Enhancing Security Measures by Tunnelling Protocol in Distributed Grid Network

T. Mohan Raj
Department of Computer Science & Engineering,
Karpagam University,
Tamil Nadu, INDIA

S. Shahul Hammed
Department of Computer Science & Engineering,
Karpagam University,
Tamil Nadu, INDIA

A. Amala Deepan
Department of Computer Science & Engineering,
Karpagam University,
Tamil Nadu, INDIA

ABSTRACT

Grid is a distributed computing architecture that integrates a large number of data and computing resources into a single virtual data management system. A Computational Grid is a natural extension of the former cluster computer where large computing tasks have to be computed at distributed computing resources. A safe registration and communication is essential in Computational Grid networks. This paper reports a secure tunnelling protocol integrated frame work, which enhances the quality of Point-to-Point Tunnelling Protocol (PPTP), Layer Two Tunnelling Protocol (L2TP) and Internet Security Protocol (IPSec). The proposed model used an encryption scheme such as Data Encryption Standard (DES) algorithm. The new packet offers a secure communication in the grid network without any time delays.

General Terms

Grid Computing, Network Protocols

Keywords

Grid networks, PPTP, L2TP, IPSec, DES

1. INTRODUCTION

Grid computing is a network increasingly capturing the attention of computing community. It uses clusters of personal computers, super-nodes or other machines. They link together to tackle complex calculations. The grid computing lets companies harness their unused computing power, or processing cycles, to create a type of supercomputer.

Grid computing is an important and developing computing initiative that involves the aggregation of network connected computers to form a large-scale, distributed system for coordinated problem solving and resource sharing [1,2]. To speedup computing workload across the distributed system of computers, grid users can take advantage of enormous computational, storage, and bandwidth resources that would otherwise only be available within traditional multiprocessor supercomputers. To give an analogy, grid computing is similar to power grids, where user does not need to know anything about what stays beyond the socket. One can absorb all the power they wants according to the agreement with electrical society. A world wide attraction is given to Grid computing due to the variety of applications ranging from Physics, Chemistry, Environmental Science, Bioinformatics, Biomedical, Aerospace and Healthcare systems [3, 4].

The term “the Grid” was made in the mid 1990’s to denote a proposed distributed computing infrastructure for advanced science and engineering. Considerable progress has since been made on the construction of such an infrastructure but the term “Grid” has also been conflated, at least in popular perception, to embrace everything from advanced networking to artificial intelligence. The Computational Grid is a novel, evolving infrastructure that provides unified, coordinated access to computing resources such as processor cycles, storage, etc. Wide variety of systems, from small workstations to supercomputers can be linked to a grid to form a powerful virtual computing environment. Much complexities involved in managing resources of a grid are hidden from the nodes to nodes, providing a seamless access to computing resources. As a great advancement towards cost reduction, computational grids can be used as a replacement for supercomputers that are presently used in many computationally intensive scientific problems like genome analysis, medical imaging, computer graphics etc.

A computational grid consists of a set of resources, such as computers, networks, communication channels, super-nodes or sensors that are tied together by a set of common services which allow users of the resources to view the collection as a seamless computing or information environment [5]. A standard grid services include security services that support user authentication, authorization and privacy. Grid offers information services, which allow users to see what resources (hardware, software and services) are available for the usage. Grid networks give high level computational job submission services. This allow user submission jobs to any compute resource that the user is authorized to use and co-scheduling services, which allow multiple resources to be scheduled concurrently. Grid gives user support services, which provide users access to "trouble ticket" systems that span the resources.

2. GRID SECURITY ARCHITECTURE

Grid architecture defines the fundamental system components of grid computing environment, specifies the purpose and function of these components. This indicates how these components interact with one another. Grid architecture is one of the first and foremost protocol architecture, with protocols defining the basic mechanisms by which Virtual Organization (VO) users and resources negotiate, establish, manage, and exploit sharing relationships. A set of individuals and/or institutions defined by such sharing rules form what we call a VO [1]. Virtual organisations are often called grids [6]. Here, VOs are connected through a high bandwidth communication medium. A standards open architecture facilitate extensibility, interoperability, portability, and code sharing between nodes in the VO. Standard protocols make use to define standard services that enhances capabilities of resource sharing and co-ordination between the nodes. So there are communication channels established between these virtual organisations.

Specialized Grids focused on close and routine interactions between people, instruments and information in support of widely distributed scientific research projects are often called collaboratories [7]. Such grid network is providing to process important and safe computational activities. So a high level safety communication is required for registration and division of jobs submitted by a user. One of the major principle should meet the safe need in grid computing is that offer the authentication solution, to guarantee to mutual verification between the subject and object. Which considering access control mechanism, visit cross-domain and visit in domain will be referred. Try one's best to guarantee but not to change existing local access control mechanism. Authentication mechanism is the foundation of the access control mechanism and each local safe tactics is built up on this foundation.

Security in grid computing is categorized into three main aspects consisting of architectural, infrastructural, and management-related issues [8]. Architecture security is concerned with information, authorization, and service security. Infrastructure security, which is the topic of interest in this paper, is related to host protection and network related issues. Here needs an important secure communication medium network for secure transmission. Trust management along with credential and monitoring issues are the components within management security.

Many security architectures in the communication have been proposed in order to offer a secure environment for grid users. But security gaps still exist when volunteering a host to a grid computing environment. In computer volunteering, the user or owner of the computing resources becomes subject to some threats which are specific to on-demand computing [9]. Such services are demanded a high level protocol with safe and secure transmission in the communication channel in a grid network. This paper discusses content related to access control mechanism in a secure communication method.

3. PROTOCOL SECURITY DESIGN

Grids provide protocols and services at five different layers as identified in the Grid protocol architecture [10]. A promising security mechanism is essential for communicating in network due to the importance of the jobs carried out in the grid network. This section discusses a new tunnelling protocol design for secure data transmission in grid network. There is design of a new pack associated with point-to-point tunnelling and a Layer two tunnelling mechanism by using the TCP/IP protocol. This security mechanism enhances basically for registration and division of the work submitted to the grid network.

Various password based schemes and Challenge Handshake Authentication Protocol (CHAP) can be used to authenticate users on a grid network and control access to network resources. Encrypting the data as it travels through the grid connection guards the privacy of corporate information. Tunnelling allows senders to encapsulate their data in Internet Protocol (IP) packets that hide an underlying routing and switching infrastructure of the Intranet or Virtual Private Network (VPN) from both senders and receivers. At the same time, these encapsulated packets can be protected against snooping by outsiders using encryption techniques.

Tunnels consist of two types of endpoints, either an individual computer or LAN with a security gateway. Only two combinations of these endpoints are considered in designing grid network. First, LAN-to-LAN tunnelling, a security gateway at each endpoint serves as the interface between the tunnel and the private LAN. In this case, users on either LAN can use the tunnel transparently to communicate with each other. Second, node-to-LAN tunnels is the type usually setup for a mobile user who wants to connect to the corporate LAN of a Virtual Organisation. There are different types of protocols used for creating communication channel across the grid network.

Point-to-Point Tunnelling Protocol (PPTP): PPTP can be used for remote access and router-to-router connections. PPTP is documented in Request for Comment (RFC) 2637 [11]. PPTP uses a Transmission Control Protocol (TCP) connection for tunnel maintenance and a modified version of Generic Routing Encapsulation (GRE) to encapsulate Point-to-Point Protocol (PPP) frames for tunnelled data [12, 13]. These PPP frames can be encrypted. The port used is TCP 1723 and Internet Protocol type 47(GRE).

Layer Two Tunnelling Protocol (L2TP): L2TP is a combination of PPTP and Layer 2 Forwarding (L2F). L2TP represents the best features of PPTP and L2F. L2TP encapsulates PPP frames to be sent over IP. L2TP can be used as a tunnelling protocol over the Intranet or a VPN. L2TP is documented in RFC 2661 [14]. The payloads of encapsulated PPP frames can be decrypted. The L2TP packet format is shown in Fig 1. The PPP payload is encrypted before encapsulation using Data Encryption Standards (DES) algorithm, a symmetric key algorithm before data is transferred. The port used for User Datagram Protocol (UDP) is 1701.

IP header	UDP header	L2TP header	PPP header	PPP payload (IP data- gram, IPX datagram)

Fig 1: L2TP Packet format

Internet Security Protocol (IPSec): Key Management IPSec uses the Internet Key Exchange (IKE) to securely establish and pass shared keys between sites [15]. Keys and security associations may also be passed. Authentication Header (AH) protocol defines methods of establishing the identity of the message originator and ensuring that, transmitted data has not been tampered with. Encapsulating Security Protocol (ESP) protocol provide same functions as the Authentication Header protocol but additionally defines encryption methods for the data [15]. All three components are designed in modular way to incorporate new algorithms and schemas, ensuring forward compatibility as new advancements in encryption or key exchange mechanisms are made. However, IPSec defines lowest level denominators to enable at least minimal interoperability between different vendors’ implementations of an IPSec gird. For instance, all IPSec grid network must include the DES (Data Encryption Standards) encryption algorithm for data encryption [16].

4. IMPLEMENTATION

Grid network provide bulk computational needs such as forensic analysis, genomics analysis, image processing etc. The data integrity during transmission in this work is very less. With the implemented grid network, the encrypted data is encapsulated, or wrapped with a header that provides routing information allowing it to traverse the shared or public transit inter network to reach its endpoints to emulate a private link. For the confidentiality an encryption scheme is used at both ends should be Data encryption standards techniques was used to visualize the security mechanism. The main phases are discussed below.

Connection Establishment: In this phase, a link is established between a super-node and a node in the grid network. The node system presents user's credentials such as username and password to the access super-node. If a valid user, a connection is established between the super-node and the node. The authentication is done using CHAP's authentication scheme. The super-node provides a file transfer service or a message transfer service to the node upon successful login. If the node is a new user, then registration is required for super-node by providing details such as IP address, subnet mask, and default gateway. After registration, the node machine will get a username and password which will be used for further login. These details are stored in the super-node so that these valid users can be listed or viewed later. The connection establishment is shown in Fig 2.

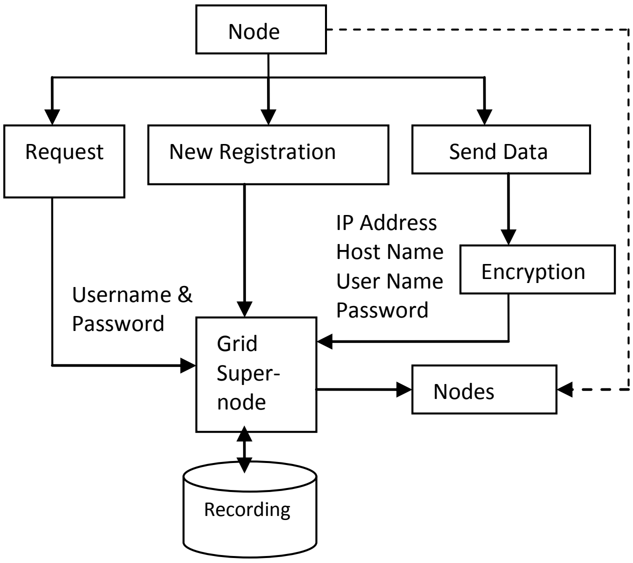

Fig 2: Request, Registration and Send data

Encryption Phase: The node can sent a file or any form of data to some other node through the tunnel super-node. The file may be a job from the user. The data of file can be encrypted using the Data Encryption Standard (DES), a symmetric encryption mechanism [17]. Both the node and the super-node use DES for encryption and decryption.

Address Specification: During this phase the super-node in the grid assigns a dynamic address to the node. There after data transferring is done, based on this dynamic or virtual IP address. When a super-node in the grid network is configured, a set or a pool of static IP addresses can be set. When the node gets connected to the network, it assigned an IP address from this address pool. In the basic configuration, when the node sends a packet to a monitoring super-node, it translates the node's IP address into some virtual or global IP address. Similarly, when an outside network sends a packet to the inside network, the user sends the packet to the virtual IP address. The super-node in the grid translates the virtual IP address into an inside address and sends the packet to the appropriate device on the node system in the private network.

Data Transfer Phase: When a division of job is to be sent to another system, the destination machine's IP address is specified. The data transferring takes place by L2TP protocol's tunnelling technology. When data reaches the receiver's end, it is first decrypted and then stored in the destination machine.

Super-node Communication: The super-node is a computer which used to register jobs and monitor its performance. A communication to the super-node is made possible by the use of Java's Remote Method Invocation (RMI). Java Remote Method Invocation Technology supports method calls between distributed Java objects [18]. RMI uses classical RPC-based interaction, precisely remote method calls. The methods and parameters must somehow be shipped to other machine. The super-node must be informed to execute a method, and return value must be shipped back. The grid system network is implemented as a system with a super-node and two or more nodes.

Node: A node comes for the first time it has to register in the super-node through a registration form. It includes textboxes for entering the IP Address, Host Name, User Name and Password of the node system. These details are sent to the super-node and the super-node will calculate a message digest for the password using Secure Hash Algorithm (SHA-1) [19]. The output of this algorithm is a 160-bit binary hash value. The super-node stores the username and this message digest value of the password. These details are used for user authentication for later time. The super-node will assign a dynamic or virtual IP address to the registering node, which will be stored in the table named, map. Later this virtual address is used for data communication.

When a machine specifies the IP address of another node to which data is to be sent, the super-node will look in the details stored to obtain the virtual address of the destination machine. The data is then send to virtual address of the destination machine. When user wants to send a file or a job, the specified file or division of job can be sent to some other system with the specified physical IP address. The static IP address of the destination machine is specified by selecting an IP address from the list box. The user has to make a selection of whether to sent text message or file to the other system. It is provide a text area for senders and receivers. The message in the textbox is send to the super-node. If the user chooses to send a file, the node interface provides a file dialog to browse and select a file from the local hard disk of the nodes with the help of an open file dialog implemented using a File Chooser component in Java. The user can directly specify the file path.

A grid communication tunnel is enabled between a node and a super-node system, through which the data can be transferred. When the node system specifies the physical IP address of the destination system, super-node in the grid network will map this address with its virtual IP address. The data to be encrypted using DES then encapsulated in an L2TP packet and is send to super-node in a network. This super-node directs this packet to the corresponding system depending on the virtual address. On receiving a file or a job, the data packet is first un-encapsulated, and then it is decrypted. When the user wants to get a connection to the super-node in the grid network, he needs to raise a connection request. The user enters username and password. The username and one way hash value of the password is sent to the super-node in the network.

The super-node receives a communication request that retrieves username and hash value of the password stored in the table during the registration process. The L2TP tunnel is established between the node and super-node, and a session is initiated for a successful authentication. An acknowledgement message is sent by the super-node to the node. The two endpoints of the tunnel created on acceptance of the tunnel are the node and the super-node. When the user intends to send data, IP address of the destination machine can be selected from the node's machine, available in the node list.

Super-node: The super-node waits for the node's requests for getting its connected. Upon authenticated connection, the super-node will provide various services to the node systems. The services include calculating the one-way hash-value of the password during the registration process and store these details. Moreover, the super-node will assign a virtual IP address to the node system. The super-node also handles the request send by the node and retrieves the username and the hash value of the password stored and compares this with that send by the node. If it matches, a communication tunnel will established between the super-node and the node. The super-node also sends an acknowledgement to the node. The super-node lists the IP address, Host Name and login time of the valid node that gets connected in the super-node interface.

The idea designed in Java objects that are integrated to GridSim package for grid enriched works [20]. GridSim is a software platform that enables users to model and simulate the characteristics of Grid resources and networks with different configurations. By using GridSim, one is able to perform repeatable experiments and studies that are not possible in a real dynamic Grid environment.

5. EVALUATION

Performance is simply evaluating the speed and bandwidth measure of packets in the network. The distributed resource usages have quantitative goals such as providing more resources simultaneously, which was registered into the super-node. The second one is qualitative goals such as effective use of network medium within a secure environment. Usually performance measurement is very hard and certain physical characteristics can be measured only. The performance aspect measures in grid network such as CPU power, network speed, cache efficiency, scheduling policies, communication scheme etc. are considered.

In this work we analysed the network speed and communication scheme policies because these parameters are affected after the introduction of new designed packets. The new data management system needs a modeller before its starts. In the initial network (before the proposed system implementation), registration of each job was carried out in the super-node without any security policy. Now we have a mechanism, so that time delay policy may change. Experiments are conducted in the GridSim package with the help of Java RMI. Fig 3. shows the initial and new time lag policies before and after the new security policy introduction. Number of packets and time in micro seconds are given in X-axis and Y-axis respectively. Note that, there is a small change in the time delay policy while registration in the super-node for heavy jobs. In Fig 3. brown line indicates initial time policies and blue line shown present (after the security scheme implementation) time policies. Remember all packets discussed in this work are in TCP/IP protocol suit and essential in super-node registration and job distribution.

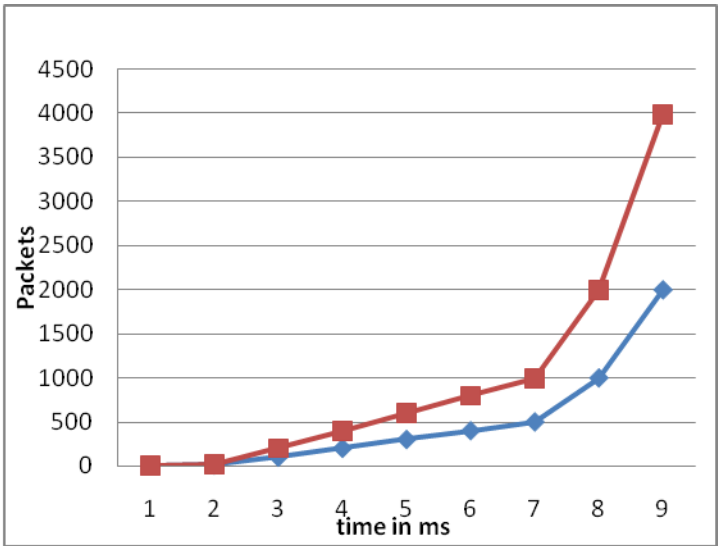

Fig 3. Network speed analysis of the proposed system

Total bandwidth of every end-to-end connection is known and dedicated reservation could be enforced. Normally in IP networks an end-to-end connection are virtual with maximum bandwidth, but offers least security. The proposed method gives to avoid major security threads and reduction in communication traffic of the grid network. This is visualised by implementing a packet oriented scheme in the grid network. We can conclude that the system is better in terms of security in the communication channel of the present grid network.

For the performance evaluation of Grid scheduling under correlated workloads we implement a case study in the evaluation section. In this case study, computing clusters with one First Come First Serve (FCFS) queue. The simulated cluster is space-shared and has 32 CPUs in total. To understand the workload characteristics we analyze trace of data representatives (Table 1). For Job 1 and Job2 we are able to roughly distinguish the Grid jobs and the locally generated jobs. Here Job 1 and Job 2 are image rendering and Job 3 is a genome analysis. By examining the ``user name" field in the traces, jobs from ``pool account" (usually a Virtual Organization name plus a unique number) are considered Grid jobs while jobs from a ``real" user name are seen as local jobs.

Different clusters with different number of CPUs have been fixed with different job arrival rates and autocorrelation structures. The arrival ratio and patterns of local jobs versus Grid jobs are also highly diversified and is shown in Table 1. The job run times have relatively better variances and are almost all long range reliant. These statistics give a good orientation on how to adjust the model parameters such as, selection of nodes in the cluster, for synthetic workload generation.

Table 1. Table captions should be placed above the table

Jobs	No. of CPUs	Mean time on Grid	Mean time on local
Job1	32	0.4 sec	40 sec
Job2	24	0.3 sec	52 sec
Job3	12	1 sec	80 sec

The workload models generate mock traces with different structures and are stored in temporary files. GridSim reads the workloads from the files and carries out the simulation. The designed interface in Java operate user console.

6. CONCLUSION

A grid network allows users to connect to corporate network for high computational work in a useful manner through a network. A secure communication channel is essential for such computation work after its registration. The secure connection appears to the user as a private network communication despite the fact that a communication occurs over an Intranet or a VPN. The system is user friendly, high reliability and work on heterogeneous networked computer systems. Scope of a new protocol exists at this point due to safe communication requirements. The tunnelling is one of the better solutions which offer a new design. The enhanced protocol out performed in a grid network environment with secure transmission by the use of TCP/IP protocol.

7. REFERENCES

[1] I. Foster, C. Kesselman, and S. Tuecke, “The anatomy of the grid,enabling scalable virtual organization”, International Journal on Supercomputer Applications, vol. 15, no. 3, 2001.

[2] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, “The physiology of the grid, an open grid services architecture for distributed systems integration”, Open Grid Service Infrastructure WG, Global Grid Forum, 2002.

[3] Pin Hu, Lingfen Sun and Emmanuel Ifeachor, “An Approach to Structured Knowledge Representation of Service-oriented Grids”, Proceedings of UK e-Science Programme All Hands Meeting, 2007

[4] Donno F. and Ronchieri E., “The impact of grid on healthcare”, System Sciences, pp. 1-9, 2009.

[5] Foster, Ian and Kesselman, Carl, “The Grid: Blueprint for a New Computing Infrastructure”, Elsevier 2nd Edn, 2004.

[6] Foster, I. and C. Kesselman, “ The Grid: Blueprint for a New Computing Infrastructure”, Morgan Kaufmann publ., 1999

[7] Agarwal, D.A., S.R. Sachs and W.E. Johnston, “The Reality of Collaboratories”, Computer Physics Communications, vol. 110, pp. 134-141,1998.

[8] Anirban Chakrabarti, Anish Damodaran, Shubhashis Sengupta, "Grid Computing Security: A Taxonomy," IEEE Security and Privacy, vol. 6, no. 1, pp. 44-51, 2008.

[9] David P. Anderson, Carl Christensen, Bruce Allen, "Designing a Runtime System for Volunteer Computing," ACM/IEEE SC 2006 Conference (SC'06), pp. 33, 2006

[10] I.Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, A. Roy., “A Distributed Resource Management Architecture that Supports Advance Reservations and Co-Allocation”, Intl Workshop on Quality of Service, 1999.

[11] K. Hamzeh, G. Pall, W. Verthein, J. Taarud, W. Little, G. Zorn, “RFC2637: Point-to-Point Tunneling Protocol”, RFC Editor, 1999.

[12] D. Farinacci, T. Li, S. Hanks, D. Meyer, P. Traina, “RFC2784: Generic Routing Encapsulation (GRE)”, RFC Editor, March 2000.

[13] G. Dommety, “RFC2890: Key and Sequence Number Extensions to GRE”, RFC Editor, September 2000.

[14] Naganand Doraswamy, Dan Harkins, “IPSec: The New Security Standard for the Internet, Intranets, and Virtual Private Networks”, Prentice Hall PTR, October 1999.

[15] S. Kent, R. Atkinson, “RFC2406: IP Encapsulating Security Payload (ESP)”, RFC Editor, November 1998.

[16] Srivaths Ravi, Anand Raghunathan, Nachiketh Potlapally, “Special session on security on SoC: Securing wireless data: system architecture challenges”, Proceedings System Synthesis, ACM Press, 2002.

[17] C. Madson, N. Doraswamy, RFC2405: The ESP DES-CBC Cipher Algorithm with Explicit IV, RFC Editor, Nov 1998.

[18] Rüdiger Kapitza, Michael Kirstein, Holger Schmidt, Franz J. Hauck, “FORMI: an RMI extension for adaptive applications”, ARM '05, ACM Press, November 2005.

[19] Mao-Yin Wang, Chih-Pin Su, Chih-Tsun Huang, Cheng-Wen Wu, “Exploration for advanced SoC design: An HMAC processor with integrated SHA-1 and MD5 algorithms”, Proceedings of ASP-DAC '04, 2004.

[20] R. Buyya, M. Murshed, Gridsim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing, Concurrency and Computation: Practice and Experience (CCPE) 14 (2002) 1175-1220.