

Investigation of Peer Grouping Methods in Peer-to-Peer Computing Networks

Jigyasu Dubey

Department of Information Technology
Shri Vaishnav Institute of Technology & Science,
Indore, India

Vrinda Tokekar, Ph.D

Institute of Engineering & Technology
Devi Ahilya Vishwavidyalaya,
Indore, India

ABSTRACT

The Peer-to-Peer (P2P) technology provides support to build virtual computing system over the Internet which is dedicated for large scale computation problems. In such systems to achieve higher scalability and decentralization participating peers are classified into different groups. In P2P computing systems each peer group is responsible to carry out certain functionality of the system. Selection of different peers in these peer groups i.e. grouping criterion is one of the issues which is to be used to improve performance of the P2P computing systems. In this paper we investigate different grouping strategies possible in P2P computing networks. To compare them parameters like reliability, scalability, execution time etc. are taken into account. This study shows that if participating peers in a peer group are spread over different geographic locations then it make system more reliable.

General Terms

Peer-to-Peer Networks

Keywords

Peers, Groups, P2P, Grouping Mechanism, P2P computing

1. INTRODUCTION

In the recent years Peer-to-Peer (P2P) technology gets the more attention from the research community as well as the industry. Decentralization is one of the main concepts of P2P networks that make it more attractive. The P2P system is a distributed system consists of a set of cooperative computers, known as peers, which share their resources like CPU cycles and memory with other peers in the system without any central authority. It is a virtual network built on top of a physical network. Initially P2P networks are used only for file sharing applications such as Napster [1] and BitTorrent[2]. Nowadays P2P networks are also used for developing large scale distributed computing applications. The SETI@home [3][4] and distributed.net [5] are the well-known examples of such kind of systems. P2P computing is a form of distributed computing that can utilizes the idle CPU cycles of PCs connected on the Internet. The P2P computing systems are dynamic in nature, i.e. peers may join or leave the system at its own decision occasionally. According to the degree of decentralization the P2P systems are classified into two categories: hybrid systems and purely decentralized system [6]. In hybrid P2P systems there is a central peer which maintains the information about all the member peers. This peer is also known as directory server or super peer. The SETI@home and distributed.net falls under this category. In pure P2P system there is no central point of control. All the peers have equal functionality in the system. Pure P2P systems are scalable in nature by avoiding need of centralized operations or servers. JNGI [7] is one of the P2P computing systems based on pure P2P architecture. JNGI divides peers into groups according to their functionalities. Division of

peers into several peer groups limits the amount of communication between the peers hence improves the scalability.

There are several issues which need to be addressed when building generalized pure P2P computing systems. One of them is organization of computational resources into groups [7]. Group is a widely used structure in distributed computing. Grouping refers to the partition of a fixed number of peers into multiple P2P communities. In pure P2P computing systems grouping decisions are important operational issue and need to be considered at the design time. In this study we investigate various grouping mechanism used in P2P computing networks. We also study the effect of grouping mechanism on performance of system on parameters like reliability, scalability, and execution time.

2. RELATED WORK

Jerome Verbeke, Neelakanth Nadgir et al. in [7] presented a decentralized P2P computing framework for large-scale computation problems named as JNGI. In this framework the computational resources are divided into groups according to their functionality. They proposed three peer groups: the monitor group, the worker group, and the task dispatcher group. The design of framework limits communication to small peer groups that enables the framework to scale to a very large number of peers. Jerome Verbeke, et al. in [8] proposed to build new types of groups called similarity groups into the JNGI system. They define a similarity groups into the JNGI system. They define a similarity group as a peer group where all the peers have common characteristics like CPU speed or memory size. These groups can be used either for qualitative (structural) or quantitative (performance) purpose. Their result shows that the uses of quantitative similarity groups increase the performance of a computation while qualitative criterion increases the homogeneity of the computation but not its performance. However peer grouping based on geographic location criteria needs to be considered to improve the reliability. Virginia Lo, et al. in [9] proposed a system named cluster computing on the fly (CCOF) which harvest the CPU cycles from ordinary users (Desktop PCs). They also proposed a wave scheduler which exploits the large blocks of ideal time at night, to provide higher quality of service for deadline-driven jobs, using a geographic based overlay to organize hosts by time zone. In this wave scheduler they explore the possibilities to capture the CPU cycles from number of machines that lie completely idle at night. It provides a higher guarantee of ongoing available cycles hence it is useful for deadline driven tasks. The system provides the higher computation performance but due to using the peers from same night time zone which belongs the same geographic location the reliability of the system decreases.

Bendikt Elser, et al. in [10] defines a concise set of requirements for general, application independent group management models in distributed systems. On the basis of

these requirements they introduced a basic set of group management style. They analyze the requirements that social groups have with regard to groups in distributed system. Here they discuss the problem of group management in fully decentralized social networking systems. This study is intended towards various group management models of social networking system to make them decentralized (P2P) systems rather than for P2P computing systems. Sing Jin Choi, et al. [11] proposed a new group based dynamic computational replication mechanism in a P2P grid computing system. Their proposed mechanism adaptively replicates computation on the basis of volunteer properties. The proposed mechanism dynamically adjusts the number of redundancy, and selects the volunteers that execute the replicated computations accordingly to the respective volunteer group properties. The replication mechanisms not only calculate the number of redundancy, but also select replicas on the basis of volunteer group. Their simulation results shows that the proposed mechanism reduce the number of redundancy and therefore complete more tasks. Yung Ming Li, et al., [12] developed an analytical model to evaluate the impacts of network scale and system parameters on the performance of P2P file sharing network. In this study by analytical and simulation analysis they shows that increasing network scale has a positive effect on the expected content availability and transmission cost, but a negative effect on the expected provision and search costs. They also propose the optimal sizing and grouping decisions. However effects on P2P computing network is not given.

Dubey J. & Tokekar V. in [13] presented an algorithm which categorized the participating peers of a P2P Computing system into eight different categories. To form these categories algorithm uses the values of various peer properties like peer availability, peer credibility, and peer computation time. Among these peer groups the group ‘A’ has the highest probability to complete the computation task into given time period and to produce correct results. They also proposed an algorithm in [14] which identifies the group of reliable peers from the available peers, for the processing of real time application’s tasks. To select the peers, the authors consider joint evaluation of peer computation time, peer credibility, along with turn-around time. The algorithm divides the peers into four peer groups among them group ‘PG1’ is best suited to carry out the real time application’s tasks. In this work authors also give method to calculate turn-around time between peers at application level.

To improve the scalability, pure P2P network architecture divides the computational resources into groups. These groups can be formed on the basis of various criteria like functioning, qualitative, quantitative, distance and geographic location of peers. Choosing one of these criteria is also an issue in design of P2P computing system. The qualitative and quantitative criterion improves the performance of the system however by using the geographic location of peers, the reliability of system increases. The table 1 gives a comparison of some existing peer grouping strategies on the basis of various performance parameters like scalability, reliability, execution time, and resource availability.

Table 1. Comparison of different grouping strategy in existing P2P computing system

S. No.	Name of P2P Computing System	Grouping Strategy	Category	Effect on various Performance Parameters			
				Scalability	Reliability	Execution Time	Resource Availability
1	JNGI	Group Peers based on functionalities in system	Random based grouping	Scalable	Unreliable	Depends on no. of peers available and transmission delay	Maximum
2	Similarity Groups In JNGI	group peers on qualitative / quantitative criteria	Random based grouping	Scalable	Unreliable	By using quantitative criteria computation performance is increased as compare to JNGI	Maximum
3	CCOF	Group the peers by using the rotation of the Earth i.e. by using the time zones.	Location based grouping	Scalable	Unreliable	Computation performance increased because in night time most of the peers ideal.	Maximum in selected time zone.

3. GROUPS IN P2P COMPUTING NETWORKS

In sociology, groups are defined by Muzafer Sherif and Carolyn W. Sherif [15]. In a decentralized P2P computing network we define a group as a computing unit consisting of a number of individual peers interacting with each other with respect to: 1) An accepted distribution of work, i.e. responsibility, 2) Common purpose and objectives; 3) Common set of service and a security level; 4) Recognized position (role, authority); 5) Established standards and principles with reference to issues related to the group; 6) Development of accepted credits (incentive and penalty) if and when Standards were respected or violated.

To formally define a group in a P2P computing system we assume that 'G' is a set of all the peers participating in a P2P computing system. The P2P computing system has 'i' different peer groups P_1, P_2, \dots, P_i . we define a peer group 'P_i' as a subset of 'G' and all the peers in 'P_i' are governed under the set of rules 'R_i' that describe the conditions required to be part of a group and formed based on a particular interest criterion.

$$P_i = \{x \mid x \in G \text{ and } x \rightarrow R_i\};$$

$$P_1 \subseteq G, P_2 \subseteq G, P_3 \subseteq G, \dots, P_i \subseteq G;$$

$$G = P_1 \cup P_2 \cup P_3 \cup P_4 \dots P_{i-1} \cup P_i \text{ where } P_i \neq \phi;$$

$$R_i = \{r \mid r \text{ is a rule or criterion to form } P_i\};$$

P_i should contain at least one peer and at most 'n' peers where 'n' is the maximum number of peers in the P2P computing system.

4. GROUPING STRATEGIES IN P2P SYSTEMS

Choosing different criteria for grouping peers is the first issue to be addressed in the formation of peer groups. In P2P computing systems we classify the grouping strategy mainly in two categories: 1) Location based grouping and 2) Random based grouping. In location based grouping the location property of a peer is used to group peers. Here by location property, we mean the geographic position of peers. In random based grouping, peer's other properties like their CPU speed; operating system, etc. are used to group the peers.

The following are the possibilities to apply location based criteria for grouping peers:

- Group those peers which are belonging to the same geographical area. The same geographical area may be a country or a town.
- Group those computing nodes which have short distance in terms of networking i.e. using the number of routers crossed between two nodes.
- Group the peers by using the rotation of the Earth i.e. by using the time zones. In this criteria group the peers together which belongs to the night time zone at that time. This criterion increases the performance of the system because in night most of the time computing nodes are unused.

To form the peer groups in a P2P computing system the location property is also important. In the above mentioned grouping criteria all the peers in the P2P computing system is chosen so that they belongs to same geographic location (same city, state or country) then in case of electric power

down or blackout situation in a particular area / region, all the peers in that area are also down and not be able to provide their resources which affects the P2P computing system. It may happen that the whole group or whole system shutoff or collapsed. The blackout situation may arise due to electric grid failure (common problem in undeveloped and underdeveloped countries) or any natural disaster like Earthquake, Cyclones, and Tsunami in that area which makes the P2P computing system unreliable.

In random based grouping the peers are grouped based on qualitative or quantitative properties of peers. To group peers on a qualitative criteria, i.e. group those peers which are having the same operating system or running same JVM versions. The qualitative criteria use the logical characteristics of peers. By using this criterion homogeneity of computation is increases but computation performance is not increases. The quantitative criteria use the physical characteristics of the peers like CPU speed, Memory size and network bandwidth. The results of Jerome Verbeke, et al. [8] show that this criterion improves the computation performance.

The above mentioned grouping strategies are based on the static properties of peers. Apart from these properties it also possible to group peers according to their dynamic properties shown by them in a P2P system like peer credibility, computation time, and distance of the peer.

- The peer credibility C_P in [13] given as the probability that the result produced by a peer is correct.

$$C_P = C_R / (E_R + C_R + I_R) \quad (1)$$

Here, E_R represents the number of erroneous results, C_R represents the number of correct results, and I_R represents the number of incomplete results. The term $E_R + C_R + I_R$ represents the number of total tasks that a peer computes.

- The peer computation time (PC_T) in [13] given as the expected Computation time when a peer processes the system's computations during IT.

$$PC_T = IT \times A_P \quad (2)$$

Where IT is peers ideal time and A_P is peer availability. The Peer computation time represent the time when a peer actually executes the system's computations in the presence of peer autonomy failures.

- The distance between peers in [14] defined as the time interval between the instance where task distributor peer T_D dispatch a task unit to task processor peer T_P and the time the result of this task unit is returned to T_D .

Suppose task distributor peer (T_D) dispatches a task unit t_{id} to a task processor (T_P) at time T_S and receives result of t_{id} at time T_C , than turnaround time $D(T_D, T_P)$ is given as:

$$D(T_D, T_P) = T_C - T_S \quad (3)$$

By using above mentioned peer properties peer groups may be formed in a P2P computing system. To achieve higher reliability, peers in a group are chosen which has higher peer credibility values. If in a peer group all the peers have higher credibility then reliability of P2P system will also higher. In case of fast computation need, peers in a group are chosen

based on their values of peer computation time and peer distance. These two factors affect the computation time in P2P system.

To increase the reliability of P2P system it is suggested that to form a peer group geographic properties of peer also considers. To form a peer group the peers are chosen so that they belong to different location, i.e. geographically and time zone. If in a particular area some disaster is happened then only few peers who belong to that area are down or not available but other members of group which are located in different geographic area will be available to continue the

computing. We say that if peers in a group are spreads over different geographic location then it makes system more reliable.

In Table 2 a comparison is given for above mentioned grouping criteria on the basis of scalability, reliability, execution time, resource availability, and transmission delay of P2P computing systems.

Table 2. Comparison of Different Possible Grouping Strategies in Pure P2P Computing Systems.

S. No.	Grouping Strategy	Category	Effect on various Performance Parameters				
			Scalability	Reliability	Execution Time	Resource / Peer Availability	Transmission Delay
1	Group those peers which are belonging to the same geographical area.	Location based grouping	Scalable	Unreliable	Depends on no. of peers available and transmission delay	Limited	Decreases
2	Group those computing nodes which have short distance in terms of networking	Location based grouping	Scalable	Unreliable	Depends on no. of peers available and transmission delay	Limited	Minimum
3	Group the peers by using the rotation of the Earth i.e. by using the time zones.	Location based grouping	Scalable	Unreliable	Computation performance increased because in night time most of the peers ideal.	Maximum in selected time zone.	Decreases
4	group peers on a qualitative criteria	Random based grouping	Scalable	Unreliable	Computation performance is not increased	Maximum	Increases
5	group peers on a quantitative criteria	Random based grouping	Scalable	Unreliable	Computation performance is increased	Maximum	Increases
6	Group peers based on functionalities in system	Random based grouping	Scalable	Unreliable	Depends on no. of peers available and transmission delay	Maximum	Increases

5. CONCLUSION

Peer groups are the fundamental building block of pure P2P systems. In this paper we discuss various possible grouping strategies which may be used in design of pure P2P computing systems. We classify these grouping strategies into two major categories – random based grouping and location based grouping. A comparison of discussed grouping strategies on the basis of various performance parameters like reliability, resource availability, and execution time, is also given here. We also compare some previously proposed P2P computing systems according to the grouping strategies used by them.

When all the members of a peer group belongs to same geographic location then execution time of a task may less due to low transmission delay, but reliability of P2P computing system will decrease due to the reasons discussed in section 4. If the member peers of a peer group are spread over different geographic areas, it has a positive effect on the resource availability and increases reliability of system but a negative effect in form of transmission delay.

6. REFERENCES

- [1] Stern, Richard, "Napster: a walking copyright infringement?" Published in IEEE Micro, Volume:20, Issue: 6, Page(s):4 - 5, 95, ISSN :0272-1732, Year Nov/Dec 2000
- [2] "BitTorrent," <http://bittorrent.com>.
- [3] D. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer. SETI@home: An Experiment in public-resource computing. Communications of the ACM, 45:56–61, 2002.
- [4] Korpela, E., Werthimer, D., Anderson, D., Cobb, J., and Lebofsky, "SETI@home: Massively Distributed Computing for SETI," In Journal on Computing in Science and Engineering, Volume 3, Issue 1, pp. 78 – 83, 2001.
- [5] distributed.net, <http://www.distributed.net>.
- [6] Jigyasu Dubey, Dr. (Mrs.) Vrinda Tokekar, Anand Rajavat, "A Study of P2P Computing Networks", Published in Proceedings of International Conference on Computer Engineering and Technology (ICCET'10), pp 623-627, 13-14 Nov. 2010, JIET group of Institution, Jodhpur, India.
- [7] Jerome Verbeke, Neelakanth Nadgir, Greg Ruetsch, Ilya Sharapov, "Framework for Peer-to-Peer Distributed Computing in a Heterogeneous, Decentralized Environment," In Proceedings of the 3rd International Workshop on Grid Computing, pp.1–12, Year 2002.
- [8] Jean-Baptiste Ernst-Desmulier, Julien Bourgeois and Francois Spies, Jerome Verbeke, "Adding New Features In A Peer-to-Peer Distributed Computing Framework," In Proceedings of the 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing (Euromicro-PDP'05), pp.34 – 41, 2005.
- [9] Virginia Lo, Daniel Zappala, Dayi Zhou, Yuhong Liu, and Shanyu Zhao, "Cluster Computing on the Fly: P2P Scheduling of Idle Cycles in the Internet", In Proc. 3rd International Workshop on Peer-to-Peer System (IPTPS 2004). San Diego, Feb. 2004.
- [10] Benedikt Elser, Georg Groh, Thomas Fuhrmann, "Group Management in P2P Networks" Proceedings of 19th International Conference on Computer communications and Networks (ICCCN) 2010, Zurich, pp. 1- 8, 2-5 Aug. 2010
- [11] SungJin Choi, MaengSoon Baik, JoonMin Gil, ChanYeol Park, SoonYoung Jung, and ChongSun Hwang, "Group-based Dynamic Computational Replication Mechanism in Peer to Peer Grid Computing," IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID 2006), Sixth International Workshop on Global and Peer to Peer Computing (GP2P), May 2006.
- [12] Yung-Ming Li, Yong Tan, and Yong-Pin Zhou, "Analysis of Scale Effects in Peer-to-Peer Networks", IEEE/ACM transactions on networking, VOL. 16, NO. 3, JUNE 2008, pp.-590 – 602
- [13] Jigyasu Dubey, Dr. (Mrs.) Vrinda Tokekar, "Identification of Reliable Peer Groups in Peer-to-Peer Computing Systems", published in proceedings of Third International Conference on Advances in Communication, Network, and Computing – CNC 2012, LNICST Volume No.108, pp. 233-237, Bangalore, India, 978-1-936968-61-9.
- [14] Jigyasu Dubey, Dr. (Mrs.) Vrinda Tokekar, "Identification of Efficient Peers in P2P Computing Systems for Real Time Applications", Published in International Journal of Peer to Peer Networks (IJP2P) Vol.3, No 6, November 2012.
- [15] Muzafer Sherif and Carolyn W. Sherif. An Outline of Social Psychology, pages 143–180. Harper & Bros, 1956.