Novateur Publication's
International Journal of Innovation in Engineering, Research and Technology [IJIERT]
ICITDCEME'15 Conference Proceedings
ISSN No - 2394-3696

DYNAMIC ASSIGNMENT OF USERS AND MANAGEMENT OF USER'S DATA IN SOCIAL NETWORK

Dipeeka V Aher
Savitribai Phule Pune University / Vishwabharti Academy's College Of Engineering,
Ahmednagar, India

Prabhudev Irabashetti
Assistant Professor Savitribai Phule Pune University / Vishwabharti Academy's College Of Engineering,
Ahmednagar, India

ABSTRACT

The issue of dynamic assignment of users to servers is studied widely but the proposed system substantiates the solution for this problem. The system - Dynamic Assignment of users to servers in social networking use community analysis algorithm (CAA) to resolve the extra load that the servers. The proposed system has capability to keep inter-server communication below required level in distributed environment (Online social network). The assessment of potential servers is made first and then proper assignment is done. CAA calculate communication degree (CD) for each user. On the basis of this CD parameter different communities are formed. This CD will help to relation between communication data. Along with previously mentioned features the proposed system will also provide user data management feature which will keep currently activated users who are linked with same event in one virtual group. In this way the proposed system will bound to give the results which will handle the overload of different distributed servers and also manage user data efficiently in system which will give more convenience to the users as well as to the system.

KEYWORDS— community analysis algorithm (CAA), Distributed Algorithms, Online Social Network, management of user data, optimal assignment of users.

INTRODUCTION

Online Social Network (OSN) like Facebook, Twitter runs the largest social networking platform that serves hundreds of millions users at peak times using tens of thousands of servers located in many data centers around the world. There are strict operational requirements on Online Social Network (OSN) platform in terms of performance, reliability and efficiency, and to support continuous growth the platform needs to be highly scalable [1]. For any Online Social Network (OSN) each user profile and its data are stored at its primary servers. Two designated servers S1 and Sn communicate with each other on behalf of their users that is Inter-server communication [2].

This indirect communication architecture enables rich application functionalities that is often difficult otherwise. This indirect communication architecture requires additional resources for handling the inter-communication among servers. In what follows, we will elaborate more on the user server assignment problem. This proposed system claims that Community Analysis algorithm CAA is more suitable for modeling this multi-user operations and multi way interactions. In the proposed system communication degree of each user is initially calculated and then these users are combined in communities.[3]

In many large scale systems, the problem of client-server assignment with a specific objective of reducing the total load while maintaining a good load balance among the servers is crucial. We assume the load at the servers is due mainly to handling messages/information passing among users. Assignment of users to servers is totally based on assessment of available servers. After assignment of users, on basis of users transactions some observations in Online Social Network (OSN) are there [4].

1) The amount of load generated by messages exchanged between two users assigned to the same server will be less than that of when the users are assigned to different servers. This is because of local communication incurred less load than non-local communications.

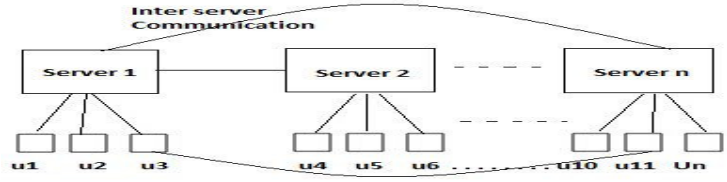

Fig 1 : Inter-Server Communication (u3 → u11 i.e. Server 1→Server n)

2) One cannot assign all users to one server, when he/her do so this server might be overloaded, resulting in degraded performance. Also, from the robustness perspective, this assignment is prone to bottleneck failures [5].

3) It is beneficial to assign two users that exchange messages with each other often to the same server in order to minimize the overall communication load. On the other hand, because of the load balance consideration, two users that rarely exchange messages with each other should be assigned to two different servers [6].

LITERATURE SURVEY

The solution for dynamic assignment of users to distributed servers in Online Social networking applications is need to be joined with user data management. When generally client server assignment problem is compare with k-way graph partitioning problem. In that G=(V,E) V is vertices and E is edges, whereas V nodes should be portioned with k subset such that weight of each server (node) must be balanced.
Dynamic assignment of users is NP Hard and any existing way is not giving exact solution, that's why the proposed system is to focus on balancing the weights on different distributed servers with node failure policy [7]. The graph node partitioning problem studied in many areas such as VLSI (very large scale integration) and many heuristic algorithms was developed similar as Kernighan-Lin Algorithm. There are many approaches used like simulated annealing, genetic algorithm, community structure based method. These all methods is used to exploit structure specific to certain application. Today's Many existing applications are also focusing on inter server message flow to reduce load on cluster. While considering quality of elements at each cluster some algorithms do not consider effect of edge cuts on weight of node. While referring different algorithmic solutions. They work on task scheduling and graph partitioning but focus on size of each partition at the time of managing weights of edge cut.
The performance of Centralized Simulated Annealing (CSA) algorithm is considered as a baseline for evaluating distributed algorithms. The CSA algorithm is based on the simulated annealing (SA) framework which is highly effective in solving many large scale combinational optimization issues.
The proposed system emphasizes on distributed algorithm execution for distribution of load at server which are affected by inter-server communication. There is no existing algorithm which balances load on distributed servers with minimum inter-server communication.

SYSTEM ARCHITECTURE

Proposed system takes a user profiles manually as a data set input. Along with data set the user communication data is taken as synthetic data. Following figure showing detailed Architecture.

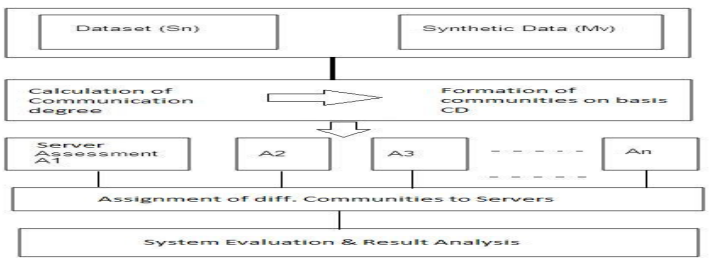

Fig 2: Architecture of Dynamic Users Assignment system

MODULES:

1) DATASET AND SYNTHETIC DATA :

This module gives user profiles as a manual input to the system and along with that user communication is also taken as a synthetic input. The manual communication is to be generated in this module.

2) COMMUNITY FORMATION MODULE :

As request is taken by central algorithm then local assessment of available server s is done by the greedy phase. After that communication degree for each user is calculated. On the basis of that communication degree parameter communities are formed. The users which cross threshold level of communication degree are taken in one community. This process will continue the assignment of users till every server reach its optimum limits [13].

3) ASSIGNMENT MODULE :

Not Only optimum assignment of users is required for maintaining a good performance objective but the real time data management for user must be done. The synthetic data set which shows simulated communication data for users. Formed communities along with their data, are assigned to designated server.

4) EVALUATION MODULE

The Community Analysis Algorithm (CAA) Algorithm is tested against DPGS (Distributed Perturbed Greedy Search) Algorithm and these results are as Follows:

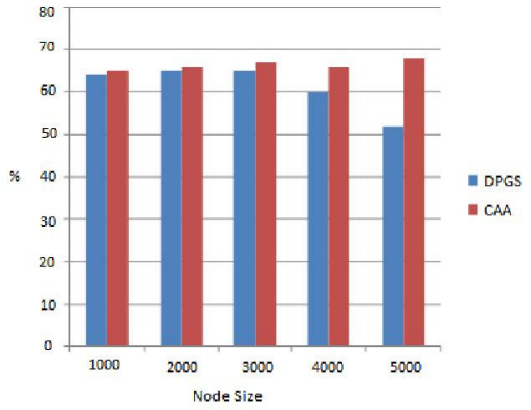

Fig 3: (a) Effect of increased load on accuracy (Accuracy in % against node size)

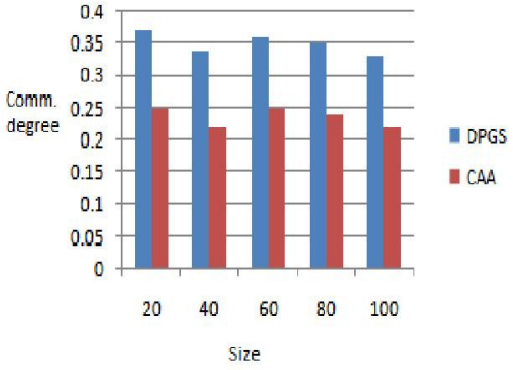

Fig 3: (b) Effect of increased load on Community degree (Size Vs Community degree: CD is difference between common communication of two users, it's Good to be low i.e. CD id lower)

In above fig. 3 (a) it shows that as user load Increases then accuracy of CA algorithm is good as compared to other. In fig 3(b) it shows that CAA works for low CD parameter.

ALGORITHMIC APPROACH

Step 1: The Users are Selected from Dataset and total load on each server is calculated (i.e. The Load equilibrium is established). Total number of active servers is termed here as span of servers.

Step 2: users are selected and their synthetic communication is established in the form of Message and calls.

Step 3: The CD is calculated for each individual user to form a community (higher Message sender is treated as virtual centre of the system). The priority is given to users on the basis of CD.

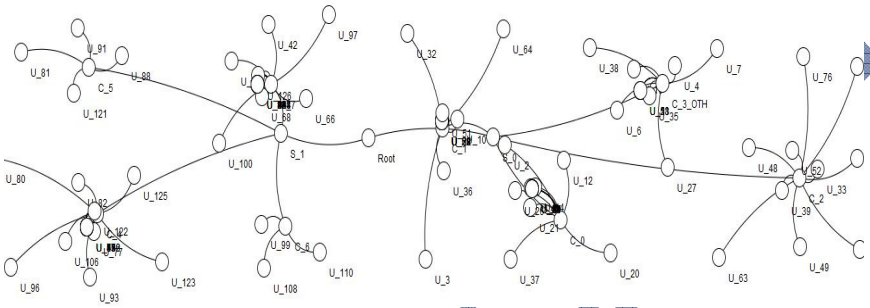

Fig 4 (a): Community Hyper graph

Different users participate in read/write events in social network system. Here one virtual Root node is decided for community. This Root contains available servers. In Fig 4 (a). We have two servers S0 and S1 these user will contain newly formed communities under them. These Communities are dynamic i.e. they can changed as the CD [20] is updated for every individual.

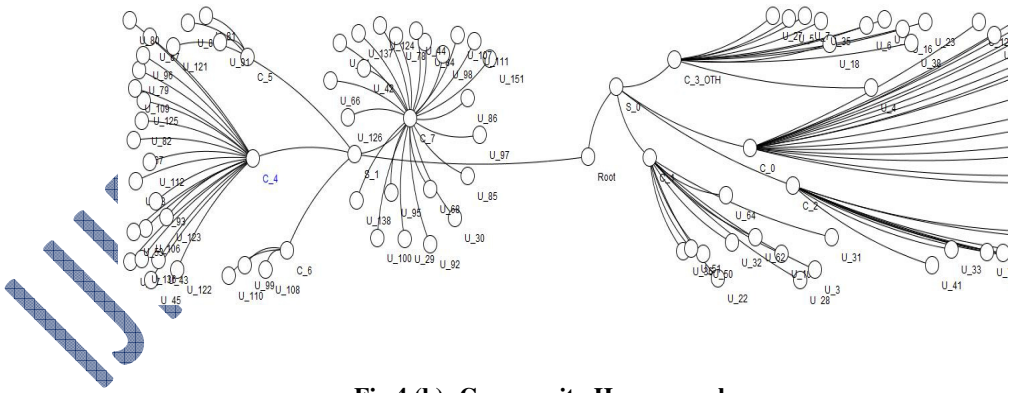

Fig 4 (b): Community Hyper graph

MATHEMATICAL MODELLING

The objective of proposed system is to find approximately optimal client assignment to servers which can maintain certain level of load balance at each server. The mathematical optimization is given below

NOTATIONS USED:

- N: Number of servers
- M: Number of users

- G = (V, E) an edge-weighted graph represents the communication pattern among all users.

V denotes the set of all users i.e. |V|=M. Vertex represents user and E represents total communication among the user. $A^{M\,X\,M}$ An adjacent matrix induced by graph G and $A_{u,v}$=0 if user u and v never communicate. The quantity $A_{u,v}$ can be re-normalized as 0<= $A_{u,v}$ <=1.
X$\epsilon^{\{0,1\}^{M\,X\,M}}$:it's an assignment matrix where $X_{u,i}$=1, if users u is assigned to server i. this is done till given user is assigned to one server exactly i.e. $\sum_{i=1}^{N} X_{u,i}$ = 1. $L^{M\,X\,M}$(X) : this describes load matrix for the assignment X.

S(X) and S_i (X) : the total load on server i from all servers.

$\bar{S}(X)$ The average server load i.e. total load from all servers divided by number of servers.

- The load imbalance at server i is defined as :

$$\Delta S_i(X) = S_i(X) - \bar{S}(X)$$

CONCLUSION
The proposed system uses Community Analysis Algorithm (CAA). It's an iterative algorithm for optimal user assignment to distributed servers. CAA is used for partitioning huge user community in social networks to reduce inter-server communication overhead. The proposed system first calculate the Communication degree for each user and put different users in different communities on basis of multi-way interactions between users in social network. Inter-server communication is basically targeted in proposed algorithm to balance the load of all servers. After efficient assignment of users the management of users data is focused. The partitioning of users also helps in managing I/O Overhead of servers. It can ultimately improves system performance.

ACKNOWLEDGEMENT
With immense pleasure, I am presenting this Project Report on "Dynamic assignment of users and management of user's data in social network" as a part of the curriculum of M.E. Computer Engineering at VACOE. It gives me proud privilege to publish this project paper for this Project Dissertation Stage-II Work under the valuable guidance of Prof. Prabhudeva Irabashatti. I am also extremely grateful to Prof. S G Joshi (HOD of ME Computer Department) and Principal for providing all facilities and help for smooth progress of Project Work. I would also like to thank all the Staff Members of Computer Engineering Department, Management, friends and my family members who have directly or indirectly guided and helped me for the preparation of this Paper and gave me an unending support right from the stage the idea was conceived.

REFERENCES
[1] A. Lakshman and P. Malik, "Cassandra: a decentralized structured storage system,"SIGOPS Oper. Syst. Rev., vol. 44, no. 2, pp. 35–40, Apr. 2010. [Online]. Available: http://doi.acm.org/ 10.1145/1773912.1773922
[2] D. Karger, E. Lehman, T. Leighton, R. Panigrahy, M. Levine, and D. Lewin, "Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web," in Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing, ser. STOC '97. New York, NY, USA: ACM, 1997, pp. 654–663. [Online]. Available: http://doi.acm.org/10.1145/258533.258660
[3] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," SIAM JOURNAL ON SCIENTIFIC COMPUTING, vol. 20, no. 1, pp. 359–392, 1998.

[4] J. Zien, M. D. F. Schlag, and P. Chan, "Multilevel spectral hypergraph partitioning with arbitrary vertex sizes," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, vol. 18, no. 9, pp. 1389–1399, 1999.
[5] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, no. 11, pp. 1101–1113, 1993.
[6] G. Sabin, V. Sahasrabudhe, and P. Sadayappan, "On fairness in distributed job scheduling across multiple sites," in Cluster Computing, 2004 IEEE International Conference on, 2004, pp. 35–44.
[7] M. Garey and D. Johnson, "Computers and intractability: A guide to the theory of np-completeness," Freeman, San Francisco, 1979.
[8] D. S. Johnson, C. Aragon, L. McGeoch, and C. Schevon, "Optimization by simulated annealing: An experimental evaluation," Operations Research, 1989.
[9] T. N. Bui and B. R. Moon, "Genetic algorithm and graph partitioning. IEEE Trans. on Computers, 1996.
[10] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors: A multilevel approach," IEEE Trans. Pattern Anal. Mach. Intell, vol. 29, p. 2007, 2007.
[11] D. Vidyarthi, "Scheduling in distributed computing systems: Analysis, design and models," 2008. [Online]. Available: http://books.google.com/books?id=HqF-OzNstX4C
[12] Y. Jiang, Y. Zhou, and W. Wang, "Task allocation for undependable multiagent systems in social networks," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 8, pp. 1671–1681, 2013.
[13] "Random partitioner," Cassandra Wiki. [Online]. Available: http://wiki.apache.org/cassandra/RandomPartitioner
[14] H. Nishida and T. Nguyen, "Optimal client-server assignment for internet distributed systems," in ICCCN 2011.

[14] H. Nishida and T. Nguyen, "Optimal client-server assignment for internet distributed systems," in ICCCN 2011.

[15] D. A. Tran, K. Nguyen, and C. Pham, "S-CLONE: Sociallyaware data replication for social networks," Comput. Netw. vol. 56, no. 7, pp. 2001–2013, May 2012. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2012.02.010

[16] K. Nguyen, C. Pham, D. A. Tran, and F. Zhang, "Preserving social locality in data replication for social networks," in Proceedings of IEEE ICDCS 2011 Workshop on

Simplifying Complex Networks for Practitioners (SIMPLEX 2011), Minneapolis, MN, June 2011.

[17] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," Decision and Control, 1985 24th IEEE Conference on, vol.

24, pp. 761–767, 1985.

[18] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in Facebook," in Proc. Workshop on Online Social Networks, 2009, pp. 37–42.