



A SECURE AND ROBUST CONGESTION CONTROL MODEL WITH FRAGMENTATION FOR CLUSTERED DISTRIBUTED DATABASE SYSTEM

Arvind Kumar Maurya
Department of Computer Science
DDU Gorakhpur University
Gorakhpur, India

Surya Pratap Singh
Department of Computer Science
DDU Gorakhpur University
Gorakhpur, India

Manish Mishra
Department of Electronics
DDU Gorakhpur University
Gorakhpur, India

Upendra Nath Tripathi
Department of Computer Science
DDU Gorakhpur University
Gorakhpur, India

Abstract: The purpose of this paper is to present a secure and robust way to adopt Distributed Database System (DDBS). Distributed database is becoming very popular now a day. Today's business environment has an increasingly need for distributed database and client/server applications. We know that the aim of current technologies is to fulfil the desire for consistent, scalable, reliable and accessible information i.e. steadily growing information. In Distributed Database System, the data fragmentation and distributed database site's clustering problems are NP-Hard in nature and difficult to solve. One of the points considered that may cause loss of data means, while upgrading from Relational Database System, data distribution over different data-sites uses fragmentation. The database queries to access the applications on the data-sites and should be performed effectively. Therefore, the fragments that accessed by queries are needed to be allocated to the distributed database sites. Thus, fragmentation should be as secure as data be lossless during query execution. And, the second one may cause expensiveness of technology changes. Thus, we present a method for grouping the sites (clustering) of Distributed Database Systems in order to achieve not only secure fragmentation but also robust and cost effective congestion control.

Keywords: Distributed Database System, Robust congestion control, Repository data-sites, Fragmentation, Clustering network sites.

1. INTRODUCTION

It is well known that during the 1970s computers were extensively used for building powerful, integrated database systems. The technology of database systems has built its theoretical foundations and has been experienced in a large number of applications. At the same time, computer networks have been extensively developed allowing the connection of different computers and exchange of data and other resources between them.

So for, the availability of databases and computer networks has given rise new field: Distributed Database System. In order to understand Distributed Database Systems, we must know more than the principles traditional databases and computer networks. We must also integrate this knowledge with the study of the particular aspect of the new technology. The most important technological problem of distributed databases derives from the issues 'co-operation between autonomous sites' occurred.

Distributed Database Systems are not simply distributed implementation of centralized database systems. Because of they allow the design of systems which present different features from traditional, centralized systems. There is various aspect of data security in Distributed Database Systems such as security at conceptual and application layer. The application layer security can be implemented by using access control policies. Here, whereas, we only concerned with conceptual layer security issues.

Very first, in Distributed Database System, we observe data independence has the same importance as in traditional

database systems. However, a new aspect is added i.e. distribution transparency. In order to do distribution transparency, we define layered reference architecture (figure-1). This architecture provides a very general conceptual framework.

The three most important objectives which motivate the features of this architecture are:

- Location Transparency, the separation of data fragmentation and allocation
- Replication Transparency, the control of redundancy, and
- Fragmentation Transparency, the independence from local data sites.

Apart from this issue, we observe another one that also affects the adaptation of this technology changes i.e. cost ineffectiveness. Here, thus, we consider a cost-effective technique i.e. clustering of data sites. To reduce the expensive change from traditional one to modern distributed environment clustering provides dominated features.

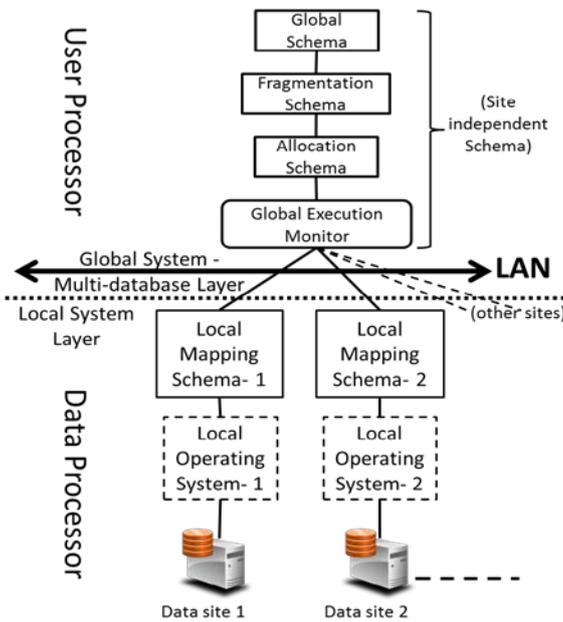


Figure 1.A reference architecture for DDDBS

2. FRAGMENTATION SYNOPSIS

Fragmentation is a design processes that allows dividing a single relation or class of a database into two or more partitions. Fragmentation must be in such way that the condition of the partitions provides the original database without any loss of data and each fragment can be stored at any data-site over a computer network of a Distributed Database system.

Fragmentation aims to improve:-

- Reliability
- Security
- Performance
- Balanced storage capacity and costs
- Communication costs

As we know, at the top level of ‘reference architecture for Distributed Database System’ is the global schema. The global schema consists of the definition of a set of global relations. And, each global relation can be split into several non-overlapping portions which are called fragmentations. The mapping between global relations and fragments is defined in the fragmentation schema. This mapping is one to many. Fragments are logical portions of global relations. Means, fragments are physically located at one or several sites of the network. And, the *allocation schema* defines at which site(s) a fragment is located. At the data processing level, it is required to map the physical images to the local DBMSs. This mapping is called a local mapping schema and depends on the type of local DBMS. So, in a heterogeneous system, we have different types of local mapping at different sites.

The decomposition of global relations into fragments can be performed by applying two different types of fragmentation [1][2]:

- Horizontal Fragmentation, and
- Vertical Fragmentation

But, there is another way a relation can be fragmented i.e. Mixed or Hybrid Fragmentation which can be obtained by using the above both fragmentations.

Horizontal Fragmentation is the subset of the relations that groups all the tuples in accordance with a selection condition on values of one or more fields. Each horizontal fragment must have all columns of the original base relation and should also confirm to the rule of re-constructiveness.

Whereas, the *Vertical Fragmentation* of a global relation is the subdivision of its fields or columns that are grouped into fragments. In order to maintain re-constructiveness, each fragment should contain the primary key field(s). Vertical fragmentation can be used to enforce privacy of data.

Correctness Rules of Fragmentation[1]:-

- Completeness.
- Reconstruction
- Disjointness

Advantages of Fragmentation: -

- Since data is stored close to the local site of usage, efficiency and performance of the database system is increased.
- Local query optimization techniques are sufficient for most queries since data is locally available. Thus, easy to manage balanced storage capacity and cost.
- Since irrelevant data is not available at the local sites, security and privacy of the database system can be maintained that increase the reliability.

Disadvantages of Fragmentation: -

- In the case of recursive fragmentations, the job of reconstruction will need expensive practices.
- Lack of back-up copies of data in different local sites may render the database ineffective in the case of failure.
- When data from different fragments are required, the access speeds may be very high.

3. CONGESTION CONTROL SYNOPSIS

The demand for more and more information both by industry and government leads to databases that will exceed the physical limitations of centralized systems. Thus, with the rise of user based on Distributed Database System, traffic congestion is one of the unavoidable situations. Even though several researchers address the congestion detection technique, its avoidance and mitigation in their research are hard to be explored for any effective solution for this problem. Congestion can be outlined as a condition that happens once the network resources are overburdened. So, resulting in such issues is measured by loss of data or delay. Congestion control is a fundamental mechanism as a troubleshooter of such problems. Congestion management could be reflected as algorithm to share network resources among competitive traffic sources. However, TCP congestion control algorithms can be interpreted as distributed primal-dual algorithms over the Internet to maximize aggregate utility. Hence, there is a need of studies that excavates the loopholes in various approaches and assists to identify a design of new congestion control technique. Congestion management mechanisms in today’s Internet enormously typical to implement. Because, the Internet continues to expand in size, diversity, and reach, taking part in an ever-increasing role within the integration of different networks (transportation, finance, etc.). It is to

be noted that distributed network system must not be thought of as associate uniting of electronic network such as the internet. Hence, we are going to discuss the particular distributed network system that may be additionally classified into a very large network of internet, wireless sensor network, and mobile ad-hoc network.

4. LITERATURE REVIEW

Many studies have been published on attempts of improving the performance of DDBS. These researches have mostly investigated fragmentation, allocation and sometimes clustering problems. In this section, we present the main contributions related to Clustered approach to Congestion Control with Fragmentation.

The authors of [3][4] present a new formulation for the problem of fragmentation and allocating those fragments with minimum cost for both structured and unstructured data, by grouping sites which are nearer to each other into one cluster, hence they have low cost. Also, a dynamic clustering method is adopted for both structured and unstructured database to reduce the movement of data between sites.

The complexity of a distributed database algorithm depends on the allocation method used. Some enhancements have been done in reallocation algorithms. The authors of [6] proposed an algorithm that reallocates fragments based on the distance between sites to minimize the number of communications and network overhead between sites and also calculate the cost for each fragment individually. The reallocation depends on finding the maximum update cost value for each fragment. This technique takes into account the network topology and set of queries frequency values employed over the network.

The authors of [7] present a biogeography-based optimization technique for no-replicated secure allocation of data fragments during database design that minimize total data transmission cost during the execution of a set of queries.

We, now, discuss the most important work dispensed to address congestion controls in networking. Seferoglu et al. [9] have conferred findings on TCP-induced packet losses of TFRC (TCP Friendly Rate Control) flows and their relation with the delay samples and their derivatives as collected/computed at TFRC senders and receivers. Mao et al. [10] have developed a hybrid traffics Active Queue Management (AQM) router with classifier and scheduler that make sure the link capacities of each traffic. The author has conferred some easily verified adequate stability conditions for the AQM policy to stabilize the TCP and UDP queues in routers. Shiang and Schaar [11] have proposed a content-aware congestion management for multimedia system streaming over TCP/IP networks achieving higher than 3dB improvement in terms of PSNR over the traditional TCP congestion control approaches, with the biggest enhancements discovered for real-time streaming applications requiring rigorous playback delays. Rahman et al. [12] have introduced a proxy transport layer protocol Datagram Congestion Control Protocol (DCCP) that's appropriate for these applications due to its exclusive characteristics. Zhou et al. [13] have presented a congestion window adaptation formula for the MPTCP (Multipath Transport Control Protocol) supply that dynamically adjusts

the congestion window for every TCP sub-flow therefore on mitigating the variety of end-to-end path delay. It addresses both wired and wireless networks.

5. CLUSTERING NETWORK SITES

First, we must understand what clustering is: - “a grouping of related items stored together for efficiency of access and resource utilization”.

The parameters considered for the proposed clustering technique are described as follows:

- *Logical Cluster C_i* Logical place that used to group network sites together based on some physical property like distance between them exist.
- *Distributed Network Sites* Set of fully connected network sites S_i, S_j, . . . S_n, of distributed database system. Each site is the place from where the transactions are triggered, and transaction results are held.
- *Distance Range DR* The maximum distance value (in Km) that is allowed between the DDS network sites for grouping into the same cluster can be decided by the network administrator. Shortest path method is used to calculate the distance between two sites.
- *Distance D(S_i,S_j)* The shortest path distance between two sites S_i and S_j in the DDS.
- *Cluster Site Matrix – CSM* Calculated matrix by which the clusters are created and their network sites are assigned.
- *Clustering Decision Value – CDV* The binary value that determines whether a pair of sites S_i and S_j can be grouped together in the same cluster. CDV is calculated using following formula:

$$CDV(S_i, S_j) = \begin{cases} 1 & : \text{if } D(S_i, S_j) \leq DR \wedge i \neq j \\ 0 & : \text{if } D(S_i, S_j) > DR \vee i = j \end{cases}$$

It is obvious that CDV for the same site is equal to zero. If the CDV(S_i, S_j) is equal to 1, then sites S_i, S_j are grouped into the same cluster, otherwise they are assigned to different clusters. Suppose the seven sites of distributed database are placed at some distance (in Km) from one another according to the site distance matrix shown in the following table.

Table-1:- Site distance matrix

Site	S1	S2	S3	S4	S5	S6	S7
S1	0	70	290	150	340	520	180
S2	70	0	640	260	725	400	540
S3	290	640	0	80	60	270	830
S4	150	260	80	0	60	510	420
S5	340	725	60	60	0	350	610
S6	520	400	270	510	350	0	50
S7	180	540	830	420	610	50	0

For setting up an efficient clustering method, it is assumed that each site is assigned to only one cluster. The clustering algorithm is described as follows:

Clustering algorithm

Input:

Site Distance Matrix showing distance between sites D(S_i,S_j)

Distance Range (DR):

Number of network sites of distributed database system (NS)

Processing:

/*Determining the sites that match the distance range in order to group them in one cluster*/

Step 1: Set i=1

Step 2: Do steps (3–12) until i>NS

Step 3: Set j=1

Set 0 to cluster site matrix CSM

Step 4: Do steps (5–10) until j >NS

Step 5: If $i \neq j$ AND $D(S_i, S_j) \leq DR$,

go to step (6) Else, go to step (7)

Step 6: Set 1 to the CSM (S_i, S_j),

go to step 8

Step 7: Set 0 to the CSM(S_i, S_j)

Step 8: End IF

Step 9: Add 1 to j

Step 10: Loop

Step 11: Add 1 to i

Step 12: Loop

Output: Cluster Site Matrix (CSM) having generated clusters and their respective network sites

End

Suppose Distance Range (DR) value is 100 Km. After using the above clustering algorithm and site distance matrix, a Cluster Site Matrix (CSM) is produced as shown in the following table.

Table-2:- Cluster Site matrix (CSM)

Site	S1	S2	S3	S4	S5	S6	S7
C1	1	1	0	0	0	0	0
C2	0	0	1	1	1	0	0
C3	0	0	0	0	0	1	1

6. RESEARCH METHODOLOGY

We observed several works related to data fragmentation and their security in Distributed Database System. In this paper, we find the adaptation of technology changes from relational database to distributed database become too expensive to use it in cost effective manner and issue of data security in Distributed Database System motivate us to apply some new to do data fragmentation. Thus, we proposed the following technique by which we able to apply secure and cost effective way for Distributed Database System.

Secure Fragment Allocation at Clusters due to Congestion

As seen above that fragmentation can be used as a base for data located on different sites. Now characteristics of these data can be used for security purpose. A data can be interrupted on one site then it needs other related data be kept on another site because actual meaning of it is useless unless this partial data is not threaten. If this possibility is to shorten and it is considered that only two sites are having precious data then also half of the data is missing so this is a secure method.

In a large-scale distributed system, different storage sites have a variety of ways to protect data. The same security policy may be implemented in various mechanisms. Data encryption schemes may vary; even with the same encryption scheme, key lengths may vary across the distributed system. The above mentioned factors can contribute to different vulnerabilities among storage sites. Although security mechanisms deployed in multiple data-sites can be implemented in a homogeneous way, besides

this different vulnerabilities may exist due to heterogeneities in computational units.

We start to address security heterogeneity issues by dividing a big storage data-site into several data-sites of different type groups. Each data-site type represents a level of security vulnerability. In a data-site type group, storage sites with the same vulnerability share the same weakness that allows attackers to reduce the data-sites' information assurance. Although it may be difficult to classify all data-sites in a system into a large number of groups, a practical way of identifying data-sites types is to organize these with similar vulnerabilities into one group. In real-world distributed systems, the fragmentation technique is usually combined with replication to achieve better performance at the cost of increased security risk to data stored in the systems. A practical distributed system normally contains multiple heterogeneous data-sites providing services with various vulnerabilities. Unfortunately, the existing fragmentation algorithms do not take the heterogeneity issues into account. Our fragment allocation solution we proposed in this paper is a bit different from the existing fragment allocation schemes discussed so far. As in the figure-2, our solution captures heterogeneous features regarding vulnerabilities of the nodes in order to improve the security and congestion control of the data stored in a distributed system.

Moreover these aspects, a technique of fragment allocation and replication at clusters are evaluated according to the performance generated by reducing the size of fragments that allocated finally at the clusters. The closest methods in the literature to the proposed technique of fragment allocation and replication are those proposed previously by some researchers. The main differences between these two methods are described as follows.

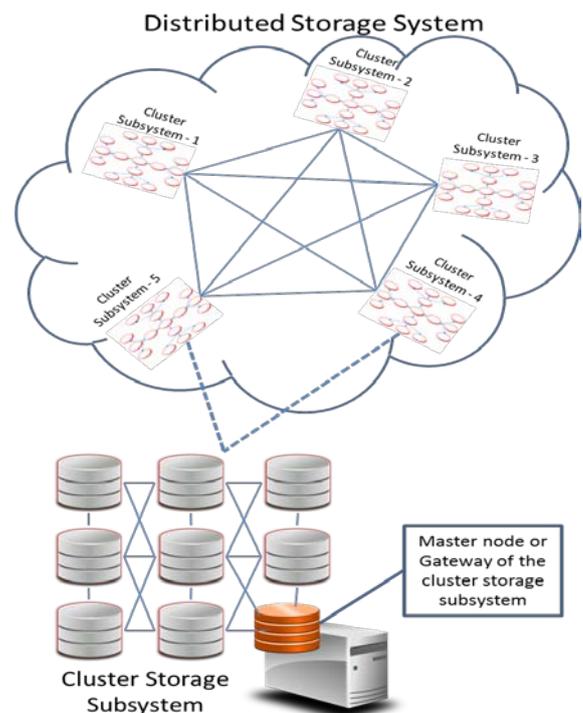


Figure 2. Architecture of Distributed Database System with secure fragment allocation at clusters due to congestion.

If we compare with the different allocation techniques introduced by some researchers as mentioned in literature review, our allocation and replication technique

considers cost effective between cluster sites for network communications such as the update and retrieval costs (i.e., it mostly represents the cost of writing operation that takes place during the execution time). Moreover, in clustering technique, data-sites are grouped according to a clustering range and not only to a specific communication cost. Our clusters can communicate with each other instead of preferring to have all fragments in their sites. This communication technique is cheaper than allocating fragments in all sites. In addition to the advancement in term of communication cost, the independency of our clusters makes our Distributed Database system more reliable and more functional.

In the above figure, a distributed storage system is involved a set of cluster storage subsystems. Multiple fragments of a file can be stored either in storage nodes within a single cluster storage subsystem or in nodes across multiple cluster storage subsystems which consist of a number of storage nodes and a gateway. Storage nodes are divided into different type groups, each of which represents a level of security vulnerability.

7. CONCLUSION

In this paper we studied relevant works done by different researchers and introduced advancement of database security as a result of technology changes. We proposed a robust congestion control model with fragmentation. This technique has applied to the Distributed Database System in the form of clusters.

In our model we proposed Clustering of different data-sites which aims to truncate expensive adaptation from Relational to Distributed database. In such database system, security of data considered to be important in case of upgradation from relational database. So, to overcome from all the problems we proposed congestion controlled fragmentation for relations in heavily loaded centralized database.

REFERENCES

[1] Özsu, M.T. and Valduriez, P. “Principles of Distributed Database Systems” Third Edition (2011) Springer New York Dordrecht Heidelberg London.

[2] Huang, Y.F. and Chen, J.H. “Fragment Allocation in Distributed Database Design” No. 17, (2001) Journal of Information Science and Engineering, pp. 491 -506.

[3] Khan, S.I. and Dr. Hoque, A.S.M.L. “A New Technique for Database Fragmentation in Distributed Systems” <http://dx.doi.org/10.5120/940-1318> Volume 5– No.9, August (2010), International Journal of Computer Applications.

[4] Hababeh, I. “Improving Network Systems Performance by Clustering Distributed Database Sites” <http://dx.doi.org/10.1007/s11227-010-0436-9> Volume 59, Issue 1 January(2012), The Journal of Supercomputing, pp. 249-267.

[5] Ceri, S., Pelagatti, G. “Distributed Databases Principles and Systems” Second Edition (1984) McGraw-Hill, New York

[6] Abdalla, H.I. “A New Data Re-Allocation Model for Distributed Database Systems” Volume 5, No. 2 (2012)

International Journal of Database Theory and Application, pp.45-60.

[7] Singh, A., Kahlon, K.S. and Virk, R.S. “Nonreplicated Static Data Allocation in Distributed Databases Using Biogeography-Based Optimization” Volume 2014 (2014), Chinese Journal of Engineering, pp.1-9

[8] Ahmad, I., Karlapalem, K., Kwok, Y.K., So, S.K. “Evolutionary algorithms for allocating data in distributed database systems” Volume 11, Issue 1, January(2002), Distributed and Parallel Databases, pp. 5-32

[9] Seferoglu, H., Ulas, C.K., Civanlar, M.R., Kempf, J. “Congestion State-Based Dynamic FEC Algorithm for Media Friendly Transport Layer”, (2009) IEEE pp. 1-10

[10] Mao, P., Xiao, Y., Qu, G. “Hybrid Traffic Congestion Control Based on 2-D Hurwitz-Schur Stability”, December (2010) Int. Conf. Control, Automation, Robotics and Vision Singapore, pp.7-10

[11] Shiang, H-P., and van der Schaar, M. “Content-aware tcp-friendly congestion control for multimedia transmission” (2011) International Conference on Acoustics, Speech, and Signal Processing

[12] Rahman, J., Saha, S., and Hasan, S.F. “A New Congestion Control Algorithm for Datagram Congestion Control Protocol (DCCP) Based Real-time Multimedia Application”, (2012) International Conference on Electrical and Computer Engineering.

[13] Zhou, D., Song, W., Shi, M. “Good put Improvement for Multipath TCP by Congestion Window Adaptation in Multi-Radio Devices” (2013) IEEE Consumer Communication and Networking Conference.



Mr. Arvind Kumar Maurya is MCA and UGC-NET qualified and pursuing Ph.D. in the department of Computer Science DDU Gorakhpur University, Gorakhpur (U.P. India) under the supervision of Dr. U.N. Tripathi. The area of research interest is Distributed Database Security, Networking. Arvind Kumar Maurya has published 5 papers in different national and international conferences/Journals



Dr. Surya Pratap Singh is MCA, UGC-NET qualified and Ph.D. from the department of Computer Science DDU Gorakhpur University, Gorakhpur (U.P. India).. The area of research interest is Database Security, Networking. Dr. Surya Pratap Singh has published

more than 30 papers in different national and international conferences/Journals.



Dr. Manish Mishra is Assistant Professor in Department of Electronics DDU Gorakhpur University, Gorakhpur (U.P. India). He has 15 years of teaching and research experience. He has published 55 papers in various National and International Journals/conferences. His area of research interest is Computer Technology, fast processor design.



Dr. Upendra Nath Tripathi is Assistant Professor in Department of Computer Science DDU Gorakhpur University, Gorakhpur (U.P. India). He has 15 years of teaching and research experience. He has published 50 papers in various National and International Journals/conferences. His area of research interest is database systems, networking.