Peer to Peer Protocols in Grid Computing

K. Shenbagavadivu
Research Scholar
Department of Computer Science
Bharathiar University, Coimbatore.

D.Ramyachitra
Assistant Professor
Department of Computer Science
Bharathiar University, Coimbatore

ABSTRACT

GRID computing is a technology for coordinating large scale resource sharing and problem solving among various groups. Grid computing is a form of network. Such networks are unlike conventional networks to concentrate on communication between devices, grid computing controls unused processing cycles of all computers in a network for solving problems too rigorous for any stand-alone machine. Grid technologies are different from other major technical trends such as internet, distributed networks and peer to peer computing. Also it has some issues in resource allocation, QoS, data management, scheduling, accounting and performance. In this paper a study on peer to peer protocols used in grid computing is given. It describes about the peer to peer protocols in grid computing and it also explains about the design approach and usage of the protocols.

Keywords: Grid computing, peer to peer protocol, file sharing, distributed Computing.

1. INTRODUCTION

Grid computing has been started as a project to link the distributed supercomputers, but now it has grown far beyond to reach its goal. The Grid technologies have been used in various applications such as data evaluation, quantity high computing and distributed supercomputing. It is well established as a prototype for distributed computing [1]. An idea is to apply peer to peer concept in the Grid systems.

A peer-to-peer computer network is one in which each system in the network can act as a client or server for the added computers in the network which allows shared access to a variety of resources such as records, peripheral devices and sensing devices without the need for a central server. Peer to peer networks can be set up within the home and is utilized for various business aspects through the Internet.

In Grid environment peer to peer technology is very important. Peer to peer protocol is a request-response protocol. It means peers or clients can send requests so that the other peers can answer for it. The transmission of requests and responses is maintained by a request and response state machine [1].

Peer to peer (p2p) is a group of technologies that allows you to put into place a data exchange system within multiple users. The advantage of a "peer-to-peer" system is that there is a network without hierarchy among users, so it doesn't require servers with extreme amounts of bandwidth existing where everything is centrally located. The exchanges occur directly from one computer to another.

Peer-to-Peer system is quite exceptional from a cost implementation view when compared to a traditional server and its ease of use; it is too quickly associated with piracy and illegal operations. Simply, peer-to-peer offers access to a easy, quick and more or less free network where you can download and transmit your data, it requires no particular technical knowledge or heavy, expensive tools. A net connection also requires a lot of hard drive space. On the other hand, your Internet bandwidth will also be used by others so as to carry on the p2p system [2].

P2P allows file sharing or computer resources and services by direct exchange involving systems, or it permits the use of devices on the Internet side-line in a non client facility. Also it is defined through three key requirements. They are (i) Operational computer of server quality, (ii) DNS independent addressing system and (iii) Manage with variable connectivity.

The remaining portion of the paper is discussed as follows. Section 2 explains the Grid infrastructure, Section 3 describes about the peer to peer protocol and its taxonomy, Section 4 describes about the protocols used in grid computing and its design approach with its usage and advantages, Section 5 explains about the application of peer to peer protocol, Section 6 describes about peer to peer protocol challenges and finally Section 7 gives the conclusion.

2. GRID INFRASTRUCTURE

Grid users generally have access to one or two clusters and have to distribute their computational time with others. So unable to run computations, it takes more time to complete because they are not allowed to use all the resources wholly for their experiments.

Simultaneously, the researchers work in labs or institutions, which have a huge number of desktop machines. Those desktops are generally not utilized and are only available to a single user. They also are extremely unpredictable (e.g. shutdown, reboot, failure).Organizing such desktop machines as a P2P network for computations or other kinds of resource sharing is now more popular. Still now existing models and infrastructures for P2P computing are limited as they support only independent workers task, usually without communications between tasks [19].

On the other hand P2P computing seems well adapted to applications with low communication/computation ratio, such as parallel search algorithms. This concept explains the P2P infrastructure of computational nodes for distributed communicating applications.

The P2P infrastructure is an unstructured P2P network, such as Gnutella [2]. In contrast to others P2P approach for computing, which are typically hierarchical or master-salve, is innovative in the way that an unstructured P2P network generally used for file sharing can be also used for computing.

The P2P infrastructure has three main characteristics. Firstly, the infrastructure is decentralized and completely self-organized. Second, it is flexible; parameters are used for adapting the infrastructure to the location where it is deployed. Finally, the infrastructure is portable since it is built on top of Java Virtual Machines (JVMs). Thus, the

infrastructure provides an overlay network for sharing Java Virtual machines.

3. PEER TO PEER PROTOCOL AND ITS TAXONOMY

Peer to peer protocol can be classified into two types namely,

(i) Peer to Peer streaming peer protocol
(ii)Peer to Peer session initiation protocol

Peer to Peer streaming peer protocol

The Peer-to-Peer Streaming Peer Protocol (PPSPP) is a transport protocol for disseminating the similar content to a group of people in a streaming fashion. PPSPP carries streaming of both pre-recorded (on-demand) and live audio/video content. It is based on the peer-to-peer model, where clients consuming the content are put on equal footing with the servers initially provides the content, to produce a system where everybody can potentially afford bandwidth. It has been intended to provide short time till playback for the end user and to prevent distraction of the streams by mean peers [3].

PPSPP has also been considered to be flexible and extensible. It can use various mechanisms to optimize peer uploading, avoid free riding, and work with different peer discovery schemes such as centralized trackers or Distributed Hash Table. It supports multiple methods for content integrity protection and mass addressing. Designed as a generic protocol that can run on top of a variety of transport protocols, it currently run on top of UDP using LEDBAT for congestion control [3].

Peer to Peer session initiation protocol

The Session Initiation Protocol (SIP) is an application layer protocol standardized by the Internet Engineering Task Force (IETF) which is used to create, modify and terminate sessions such as Voice over Internet Protocol (VoIP) and multimedia conferences [4]. Proposed as a member of a larger multimedia construction. Session Initiation Protocol was designed with the native ability to incorporate a range of IETF protocols such as Real-time Transport Protocol (RTP) [5], Real time Streaming Protocol (RTSP) and Media Gateway Control Protocol (MEGACO) [6].

Peer to Peer Taxonomy

Distributed Computing

In recent times there are more tools than ever to help companies unused computing power in the hundreds of PC being used by their employees. The massive power of P2P computing is increasingly being blended with its more complex distributed (or grid) computing [7].

Traditionally, there have been three categories of "distributed" computing:

(i) Cluster computing: Related machines (usually servers of related power and configuration) are attached to form a necessary machine.

(ii)Peer-to-peer: Many desktop computers are connected to collective process powers. The distinguishing characteristic is the machine itself, which almost entirely is a low-power client PC. Often, the link is through the Internet.

(iii)Distributed computing: Mostly known as grid computing, it connects a large variety of computer types and its resources such as storage area networks, to produce vast virtual reservoirs of computers serving geographically among separated users.

The clustering retains a distinct position, although there is a link between P2P and distributed computing. One reason is that many distributed computing software vendors are building it easier to integrate a mix of ordinary servers and even high-end multiprocessor servers as nodes in a peer-to-peer computing system.

File sharing

File sharing is a popular method for exchanging large files in the internet world, mainly audio's and video's. Most Peer to peer file sharing systems does not use any central servers but instead it permits all computers on the network to function both as a client and a server.

Several free P2P software programs exist each with their own technical advantages and loyal community following. Instant messaging systems are a type of P2P application most commonly used for conversation, but all trendy software also supports sharing files [9].

Collaborative Computing

Collaborative Computing is a range of activities where people interact with everyone using desktops, laptops, palmtops, and complicated digital cellular phones. As computers are best at managing data and representing information, human communication is enriched by the capability to share the resources, modify them and collaboratively create data and information. So, collaborative computing is the future of the Internet.

The Internet will develop from its current role as a channel for information distribution to a person-to-person communication medium. This is exactly constructed, deployed and supports the tools for personal interaction over the Internet [8].

Grid computing platform

When there is a need to run more iteration, simulations, analysis and get business results sooner and when there is either big data and/or compute-intensive problems to solve. So there is a need to maximize the potential of the computing power and the supporting infrastructure to speed up the application at a particular scale, extract insight from the data and make better decisions quicker [10].

It requires IBM Platform Computing of high performance with low latency systems containing management solutions and services to pool the technical computing resources. Managed them powerfully across several groups and get the most prevalent as IT investment.

At present even faster, with improved security, more flexibility and scale to over 150,000 cores, IBM helps to optimize and manage the first departmental cluster through highly protected multi-site grids and HPC clouds [11].

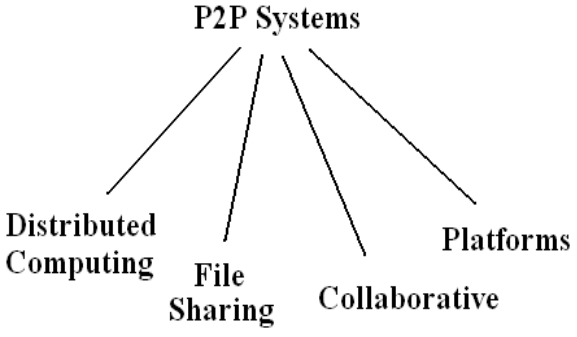

Fig. 1: Peer to Peer System Taxonomy

4. PROTOCOLS USED IN GRID COMPUTING AND PROTOCOL DESIGN APPROACH

The protocols associated with each layer in the grid architecture are (i) Grid security infrastructure (ii) Grid resource allocation management (iii) Grid file transfer protocol and (iv) Grid information services.

Grid security infrastructure

The grid security infrastructure for grids have been defined by creating extensions to standard and well known protocols and APIs. The Grid security protocol should have a robust API/SDK that allows direct calls to the various security functions [12].

Grid resource allocation management (GRAM)

Grid resource allocation management protocol and client API allows programs to be started on remote resources. [12]. GRAM protocol is a simple, HTTP based remote procedure call. It sends messages such as job request, job cancel, status and signal. Resource Specification Language is a common notation for exchange of information between application, resource brokers and resource managers.RSL provides two types of information. They are (i**)** Resource requirements such as System type, types of nodes, memory status etc (ii) Job configurations such as directory, executable, arguments and environment.

Grid file transfer protocol

For data transfer, at first the data's are being collected at one location while the researchers who need access to the data are distributed across the globe. One of the key requirements for the data intensive grids is high speed and reliable access to remote data. The standard FTP protocol has been extended at the same time as preserving interoperability with existing servers to develop GridFTP [11].The extensions provide for striped/parallel information channels, partial records, automatic and manual TCP buffer size setting, evolution monitoring and extended restart functionality. The protocol extension to FTP for the grid (GridFTP) has been submitted as a draft to the Global Grid Forum Data Working group [12].

Grid information services

A set of protocols and APIs are defined in the resource layer that provides key information about the grid infrastructure. Grid information service (GIS) provides access to static and dynamic information regarding a grid's various components and includes the type and state of available resources [11].

There are two types of Grid information services. The Grid Resource Information Service (GRIS) and the Grid Index Information Service (GIIS).The GRIS supplies information about specific resource while the GIIS is a combined index service. GIIS provides a collection of information that has been collected from multiple GRIS servers.

The Grid Resource Registration Protocol is used by resources to register with the GRIS servers. The Grid Resource Inquiry Protocol is used to query a resource description server for information and also query the aggregate server for information [12].

Protocol Design Approach

There are three high-level requirements for a peer-to-peer protocol [13].The requirements are (i) Resource publishing and lookup (ii) P2P network maintenance (iii) Heterogeneous connectivity.

Resource publishing and lookup

The peer to peer protocol provides a mechanism for a peer to distribute a resource-object and a methodology to look at the resource-object and the node offering a service.

P2P network maintenance

For maintaining connectivity and resource availability the protocol has provided a mechanism in a peer-to-peer network.

Heterogeneous connectivity

Nodes should be able to form an overlay in various network environments and exchange information about their uptime and capacity.

Usage of Peer to Peer Protocols

Peer to Peer Protocol allocates nodes to form a structured or an unstructured overlay. Nodes that take part in the routing and overlay maintenance decisions are called peers (super-nodes in conventional P2P terminology). Nodes that do not take part in an overlay are called clients (ordinary-nodes in conventional P2P terminology). Clients can use overlay services throughout peers. Multiple clients can take part in the overlay through one peer. Alternatively, a client can communicate with multiple peers at the same time. Client and peers transmit using the same protocol as peers do, i.e., there is no different protocol for client-to-peer and peer-to-peer communication [14]. The process to discover a peer already in the overlay is outside the scope of this specification. A number of techniques such as cache of nodes during last connection, multicast, or a suitable rendezvous method can be used. Once a node already in the overlay has been revealed, a peer may be admitted into an overlay. The requirement does not specify the conditions under which a peer may or may not be admitted into the overlay and leaves policy decisions such as admission control to the implementers of an overlay. This presents P2PP its desired flexibility.

The protocol permits nodes to CPU and bandwidth utilization, exchange uptime, network bandwidth and network connectivity in order with other nodes on the other hand, it does not mandate the exchange of this information.

The protocol can be used in different network atmosphere. In an Internet deployment, peers can have various connectivity. An overlay designer may organize a policy such that only peers with a public IP address can contribute in the structure overlay. In a business deployment, all of the nodes may be behind firewalls which can participate in the overlay. The protocol does not insist on any restriction on which nodes may participate in the overlay and leaves this decision to the overlay implementer [15].

Advantages

The advantages of Peer-to-Peer Networking over Client Server networking are [16].

(i) In peer to peer networking, the configuration of computers on the network is easy.

(ii) The distributed resources and contents are shared by all the peers, not like server-client structural design where Server shares all the contents and resources.

(iii) P2P is more reliable as central dependency is detached. Breakdown of one peer doesn't affect the implementation of other peers.

(iv)There is no need for full-time System Supervisor. Every user is the supervisor of his device. User can control their shared resources.

(v) The overall cost for construction and maintenance for this type of network is comparatively very less.

(vi) Well tested simplicity

(vii) Easy to install.

5. APPLICATIONS

P2P applications organize and manage resources at the edges of the Internet in a decentralized approach with little or no communication with central servers. The resources may be storage or content (e.g. file sharing applications), processor cycles, or human presence. The edges of the Internet are machines such as the average home PC which has just a single line to connect it to the vast Internet cloud. With the spirit of P2P network, applications of it grow constantly, from file sharing to real time applications. [16]

KaZaA

KaZaA is one of the few applications that address this issue by attempting to stream downloads from multiple sources and automatically resumes downloads on failure. This technology is under developed and so it needs to be developed further before it can be considered a successful solution to the problem.

GNUtella

The networks such as Napster and GNUtella downloads of files can quite commonly be interrupted or cancelled entirely by clients becoming unresponsive or logging off the network and dropping all their connections. This can make downloading of resources a complicated one and for P2P applications to be widely accepted must be addressed first [17].

Emule

Content storage and exchange is one of the areas where P2P technology has been most successful. File sharing applications focus on storing and retrieving information from various peers in the network. One of the best known examples of such P2P systems is Emule, KaZAa.

SETI@home

These applications use resources from a member of network computers. The general idea behind these applications is that idle cycles from any computer connected to the network can be used for solving the problem of the other computers that wants more computation. SETI@home is an example of Distributed Computing **[17].**

Groove and Jabber

Collaborative P2P applications aim to allow application level collaboration among users. These applications range from instant messaging, chatting, online games and shared applications that can be used in business, education and home environments. Groove and Jabber are the examples of such systems.

6. CHALLENGES IN PEER TO PEER PROTOCOL

The main challenge in P2P computing is to design and implement a tough distributed system composed of distributed and a variety of peer nodes, located in distinct administrative domains. In a typical P2P system, the participants can be domestic or enterprise terminals connected to the Internet.

Powerful PCs and the Internet enable people to come together through a common atmosphere. Many people are taking information sharing into their own hands, and companies are verdict themselves in a situation of trying to control people and influence the use of the Internet [18]. Future trends and challenges include:

(i)Record Companies: It generates a copy of protected media format for music files designed to avoid people from sharing prohibited copies of music files.

(ii)Corporations: growing network security levels on corporate networks due to employees granting illegal user's access to their PC to enable file sharing.

(iii)Education Organizations: changing the learning surroundings from traditional classrooms to effective classrooms and developing/implementing effective peer-to-peer learning opportunities.

7. CONCLUSION

Peer-to-peer computing is becoming more and more prevalent as people are using their own PCs to bypass central servers to connect directly among all. This is changing the way people shares information; act as a team to learn properly. Although peer-to-peer computing gives incredible control to the individual, the outcome is somewhat unpredictable. From this survey it can be concluded that a peer to peer protocols have vast range of computing in grid environment. Thus new innovations are encouraged in the challenging area of grid computing.

REFERENCES

[1] Foster and Kesselman, C.1999, Computational Grids; The Grid: Blueprint for a New Computing Infrastructure.

[2] Godwin and Anand, P.2012, Peer to Peer Networks Management Survey "IJCSI International Journal of Computer Science" (January 2012) .

[3] http://datatracker.ietf.org/doc/draft-ietfppsp-peer-protocol

[4] http://eprints.ru.ac.za/1115/1/MosiuoaTsietsi.pdf

[5] Rosenberg, J. Schulzrinne, H.Camarillo, G. Johnston, A. Peterson, J. Sparks, R. Handley, M. and Schooler, 2002. RFC 3261: SIP: Session Initiation Protocol.

[6] Schulzrinne, H. Casner, S. Frederick, R. and Jacobson.,V.1996 , RFC 1889: A Transport Protocol for Real-Time Applications, (January 1996).

[7] Jim pinto and San Diego,2001, Distributed and Grid Computing.

[8] http://www.collabworx.com

[9] http://www.druva.com/insync/file-sharingcollaboration.

[10]http://www03.ibm.com/systems/technicalcomputing/ platfomcomputing/index.html.

[11] http://compnetworking.about.com

[12] Ahmar Abbas, 2009, Grid Computing: A Practical guide to technology and applications

[13] http://www1.cs.columbia.edu/~salman drafts/draft-baset-p2psip-p2pcommon01.html

[14] www.cse.hcmut.edu

[15] Schulzrinne, H. Rao, R,.and Lanphier. R.1998RFC 2326: Real Time Streaming Protocol,(April1998).

[16] http://www.ianswer4u.com/2011/05/peer-to-peer-network-p2p-advantages.

[17] Cuervo, R. Greene, N. Rayhan, A. Huitema, C. and Rosenberg, J. 2010,RFC 3015: Megaco Protocol Version.