ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(29), DOI: 10.17485/ijst/2015/v8i29/79163, November 2015* ISSN (Online) : 0974-5645

A Recommendation System for Dynamic Replication and Securing Data Process

Veeramalai Sankaradass[1*] and G. Nalinipriya[2]

[1]Deptartment of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai - 600062, Tamil Nadu, India; veera2000uk@gmail.com
[2]Department of Information Technology, Saveetha Engineering College, Thandalam, Chennai - 602105, Tamil Nadu, India; nalini.anbu@gmail.com

Abstract

This research work focus on distributed system based network issues. The networked servers and computers organize their actions only by passing the proper messages. This networked system having especially significant characteristics like concurrency components and independent failures components. While we share the files through by passing message, there will be security problems because of connecting the entire computers in network. For High performance in data accessing, data surviving and also secure storage designing, the data partitioning and replication techniques can be combined. In this work, two heuristic algorithms developed for two sub problem in which the Authentication of the data share can be done using network manager to improve security. The replicated data are authenticated and send dynamically in the distributed system. The newly developed algorithms provide maximum performances in data sharing through distributed system and in achieving maximum security.

Keywords: Data process, Data sharing, Dynamic replication, Partitioning, Secret sharing

1. Introduction

Distributed computing refers to the use of distributed systems to solve computational problems. In distributed computing, a problem can be divided into different levels, each of which is processed and finalized by one or more computers[1]. The independent data objects are considered initially in the system. The data are secretly shared later, the full set of shares are statically distributed over the network. The most commonly used schemes for data partitioning include secret sharing. As the data are statically distributed information leakage can occur[2]. When it is used in centralized system, single point failure can also occur. The major issue is that the data survivability and security are not present. To overcome this hazard, data partitioning and the replication techniques are combined through which performance growth rate can be increased. To increase the availability and improve efficiency and consistent information, the replication techniques are often applied. The complete network infrastructure will be analysed based on cluster and peer to peer system. Through which, the information originality will be justified.

Information replication techniques are mostly preferred for efficient access. This is performed by encrypting the replicas and sending the data dynamically in distributed system over the network. On implementing this, an additional security process can be provided. It has two functions to find which system is in active peer list state and which is in inactive peer list state. Decentralized adaption identifies number of available paths in particular network then file replication is used to make copies of sent data and finally given to distributed system over the network. The aim of this work is to reduce the replication and communication cost.

An Intrusion tolerant distributed system is designed so that any intrusion into a part of the system will not endanger integrity, confidentiality and availability. [3]The

*Author for correspondence

intruders can be external intruders or internal intruders. The intrusion targets may be information integrity and services availability. Partitioning, fragmentation and scattering are the approaches which are used to increase confidentiality[4]. Eager replication keeps all replicas exactly synchronized at all nodes as a part of one atomic translation. The eager replication uses a locking scheme to detect and regulate concurrent execution[5]. The lazy replication may asynchronously propagate replica updates to other nodes after the updating transaction commits .Lazy replication uses a multi-version concurrency control scheme to detect non-serializable behaviour. Security and performance were the issues considered[5]. To support the evolution of commands & control system there are two major components were designed and developed, they are the infrastructure and data manager.

The infrastructure, data manager were integrated by using the distributed object management technology. The major goals of the initiative included determining the system infrastructure requirements and to identify the migration path of legacy systems. This infrastructure provided the software backplane for applications and insulates software and hardware. Infrastructure had components that included operating system services as well as other services such as inter process communication and scheduling[6]. The application used to host on the infrastructure was the multi sensor integration application. The requirements of these included fault tolerance in real-time scheduling, over load management and admission control. The security and real-time processing were the major issue[7]. The Grid research field is divided into computational grid and data grid. Data grid enables fast and efficient file transfer and replica catalogue is used for managing files. The object location table is used for managing objects and directory services for files. Data Replication Options (DRO) can be used in two ways data can be written first and then replicated or multiple replicas are created and then data can be written.

The consistency of replica data can be done using read only data. The problem is that DRO is used in local area network and is not optimized for use in a wide area network[8]. Dynamic replication can be used to reduce band width consumption and access latency in high performances data grid where user requires remote access to files. The aspects of a grid are sharing of data and sharing of resources. The replication can be static or dynamic, static replication can be used to achieve load balancing but it cannot adapt to user changes[9]. In dynamic replication the replica creation, deletion and management are done automatically to adapt changes in user behaviour. The simulator has three parts, the basic node core simulates various nodes the replication is built above this and the final has trigger file request. The simulation working consists of topology specification, starting simulation, locating nearest replica, file transfer, record keeping, scaling and access patterns[10]. A novel job scheduling strategy called Weighted Scheduling Strategy (WSS) that uses hierarchical scheduling to reduce the search time for an appropriate computing node[11]. It considers the number of jobs waiting in a queue, the location of the required data for the job and the computing capacity of the sites.

Replication decision are made based on cost model that evaluates data access cost and performance gain of each replica[12]. Replicas are created at nodes that receive high number of client request. In hierarchical model the replicas are placed at different levels and communicate in client server scheme and in flat propagation scheme any replica can synchronize with other replicas[13]. Replica Placement Algorithm which uses index list of nodes and list of locations is used[14]. It uses local replica indexes and global replica indexes. Here replication improves access performance in terms of response time and response time is considered as a performance metric as it represents both data transfer cost and gain. The problem is that, it is based on synthetic workload and simplified scenario and it is interested in exploring different adaptive replication algorithm[15]. The static systems have no read or write operation during the period and dynamic system have fragments sent over network during some period[16].

Dynamic allocation algorithm is proposed which assumes stable read-write frequency patterns which moves fragments between servers in such a way to converge to a mapping that will give maximal dynamic assurances. A static assurance does not consider the risk involved in fragment movements and data loss is also to be taken into an account[17]. Replication improves performances and secret sharing offer better data confidentiality[18]. Requests are authorized using access control list at each server which is updated securely in a timely fashion by a system administrator[19]. TCP doesn't impart fastness and UDP doesn't implement security and doesn't guarantee the delivery of packets[20,21]. The share renewal protocol is used for data items that need long term secrecy[21]. Share renewal protocol can be used for renewal[22]. The problem is that

the system provides same level security for all objects and secures authorization services providence. In this paper, we consider the replicas are dynamically allocated to minimize the cost.

2. System Architecture

A collection of independent computers appear to its users as a single coherent system. A distributed system consists of multiple autonomous computers that communicate through a computer network. The Figure 1 shows the communication between distributed systems through the network with the help of active server. When a particular user in the distributed system wants to send the data, he enters into the Network manager. With the help of network manager the authentication is provided for the data. The request enters into active server, whose database gives details of the peer list of systems that are connected to the network; it displays the transaction made in that distributed system. It selects the system for which the transaction is needed. The authenticated data is send to the active system. Then the system replicates the data dynamically to the particular sub system which needs the authenticated data. The data is downloaded by the client who wants it. When the data is dynamically sent, it will be highly secure. As replications are made there is no loss in the data they can be regained by some resources. The request is made from the one of the distributed system to another. For connection it searches the database.

3. Proposed algorithm

In the existing system, distributed system sends the data to the subsystem of a particular distributed system that makes request and also found to be active in the active server. As the number of replications in the system is increased and distributed statically. Thus the number of replications can be lessening by the proposed algorithm as given below.

- Define the sub system in the distributed system.
- Read the data to be send by authenticating it using network manager.
- Compute by selecting the active system from DB of active server.
- Set the system for which the transaction has to be done.
- Send the authenticated data to the system.
- The system send the data to the sub system by replicating the data dynamically.

The algorithm that is described above increases the efficiency by decreasing the number of replicated copies of data.

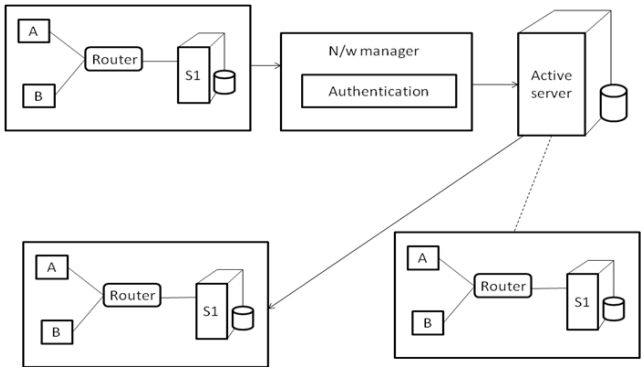

Figure 1. System Architecture.

4. Experimental Results

The proposed recommendation system performance is analysed in terms of replication and cost. Initially, the cost of the system will be gradually increased when the replication rate is increased. After some determined level, the increasing rate of cost will be slowly controlled. The replication level will be random. And the same way it will be allocated and shared with in a group. In the no replication allocation strategy, there is no replication in any cluster. Generally, the cost always will be high when the rate of replication data increased. When the cost became high, then the total system complexity will became more complex. All the time, we have to be very careful in replication as well as cost of the system. Finding error and avoiding complexity is very important in performance analysis. At the same time, the number operation at the particular will be choosing randomly and following a uniform distribution.

In Figure 2 on comparing the above graph the replications that are made dynamically decreases the communication cost. Compared to the complete replication strategy, the performance of our heuristic algorithm is much better when the graph degree is small.

In the above Figure 3 shows the security level between propsed and exisiting system. From this we found that the 10-15 % security level is improved in propsed system.

Table 1 shows the comparison of static and dynamic performance based on the number of replications.

Figure 4 shows the performances of both existing and proposed system. From the above fig we found that the proposed system efficiency is improved 5% compared with existing system.

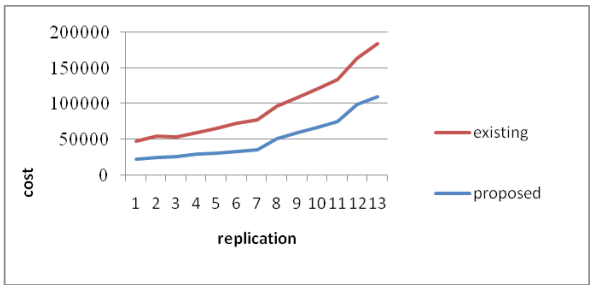

Figure 2. Performance analysis of proposed system.

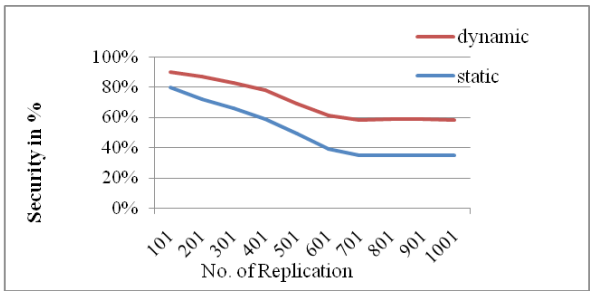

Figure 3. Security level comparsion.

Table 1. Comparison of static and dynamic performance

No.of.Replication	Performance (%)	
	Static	Dynamic
10	80	85
20	70	75
30	60	65
35	50	58
40	45	48

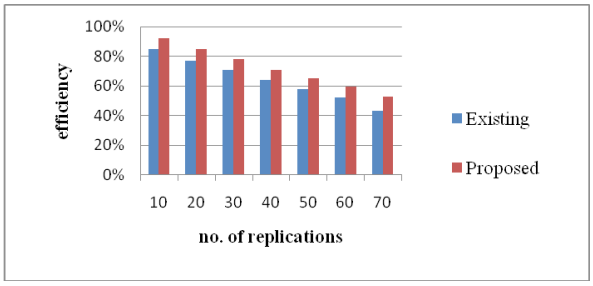

Figure 4. Performance Analysis.

5. Conclusion

The proposed system focused on the replication and sharing cluster. This system determines the number of clusters needs to be maintained to share the replication through which the information can be sent with different nodes without any disturbance. The proposed dynamic replication will helps in improving the performance in data grid and also in providing security for the data. For data security, the symmetric key algorithm is used to improve the security.

6. References

1. Krawczyk H. Secret Sharing Made Short. Proceedings 13th Annual International Cryptology Conference (Crypto'93). 1993. p.136–46.
2. Chervenak A, Deelman E, Foster I, Guy L, Hoschek W, Kesselman C, et al. Giggle: A Framework for Constructing Scalable Replica Location Services. Proceedings ACM/ IEEE Conference Supercomputing (SC). 2002. p. 1–7.
3. Lakshmanan S, Ahamad M, Venkateswaran H. Responsive Security for Stored Data. IEEE Transactions on Parallel and Distributed Systems. 2003; 14 (9):818–28.
4. Nagaratnam N, Janson P, Dayka J, Nadalin A, Siebenlist F, Welch V, et al. The Security Architecture for Open Grid Services, Version 1. 2002; 82–93.
5. Ranganathan K, Foster I. Identifying Dynamic Replication Strategies for a High Performance Data Grid. Proceedings Second International Workshop Grid Computing, 2001; 75–86.
6. Arora S, Raghavan P, Rao S. Approximation Schemes for Euclidean k-Medians and Related Problems. Proceedings 30th ACM Symposium Theory of Computing (STOC'98). 1998; 106–13.
7. Tu M. A Data Management Framework for Secure and Dependable Data Grid. PhD dissertation. University of Texas at Dallas. Available from: http://www.utdallas.edu/~tumh2000/ref/Thesis-Tu.pdf. 2006 Jul.
8. Thuraisingham BM, Maurer JA. Information Survivability for Evolvable and Adaptable Real-Time Command and Control Systems. IEEE Transactions on Knowledge and Data Engineering, 1999; 11(1):228–38.
9. Rabin M. Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance. Journal of the ACM, 1989; 36(2):335–48.
10. Samar A, Stockinger H. Grid Data Management Pilot (GDMP): A Tool for Wide Area Replication. Proceedings IASTED International Conference Applied Informatics (AI). 2001; 7: p. 50–64.

11. Mansouri N, Dastghaibyfard GH. Enhanced Dynamic Hierarchical Replication and Weighted Scheduling Strategy in Data Grid' Journal of Parallel and Distributed Computing, 2013; 73(4): 534–43.
12. Singh G, Bharathi S, Chervenak A, Deelman E, Kesselman C, Manohar M,et al. A Metadata Catalog Service for Data Intensive Applications. Proceedings ACM/IEEE Conference Supercomputing (SC). 2003; 33.
13. Stockinger H. Distributed Database Management Systems and the Data Grids. Proceedings 18th IEEE Symposium Mass Storage Systems and Technologies. 2001; 1.
14. Baker M, Buyya R, Laforenza D. Grids and Grid Technology for Wide-Area Distributed Computing. Software- Practice and Experience. 2002; 32(15):1437–66.
15. Foster I, Lamnitche A. On Death, Taxes, and Convergence of Peer-to-Peer and Grid Computing. Proceedings Second International Workshop Peer-to-Peer Systems (IPTPS). 2003; 118–28.
16. Kalpakis K, Dasgupta K, Wolfson O. Optimal Placement of Replicas in Trees with Read, Write, and Storage Costs. IEEE Transactions on Parallel and Distributed Systems. 2001; 12(6): 628–37.
17. Kubitowicz J, Bindel D, Chen Y, Czerwinski S, Eaton P, Geels D, et al. Ocean Store: An Architecture for Global-Scale Persistent Storage. Proceedings Ninth International Conference Architectural Support for Programming Languages and Operating Systems (ASPLOS). 2000; 35(11). p.199–201.
18. Matossian V, Parashar M. Enabling Peer-to-Peer Interactions for Scientific Applications on the Grid. Proceedings Ninth International Euro-Par Conference. (Euro-Par). 2003. p.1240–7.
19. Mei A, Mancini LV, Jajodia S. Secure Dynamic Fragment and Replica Allocation in Large-Scale Distributed File Systems. IEEE Transactions on Parallel and Distributed Systems. 2003; 14(9):885–96.
20. Reiter M, Rohatgi P. Homeland Security. IEEE Internet Computing, 2004; 8(6):16–7.
21. Rajkumar K, Swaminathan P. Combining TCP and UDP for Secure Data Transfer, Indian Journal of science and technology. 2015 May; 8(S9).
22. Wolfson O, Jajodia S, Huang Y. An Adaptive Data Replication Algorithm. ACM Transactions on Database Systems (TODS). 1997; 22 (2):255–14.