International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Secure Information Hiding Approach in Cloud Using LSB

Mudasir Ahmed Muttoo[1], Pooja Ahlawat[2]

[1]M. Tech. Scholar, Department of Computer Science and Engineering, R.N. College of Engineering and Management, Rohtak.

[2]Assistant Professor, Department of Computer Science and Engineering, R.N. College of Engineering and Management, Rohtak.

Abstract: ***The continuous development of cloud computing is giving way to more cloud services, due to which security of cloud services, especially data privacy protection, becomes more critical. This research work explores the basic features of data mining techniques in cloud computing and securing the data. The status of the development of cloud computing security, the data privacy analysis, security auditing, data monitoring and other challenges that the cloud computing security faces have been explored. The recent researches on data protection regarding security and privacy issues in cloud computing have partially addressed some issues. The implementation of data mining techniques through cloud computing encourages the users to extract meaningful hidden predictive information from virtually integrated data warehouse that reduces the costs of storage and infrastructure.***

Keywords: Cloud Computing, data security, Knowledge Discovery

1. Introduction

The constant development of information technology in different fields of human life has provoked the broad volumes of information storage in various formats like records, documents, images, sound recordings, videos, scientific data, and many new data formats. The information assembled from diverse applications requires legitimate knowledge/information extraction system to contribute in better decision making. Knowledge discovery in databases (KDD) goes for the exposure of important data from huge accumulations of information. Data mining incorporates numerous methods and algorithms to discover and extract patterns of stored data. From the most recent two decades data mining and knowledge discovery applications have got much attention due to its significance in decision making and it has turn into a vital segment in different associations [1].

Association rule mining is an important research topic of data mining; its task is to find all subsets of items which frequently occur, and the relationship between them. Association rule mining has two main steps: the establishment of frequent item sets and the establishment of association rules [11].

2. Data Mining

In this technological era, the data is being generated at an enormous rate. As the advancements in electronics and computer technologies have empowered almost unlimited storage resources, virtually every bit of new data is stored, preserved, and made available. The Internet hosts an almost unimaginable amount of human-generated data across the globe.

Modern computer technology and software design takes into account the data acquisition and storage as well as large scale data analysis. The endeavours in this direction have brought about various elite data mining libraries, online assets for data search and investigation, cloud services for data storage and analysis, open standards for data exchange and descriptions of data analysis models which portrays the semantics of data, devices and administrations with the mean to empower their interoperability.

The term data mining signifies the action of extracting new, valuable and nontrivial information from extensive volumes of data. The aim is to discover patterns or fabricate models using particular algorithms from various scientific disciplines including artificial intelligence, machine learning, database systems and statistics. The data mining tasks can be classified into two categories with respect to this definition:

1. **Predictive data mining** where some variables or fields in the database are used to predict unknown or future values of other variables of interest and the goal is to build an executable model from data which can be used for classification, prediction or estimation.

2. **Descriptive data mining** where the focus is on finding human-interpretable patterns describing the data and relationships in data. The KDD process includes an iterative sequence method [3], [4]:

- **Selection:** The principal step begins with gathering the vital and important information about the domain and setting the objectives to be accomplished. This data is then utilized as a part of the planning of a data set which incorporates selecting a proper (sub) set of information tests and/or variables.
- **Cleaning and Pre-processing**: It incorporates discovering mistaken or missing information. It additionally incorporates removal of noise or exceptions along with gathering important data to model or record for noise, representing time arrangement data and known changes.
- **Transformation**: It is changing over the data into a typical configuration for processing. Some data may be encoded or changed into more usable format. Data reduction, dimensionality reduction & data transformation method may be used to reduce the number of possible data values being considered.
- **Data Mining:** This is the most elaborate step as it consists of choosing the function of data mining, choosing the right data mining algorithm and its application. Choosing the

function includes deciding the purpose of the resulting data mining model, such as classification, regression, clustering and summarization. The selection of the data mining algorithm encompasses the decision which models and parameters are appropriate and matching with the criteria of the process.

- **Interpretation/Evaluation:** In the last step the discovered patterns are evaluated and their validity and relevance are assessed. Redundant and irrelevant patterns are removed while the remaining, relevant patterns are studied and interpreted. Application of the discovered knowledge includes resolving potential conflicts with existing knowledge, taking actions based on the obtained knowledge, such as aiding, modifying and improving existing processes and procedures, especially those involving human experts, and storing, documenting and reporting to interested parties.

3. Cloud Computing

Cloud Computing is a booming era which guarantees reliable, scalable, pay-per-use, customized and dynamic computing environments for end-users. The quickly evolving technology is subsequently leading to the rise in the requirements of the clients. This new paradigm of cloud computing is appealing vendors and various associations have begun understanding the profits by putting their applications and data into the cloud. This helps in cheaper and efficient utilization of available resources and easier handling of larger computational problems.

Cloud Computing is an aggregation of two technological terms - Cloud and computing. Cloud is a pool of heterogeneous resources and a mesh of huge infrastructure including both these applications to be delivered as services over the Internet and the hardware and system software in data centers required for providing those services. Computation in cloud is done based on SLA with the aim to achieve maximum resource utilization with higher availability at minimized cost.

Cloud service delivery is divided among three service models- Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS).

The main goal of cloud computing is to combine the distributed resources to achieve higher throughput, high resource utilization and be able to solve large scale computation problems. The cloud computing has many potential advantages in comparison to traditional IT model. But the major barrier for the adoption of cloud computing are the security concerns. Security control measures in cloud are similar to ones in traditional IT environment.

4. Security of Data in Cloud

Security is a key barrier to the broader adoption of cloud computing. The real and perceived risks of providing, accessing and controlling services in multitenant cloud environments can slow or hinder the migration to services by IT organizations [12]. Although cloud computing promises lower costs, rapid scaling, easier maintenance, and service availability anywhere, anytime, a key challenge is how to ensure and build confidence that the cloud can handle user data securely. To make the cloud computing be adopted by users and enterprise, the security concerns of users should be rectified first to make cloud environment trustworthy.

The development of new services bring along new opportunities and challenges. At present, almost all IT enterprises are involved in cloud storage by services provision. But while provision of services, we must take into account the problems emerging from the storage operations in cloud. When the data store on personal devices, users have the highest privilege to operate on them and ensure its security. But once the users choose to put data into cloud, they lose their control over the data [9].The user's authentication and authorization is needed to access the data so as to prevent stealing other user's data through service failure or intrusion.

The data in the cloud may be divided into the data in IaaS environment and the data in PaaS or SaaS environment related to cloud based applications. The data stored in the cloud storages is similar with the ones stored in other places and needs to consider three aspects of information security: confidentiality, integrity and availability. The common solution for data confidentiality is data encryption. To ensure the effect of encryption, the use of both encryption algorithm and key strength are needed to be considered. As the cloud computing environment encompasses large amounts of data transmission, storage and handling so there also needs to consider processing speed and computational efficiency of encrypting large amounts of data. In such cases, symmetric encryption algorithm is more suitable than asymmetric encryption algorithm. The major issue about data encryption is key management. The major issue considered in key management is as who will be responsible for key management. Ideally, the data owners are responsible for managing the key. As the cloud providers need to maintain keys for a large number of users, key management become more complex and difficult [6].

5. Steganography

The scope of the work is to extract the useful information from large amount of data and store at cloud in secure fashion and then make inferences required by the organization. But the predictions that are generated as a result of mining should be secure from any kind of interception. In this sense, steganography is the best option for sending information secretly because it hides the existence of secret message and provides more security. The security module which is used is image steganography as images are the most popular because of their frequency on the Internet. So the prime focus is to increase the capacity to provide better security during transmission.

Steganography is the process of hiding the one information into other sources of information like text, image or audio file, so that it is not visible to the natural view. There are varieties of steganography techniques available to hide the data depending upon the carriers we use. In steganography the message is kept secret without any changes but in cryptography the original content of the message is differed in different stages like encryption and decryption.

Steganography supports different types of digital formats that are used for hiding the data. These files are known as carriers. The main file formats that are used for steganography are text, images, audio, video, protocol. Images are the most popular cover objects used for steganography.

6. Apriori Algorithm

Since there are usually a large number of distinct single items in a typical transaction database, and their combinations may form a very huge number of item sets, it is challenging to develop scalable methods for mining frequent item sets in a large transaction database. The Apriori algorithm is the most general and widely used association rule mining algorithm. Apriori is designed to operate on databases containing transactions. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first-search and a tree structure to count candidate item sets efficiently. It uses an iterative method called layer search to generate (k+1) item sets from k item sets. A k-item set is frequent only if all of its sub-item sets are frequent. This implies that frequent item sets can be mined by first scanning the database to find the frequent 1-itemsets, then using the frequent 1-itemsets to generate candidate frequent 2-itemsets, and check against the database to obtain the frequent 2-itemsets. This process iterates until no more frequent k-item sets can be generated for some k. This is the essence of the Apriori algorithm [14]-[15].

Apriori Algorithm is based on rule parameters – support, confidence and number of cycles used, but these rule measures are not considered for Predictive Apriori Algorithm. The default number of best rules in Apriori Algorithm and in Predictive Apriori Algorithm is respectively 10 and 100. In case of Apriori Algorithm, the number of cycles required to generate best rules has inverse relation with the value of minimum support being used and is independent of number of attributes and instances. The minimum support threshold has great effect on the best rules produced and also on the average size of frequent item sets. In Apriori Algorithm, the number of best rules generated are independent of the number of instances and attributes but are dependent on the value minimum support taken.

In Predictive Apriori Algorithm, the best rules depends on the dataset being used and the number of selected attributes .The greater the number of best rules, the greater the expected accuracy. A rule is added if the expected predictive accuracy of the particular rule is among 'n' number of best rules and it is not a part of another rule with at least the same expected predictive accuracy.

7. Motivation

The major concerns of users or companies, which put their information on the cloud is they are having no idea what's happening to it. When they will have audit of when their information is approached, who access the data increase to strengthen the confidence that their information is being handled properly. Cloud repository purposes an on-demand information service model, and its reputation increasing because of its scaling down and less repair capital properties. Even, safety measure involvement arises when information repository is overcome to third-party cloud companies. This is essential to able cloud users to check their integrity of the important information on cloud, if the information has corrupted or attacked [9]. Cloud infrastructure is multi-holder, with various applications which are sharing physical framework. That gives aid of much capable resource using. Even there is no physical barriers between them, it is necessary to create and maintain balance safety measure controls to lesser the effect of malwares to distribute via cloud [14].Companies taking cloud services need to understand the involvement for maintaining the confidentiality of owners or other critical business information. The major attention is how the physical location of information affects its use. Ensure only specific users and devices can see sensitive information. One of the biggest concerns for companies coming to contact with cloud computing is confidentiality. In fully-managed public cloud service, confidentiality and aloofness risks are often likely to change accordingly to the provider's aloofness policy.

8. Proposed Scheme

The paper aims to:

1) Implement secure cloud system using CloudSim simulator and java Eclipse
2) Collection and preprocessing of data for mining
3) Encrypt the dataset into an image and securely migrate that image to cloud
4) Apply data mining in cloud and secure the generated mining report.

The proposed methodology is as:

- Request for Data mining report
- Select an image to hide the data mining request
- Encrypt the dataset into an image using edge detection method
- Find the edges of the selected image
- Use these edges as a Pixel keys pattern
- Randomized the pixels and generate the sequence of pattern positions to hide the request
- Convert the request into bits and replace them with the pattern positions
- Send this encrypted image to cloud storage
- Hidden dataset in an image is decoded at the cloud.
- Data mining is performed i.e. from the collections of files and data-sets the mining report is generated.
- For the security purpose, again the mining report is encoded into an image and that image is send to user end
- At user end, image is decoded, and the desired secure mining report is generated.

The proposed solution is to be implemented in CloudSim simulator and Java Eclipse.
The Eclipse Platform [15] is designed and built to meet the following requirements

- Support the construction of a variety of tools for application development.
- Support an unrestricted set of tool provider s, including independent software vendors (ISVs).

- Support tools to manipulate arbitrary content types (e.g., HTML, Java, C, JSP, EJB, XML, and GIF).
- Facilitate seamless integration of tools within and across different content types and tool providers.
- Support both GUI and non-GUI-based application development environments.
- Run on a wide range of operating systems, including Windows and Linux
- Capitalize on the popularity of the Java programming language for writing tools.

9. Results and Discussion

This section presents the simulation results of the proposed system implemented in CloudSim simulator and Java Eclipse.

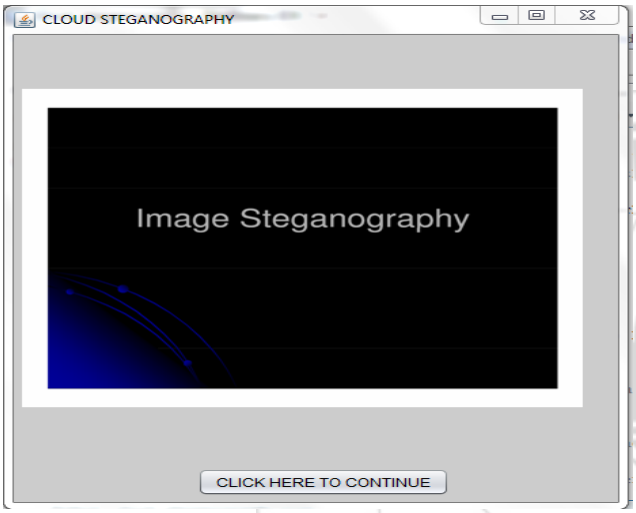

Figure 1: shows the home page of the Image Steganography system.

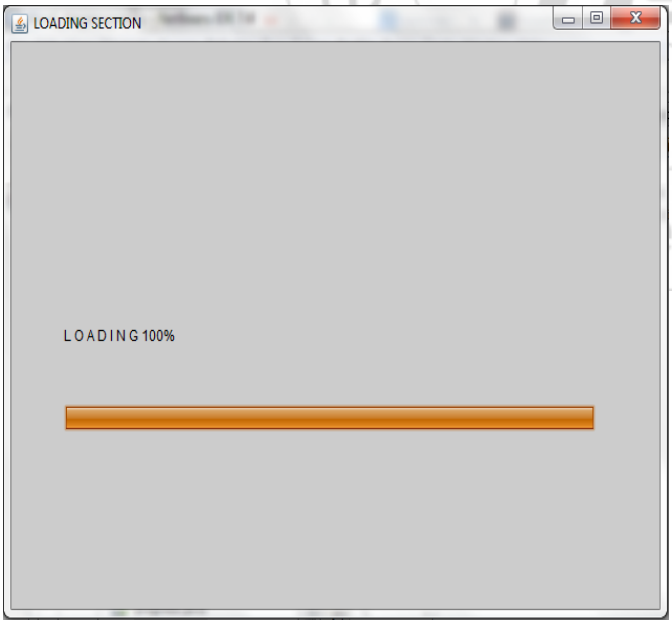

Figure 2: shows the loading process of the data mining report to be hidden and made secure. Here, the data set is being loaded and the apriori algorithm works here and the generated data mining report is loaded. This data mining report is made secured by embedding it into an image.

Figure 3: shows the hiding of the mining report inside an image using edge detection method. Here the encoding is done to keep the file secured from adversaries. The name of the output image file is given here which we actually going to store the critical data into itself.

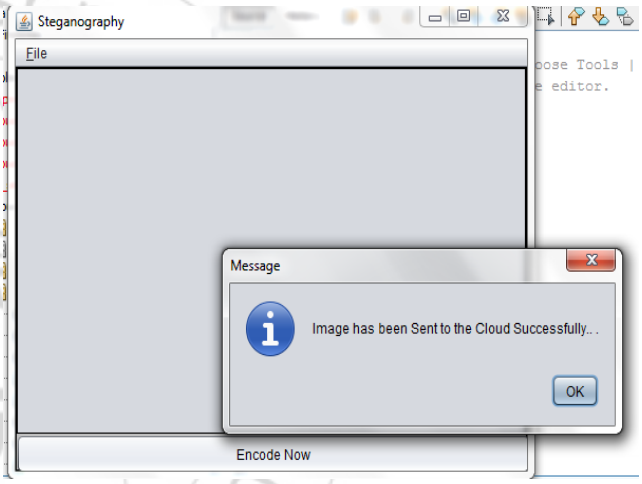

Figure 4: shows the successful encryption of the data mining report into an image and the image is further sent to cloud.

Figure 5: shows the encrypted image. This encrypted image is containing the data mining report inside it but it appears as a mere image to other users and they fail to judge the presence of crucial data inside it.

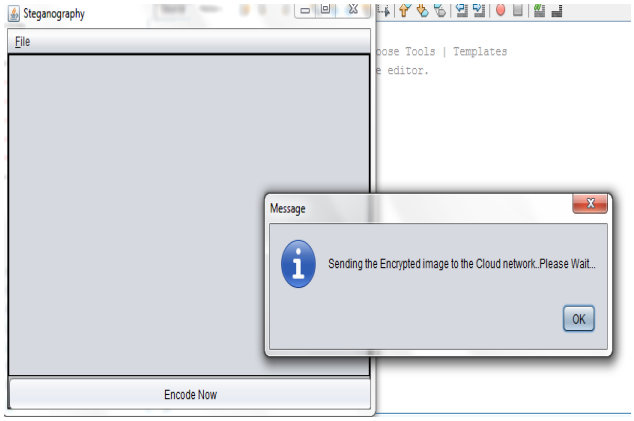

Figure 6: shows that the encrypted image has been now sent to the cloud network

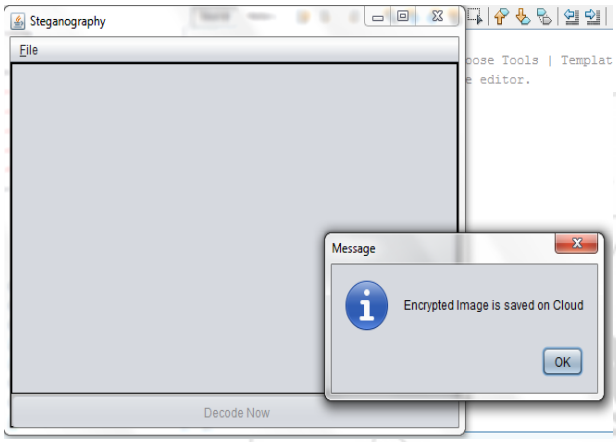

Figure 7: depicts that the moved encrypted image is saved on cloud.

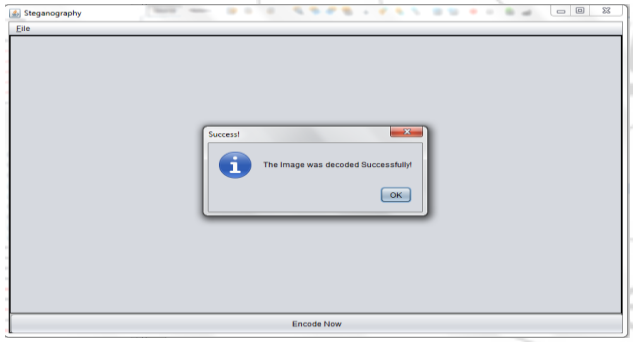

Figure 8: shows that the encrypted image is decoded successfully to retrieve the data report from it.

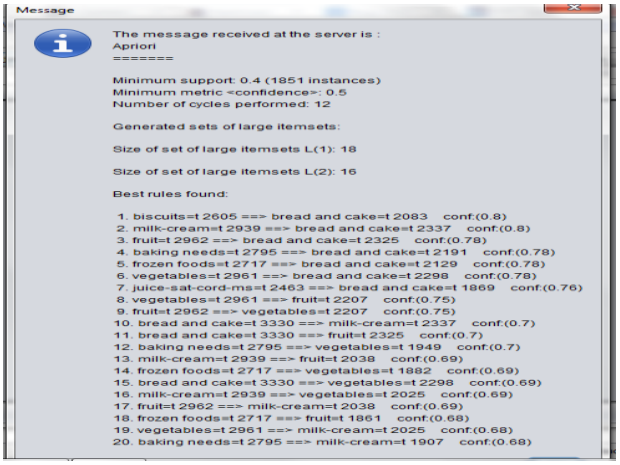

Figure 9: shows the decrypted data mining report

```
Apriori
=======

Minimum support: 0.25 (19 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 12

Size of set of large itemsets L(3): 1

Best rules found:

 1. outlook=overcast 20 ==> play=1 20     conf:(1)
 2. temperature=cool 20 ==> humidity=medium 20     conf:(1)
 3. humidity=medium windy=FALSE 20 ==> play=1 20     conf:(1)
 4. humidity=medium 35 ==> play=1 30     conf:(0.86)
 5. play=0 30 ==> humidity=high 25     conf:(0.83)
```

Figure 10: Apriori Algorithm

```
PredictiveApriori
=================


Best rules found:

 1. outlook=overcast 20 ==> play=1 20     acc:(0.9948)
 2. temperature=cool 20 ==> humidity=medium 20     acc:(0.9948)
 3. humidity=medium windy=FALSE 20 ==> play=1 20     acc:(0.9948)
 4. outlook=sunny humidity=high 15 ==> play=0 15     acc:(0.9944)
 5. outlook=sunny play=0 15 ==> humidity=high 15     acc:(0.9944)
 6. outlook=rainy humidity=high 15 ==> temperature=mild 15     acc:(0.9944)
 7. outlook=rainy play=1 15 ==> windy=FALSE 15     acc:(0.9944)
 8. windy=FALSE play=0 15 ==> humidity=high 15     acc:(0.9944)
 9. outlook=sunny temperature=hot 10 ==> humidity=high play=0 10     acc:(0.99221)
10. outlook=sunny humidity=medium 10 ==> play=1 10     acc:(0.99221)
11. outlook=sunny play=1 10 ==> humidity=medium 10     acc:(0.99221)
12. outlook=overcast temperature=hot 10 ==> windy=FALSE play=1 10     acc:(0.99221)
13. outlook=overcast windy=FALSE 10 ==> temperature=hot play=1 10     acc:(0.99221)
14. outlook=rainy windy=TRUE 10 ==> play=0 10     acc:(0.99221)
15. temperature=hot play=1 10 ==> outlook=overcast 10     acc:(0.99221)
16. temperature=hot play=0 10 ==> outlook=sunny 10     acc:(0.99221)
17. temperature=mild humidity=medium 10 ==> play=1 10     acc:(0.99221)
```

Figure 11: Predictive Apriori Algorithm

10. Conclusion

In an emerging discipline, like cloud computing, security needs to be analyzed more frequently. With advancement in cloud technologies and increasing number of cloud users, data security dimensions will continuously increase. Cloud computing security needs consider both technology and strategy, including: audit, compliance and risk assessment. Both the Service providers and the clients must work together to ensure safety and security of cloud and data on clouds. Mutual understanding between service providers and users is extremely necessary for providing better cloud security. In our paper we are laying stress on the security issue in the cloud. The paper presents the simulation results

and the comparison of the Apriori and Predictive Apriori algorithm.

11. Acknowledgment

The paper has been written with the kind assistance, guidance and active support of my department who have helped me in this work. I would like to thank all the individuals whose encouragement and support has made the completion of this work possible

References

[1] Uppunuthula Venkateshwarlu, Puppala Priyanka, "Survey on Secure Data mining in Cloud Computing", ISSN : 2347 - 8446 (Online) ISSN : 2347 - 9817 (Print), International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014), Vol. 2, Issue 2, Ver. 1 (April - June 2014)

[2] Juan Li1, Pallavi Roy1, Samee U. Khan1, Lizhe Wang2, Yan Bai3, "Data Mining Using Clouds: An Experimental Implementation of Apriori over MapReduce"

[3] Pramod Kumar Joshi1 and Sadhana Rana, "Era of Cloud Computing", High Performance Architecture and Grid Computing Communications in Computer and Information Science Volume 169, 2011, pp 1-8 ISSN 1865-0929, DOI 10.1007/978-3-642-22577-2_1, Springer-Verlag Berlin Heidelberg 2011

[4] Eman Elghoniemy, Othmane Bouhali, Hussein Alnuweiri, "Resource Allocation and Scheduling in loud Computing", 978-1-4673-0009-4/12, IEEE 2012

[5] Rabi Prasad Padhy, Manas Ranjan Patra, Suresh Chandra Satapathy, "Cloud Computing: Security Issues and Research Challenges", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS) Vol. 1, No. 2, December 2011

[6] Prof. G. Thippa Reddy, 2Prof. K. Sudheer , 3Prof. K Rajesh, 4Prof. K. Lakshmanna, "Employing Data Mining On Highly Secured Private Clouds For Implementing A Security-As a- Service Framework", Journal of Theoretical and Applied Information Technology 20th January 2014. Vol. 59 No.2 © 2005 – 2014, ISSN: 1992-8645

[7] Ramadhan Mstafa1, Christian Bach, "Information Hiding in Images Using Steganography Techniques", 2013 ASEE Northeast Section Conference Norwich University,Reviewed Paper, March 14-16, 2013

[8] Cong Wang, Qian Wang, and Kui Ren,"Ensuring Data Storage Security in Cloud Computing"

[9] C.P.Sumathi1, T.Santanam2 and G.Umamaheswari, "A Study of Various Steganographic Techniques Used for Information Hiding", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.4, No.6, December 2013

[10] Monjur Ahmed1 and Mohammad Ashraf Hossain, "Cloud computing and security issues in the cloud" , International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.1, January 2014

[11] Lingjuan Li Min Zhang, " The Strategy of Mining Association Rule Based on Cloud Computing", 2011 International Conference on Business Computing and Global Informatization

[12] Zhangn Chun-sheng Li yan, "Extension of Local Association Rules Mining Algorithm Based on Apriori Algorithm", 978-1-4799-3279-5 /14/$31.00 ©2014 IEEE

[13] B. Kamala, "A study on integrated approach of data mining and cloud mining", International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058 Volume- 1, Issue- 2, Nov-2013

[14] Zeba Qureshi1, Jaya Bansal2, Sanjay Bansal3, "A Survey on Association Rule Mining in Cloud Computing", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013)

[15] Jiawei Han, Hong Cheng, Dong Xin, "Xifeng Yan Frequent pattern mining: current status and future", , Data Min Knowl Disc (2007) 15:55–86, DOI 10.1007/s10618-006-0059-1

[16] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, " From Data Mining to Knowledge Discovery in Databases", American Association for Artificial Intelligence. All rights reserved. 0738-4602-1996

[17] Vahid Ashktorab2, Seyed Reza Taghizadeh , "Security Threats and Countermeasures in Cloud computing", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 1, Issue 2, October 2012 ISSN 2319 - 4847

[18] Garima Saini Naveen Sharma, "Triple Security of Data in Cloud Computing", International Journal of Scientific and Research Publications, Volume 4, Issue 6, June 2014 1,ISSN 2250-3153

[19] Jijo.S. Nair, BholaNath Roy, " Data Security in Cloud", International Journal of Computational Engineering Research (IJCER) ISSN: 2250-3005, National Conference on Architecture, Software system and Green computing

[20] 1T.V.Mahendra 2N.Deepika 3N.Keasava Rao, "Data Mining for High Performance Data Cloud using Association Rule Mining", Volume 2, Issue 1, January 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering .

[21] Sneha Arora1, Sanyam Anand2, "A New Approach for Image Steganography using Edge Detection Method", International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 3, May 2013, ISSN (Print) : 2320 – 9798 ISSN (Online): 2320 – 9801