Identification of Reliable Peer Groups in Peer-to-Peer Computing Systems

Jigyasu Dubey[1] and Vrinda Tokekar[2]

[1] Shri Vaishnav Institute of Technology & Science, Indore, India
[2] Devi Ahilya Vishwavidyalaya, Indore, India
jigyasudube@yahoo.co.in, vrindatokekar@yahoo.com

Abstract. The Peer-to-Peer (P2P) computing systems are based on desktop computers (PCs) exists at the edge of Internet. This type of frameworks contains several thousands of computing nodes which spread all over the world and need to be organized. To achieve decentralization the P2P computing architectures classify the peers into different groups based on different peer properties. In order to improve the reliability, scalability, and performance of decentralized P2P computing systems efficient peer grouping strategy is required. Here we proposed an algorithm to identify the reliable peer groups in P2P computing systems by using the peer properties like peer availability, credibility and computation time.

Keywords: Peer, P2P, Group.

1 Introduction

A peer-to-peer computing system is a platform that achieves a high throughput computing by utilizing the CPU power of numbers of idle desktop/PCs which are known as peers and exists at the edge of the Internet [1]. JNGI [2] is one of the first P2P computing systems based on pure P2P architecture (decentralized). It considers the dynamism of the P2P environment and maximizes the utilization of unpredictable resources. It divides computational resources in to groups to increase the scalability of system and to achieve decentralization. Division of computational resources into several peer groups limits the amount of communication between the peers and avoids the bottleneck in system. This in turn improves scalability. Currently there are several issues which need to be addressed when building a generalized framework for pure P2P computing systems such as decentralizing the task of job submission and result retrieval, fair allocation of resources, problem of free riding and security [3]. Most of the research work in the area of P2P computing system is focused around these issues. Apart from these issues one important issue, peer grouping criterion is there in design of pure P2P computing systems which requires more attention from the research community. In pure P2P computing systems peer groups will be formed such that they can improve the performance of the P2P computing system. Jerome Verbeke, Neelakanth Nadgir et al. in JNGI [2] divide computational resources into three peer

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 233–237, 2012.
© Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2012

groups according to their functionalities. Jerome Verbeke, et al. in [4] builds similarity groups in JNGI. In similarity group all the peers have common characteristics like CPU speed, memory size, operating system, or JVM versions. These groups can be used either for qualitative (structural) or quantitative (performance) purpose. However peer grouping based on geographic location criteria needs to be considered to improve the reliability. CCOF [5] harvest the CPU cycles from ordinary users (Desktop PCs) at night time, to provide higher quality of service for deadline-driven jobs, by organizing the host according to their time zones.

The grouping criterion plays an important role in order to get the maximum throughput from the system. The bad strategy to group peers leads the P2P computing system towards the performance degradation. Here we present an algorithm to identify the reliable and efficient peer groups based on peer properties in P2P systems.

2 Identification of Reliable Groups

A P2P computing system is based on desktop PCs which are connected to the Internet. The peers in the system can freely join and leave the system, in between the computation, without any constraints. The peers are not totally dedicated to the P2P computing system. The systems computations get temporarily stopped by the execution of a private job of PCs personal user. These situations are called as "peer autonomy failures" because it leads to the delay and suspension of computation and may partial or entire loss of the computations. The performance of a P2P computing system is strongly affected by the peer's computation time, peer's availability, and peer's credibility. In P2P computing system the task completion time is strongly dependent on above mentioned properties of peers. These three properties can also be used to form the different peer groups in pure P2P computing systems. SungJin et al. [6] defines the peer ideal time, peer availability, and peer computation time as follows:

- Peer's Ideal Time: The Ideal time (IT) of a peer is the total time period when a peer is supposed to perform the computing work or active in the group.

  $$\mathrm{IT} = \mathrm{IT_S} + \mathrm{IT_P} \tag{1}$$

  Here, IT_S represent the idle time. It is defined as the time period when a peer is supposed to provide its computing resource to the system. A peer mostly performs system's computations during IT_S and rarely perform PC user's personal tasks. The IT_P represents the unexpected personal computation time. The peer usually performs PC user's personal computations during IT_P and rarely performs the system's computations.

- Peer Availability: The peer availability (A_P) is the probability that a peer is operating correctly and is able to perform the computations during the Ideal time (IT). In a P2P computing system, the computation is more frequently delayed and suspended by peer autonomy failures. The availability must reflect peer autonomy failures.

$$A_P = MTTPAF / (MTTPAF + MTTR) \quad (2)$$

Here, the *MTTPAF* represents mean time to peer autonomy failures and the *MTTR* represents mean time to rejoin. The *MTTPAF* represents the average time before a peer autonomy failures happen, and the *MTTR* represents the mean duration of peer autonomy failures. The A_P reflects the degree of peer autonomy failures.

- Peer Credibility: The peer credibility C_P is the probability that the result produced by a peer is correct.

$$C_p = C_R / (E_R + C_R + I_R) \quad (3)$$

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.

- Peer Computation Time: The peer ideal time (*IT*) does not reflect the peer autonomy failure. If a peer suffers from peer autonomy failures, the time duration for which peer computes the system task is decreases thus peer computation time (PC_T) is more important. The peer computation time (PC_T) is the expected Computation time when a peer processes the system's computations during IT.

$$PC_T = IT \ \text{X} \ A_P \quad (4)$$

It represent the time when a peer actually executes the system's computations in the presence of peer autonomy failures.

The peer groups are constructed by the algorithm of peer group construction as given below in figure 1(a) , the peers are classified into *A, B, C, D, E, F, G,* and *H* classes depending on the peer availability(A_P), peer computation time (PC_T), and peer credibility(C_P).

In figure 1(b) we show a unit cube. The three dimensions of the cube correspond to the three important peer characteristics which affect the performance of a peer group. The vertical dimension represents the peer availability (A_P), horizontal dimension represents peer computation time (PC_T), and dimension perpendicular to plan represents the peer credibility (C_P). We divide this cube into eight equal volume sub-cubes *A, B, C, D, E, F, G,* and *H* as shown in the figure 1(b) which corresponds to the peer groups constructed by the algorithm.

- The group '*A*' (sub cube A in fig 1(b)) represents a peer groups in which all the peers have high values of A_P, PC_T, and C_P. In group '*A'* all the peers have high possibilities to execute task reliably because they have high credibility as well as availability.
- The group '*B*' (sub cube B in fig 1(b)) represents a peer group in which all the peers have high values of A_P, and PC_T but low values of C_P. it means that the peer group has high possibility to complete the task; however its results might be incorrect.

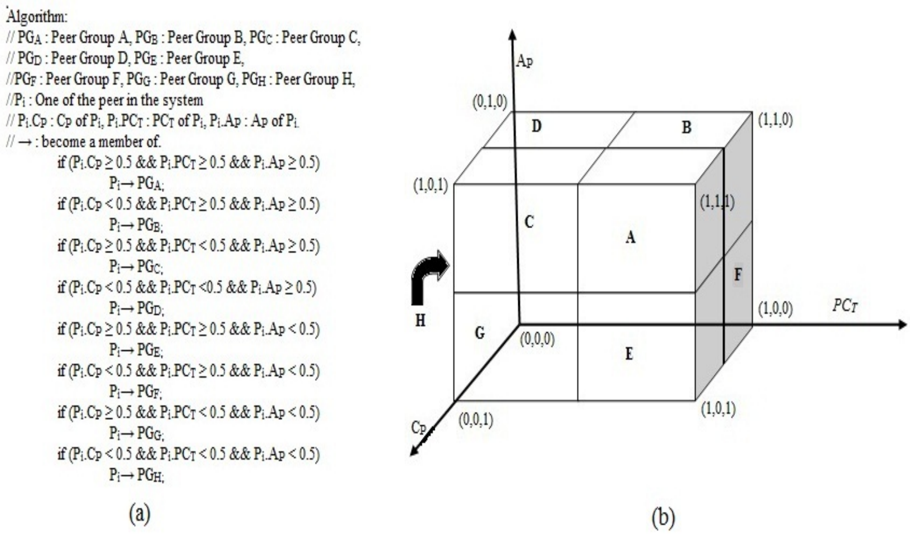

Fig. 1. (a) Algorithm for peer group construction (b) Categorization of peer groups

- The group '*C*' (sub cube C in fig 1(b)) represents a peer group in which all the peers have high values of A_P, C_P, but low values of PC_T. The peer group has the high possibility to produce correct results, however it cannot complete the assigned task because lack of peer computation time.
- The group '*D'* (sub cube D in fig 1(b)) represents a peer group in which all the peers have high values of A_P, but low values of PC_T, and C_P. this peer group has low probability that it can complete the task due to lack of peer computation time and also results produced by it might be incorrect.
- The group '*E*' (sub cube E in fig 1(b)) represents a peer group in which all the peers have high values of PC_T, and C_P but low values of A_P. In this peer group peers have small availability but high peer computation time so there is possibility to complete the computation task with correct results.
- The group '*F*' (sub cube F in fig 1(b)) represents a peer group in which all the peers have high values of PC_T, but low values of C_P and A_P. this peer group has less peer availability and credibility hence it cannot complete the task and not recommended to use for computations.
- The group '*G*' (sub cube G in fig 1(b)) represents a peer group in which all the peers have high value of C_P, but low values of A_P, and PC_T. this peer group has least probability to complete the task and it is not recommended to use for the computations.
- The group '*H*' (sub cube H in fig 1(b)) represents a peer group in which all the peers have low values of A_P, PC_T, and C_P. this group is also not recommended to use for the computations.

3 Conclusion

In pure P2P computing system peer's various properties like peer availability, peer credibility, and peer computation time may also be used to group the peers. Here we proposed an algorithm which categorized the existing peers in a P2P computing system into eight different categories according to the values of above mentioned peer properties. In group '*A*' all the peers have high values for reliability, credibility, and computation time hence this group has the highest probability to complete the computation task into given time period and to produce correct results. The group '*A*' may be very useful for the real time and dead line driven computations. The group '*B*' may be useful in such computations were time deadline is important and fraction of error in computation is acceptable. The group '*C*' and '*E*' may be used for those computations in which accuracy is must but having no time dead line. The group '*D*' can be used for those computations were fraction of error is acceptable and also having no time deadline for the completion of computation. The group '*F*', '*G*', and '*H*' are not recommended to use for the computation purpose.

References

1. Dubey, J., Tokekar, V.: P2PCS – A Pure Peer-to-Peer Computing System for Large Scale Computation Problems. In: IEEE International Conference on Computational Intelligence and Communication Networks (CICN), October 7-9, pp. 582–585 (2011)
2. Verbeke, J., Nadgir, N., Ruetsch, G., Sharapov, I.: Framework for Peer-to-Peer Distributed Computing in a Heterogeneous, Decentralized Environment. In: Parashar, M. (ed.) GRID 2002. LNCS, vol. 2536, pp. 1–12. Springer, Heidelberg (2002)
3. Dubey, J., Tokekar, V., Rajavat, A.: A Study of P2P Computing Networks. In: International Conference on Computer Engineering and Technology (ICCET 2010), November 13-14, pp. 623–627. JIET group of Institution, Jodhpur (2010)
4. Ernst-Desmulier, J.-B., Bourgeois, J., Spies, F., Verbeke, J.: Adding New Features In A Peer-to-Peer Distributed Computing Framework. In: 13th Euromicro Conference on Parallel, Distributed and Network-Based Processing (Euromicro-PDP 2005), pp. 34–41 (2005)
5. Lo, V., Zappala, D., Zhou, D., Liu, Y.-H., Zhao, S.: *Cluster Computing on the Fly*: P2P Scheduling of Idle Cycles in the Internet. In: Voelker, G.M., Shenker, S. (eds.) IPTPS 2004. LNCS, vol. 3279, pp. 227–236. Springer, Heidelberg (2005)
6. Choi, S.J., Baik, M.S., Gil, J.M., Park, C.Y., Jung, S.Y., Hwang, C.S.: Group-based Dynamic Computational Replication Mechanism in Peer to Peer Grid Computing. In: IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID 2006), 6th International Workshop on Global and Peer to Peer Computing (GP2P) (May 2006)