IMPOTANCE OF DISTRIBUTED SYSTEMS AND ENHANCEMENT OF SECURITY ISSUES IN DISTRIBUTED SYSTEMS

Mr Subramanya .G. Bhagwath
Department of Computer Science & Engineering.
Anjuman Institute of Technology and Management, Bhatkal.

Abstract---A distributed system is piece of software that serves to coordinate the actions of several computers. This coordination is achieved by exchanging messages, i.e., pieces of data conveying information. The system relies on a network that connects the computers and handles the routing of messages. From the last two decades the trends in the Computing Industry is towards Distributed, low cost unit, and high unit volume product. Moreover software development activity is becoming more decentralized thereby expanding more development efforts for organizations. The performance of computing system has increased drastically with the inclusion of Multiprocessing and Multicomputing. This paper presents a systematic literature review on the issues, Recent Innovations,Challenges and Benefits of Distributed Systems. With An example . This paper presents a the security issues associated with those systems. Four commonly used distributed systems were considered for detailed analysis in terms of technologies involved, security issues faced by them and solution proposed to circumvent those issues. Finally the security issues and the solutions were summarized and compared with each other.This paper is also aimed to give a report on the Distributed Computing Technologies in Practice.

Keywords--- Alchemi, Computing Clusters,Distributed systems, Distributed Storage System Security, Grid System Security

I. INTRODUCTION

THIS document 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.

There are several definitions and view points on what distributed systems are. Coulouris defines a distributed system as "a system 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]. Leslie Lamport – a famous researcher on timing, message ordering, and clock synchronization in distributed systems once said that "A distributed system is one on which I cannot get any work done because some machine I have never heard of has crashed" reflecting on the huge number of challenges faced by distributed system designers. Despite these challenges, the benefits of distributed systems and applications are many, making it worthwhile to pursue.

Various types of distributed systems and applications have been developed and are being used extensively in the real world. In this article, we present the main characteristics of distributed systems and look at some of the challenges that are faced by designers and implementers of such systems, and also introduce an example distributed system.

II. MAIN FEATURES AND BENEFITS OF A DISTRIBUTED SYSTEM .

A common misconception among people when discussing distributed systems is that it is just another name for a network of computers. However, this overlooks an important distinction. 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 (using an IP address, for example).

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 Units

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 decentralised distributed systems, which enable applications such as file-sharing, instant messaging, online multi-user 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

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. This information could be generated, stored, analysed and used by different systems or applications which may or may not be aware of the existence of the other entities in the system.

*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-ordinated at well-defined stages by exchanging messages. Also, entities are heterogenous, 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.

*Fault tolerance and failure management

Failure of one or more components should not bring down the entire system, and should be isolated.

*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

Allow 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.

III.DISTRIBUTED COMPUTING TECHNOLOGIES IN PRACTICE

Over the years, technologies such as CORBA and DCOM 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. However, 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.

Web Services are based on XML and provide a means to develop distributed systems that follow a Service Oriented Architecture (SOA). Services are described in an XML-based dialect (WSDL). In a similar fashion, the request and reply messages exchanged in such systems are formatted according to the Simple Object Access Protocol (SOAP). SOAP messages can be encoded and transmitted by using Web protocols such as the Hypertext Transfer Protocol (HTTP). Various industrial technologies and application platforms such as .NET from Microsoft, J2EE from Sun, WehSphere from IBM are targeted at supporting the development of applications based on Web Services

. Along with Web Services, Grid computing is another emerging paradigm for creating wide-area distributed applications. Web Services are foundation technologies that can be used in building many types of distributed systems and applications including Grid systems. Web Services are in the core of the current implementations of Grid technologies such as Globus from Argonne National Laboratory in USA and the Gridbus from the University of Melbourne, Australia. Grid computing scales from an enterprise/organisation to a global level. Global Grids are established over the public
Internet infrastructure, and are characterized by a global presence, comprise of highly heterogeneous resources, present sophisticated security mechanisms, focus on single sign-on and are mostly batch-job oriented.

To enable global Grids, one requirement is that current enterprise and campus Grids are able to interoperate. Enterprise and campus Grids consist of resources spread across an enterprise and provide services to users within that organisation and are managed by a single administrative domain. Such Grids are more concerned with cycle stealing from unused desktops and use virtualization of resources in order to provide better means to manage and utilize them within an enterprise. For example, Oracle 10g uses a virtualization approach to split data storage from the database transaction and process layer. However, scalability and the design of security mechanisms are not as difficult as they are for global Grids.

IV.ALCHEMI: AN EXAMPLE DISTRIBUTED SYSTEM.

In a typical corporate or academic environment there are many resources which are generally under-utilised for long periods of time. A “resource” in this context means any entity that could be used to fulfil any user requirement; this includes compute power (CPU), data storage, applications, and services. An enterprise grid is a distributed system that

dynamically aggregates and co-ordinates various resources within an organisation and improves their utilisation such that there is an overall increase in productivity for the users and processes. These benefits ultimately result in huge cost savings for the business, since they will not need to purchase expensive equipment for the purpose of running their high performance applications.

The desirable features of an enterprise grid system are:

*Enabling efficient and optimal resource usage.
*Sharing of inter-organisational resources.
*Secure authentication and authorization of users.
*Security of stored data and programs.
*Secure communication.
*Centralised / semi-centralised 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:

*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.

The Manager node is the master / controller whose main function is to service the user requests for workload distribution. It receives a user request, authenticates the user, and distributes the workload across the various Executors that are connected to it. The Executor node is the one which actually performs the computation. Alchemi uses role-based security to authenticate users and authorize execution. A simple grid is created by installing Executors on each machine that is to be part of the grid and linking them to a central Manager component.

V. SECURITY IN DISTRIBUTED SYSTEMS

Security is one of the most important issues in distributed systems. When data is distributed across multiple networks or information is transferred via public networks, it becomes vulnerable to attacks by mischievous elements. Similarly other computing resources like processors, storage devices, networks etc., can also be attacked by hackers.

A. Security for Computing Clusters

When the computing clusters are made available to the public or networks are setup using public resources such as the Internet, they become subject to various kinds of attacks. The most common types of attacks on the clusters are computation-cycle stealing, inter-node communication snooping, and cluster service disruption [17]. Hence the clusters have been protected by security mechanisms that include services like authentication, integrity check, and confidentiality. The main purpose of the security mechanisms is to protect the system against hackers as well as to meet the security requirements of the applications.

Li and Vaughn have studied the security vulnerabilities of computing clusters using exploitation graphs (e-graphs). They have modeled several attacks that can be carried on all three pillars of security namely, confidentiality, integrity and availability. They

have shown that e-graphs can be simplified based on domain knowledge such as cluster configurations, detected vulnerabilities, etc. they further state that this technique could be used for certification of clusters with the help of a knowledge base of cluster vulnerabilities[18].

Xie and Qin have developed two resource allocation schemes named Deadline and Security constraints (TAPADS) and Security-Aware and Heterogeneity-Aware Resource allocation for Parallel jobs (SHARP). These two schemes ensure that parallel applications executed on computing clusters meet the security requirements while meeting the deadline of executions [17]. Hence it could be seen that if these schemes ensure mainly the availability of the system as timely execution of an application is an indication of the availability of the resources.

Denial of Service (DoS) attack is one of the common attacks on distributed systems. These attack mainly target resources in such a manner that the resources are prevented from carrying out their legitimate operations. A method that uses services and markov chain to mitigate the effects on the DoS attack on a cluster based wireless sensor network has been presented in [19].

Hence it can be seen that computing clusters are vulnerable to attacks by mischievous elements like hackers and crackers due to its open nature and use of public resources such as the internet. Extensive research has been carried out by several researchers on the security of clusters and they have proposed several methods that can be made used to protect the clusters from these attacks.

B. Grid System Security

Grid computer systems provide several security mechanisms to protect the grid resources against attacks. Middleware is one of the critical system software in the grid infrastructure as it provides the common communication infrastructure and makes the grid services available to applications. Middleware also allows for a uniform security configuration at the service container or messaging level. Grid authentication is based on Public Key Infrastructure (PKI) and capable of handling different types of user credentials such as PKI, SAML, Kerberos tickets, password, etc., Delegation is one of the necessary mechanisms in grid service delivery and is implemented using X.509 Proxy Certificate. Authorization to access grid resources is based on Virtual Organization (VO) attributes assigned to a user and managed by Virtual Organization Membership Service (VOMS). Trust management in grid systems are handled using certificates and trust relations are represented by a certificate chain that include Grid Certification Authority (CA) certificate and other successively generated proxies [20].

Grid authentication module is one of the critical components in preventing external users from randomly accessing internal grid and protecting the grid system from unauthorized users. This module handles security threats from internal network, when certificated grid users carry out illegal (unauthorized) operations within the grid [21].

These grid security mechanisms are all implemented on almost all grid systems available today. There several grid community initiatives going on in the area of grid middleware interoperability which would finally unify the grid security as a single coherent security platform and scheme.

C. Distributed Storage System Security

Several active researches are going on in the area of threat modeling and developing security model for protecting distributed storage systems. The most important resource in the distributed storage system is the data stored in the storage devices of the system. This data needs to be properly labeled and protected. Also any protection system introduced must be backward compatible in other words; it not only should protect the data stored after the security scheme is installed but also the data that had been there prior to the introduction of that scheme.

Hasan et al., have introduced a threat model named CIAA threat model. This model addresses all the security issues namely, Confidentiality, Integrity, Availability and Authentication. In arriving at this model, authors have organized the threats on a

distributed storage system under each category of the CIAA pillars of security and provided techniques that can be used to circumvent the threats. The other security model discussed by the authors is the Data Lifecycle Model that examines the types of threats that may

extinction. Under this model threats have been organized under six groups and solutions have been proposed [22].

Dikaliotis, Dimakis and Ho have proposed a simple linear hashing technique that can detect errors in the storage nodes in the encoded distributed storage systems [23]. Mutually Cooperative Recovery (MCR) mechanism enables the system to recover data in situations of multiple node failures. The transmission scheme and design a linear network coding scheme based on (n, k) strong-MDS code proposed help recover systems from failure with relative ease [14].

Hence it can be seen that the security schemes in the distributed storage systems mainly concentrate on data security in terms of integrity and failure management (availability).

D. Distributed Database Security

Distributed database management systems face more security threats compared to their counterpart centralized database systems. The development of security for distributed database systems have become more complicated with the introduction of several new database models such as object-oriented database model, temporal database model, object relational database model etc.

In traditional security model, all the data stored in database and the users who access that data belong to the same security level. A multilevel secure database system assigns security level to each transaction and data. Clearance level of a transaction is represented by security level assigned to it and the classification level of data is given by the classification level. A multilevel secure database management system (MLS/DBMS) restricts database operations based on the security levels [24]. From the above discussion, it can be seen that by introducing the military information classification and access control security of distributed databases can be enhanced.

Zubi has presented a design that would improve the scalability, accessibility and flexibility while accessing various types of data in a distributed database system. He has also proposed multi level access control, confidentiality, reliability, integrity and recovery to manage the security of a distributed database system [25].

CONCLUSION

we have noted thus far, distributed systems have been an important part of peoples' lives as a result 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!

From the above discussion, it can be seen that security becomes more prominent when the systems have been distributed across over multiple geographic locations. Each type of distributed system has its own peculiar security requirements. But, all the systems have the common CIA triad as the heart of any security implementation. In computing clusters and grids the security mainly concentrates on protecting the data in transit and access to distributed resources. Security in clusters is somewhat simpler compared to grid due to homogeneous nature of

clusters. One of the main attacks that has been carried out on clusters is the Denial of Service (DoS) attack. Researchers have proposed novel methods based on markov chain to mitigate the impact of DoS attacks.

In grid the middleware layer provides the platform for the implementation of security on the entire grid system. Grid system use strong security based on PKI and X.509 certificates. The user authentication module in the grid provides security against threats by external sources and illegal actions by internal users.

Security of distributed storage systems mainly concentrate on securing data. The main areas concentrated on distributed storage are protection against data corruption and protection of data in situations of node failures. Researchers have proposed various models and schemes to protect the storage system against attacks and node failures.

In distributed database system, the security implementation has been made more complicated due to the availability of different kinds of database models. But researchers have shown that by applying multi level security based on military information classification and access control, distributed database security can be enhanced.

In this paper, the development of distributed systems was discussed in terms of what a distributed system is and the objectives of setting up a distributed system. From all the available distributed systems, four most commonly used distributed systems were discussed in depth and then the security issues faced by these systems and the solutions proposed by various researchers were discussed in depth. Finally the security issues and solutions proposed for different systems were summarized and compared with each other.

References

1. G. Couloris, J. Dollimore, and T. Kinberg, Distributed Systems - Concepts and Design, 4th Edition, Addison-Wesley, Pearson Education, UK, 2001.
2. A. Tanenbaum and M. Van Steen, Distributed Systems: Principles and Paradigms, Prentice Hall, Pearson Education, USA, 2002.
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. Darlan K. E De Carvalho, Paulo R.M Lyra, and Ramiro B Willmersdorf, "A First Step towards a Petroleum Reservoir Simulator Using an Edge-Based Unstructured Finite Volume Formulation," in 2nd Brazilian R & D Congress on Oil and Gas, Rio de Janeiro, Brazil, 2003.
8. Ronald Scrofano, Maya B Gokhale, Frans Trouw, and Viktor K Prasanna, "Accelerating Molecular Dynamics Simulations with Reconfigurable Computers," IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 6, pp. 764-778, June 2008.
9. Weitao Sun, Jiwu Shu, and Weimin Zheng, "Parallel Seismic Propagation Simulation in Anisotropic Media by Irregular Grids Finite Difference Method on PC Cluster," in Computational Science and Its Applications – ICCSA 2005, Osvaldo Gervasi et al., Eds. Berlin / Heidelberg, Germany: Springer, 2005, pp. 762-771.
10. KeJing Zhang, Ping Jun Dong, Biao Ma, Bing Yong Tang, and Hong Cai, "Innovation of IT Service in Textile Industrial Clusters from the Service System Perspective," in International Conference on Systems and Intelligent Management Logistics, Harbin, China, 2010, pp. 1819 - 1822.
11. Rajkumar Buyya and Srikumar Venugopal, "Market Oriented Computing and Global Grids: An Introduction," in Market Oriented Grid and Utility Computing, Rajkumar Buyya and Kris Bubendorfer,

Eds. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2010, ch. 1, pp. 3-27.

12. Huang Ye et al., "Using Metadata Snapshots for Extending Ant-Based Resource Discovery Service in Inter-cooperative Grid Communities," in Proceedings of the First International Conference on Evolving Internet (INTERNET 2009), Cap Esterel, France, 2009, pp. 89-94.

13. Yuchong Hu, Yinlong Xu, Xiaozhao Wang, Cheng Zhan, and Pei Li, "Cooperative Recovery of Distributed Storage Systems from Multiple Losses with Network Coding," IEEE Journal on Selected Areas in Communications, vol. 28, no. 2, pp. 268-276, February 2010.

14. Xiao Gao Yu and Wei Xing Li, "A new network storage architecture based on NAS and SAN," in 10th International Conference on Control, Automation, Robotics and Vision (ICARCV 2008), Hanoi, Vietnam, 2008, pp. 2224 - 2227.

[15] Ali Safari Mamaghani, Mostafa Mahi, Mohammad Reza Meybodi, and Mohammad Hosseinzadeh Moghaddam, "A Novel Evolutionary Algorithm for Solving Static Data Allocation Problem in Distributed Database Systems," in Second International Conference on Network Applications, Protocols and Services (NETAPPS), Alor Setar, Kedah, 2010,pp. 14-19.

[16] Tao Xie and Xiao Qin, "Security-Aware Resource Allocation for Real-Time Parallel Jobs on Homogeneous and Heterogeneous Clusters," IEEE Transactions on parallel and distributed Systems, vol. 19, no. 5, pp. 682-697, May 2008.

[17] Wei Li and Rayford B Vaughn, "Cluster Security Research Involving the Modeling of Network Exploitations Using Exploitation Graphs," in Sixth IEEE International Symposium on Cluster Computing and the Grid Workshops (CCGRIDW'06), Singapore, 2006, pp. 26-36.

[18] Zhongqiu Jiang, Shu Yan, and Liangmin Wang, "Survivability Evaluation of Cluster-Based Wireless Sensor Network under DoS Attacks," in 5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09. , Beijing, China, 2009, pp. 1-4.

[19] Yuri Demchenko, Cees de Laat, Oscar Koeroo, and David Groep, "Re-thinking Grid Security Architecture," in IEEE Fourth International Conference on eScience, 2008 (eScience '08), Indianapolis, IN, USA, 2008, pp. 79-86.

[20] Quowen Xing, Shengjun Xue, and Fangfang Liu, "Research of Grid Security Authentication Model," in International Conference on Computer Application and System Modeling (ICCASM), vol. 1, Taiyuan, Shanxi, China, 2010, pp. 78-80.

[21] Ragib Hasan, Suvda Myagmar, Adam J Lee, and William Yurcik, "Toward a threat model for storage systems," in Proceedings of the 2005 ACM Workshop on Storage Security and Survivability (StorageSS '05), FairFax, VA, USA, 2005, pp. 94-102.

[22] Theodoros K Dikaliotis, Alexandros G Dimakis, and Tracey Ho, "Security in distributed storage systems by communicating a logarithmic number of bits," in IEEE International Symposium on Information Theory Proceedings (ISIT), , Austin, TX, USA, 2010, pp. 1948 - 1952.

[23] Navdeep Kaur, Rajwinder Singh, A K Sarje, and Manoj Misra, "Performance evaluation of secure concurrency control algorithm for multilevel secure distributed

database system," in International Conference on Information Technology: Coding and Computing (ITCC 2005), Las Vegas, NV, USA,International Journal of Computer Information Systems, Vol. 2, No. 2, 2011 2005, pp. 249-254.

[25] Zakaria Suliman Zubi, "On distributed database security aspects," in International Conference on Multimedia Computing and Systems, Ouarzazate, Morocco, 2009, pp. 231-235.