides technique called clustering and parallel and distributed approach is good to handle. Clustering creates the group of same types of items. In the proposed work clustering is used in the construction of overlay network firstly and then the incoming data is to be divided in particular cluster, because of this data is to be easily available when required.

An overlay network is the network on the top of another network or on the top of existing network. An overlay network is actually robust it means very strong network. An overlay network could not be easily connectionless, it always with strong connection. An overlay based parallel data mining is good option to increase the scalability. By using overlay network [9] all nodes can execute processing and manageable tools, these can balance the management load and can achieve availability of services against breakdown of server [15]. In the proposed work parallel and distributed architecture is used in construction of an overlay network. It is easy to manage and handle data from network with parallel and distributed architecture.

Node location is the concept used in construction of an overlay network. Node location means location of the node in its group. Node location is useful when there is need of replacing node in any critical situation. Node location is useful in neighbour selection scheme. Neighbour selection is the selection of node from same group or same cluster. The neighbour selection method that manage the connectivity of nodes because of that there is increase in availability of nodes against the physical network tolerance. That's why there is possibility of failure of processing result due to more nodes. The neighbour selection and task allocation overcomes the above problem and it extracts the data even in physical network disturbance [1].

The proposed work used node location when there is failed super node is occurred due to any reason such as node crash. Proposed work uses two algorithms failed node removal algorithm and network reconstruction algorithm to remove failed node and reconstructs network to make connection active again to maintain data and service availability.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Reconstructed network takes new super node from same cluster at the place of failed or removed super node. This work makes the high service availability and manages data easily.

II. REVIEW OF RELATED WORK

Review of related work includes some past methods of construction of an overlay network some provides high scalability, some provides good connectivity, some have critical connectivity and so on.

1 Physical Network Aware Neighbour Selection Scheme

The neighbor selection method effects to connectivity of an overlay network. That's why Author k. Suto, N. Hiroki, N. Kato [1] this method to maintain the availability against disturbance of physical network. In this scheme servers are located at same place in physical network. To construct the neighbor node selection there are methods called node joining procedure, and the method is independently executed by a newly joined node (NJN) to deal with management load to all nodes and another is network maintenance procedure.

The merit of neighbour selection node is that it provides high availability of system. They keep good connectivity of network. Its disadvantage is the neighbour selection increases number of nodes available in the oppose of physical network disturbance because better connectivity. But overlay based data mining architecture become not successful to produce processing result when there are mappers have similar data block are removed.

2 Physical Network Aware Task Allocation Method

The task allocation scheme [1] still mines data even if there is physical network disturbance is occurred. In neighbor selection nodes are removed from the same area of group by physical network disturbance in the overlay network. From this it is getting to know that choosing nodes from the longest groups as mappers make definitely the existence of at least one unnecessary data. That's why the task allocation scheme divides each and every unnecessary block of data to definite nodes in diagonally cornered groups. This method has a reception node that grt request from a client which can start a task allocation procedure.

The advantage of ask allocation scheme is there is 100% success probability of data mining with minimum replications. And it has disadvantage that the task allocation scheme is the heavy computation. And it has complex architecture.

Both of above methods are proposed by authors k. Suto, N. Hiroki, N. Kato [1] which have greater scalability but in that method when there is failed super node is occurred then it removes whole cluster along with that failed super node. Then reconstruct the network again with distributing other clusters.

3 Bimodal Degree Distribution

Since distributed networks [3] are able to allow nodes to join or leave easily, still there remains issue of lack of resilience of both faults and attacks. The unstructured P2P network with power law degree distribution is harmful to attacks such as DOS. To resolve and address this issue there is a method of construction of bimodal degree distribution. It is possibility of network collapse because of network failure which are faults and attacks. It is because of connectivity problem. The bimodal network is used to achieve high robustness against faults and attacks without increasing average degree in network. There are two methods in bimodal degree distribution [4] and from these methods performance is achieved by varying parameters of high degree nodes and average degree. Here bimodal distribution evaluates the performance through computer simulations using MATLAB.

The advantage of bimodal degree distribution is that which is robust to deal with both faults and attacks. But this modal cannot ensure the service availability against physical network tolerance.

4 THUP(Churn/DOS Tolerance, Hierarchical, Unstructured, Peer to Peer Network)

Hierarchical Unstructured P2P Network Tolerant to Churn and DOS Attack [5]. In hierarchical unstructured peer to peer (P2P) network [11] is for file sharing, Gnutella and Kezaa which are tolerance to departures and communication efficiency are high because there are ultra peer which increases network connectivity. The hierarchical unstructured peer to peer networks tolerant to churn and DOS attack includes three methods which are bimodal degree distribution, optimal topology for Dos attack tolerance, and peer joining procedure in THUP [5]. In the performance of THUP they

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

performed neighbor selection in THUP to investigate impacts of number of inter group lists on tolerance of churn and DOS attack. Then they performed random neighbor selection based on bimodal degree distribution.

The advantage of THUP is that THUP increases communication efficiency and stability. And the shortcoming is that it has multiple layers.

5 A Distributed Framework for Prioritizing Iterative Computation

Iterative computation is pervasive among web search, online social network analysis, recommendation system such data analysis applications. A distributed computing framework 'Prlter' helps to the prioritized execution of iterative computation. PrIter stores either intermediate data in files for scaling to larger data sets or intermediate data in memory for fast convergence. This method evaluates PrIter on a local cluster of machines and on Amazon EC2 Cloud. This method gives results of that PrIter achieves up to 50x speedup over Hadoop for a series of iterative algorithms. Also, PrIter gives better performance for iterative computations than other state-of-the-art distributed frameworks such as Spark and Piccolo [6].

The advantage of this method ensures fast convergence to perform updates, rather than performing updates on all data.

6 Availability aware Data PlacemenT (ADAPT)

This strategy dispatches data on the bases on the availability of the each node, improve data locality, optimize the application performance, and reduce network traffic. It implements the prototype of ADAPT within the haddop framework, an open source implementation of MapReduce [10] model. ADAPT evaluated in non dedicated distributed environment which is emulated. And from this authors Hui Jin, Xi Yang, Xian-He Sun, Ioan Raicu [7] evaluated result, which shows that ADAPT is able to improve performance more than 30% and also achieves high reliability

The advantage of this strategy is that, this strategy improves the performance of application without extra cost. This strategy is not able yet to provide strong efficiency. And also there may be possibility of the capability of data mining may dramatically decrees.

7 Node Isolation under Churn in Unstructured Peer-To-Peer (P2P) Networks: General Node Isolation Mode

The general node isolation model analyzes two age based neighbor selection strategy. This model includes methods such as Hyper-Exponential Approximation, Isolation Probability, Necessity of Neighbor Replacement, and Verification of Isolation Model [8]. These methods are able to generate the result which may generalizes to all heavy-tailed distributions in which the expected remaining lifetime increases and residual lifetime becomes larger with age.

The advantage of this method is that it dramatically reduces the probability of user isolation. And also make graph partitioning compared with uniform neighbor selection.

8 Node Isolation under Churn in Unstructured Peer-To-Peer (P2P) Networks: Age-Based Random Walk Selection

Age based random selection model introduces a neighbour selection strategy based on random walks over weighted directed graphs and also deals with distribution of neighbour residual lifetimes and calculates the corresponding isolation probability. The age weighted graphs indicates that for lifetimes with infinite deviation, the system increases its flexibility as its age and size grow [8].

The advantage of this method is that it has lifetimes with infinite flexibility. This method is not able yet for distributed finite systems.

9 A Novel Based P2P Overlay Network

A novel base P2P overlay network also called as CloudCCC [12] that achieves a good balance with regard to the aforementioned qualities. This method is very much useful in distributed computing, under relatively stable conditions. CayleyCCC helps explicit grouping of peers, so facilitating effective resource browsing. The design of this method offers simpler searching routing and many other properties compared with many previously proposed overlay networks, such as Chord [13] and Ulysses [14].

The advantage of this method is that it has many desirable features, including short query paths, compact routing tables, high clustering, and robustness. But this method yet does not provide multicast and other applications.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

III. PROPOSED WORK

Author k. Suto, N. Hiroki, N. Kato [1] proposed two methods an overlay network construction scheme based on node location in physical network, and distributed task allocation strategy using overlay network technology. This existing work removes whole cluster along with failed node in the construction of an overlay network. Our goal is to remove failed or faulty nodes from an overlay network which are reducing the performance of network. Here there is removal of failed super node of a cluster and then reconstruction of network with that broken cluster by taking new super node to manage that cluster keeping connection active. To do this Proposed work of this paper contains two methods one is failed node removal and another is reconstruction of network. These methods are uses node location in neighbour selection to construct an overlay network. Neighbour selection with node location lays an important role in the construction of an overlay network.

In the description of the proposed work we considered the scenario where super node in the network fails during the operation. Other leaf node (LN) from same cluster keeps capability to hold the responsibility of the super node (SN). We proposed two algorithms to do this proposed work. These are Failed Node Removal Algorithm which avoids reducing robustness of an overlay network and Network Reconstruction Algorithm to maintain high connectivity, explain as follows.

1. **Node Removal Strategy**:

The existing system does node removal strategy on the basis of neighbor node selection for an overlay network. In our proposed work we implemented broken server removal with the proposed algorithm called selective failed node removal algorithm. In this approach firstly we select the node which is broken during the operation it means firstly algorithm searches for broken or failure super node. Then perform the failed node removal approach to remove that failed super node from cluster.

Algorithm 1: Selective Failed Node Removal
Input: SN[] Set of Super nodes in the network.
Output: Failed SN[i]

1. Check For all SN[i] in SN[] ;
2. if(SN[i]==Failed)
3. Remove SN[i];
4. Else
5. Check for other in SN[]
6. End if
7. End for
8. Return Failed SN[i];

The above algorithm1 is proposed selective failed node removal algorithm firstly checks for all super nodes from the provided set of super nodes. Then it checks condition for if super node is failed or getting fail or not if yes then it removes that failed super node from the network. To overview this scenario the figure 1 shows that the overview of failed node removal. As shown in figure 1, Dark nodes are super nodes which are connected each other and leaf nodes are connected to each super node with one extra leaf node in each cluster to use in critical situation at the place of any required node. Leaf node are connected each other as requirement also each leaf node is connected to respected super node. The dark dotted circle shown in the faint line cluster is the broken cluster for a while. Dotted dark node is the failed super node which is getting removed from the network. Then to reconstruct network with new super node is shown in figure 2.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

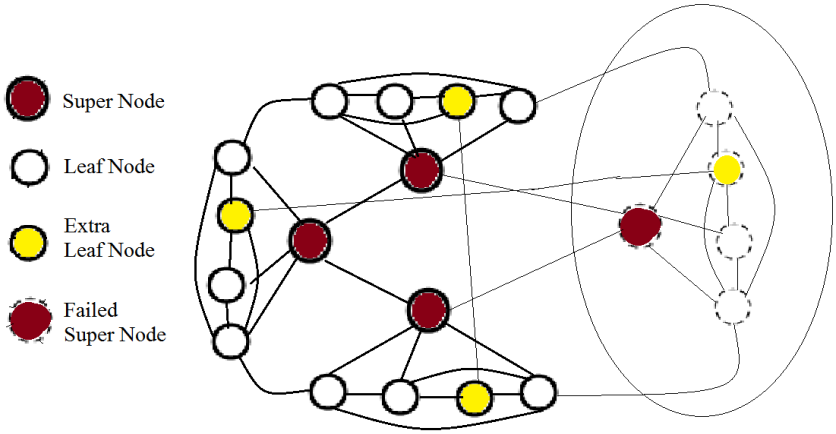

Figure 1: Review of Failed Super Node Removal

2. Network Reconstruction Strategy:

After network node removal strategy we are proposing network reconstruction strategy. As overlays network assure the higher connectivity in the network by keeping the connection active, it is necessary to reconstruct the network as the existing system removes all nodes which belongs to the server being removed.

After failed node getting removed the reconstruction of network is need to do. To reconstruction of network the second algorithm Network reconstruction is used and steps shown in algorithm 2. In that after getting removed failed super node then it calculates the degree of each leaf node from the broken cluster. Then there is the sorting process of calculated degrees of leaf nodes and takes highest degree leaf node to make it as new super node for that broken cluster to reconnect it in the network.

Algorithm 2: Network Reconstruction
Input: All SN[], LN[];
Output: NewSN (New super node);

1. Checking all SN[i] in network from SN[]
2. If(SN[i]==Failed)
3. For all LN[]
4. GetDegree();
5. End for
6. SortedNode[]=DecendingSort(Degree[]);
7. NewSN=GetHighestDegreeNode(SortedNode[]);
8. Else
9. Check for next SN[i]
10. Return (NewSN)

Figure 2 show that the review of reconstructed network with new super node by removing failed super node. In that broken cluster shows that it contains failed super node and replaces it as new highest degree leaf node as new super node and reconstructs the network again. It is helpful to maintain data again in the network with high connectivity and more number of nodes availability.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

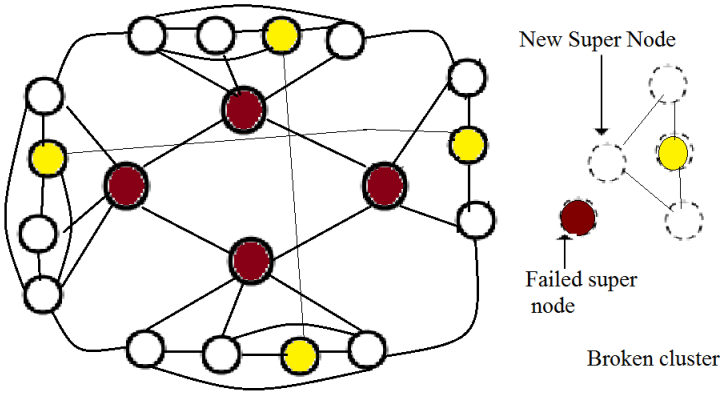

Figure 2: Overview of reconstructed network with new super node

Figure 3 shows the basic architecture of proposed work. Proposed work used the parallel and distributed architecture to construct an overlay network. In that super nodes are parallel connected to each other. Leaf nodes are distributed in each cluster along with its super node. Again these leaf nodes are parallel connected to each other which are distributed in different clusters along with respected super nodes. This parallel and distributed architecture is easy to construct the network. An overlay network is already robust network that's why there is very less possibility of disconnection.

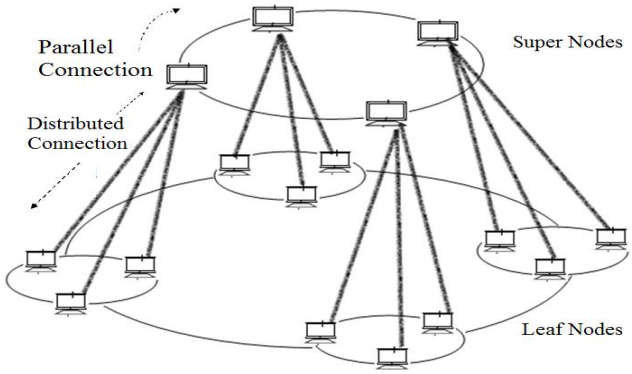

Figure 3: Basic architecture

IV. PERFORMANCE EVALUATION

This section evaluates the experimental results. Experiments are conducted to evaluate the performance of proposed work with existing work [1]. Implementation is done with java language and for database MySql is used with Sqlyog tool. Performance is done on a computer with Intel core i5 2.10GHz processor and 3GB of RAM on 32bit Microsoft Windows 7 operating system used. The comparison is done to analyze number of total nodes, number of remaining nodes, and number of removed nodes in both methods existing and proposed. The experimental result shown in figure 4.The comparison shown in graph of figure 4 explains that Existing column and proposed column are total number of nodes contains each in the network and are 20 nodes each in network. ERemaining is the remaining nodes(15) and EFails column is the failed number of nodes (5) in the existing network after failed node removal and reconstruction process is over. PRemaining column is the remaining number of nodes(19) and PFailes column is the failed number of nodes(1) in the proposed system. That's y the experimental result shows that there are more number of nodes remain in the proposed overlay system as compare to existing overlay system.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

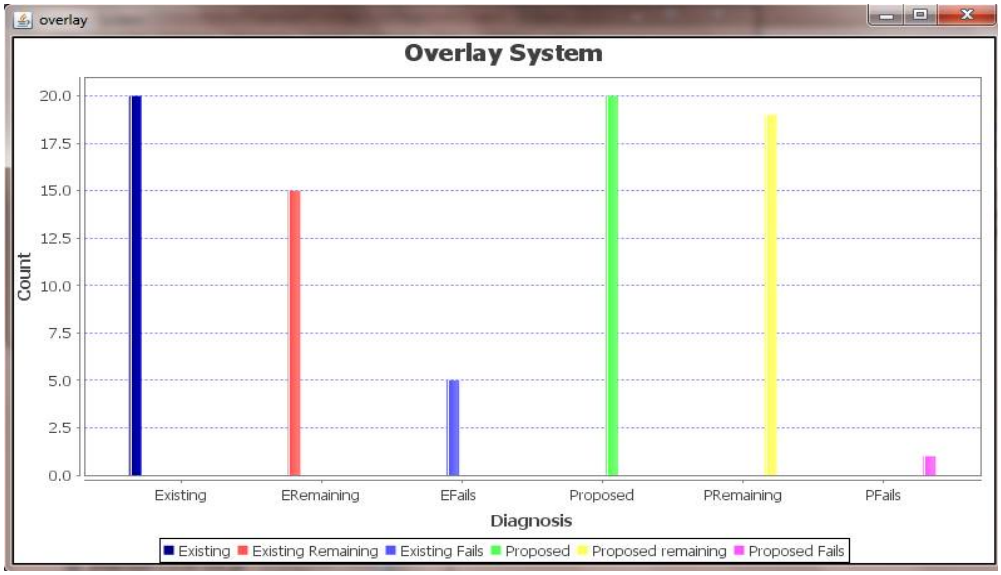

Figure 4: Comparison of existing and proposed system

V. CONCLUSION AND FUTURE WORK

The proposed work of this paper is that the construction of an overlay network with parallel distribute architecture. Neighbour selection with node location is useful for selection of node to remove the failed super node. In the existing system when failed node is occurred it removes whole cluster along with failed super in from the network, that's why there are more number of unnecessary nodes also getting removed. As remedy proposed work firstly removes failed super node and then reconstructs network by connecting new super node. This new super node is taken from same cluster which is the highest degree leaf node. The purpose of removing failed super node is that the failed or faulty super node may reduce the robustness of an overlay network. That's why it is necessary to remove the failed super node and then reconstruct network again to make connection active again, it is useful to maintain high availability of nodes in the network. Additional work is done in this paper is that maintenance of data by applying some database queries to handle data easily. The experimental result shows that there is more number of nodes available in proposed work and it provides good maintenance of data, high service availability and robust connectivity.

For the future work there will be the construction of an overlay network with task allocation scheme of big data mining in real time systems.

REFERENCES

1 Katsuya Suto, Hiroki Nishiyama, Nei Kato, Kimihiro Mizutani, Osamu Ahashi, and Atsushi Takahara, "Overlay Based Data Mining Architecture Tolerant to Physical Network Disruption", IEEE transaction on Emergence topics on computing, Publication- 11-Jun 2014.

2 F. Gebara, H. Hofstee, J. Hayes, and A. Hylick, ``Big data text-oriented benchmark creation for hadoop," IBM Journal of Research and Development, vol. 57, no. 3/4, pp. 10:1_10:6, May-Jul. 2013.

3 K. Suto, H. Nishiyama, X. S. Shen, and N. Kato, ``Designing P2P networks tolerant to attacks and faults based on bimodal degree distribution," Journal of Communications, vol. 7, no. 8, pp. 587_595, Aug. 2012.

4 T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H. E. Stanley, "Optimization of network robustness to waves of targeted and random attacks," Physical Review E, vol. 71, no. 4, p. 047101, Apr. 2005.

5 K. Suto, H. Nishiyama, N. Kato, T. Nakachi, T. Fujii, and A. Takahara, ``THUP: A P2P network robust to churn and dos attack based on bimodal degree distribution," IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, pp. 247_256, Sept. 2013.

6 Y. Zhang, Q. Gao, L. Gao, and C. Wang, ``Priter: A distributed framework for prioritizing iterative computations," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 9, pp. 1884_1893, Sept. 2013.

7 J. Dean, ``Adapt: Availability-aware MapReduce data placement for non-dedicated distributed computing," in Proc. of 32nd IEEE International Conference on Distributed Computing Systems, Macau, Jun. 2012, pp. 516_525.

8 Z. Yao, X.Wang, D. Leonard, and D. Loguinov, ``Node isolation model and age-based neighbor selection in unstructured P2P networks," IEEE/ACM Transactions on Networking, vol. 17, no. 1, pp. 144_157, Feb. 2009.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

9 M. E. J. Newman, ``Assortative mixing in networks," Physical Review Letters, vol. 89, no. 208701, May 2002.
10 J. Dean and S. Ghemawat, ``MapReduce: Simplified data processing on large clusters," in Proc. of 6th Symposium on Operating Systems Design and Implementation, San Francisco, USA, Dec. 2004, pp. 137_150.
11 P. Flocchini, A. Nayak, S. Member, and M. Xie, ``Enhancing peer-to-peer systems through redundancy," IEEE Journal on Selected Areas in Communications, vol. 25, no. 1, pp. 15_24, Jan. 2007.
12 W. Xiao, M. He, and H. Liang, ``Cayleyccc: A robust P2P overlay network with simple routing and small-world," Academy Publisher Journal of Networks, vol. 6, no. 9, pp. 1247_1253, Sept. 2011.
13 I. Stoica,, R. Morris, D. Liben-Nowell, et. al., "Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications," IEEE/ACM Trans. Networking, Vol. 11, No. 1, pp. 17-32, 2003.
14 A. Kumar, S. Merugu, J. Xu, and X. Yu, "Ulysses: A Robust, Low-Diameter, Low-Latency Peer-to-Peer Network," European Trans. Telecommunications, Vol. 15, No. 6, pp. 571-587, 2004.
15 Namrata Jadhav, S. N. Shelke "An Overlay network construction scheme based on node location in physical network", IJAIEM vol. 4, November 2015

BIOGRAPHY

Namrata B. Jadhav is a Student of Master of Engineering receiving Computer Engineering degree in Sinhgad Academy of Engineering from Pune university 2016, pune. Received Bachelor of Computer Science and Engineering degree in 2014 from Dr. Babasaheb Ambedkar Marathwada University Aurangabad.