Research Journal of Applied Sciences, Engineering and Technology 10(3): 288-292, 2015
DOI: 10.19026/rjaset.10.2489
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: December 14, 2014 Accepted: February 22, 2015 Published: May 30, 2015

Research Article

Efficient Method for Secure Key Matching Process of Large Data Set Integration in Grid Computing

[1]K. Ashokkumar, [2]C. Chandraseakr, [1]M.K.S. Akshya and [1]Leona Vincy
[1]Department of Computer Science and Engineering, Sathyabama University, Chennai-600119, India
[2]Department of Computer Science, Periyar University, Salem, India

Abstract: To efficient process of large data set and quick integration in grid computing using the key matching, K-NN and cladogam method, that deals with lot of computing resources and each resource will locate in different locations. Data keep in various localities are searched in parallel mode and firmly accessed and permit to share the info between numerous grids by exploitation searching, key matching techniques. Key matching technique is the process in which two dataset are mapped together by matching the key. The key generated by the algorithm will be used to match the datasets. And Data integration is used to integrate different datasets from various localities into one large dataset. By using key matching algorithm the each dataset can assign different key it helps to integrate different dataset which lies in different location. This helps to integrate single level clusters into multiple clusters. From the general finding of this analysis study conjointly well-tried that our planned frame work might perform higher within the handling of huge knowledge storage and access than different existing frame work utilized in grid computing environments.

Keywords: Cladogram algorithm, data integration, grid computing, key matching, K-NN

INTRODUCTION

The grid computing aims to provide an infrastructure allowing access to a wealth of sharable resources such as processing power, storage, databases, applications and any other devices (hardware) or components (software); all of which can be reached computational science and engineering, experimental science, industrial engineering, corporate communications.

In this grid system we have multiple clusters/grid resources in each and every data set. So, execution in the form of searching, matching everything is possible in data set. Because, each dataset has their own cluster system. Normally in grid system the datasets are situated in different clusters. So it is difficult to access the datasets because these dataset will reside in different location and it will be in single clusters (Kumar and Sekar, 2010; Michael, 2005). In order to avoid this we going for multiple clusters system. In multiple cluster system the similar type of datasets are integrated together using the key matching and integration algorithm.

Key matching process is the technique in which the key is assigned for every data set. That dataset have unique id which is termed as a key (Ashok Kumar and Chandra Sekar, 2014). The key match with assigned key which is presented in each data set. Once matching occurs, corresponding data in the dataset will display for process and integration (Ashok Kumar and Sankar, 2014). And the security of the system will be high and after performing this function the integration of the dataset take place and it will protect any unwanted usage of data due to framework.

MATERIALS AND METHODS

Related work: This approach uses an authentication protocol in order to improve the authentication service in grid environment. Secure group communication is brought about by effective key distribution (Sudha Sadasivam *et al*., 2010). To authenticated users of the channels serviced by resources and facilitates reduced computation and efficient group communication (Archana *et al*., 2011; Xukai *et al*., 2007). The main drawback of this study is the scheme includes manual key distribution, hierarchal trees and secure lock as result the generation of key takes more time.

This study focuses on an extensible and open Grid architecture, within which protocol, services, application programming interfaces and software system development kits are categorized according to their roles in enabling resource sharing (Ian *et al*., 2001). The drawback herer is Grid Problem and Class Problem.

Corresponding Author: K. Ashokkumar, Department of Computer Science and Engineering, Sathyabama University, Chennai-600119, India

This study proposes an elegant Dual-Level Key Management (DLKM) mechanism using an innovative concept/construction of Access Control Polynomial (ACP) and one-way functions (Xukai *et al*., 2007). Internet is not security-oriented deliberately; there exist with some attacks, especially malicious internal and external users (Archana *et al*., 2011).

Focuses on virtualization comes in many types focusing on control and usage schemes that emphasize efficiency. Data from different hard drives can be coalesced into a central location. It does not focus on the Performance and the efficiency (Ashok Kumar and Sankar, 2014).

Here Identity-Based Cryptography (IBC) has the properties which seem to align well with the demands of grid computing. The problem facing is the Key generation time consumption (Lim and Paterson, 2005).

Existing method: In the existing system, the data set is configured in single level clusters and each and every data set is located in different location. And in this system there is no proper method for maintaining the dataset. In order to access any dataset we have to access the entire dataset concurrently to get information which leads to high accessing time and this process is costly in nature. Moreover the accuracy of the existing system is low because the giving the user reasonable size response with high precision can mean missing several relevant texts. The consistency of this system is low because many information retrieval environments require indexing of the text by the groups of indexers by the users.

Proposed method: The proposed method is aimed at integrating multiple data set of database efficiently and that multiple data sets will compute by the grid infrastructures in key level. This can be providing faster search of data set and its integration to required area. The searching and integration is based on the process of looking within the NB-Tree initiates by shrewd the question purpose norm. Currently one Dimensional B+-tree will be searched against question points. The search procedures square measure supported the question sort. Grid assortment ways support 3 forms of queries as follows:

- Point question
- Vary question
- k-NN question (Here K-NN represent k nearest neighbor)

This study uses purpose question since the administrator and user needs specifically matched records from the information. Vary question and k-NN question square measure utilized in content-based retrieval. This system is also mainly focused on large data sets based service of ECC key matching algorithm is used for secure and resolving the service discovery that can deal with large data set uncertainty of service providers in Fig. 1.

Algorithm used in the proposed system are k-NN, Elliptic curve cryptography algorithm, Cladogram algorithm.

Key generation: Key generation is an important part where we have to generate both public key and private key. The sender is going to be encrypting the message with receiver's public key and the receiver will decrypt its private key. And key for matching the data sets. These keys will use for fetching data from the different localities databases and its data set. Also it will be involve for concurrent parallel process.

Now, we have to select a number's' within the range of 'n'. Using the following equation we can generate the public key Q = d * Where d = the random number that we have selected within the range of (1 to n-1). P is the point on the curves' is the public key and 's' is the private key. Where n is number of resources:

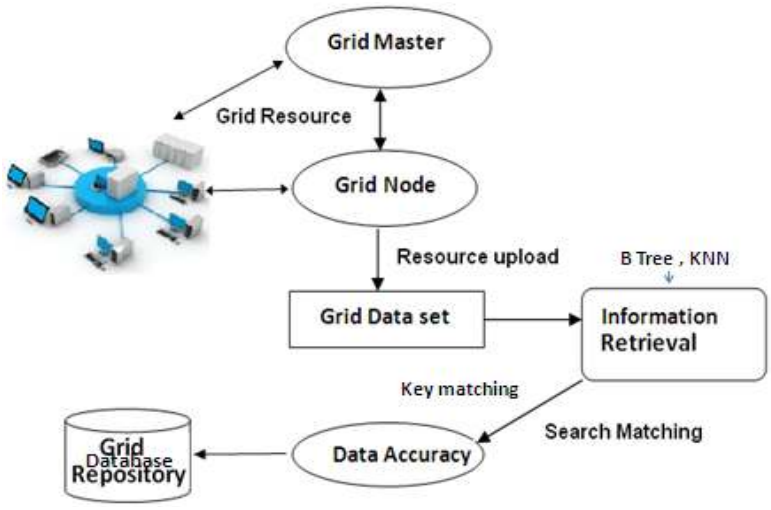

Fig. 1: Proposed architecture

```
leaf = search + Tree (dist (q))
Do
leaf = leaf.PreviousSearchToTheRight
upperLimit = upperLimit + delta
While (leaf.key<= upperLimit)
If (dist (leaf.point, q) <list.LastElement Or list.size
<k)
list.Insert (leaf.point)
End If
leaf = leaf.right
End While
leaf = leaf.PreviousSearchToTheLeft
lowerLimit = lowerLimit + delta
While (leaf.key>= lowerLimit)
If (dist (leaf.point, q) <list.LastElement Or
list.size<k)
list.Insert (leaf.point)
End If
leaf = leaf.left
End While
While (dist (list.LastElement (), q) >radius)
kNN = list [0, k-1]
```

Searching nearest dataset: Nearest Search (used in allication for search efficiency):

- Multistep NN (Q, K)
- Retrieve the k NNs (P1, ...Pk) of Q According to DST
- RS = {P1,....., Pk), sorted according to DST
- DST max = DST (Q1, Pk) // the current k th NN DST
- P = next NN of Q according to dst
- While DST (Q, P) <DST max
- If DST (Q, P) <DST max
- .Insert P into RS and remove previous k th NN
- Update DST max over RS
- P = next NN of Q according to dst

Psuedo code.

Encryption: Let 'm' be the message that we are causation. We have to represent this message on the curve. Consider *'m'* has the point *'M'* on the curve *'E'*. Randomly select 'k' from [1- (n - 1)]. Two cipher texts will be generated let it be C1 and C2:

$C1 = k*P$
$C2 = M + k*Q$
C1 and C2 will be send

Decryption: We have to get back the message 'm' that was send to us, M = C2 - d * C1 m is that the original message that we have send.

For subgroup number $h = \frac{|E(IF_p)|}{n}$ domain parameters are (p, a, b, G, n, h) and in the binary case they are (m, f, a, b, G, n, h).

Table 1: Relationship between the nodes and datasets

	Nodes (resources, its DB)				
Out group	1	2	3	4	5
Dataset A	1	0	0	0	1
Dataset B	1	1	0	1	0
Dataset C	1	0	1	1	0

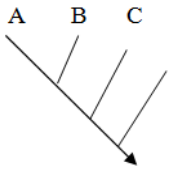

Fig. 2: Integration of dataset to user view for process after match

Cladogram algorithm for data integration: A cladogram is a diagram used in cladistics which shows relations among datasets here. A cladogram is not, however, an evolutionary tree, this makes the relationship between the nodes and datasets. Here for experiment, 5 nodes used and three different and related data set used. And due to this classification, the system shows the exact and relevant match among relations and integrations (Table 1 and Fig. 2).

This is to understand with a data set where A, B, C are Datasets, 1, 2, 3, 4, 5 are computing nodes/grid resources, respectively.

Clad gram algorithm works as:

- Gather and organize the set of resources.
- Identify the best candidates that are most consistent with the characteristic data.
- Create additional candidates by creating several variants of each of the best candidates from the prior step.
- Use heuristics to create several new candidate clad grams unrelated to the prior candidates.
- Repeat these steps until the cardiograms stop getting better.
- Select the best clad gram.

RESULTS AND DISCUSSION

User authentication: In the user authentication module, there are two users involved in the system. The Admin user mainly performs the operation of the Grid server. The user uses the valid credentials in order to connect to the Grid server.

Grid node: In the Grid Node, the inputs are taken from the user and are checked with the grid server. The inputs after the validation perform the user to make further operations. The operations performed by the user are updated to some of the functionalities of the grid server.

Grid dataset: The data is created and loaded by the grid server. The created tax_id GeneID, Protein-coding

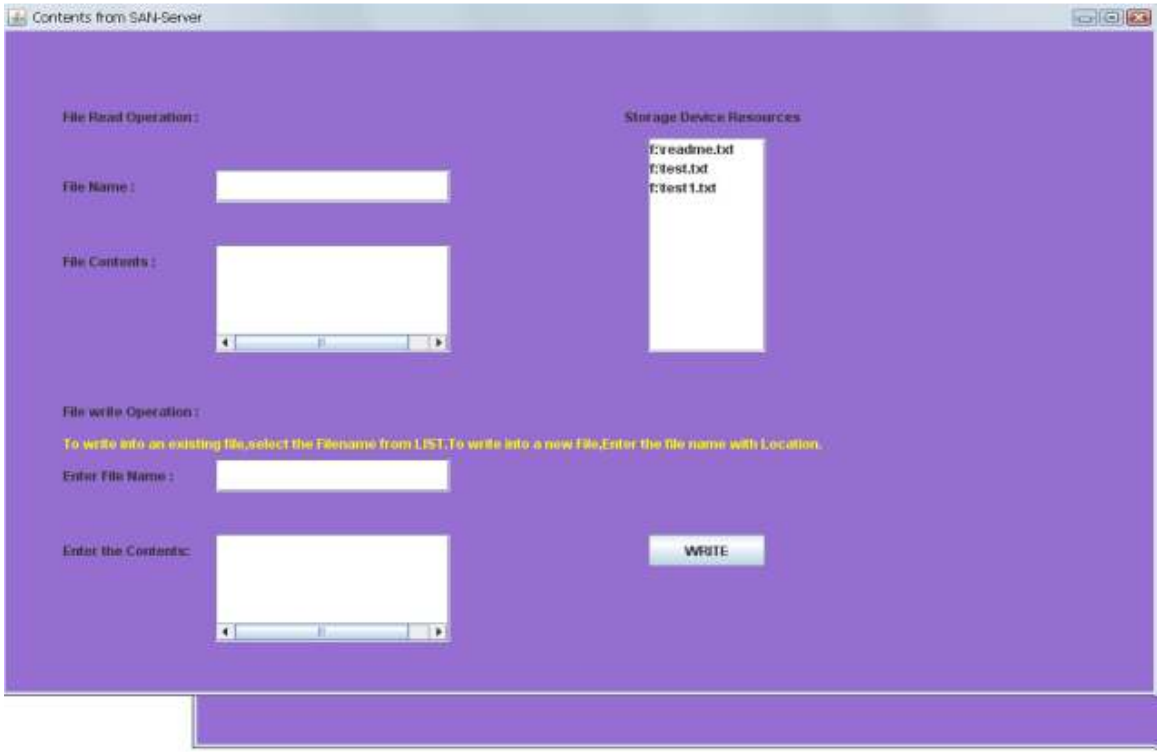

Fig. 3: Software key matching user interface

data is stored in the text file that is considered as the dataset. The prepared dataset needs to be used.

Key matching: The data needs to be secured in the network. To maintain the security, the ECC algorithm is used in which the private key and the public keys are generated automatically. This enables the data to be protected by means of the encryption technique performed by the ECC algorithm:

```
Int Key id;
If (key id = = key id in db) then Match found;
Else Search goes until match found;
End;
```

Data analysis: The key matching process will be performed by the ECC algorithm by matching the public key with the private key. After the matching of the keys, the data will be decrypted and analyzed.

Experimental design: In the experimental analysis, the credentials that are taken from the user will be some of the attributes like the account no, user id and the password. The user is created by the administrator that will be recorded in the dataset by means of a text file. The text file will act as the dataset using which the key matching process will be performed. For the key matching process, the ECC algorithm will be implemented. The dataset will be encrypted by creating the public key which will be stored in the ECC file. The encrypted file will be given as the input to the key matching process in which the public key will be generated. The generated public key and the private key will be matched and the decryption process will be take place in Fig. 3.

CONCLUSION

The Efficient method for key matching for large dataset in grid computing aims to match the different datasets first using the key generated by the elliptical curve cryptography algorithm and k-NN. In second step it integrates different datasets residing in different location using the cladogram technique. Hence single level clusters can be clustered into multilevel clusters towards faster integration and it improves the security of the data set as well.

REFERENCES

Archana, M., N. Suman, K. Gaurav and K. Sandeep, 2011. High performance architecture and grid computing. Proceeding of International Conference, HPAGC 2011. Chandigarh, India, July 19-20, 2011. Communications in Computer and Information Science 169, Springer 2011, ISBN: 978-3-642-22576-5.

Ashok Kumar, K. and E. Sankar, 2014. Efficient method for parallel process and matching of large data set in grid computing environment. J. Eng. Sci. Technol. Rev., 7(4): 109-113.

Ashok Kumar, K. and C. Chandra Sekar, 2014. Optimized method for heterogeneous integration and process of large data set in grid computing environment. Proceedings of the ICMS 2014-Elsevier Conference.

Ian, F., K. Carl and T. Steven, 2001. The anatomy of the grid: Enabling scalable virtual organizations. Int. J. High Perform. C., 15(3): 200-222.

Kumar, K.A. and C.C. Sekar, 2010. Data management and heterogeneous data integration in grid computing environments. Proceeding of the 2010 International Conference on Communication and Computational Intelligence (INCOCCI, 2010), pp: 437-442.

Lim, H.W. and K.G. Paterson, 2005. Identity-based cryptography for grid security. Proceeding of the 1st International Conference on e-Science and Grid Computing. Melbourne, pp: 395-404.

Michael, D.S., 2005. Distributed data management for grid computing. John Wiley & Sons, Hoboken.

Sudha Sadasivam, G., V. Ruckmani and K. Anitha Kumari, 2010. Secure group communication in grid environment. Int. J. Secur., 4(1).

Xukai, Z., D. Yuan-Shun and R. Xiang, 2007. Dual-level key management for secure grid communication in dynamic and hierarchical groups. Future Gener. Comp. Sy., 23(6): 776-786.