

Overview of Distributed Computing Technology

Haikui Wang^{1 a}

¹Zhengzhou University of Industry Technology, Department of Information Engineering,
Zhengzhou, China

^aHouyanyang521@sina.com

Keywords: Distributed computing; Middleware; Grid; Mobile agent; P2P; Web service

Abstract: In the past 20 years, there have been a lot of distributed computing technologies, such as middleware technology, grid technology, mobile agent technology, P2P technology and recently introduced Web Service technology. Each technology has been recognized to a certain extent, solving the problem of distributed computing in a specific range. But there are some unsolved problems in the existing distributed computing technology, which affect the application and popularization of the distributed computing technology. The integration of a variety of distributed computing technologies to meet the needs of distributed computing is the research direction in the future. It is also the key to the smooth development of distributed computing.

Overview

Distributed computing is using the network to connect thousands of computers, to form a virtual supercomputer, to complete the solutions that a single computer cannot complete. The earliest form of distributed computing came to Intel in the late 1980s. Intel uses their workstation's free time to design and compute data sets for the chip, using LAN adjustment research. With the rapid development and popularization of Internet, the research of distributed computing has reached its climax after 90 's. At present, distributed computing is very popular on Internet. The research of distributed computing mainly focuses on the research of distributed operating system and distributed computing environment. In the past 20 years, there have been a lot of distributed computing technologies, such as middleware technology, grid technology, mobile agent technology, P2P technology and recently introduced Web Service technology. Each technology has been recognized to a certain extent and has been widely used in certain areas. However, all distributed computing technology has not been recognized around the world. No technology can be shown to be the mainstream of distributed computing technology. The complexity and diversity of technology make the research of distributed computer very active, and also make the popularization of distributed computing technology very difficult. This paper tries to introduce the development history and technical thinking of distributed computing technology, summarize the common problems in the research of distributed computing technology, and deduce the future development direction.

Typical Distributed Computing Technology

Middleware Technology.

The middleware is the basic software, is between the operating system (or network protocol) and the distributed application, thus shielding the operating system (or network protocol) the difference, realizing the distributed heterogeneous system interoperation. At present, there is not a unified definition of middleware. The commonly-accepted definition of IDC is that middleware is an independent system software or service program, and distributed application software shares resources among different technologies. The middleware is located on top of the client server's operating system and manages computing resources and network communications[1].

Early middleware was relatively simple, with only messaging and transaction management capabilities. In the 1990s, because of the diversity of middleware requirements, middleware technology developed rapidly, resulting in different types of middleware products, mainly

including: MOM (Message-Orientated Middleware), database middleware, RPC (Remote Processor Call), ORB (Object Request Processing Middleware) In the late 1990s. Due to the impact of demand growth and multi-technology integration, the connotation and extension of middleware further expanded, and the category of middleware further increased, such as the emergence of reflective middleware, mobile middleware, adaptive middleware and so on[2].

Grid Technology.

The concept of Grid originated in the mid-1990s and originally meant to provide a distributed computing infrastructure for cutting-edge research and engineering[3]. With the rapid development and popularization of Internet, grid technology has received great attention, and a large number of influential research projects have emerged, such as Globus of Argonne National Laboratory and Legion of the University of Virginia, which is an object-based meta-system software project. The IPG project led by NASA's Advanced Supercomputing team, EuroGrid, a research and technology development project co-funded by the European Economic Community, and TeraGrid, funded by the National Science Foundation, have developed well-known grid computing tools such as Globus Toolkit. Recently, Grid computing has been supported by more and more famous companies, such as IBM, SUN, Microsoft, HP, etc., which have announced their own grid research plan.

Grid computing does not yet have a universally accepted definition. Ian Foster, a prominent grid expert and leader of the Globus project, says that: Grid implementations share resources and work together to solve problems across a number of dynamic virtual agencies[4]. But with the development of grid technology, especially the introduction of OGSA architecture, Ian Foster has extended the definition of grid: Grid is an extensible set of grid services, which are combined in different ways to meet the needs of virtual institutions. These virtual institutions themselves are defined as part of the services they operate and share.

Mobile Agent Technology.

In the early 1990s, General Magic firstly introduced the concept of mobile Agent, a software entity that can independently migrate from one host to another in a heterogeneous network environment and interact with other agents or resources, when it introduced its commercial system Telescript. Mobile Agent is a special kind of software agent, which can be regarded as a combination of software agent technology and distributed computing technology. It has not only the basic characteristics of software agent-autonomy, responsiveness, initiative and reasoning, but also the mobility. That is, it can automatically move from one host to another on the network, representing the user to complete the assigned task. Because mobile agent can move freely in heterogeneous software and hardware network environment, this new computing mode can effectively reduce network load in distributed computing, improve communication efficiency, dynamically adapt to changing network environment, and has good security and fault tolerance ability. Since the definition of Agent is not yet clear, there is not yet an exact definition of Mobile Agent. Mobile Agent is a kind of autonomous program which can move from one computer to another under its own control. It can provide convenient, efficient and robust execution framework for distributed applications. At present, all mobile Agent systems are still immature and have various defects. Many agent systems can be regarded as laboratory systems, and there is still a long way to go from real practical products.

P2P technology.

P2P originated from the initial way of network communication. It is an old technology, such as USENET in 1979, FidoNet in 1984 are based on P2P technology[5]. But now, P2P has been given a new meaning. It is a new application mode of the old technology. The rise of P2P technology can be attributed to a lawsuit between Napster software and U.S. record labels, although research on P2P methods has never stopped before. At present, a large number of products and technologies have been produced, such as Microsoft's distributed file system Farsite[6], file exchange software Napster, Gnutella's information retrieval software Intrasearch, Intel's collaborative computing software Netbatch[7], desktop collaboration software Groove and so on.

Because P2P applications are widely distributed and the problems solved are very different, a unified P2P definition has not been formed so far. The original meaning of P2P is a communication

mode in which each part has the same ability and any part can start a communication. Now, the concept of P2P has been extended, such as International Business Machines Corp that the P2P system consists of a number of interconnected computers, and at least one of the following features: the system depends on the active collaboration of devices that are marginalized (non-central servers). Each member benefits directly from the participation of other members rather than the server; The members of the system act as both server and client. The users of the system application can realize each other's existence and form a virtual or actual group. P2P technology has developed into an important distributed computing technology, although it cannot solve all distributed computing problems.

Web Service Technology.

Web Service is a new distributed computing technology that has emerged in recent years. It is an extension of object/ component technology in Internet and an object/ component deployed on the Web. At present, there is no consistent definition of Web Service. The definition of ISO W3C is that Web Service is a software application identified by URL. Its interface and binding can be defined, described and discovered by XML documents, and is based on messaging on Internet protocol. Direct interaction with other software applications using XML support[8]; Microsoft defines Web Services as an application logical unit that provides data and services to other applications, and how each Web Service is implemented is completely independent of the standard Web protocol and data format for Web Services, such as HTTP, XML, and SOAP. How each Web Service is implemented, such as HTTP, XML, and SOAP, is completely independent. Web Services, with the advantages of component-based development and Web, is the core of Microsoft's. NET programming model. International Business Machines Corp is defined as: Web Services are self-contained, modular applications that can be described, published, searched, and invoked on a network, typically on the Web.

Common Problems Existed

The problem of standards.

At present, almost no distributed computing technologies have a complete uniform standard, although this work has begun, such as the Web Service has developed some standards, it is far from mature. The lack of standards makes it difficult for distributed computing research to form a stable research direction and to develop widely accepted products. Therefore, it restricts the development of distributed computing technology to a great extent.

The problem of software method.

Software methodology is the premise that software can be industrialized. At present, software methodology is mainly aimed at single-machine system environment, such as structured program design method, object-oriented programming method, etc. There is no software method for distributed computing system, and there is little research in this field. The lack of feasible software methodology makes it difficult to guarantee the quality and development schedule of distributed computing software. Without industrial production mode, the popularization of distributed computing system will be very difficult.

The problems of quality and availability.

Because distributed computing system has no unified control means, it is difficult to guarantee the quality and availability of distributed computing system. How to provide high-quality services in distributed computing systems, how to use services, how to guarantee the security and reliability of services, have become difficult problems in distributed computing research. It is difficult to accept this technology until people are guaranteed satisfactory quality of service and availability.

The problem of heterogeneous structure.

Now the network is a heterogeneous environment. Distributed computing technology, firstly, needs to solve the interoperability problem of heterogeneous environment. To solve the interoperability problem of heterogeneous environment, the primary task is how to identify each other. At present, it is not possible to require all resources to be described in the same way (although Web Service technology wants to do so, it is difficult to do so in practice), and there is no

way to intelligently identify these resources (semantic research is far from practical). This causes any kind of distributed computing technology to be used only within a certain range (typically in the same environment).

Development Trend

Recently, many countries, especially developed countries, pay more and more attention to the research of distributed computing technology, and the scale of investment is increasing. Almost all the well-known IT manufacturers in the world have shown a strong interest in distributed computing technology, announced their own research plans, and constantly release a variety of related products, to seize the future development of computers at the top[9]. But so far, all distributed computing technologies have some unsolved problems more or less. No technology has been recognized by all researchers as the direction of distributed computing research, and no technology has been able to achieve a full sense of distributed computing to meet the needs of all distributed computing. With the development of distributed computing technology, many researchers find that the limitation of single technology is more and more obvious in technology itself, application field and so on. A satisfactory result has been achieved.

Conclusion

Distributed computing technology is the product of computer network and the future development direction of computer network application. Although a lot of research has been done and some technologies such as middleware, grid, mobile agent, P2P and Web Service have been produced, it is still far from perfection. The research direction of distributed computing technology is to organically integrate existing technologies, absorb the advantages of all technologies, and reasonably solve the needs of distributed computing. It is believed that in the near future, there will be a perfect distributed computing system.

References

- [1] Middleware white paper. International Systems Group, Inc, 1997.2.
- [2] Capra, Licia, et al. "Towards a mobile computing middleware: Synergy of reflection and mobile code techniques." Proceedings Eighth IEEE Workshop on Future Trends of Distributed Computing Systems. FTDCS 2001. IEEE, 2001.
- [3] Foster, Ian, and Carl Kesselman. "The Grid: Blueprint for a New Computing Architecture." (1999).
- [4] Foster, Ian, Carl Kesselman, and Steven Tuecke. "The anatomy of the grid: Enabling scalable virtual organizations." The International Journal of High Performance Computing Applications 15.3 (2001): 200-222.
- [5] Tucker, Peter A., David Maier, and Tim Sheard. "Applying punctuation schemes to queries over continuous data streams." IEEE Data Eng. Bull. 26.1 (2003): 33-40.
- [6] streams." Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001.
- [7] Yi, B-K., et al. "Online data mining for co-evolving time sequences." Proceedings of 16th International Conference on Data Engineering (Cat. No. 00CB37073). IEEE, 2000.
- [8] Guha, Sudipto, et al. "Clustering data streams." Proceedings 41st Annual Symposium on Foundations of Computer Science. IEEE, 2000.
- [9] Korn, Flip, S. Muthukrishnan, and Divesh Srivastava. "Reverse nearest neighbor aggregates over data streams." VLDB'02: Proceedings of the 28th International Conference on Very Large Databases. Morgan Kaufmann, 2002.