Recent innovations in Distributed Systems: Challenges and Benefits

Amandeep Kaur[1], Madhurima [2] and Madhulika[3]

**[1,3] Department of CSE, AmityUniversity,
Noida,Uttar Pradesh, India**

**[2] Department of IT, AmityUniversity,
Noida,Uttar Pradesh, India**

Abstract

The World Wide Web is used by millions of people everyday for various purposes including email, reading news, downloading music, online shopping or simply accessing information about anything. Using a standard web browser, the user can access information stored on Web servers situated anywhere on the globe. This gives the illusion that all this information is situated locally on the user's computer. In reality, the Web represents a huge distributed system that appears as a single resource to the user available at the click of a button. This paper gives an overview of distributed systems in current IT sector. Distributed systems are everywhere. The internet enable users throughout the world to access its services wherever they may be located [1]. Each organization manages an intranet, which provides local services for local users and generally provides services to other users in the internet. Small distributed systems can be constructed from mobile computers and other small computational devices that are attached to a wireless network.
Keywords: distributed systems, mainframes, multicomputer, Alchemi.

1. Introduction

A distributed system is a collection of independent computers that appear to the users of the system as a single computer. This definition has two aspects. The first deals with hardware that is autonomous in nature and second deals with software that seems to users a single computer [2]. A distributed system consists of multiple autonomous computers that communicate through a computer network. The computers interact with each other in order to achieve a common goal. A computer program that runs in a distributed system is called a distributed program, and distributed programming is the process of writing such programs. Distributed computing also refers to the use of distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers.

There are several definitions and view points on distributed systems. Coulouris [1] defines a distributed system are those systems in which hardware or software components located at networked computers communicate and coordinate their actions only by message passing" [1] and Tanenbaum defines it as "A collection of independent computers that appear to the users of the system as a single computer" [2].

2. Benefits and features of Distributed system

When we talk about a network of computers and distributed system, they both are different A distributed system is built on top of a network and tries to hide the existence of multiple autonomous computers. It appears as a single entity providing the user with whatever services are required. A network is a medium for interconnecting entities (such as computers and devices) enabling the exchange of messages based on well-known protocols between these entities, which are explicitly addressable.

There are various types of distributed systems, such as Clusters [3], Grids [4], P2P (Peer-to-Peer) networks [5], distributed storage systems and so on. A cluster is a dedicated group of interconnected computers that appears as a single super-computer, generally used in high performance scientific engineering and business applications. A grid is a type of distributed system that enables coordinated sharing and aggregation of distributed, autonomous, heterogeneous resources based on users' QoS (Quality of Service) requirements. Grids are commonly used to support applications emerging in the areas of e-Science and e-Business, which commonly involve geographically distributed communities of people who engage in collaborative activities to solve large scale problems and require sharing of various resources such as computers, data, applications and scientific instruments. P2P networks are decentralized distributed systems, which enable applications such as file-sharing, instant messaging, online multiuser gaming and content distribution over public networks. Distributed storage systems such as NFS (Network File System) provide users with a unified view of data stored on different file systems and computers which may be on the same or different networks.

The main features of a distributed system include [1] [2]:
• Functional Separation: It is based on the functionality/services provided, capability and purpose of each entity in the system.
• Inherent distribution: Entities such as information, people, and systems are inherently distributed. For

example, different information is created and maintained by different people.

• Reliability: Long term data preservation and backup (replication) at different locations.

• Scalability: Addition of more resources to increase performance or availability.

• Economy: Sharing of resources by many entities to help reduce the cost of ownership.

As a consequence of these features, the various entities in a distributed system can operate concurrently and possibly autonomously. Tasks are carried out independently and actions are co-ordinate at well-defined stages by exchanging messages. Also, entities are heterogeneous, and failures are independent. Generally, there is no single process, or entity, that has the knowledge of the entire state of the system.

Various kinds of distributed systems operate today, each aimed at solving different kinds of problems. The challenges faced in building a distributed system vary depending on the requirements of the system. In general, however, most systems will need to handle the following issues [1] [2]:

• Heterogeneity: Various entities in the system must be able to interoperate with one another, despite differences in hardware architectures, operating systems, communication protocols, programming languages, software interfaces, security models, and data formats.

• Transparency: The entire system should appear as a single unit and the complexity and interactions between the components should be typically hidden from the end user.

• Scalability: The system should work efficiently with increasing number of users and addition of a resource should enhance the performance of the system.

• Concurrency: Shared access to resources should be made possible.

• Openness and Extensibility: Interfaces should be cleanly separated and publicly available to enable easy extensions to existing components and add new components.

• Migration and load balancing: It allows the movement of tasks within a system without affecting the operation of users or applications, and distribute load among available resources for improving performance.

• Security: Access to resources should be secured to ensure only known users are able to perform allowed operations.

Several software companies and research institutions have developed distributed computing technologies that support some or all of the features described above.

Thus the properties of distributed systems can be summarized as:

a) The system has to tolerate failures in individual computers.

b) The structure of the system (network topology, network latency, number of computers) is not known in advance, the system may consist of different kinds of computers and network links, and the system may change during the execution of a distributed program.

c) Each computer has only a limited, incomplete view of the system. Each computer may know only one part of the input.[7]

3. Advantages of distributed system over centralized system and isolated (personal) computers

The main advantage why we use distributed system as compared to centralized system is that the main driving force will be the existence of large numbers of personal computers and the need for people to work together and share information in a convenient way without bothered by geography or the physical distribution of people, data and machines[9]. The advantages are summarized below:

Table 1: Advantages of distributed system over centralized system

Sr. no.	Items	Description
1	Economics	Distributed systems offer a better price/performance than mainframes.
2	Speed	A distributed system may have more total computing power than mainframes.
3	Inherent distribution	Some applications involve spatially separated machines.
4	Reliability	If one machine crashes, the system as a whole can still survive.
5	Incremental growth	Computing power can be added in small increments.

Table 2: Advantages of distributed system over isolated computers

Sr. no.	Items	Description
1	Data sharing	Allow many users access to a common data base.
2	Device sharing	Allow many users to share expensive peripherals like a color printers.
3	Communication	Make human – to - human communication easier, for example by electronic mail.
4	Flexibility	Spread the workload over the available machines in the most cost effective way

4. Disadvantages of distributed system

Although distributed systems have their strengths, they also have their weakness. Following are the disadvantages of distributed system:

Table 3: Disadvantages of distributed system

Sr. no.	Items	Description
1	Software	Little software exists at present for distributed systems.
2	Networking	The network can saturated or cause other problems.
3	Security	Easy access also applies to secret data

4. Current Distributed System technologies

Technologies such as CORBA (Common object request broker architecture) and DCOM (Distributed component object model) have provided the means to build distributed component-based systems. Such technologies allow systems to interoperate at the component level, by providing a software layer and protocols that offer the interoperability needed for components developed in different programming languages to exchange messages. Such technologies present scalability issues when applied to, for instance, the Internet and some restrict the developer to a specific programming language. Hence, approaches based on Web protocols and XML (extensible Markup Language) have been proposed to allow interoperable distributed. Systems irrespective the programming language in which they are developed.

Alchemi: An example distributed system

In a typical corporate or academic environment there are many resources which are generally under-utilized for long periods of time. A "resource" in this context means any entity that could be used to fulfill any user requirement; this includes compute power (CPU), data storage, applications, and services.
The desirable features of an enterprise grid system are:

• Enabling efficient and optimal resource usage.
• Sharing of inter- organizational resources.
• Secure authentication and authorization of users.
• Security of stored data and programs.
• Secure communication.
• Centralized / semi-centralized control.
• Auditing.
• Enforcement of Quality of Service (QoS) and Service Level Agreements (SLA).
• Interoperability of different grids (and hence: the basis on open-standards).
• Support for transactional processes.

Alchemi [6] is an Enterprise Grid computing framework developed by researchers at the GRIDS Lab, in the Computer Science and Software Engineering Department at the University of Melbourne, Australia. It allows the user to aggregate the computing power of networked machines into a virtual supercomputer and develop applications to run on the Grid with no additional investment and no discernible impact on users. The main features offered by the Alchemi framework are the main features offered by the Alchemi framework are:

• Virtualization of compute resources across the LAN/ Internet.
• Ease of deployment and management.
• Object-oriented "Grid thread" programming model for grid application development.
• File-based "Grid job" model for grid-enabling legacy applications.
Web services interface for interoperability with other grid middleware.
• Open-source .Net based, simple installation using Windows installers.

Alchemi Grids follow the master-slave architecture, with the additional capability of connecting multiple masters in a hierarchical or peer-to-peer fashion to provide scalability of the system. An Alchemi grid has three types of components namely the Manager, the Executor, and the User Application itself [8].

5. Conclusions

Distributed systems have been an important part of everyone's lives because of innovations in the recent past in the area of Web-based applications, and will continue to make a serious impact in the future. Emerging technologies such as Grids will drive the next wave of innovation enabling the creation of applications that deliver IT as the 5th utility after water, electricity, gas, and the telephone. In conclusion, distributed computing is a very broad area with vast potential to improve efficiency of business processes and quality of life.

References

[1] Georage Coulouris and Jean dollimore, Distributed systems concepts and design, Pearson education, pg 39-40, 2006
[2] Andrew S. Tanenbaum, Distribted Operating System, Pearson Education, 3rd edition, 2006
[3] R. Buyya (editor), *High Performance Cluster Computing*, Prentice Hall, USA, 1999
[4] I. Foster and C. Kesselman (editors), *The Grid: Blueprint for a Future Computing Infrastructure*, Morgan Kaufmann Publishers, USA, 1999.
[5] R. Subramanian and B. Goodman (editors), *Peer-to-Peer Computing: Evolution of a Disruptive Technology,* Idea Group Inc., Hershey, PA, USA, 2005.
[6] A. Luther, R. Buyya, R. Ranjan, and S. Venugopal, Peer-to-Peer Grid Computing and a .NET-based Alchemi Framework, In High Performance Computing: Paradigm

and *Infrastructure*, L Yang and M. Guo (eds), Wiley Press, New Jersey, USA, June 2005.

[7] Notes on Distributed system, http://en.wikipedia.org/wiki/Distributed_computing#cite_note-0

[8] http://www.alchemi.net/

[9] Harish Rohil, Sanjna Kalyan, "A Heuristic Based Load Balancing Algorithm" International Journal of Computational Engineering and Management IJCEM, Vol. 15 Issue 6 Nov. 2012.

Amandeep kaur has received B.tech degree in computers from SSIET- DeraBassi, Punjab, India in 2007, M.E. in Computers Science and Engineering from UIET, Panjab University, Chandigarh, India in 2010. Currently She is working as Assistant Professor in department of computer science and engineering,Amity University, Noida, U.P., India. Her interest areas include image processing, pattern recognition and data mining.

Ms. Madhurima received her Master's Degree in Information Technology from Guru Gobind Singh Indraprastha University, New Delhi, B.Engg. in Information Technology from B.S. Anangpuria Institute of Technology and Management, Faridabad and Diploma in Information Technology from Government Polytechnic for Women, Faridabad. After working as a lecturer in B.S.A.I.T.M, Faridabad for 1.3 years, she was associated with the Apeejay group for approximately 4.9 years. She has joined Amity University as Assistant Professor Grade-I in August, 2012. She has published approximately sixteen research papers in various international and national journals and conferences. Her research interests are in Software Testing, AJAX, Object Tracking and Object Oriented Technologies. She has published a book "Video Object Tracking" with LAP LAMBERT Academic Publishing GmbH & Co. KG Germany and another book with the title "Computer Networks" with Laxmi Publications. She is a member of ACM.

Ms. Madhulika is pursuing Ph.D in Video object tracking using soft computing Techniques from Amity University, Noida. She received her Master's Degree in Computer Science Engineering from Lingayas Institute of Management and Technology, Faridabad, Masters in Business Administration in Information Technology from GJU,Hisar ,B.Engg. in Computer Science Engineering from B.S. Anangpuria Institute of Technology and Management, Faridabad and Diploma in Information Technology from Government Polytechnic for Women, Faridabad. After working as a lecturer in B.S.A.I.T.M, Faridabad for 2 years, she was associated with the Apeejay group for approximately 3 years. Then she worked with Rama groups of Institutions for 1 year. She has joined Amity University as Assistant Professor Grade-I in June, 2012. She has published approximately eighteen research papers in various international and national journals and conferences. Her research interests are in Object Tracking and Brain Mapping. She has published a book "Facial Expression Recognition System" with LAP LAMBERT Academic Publishing GmbH & Co. KG Germany and another book with the title "Computer Networks" with Laxmi Publications. She is a member of ACM, ISO, and CSI.