

# A Novel Hybrid Approach for Secure Cloud Mining using Lossless Image Format

Sanjima Manocha

School of Computer Science and Engineering,  
Lovely Professional University,  
Jalandhar - Delhi G.T. Road (NH-1), Phagwara,  
Punjab (India) –144411

Sheveta Vashisht

School of Computer Science and Engineering,  
Lovely Professional University,  
Jalandhar - Delhi G.T. Road (NH-1), Phagwara,  
Punjab (India) –144411

## ABSTRACT

This research work explores the basic features of data mining techniques in cloud computing and securing the data using edge detection method. This research work tries to integrate data mining techniques into cloud computing and image processing making it a hybrid approach. 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. Centralize management of software and data storage, with assurance of efficient, reliable and secure services for their users through Edge detection based approach for image steganography.

## General Terms

Data Mining, Cloud Computing, Image Steganography, Canny Edge Detection.

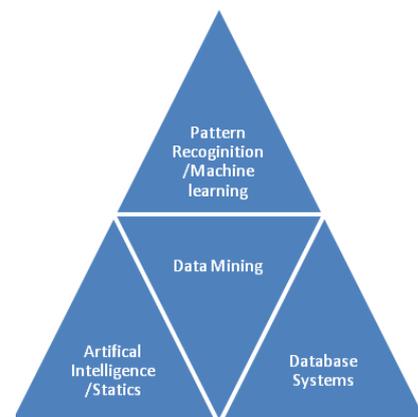
## 1. INTRODUCTION

In this era of internet, e-Commerce and social activities information and data are growing at a phenomenal rate. With the rapid growth of a variety of Internet services and applications, there are usually huge amounts of data; the need for quickly and efficiently manipulating the datasets in a scalable and reliable way is exceptionally high. Data mining applications and techniques are very much useful in the cloud computing model. Cloud computing denotes the new trend and practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. Data mining in cloud computing is the process of extracting structured information from unstructured or semi-structured web data sources. The data mining in Cloud Computing allows organizations to centralize the management of software and data storage, with assurance of efficient, reliable and secure services for their users.” The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage [1]. Secure data transformation on internet has been a dream since the emergence of internet. Steganography [2] is one of the solutions to securely transmit data by hiding data in data. Data used to hide data in Steganography can be text or image.

## 2. DATA MINING

With the onset of 21st century it has been an established fact that we are in an information technology driven society, where knowledge has proved to be an invaluable asset to any individual, organization or government. Data Mining provides that boost to enable organizations to turn their huge amounts of data into valuable information or knowledge that not only

could meet customer demands but also keep the future prospects in check. Data mining [1] is defined as a “type of database analysis that attempts to discover useful patterns or relationships in a group of data.



Data mining tracks these patterns and relationships using data analysis tools and techniques to build models. The business environment faced by all organizations has changed a lot with customers becoming more demanding in terms of their needs and in terms of products and services that they require. Organizations that wish to survive to these changes should be proactive and have to anticipate what their current and prospective customer desire. Association rule mining addresses this need. Association is the bonding of togetherness or connection of objects also known as association-rule. It finds rules associated with frequently co-occurring items like objects of one database with another database objects. Mostly used for market basket analysis, cross-sell, root cause analysis, product bundling, in-store placement, and defect analysis. Association rule mining is usually split up into two separate steps: First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules. There are a lot of different association rule algorithms. Apriori algorithm is the first, simplest and best-known for association rules mining. It is an influential algorithm for mining frequent itemsets for Boolean association rules. The core principles of this theory are the subsets of frequent item sets are frequent item sets and the supersets of infrequent item sets are infrequent item sets. It cut down candidate item sets using the principle that all non empty subsets of frequent item sets are frequent too.

### 3. NEED OF CLOUD COMPUTING AND STEGANOGRAPHY

In cloud computing, the term "cloud" is used as a metaphor for the Internet and cloud computing is a type of distributed computing paradigm where different services such as servers, storage and applications collectively known as configurable computing resources are rapidly equipped and released with minimal management efforts. The cloud computing model allows access to information and computer resources from anywhere provided a network connection is available. With cloud environment data can be replicated in other locations so it provides disaster recovery servers and data storage at low cost. Cloud is best solution to match uncertain Scaling demands. Small size companies can reduce their capital and operational expenditure for their increasing computing needs as no longer do IT professional need to worry about keeping software up to date. Organizations can store ample amount of data, even more than what they can store on private computer systems or servers.

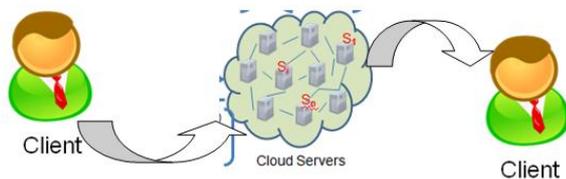


Fig 2: Overview of Cloud Computing

The scope of the research work is to extract the useful information from large amount of data and store at cloud in secure fashion and then make predictions or inferences required by the organization. But the predictions that are generated as a result of mining should be secure from 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 prime focus is to increase capacity to provide better security during transmission. In this research, Edge detection based approach for image steganography is discussed. Canny edge detection method produces single pixel thick, continuous edges even in noise conditions i.e ability to detect true weak edges. In this approach, Advantage of edge detection technique will be taken to increase capacity. Because editing in edge areas cannot be detected well by human eye, but editing in smooth areas can be detected easily.

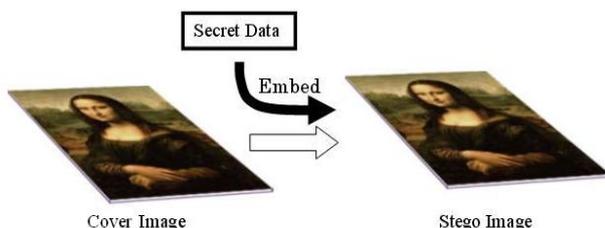
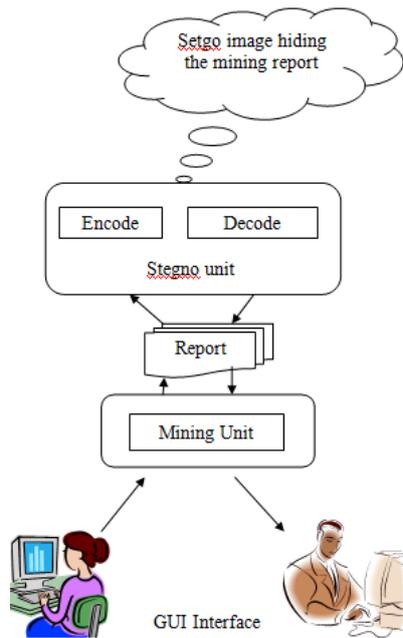


Fig 3: Basic Process of Image Steganography

### 4. PROPOSED SOLUTION

The proposed system consists of a User who needs the mined information for his business analysis. Data mining is performed on industry data continuously being collected and

stored using Apriori algorithm to mine frequent itemsets for discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes and present the same in the form of report. Association rule mining is one of the most well studied data mining tasks and important data mining's functionalities. It discovers relationships among attributes in databases, producing if-then statements concerning attribute values [8]. In association rules we find the co-occurrences among item sets through finding the large item sets. Association rule mining [9] is usually split up into two separate steps: First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules. There are a lot of different association rule algorithms. Apriori algorithm is the first and best-known for association rules mining. It is an influential algorithm for mining frequent itemsets for Boolean association rules. This algorithm uses a level-wise search, where k-itemsets are used to explore (k+1) itemset, to mine frequent itemsets from transactional database. It is an iterative algorithm to calculate the specific length of item collection of given database to produce frequent item sets. The core principles of this theory are the subsets of frequent item sets are frequent item sets and the supersets of infrequent item sets are infrequent item sets. It cut down candidate item sets using the principle that all non empty subsets of frequent item sets are frequent too. Apriori algorithm basically works in two steps. In first step candidate itemset is generated using linking process and in next step frequent itemset from those candidate itemset is found based on minimum support count by scanning the database. Apriori algorithm is easy to execute and very simple, is used to mine all frequent itemsets in database. Edge detection based approach for image steganography is used. The Canny edge detector is known to many as the optimal edge detector. Canny's intentions were to enhance the edge detectors already out at the time he started his work. The most obvious is low error rate. It's important that edges occurring in images should not be missed and that there should be no responses to non-edges. The 2nd criterion is that the edge points should be well localized or we can say that the distance between the edge pixels as found by the detector and the actual edge is to be at a minimum. A 3rd criterion is to have only one response to a single edge and completely eliminate the possibility of multiple responses to an edge. Based on these three criteria, the canny edge detector first smoothes the image to eliminate the noise. After that the edges should be marked where the gradients of image has large magnitudes and only local maxima should be marked as edges. Potential edges are determined by thresholding and then final edges are determined by suppressing all the edges which are not connected to certain edge. Edges will be detected first by using canny edge detection method and then data mining report is encoded into the cover image for data hiding resulting into Stego image. After hiding the data into image the Stego image is send to cloud so that both customer data and business information should be revealed to authorize parties easily for market based analysis. Other branches of the organization who participate in the decision making process will decode the Stego image available on the cloud using image steganography and extract the report inorder to fetch the meaningful information. Secure business information report is circulated to all branches of the organization.



#### 4.1 Algorithm for the proposed system

START

1. Select the Sample data
  2. Apply the Apriori algorithm and the desired report is formed
  3. Through GUI interface user interact with system
  4. Choose the cover image for data hiding
  5. Detect the Edges using Edge detection method.
  6. After edge detection, Encode the generated report using Least significant bit image Steganography
  7. Send the Stego image to cloud
  8. Decoding the Stego image on cloud using image Steganography
  9. Secure business information report is available
- END

#### 4.2 Flow of the system

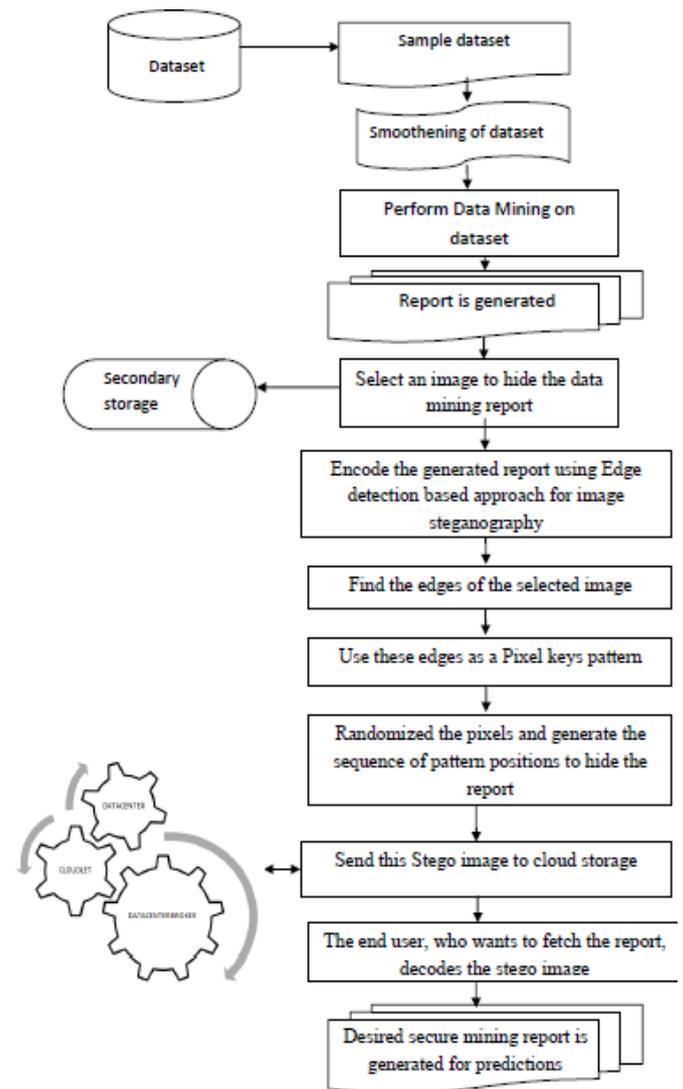


Fig 5: The overall process

#### 4.3 Implementation snapshots

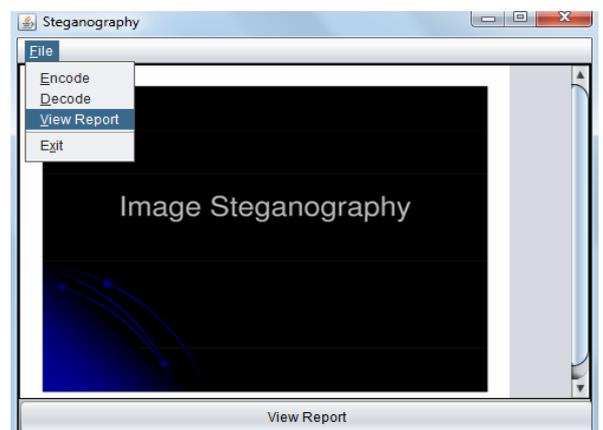


Fig 6: GUI interface of proposed system

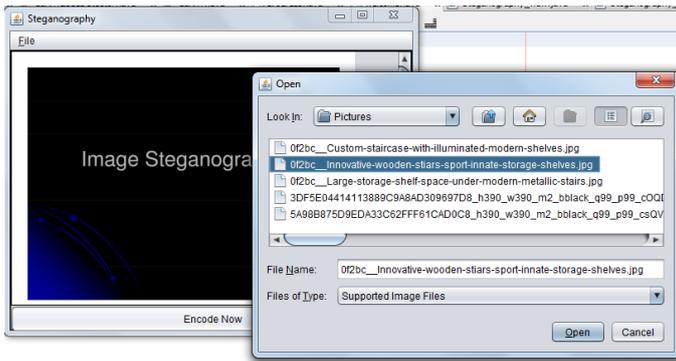


Fig 7: Open file dialog box work selecting cover image

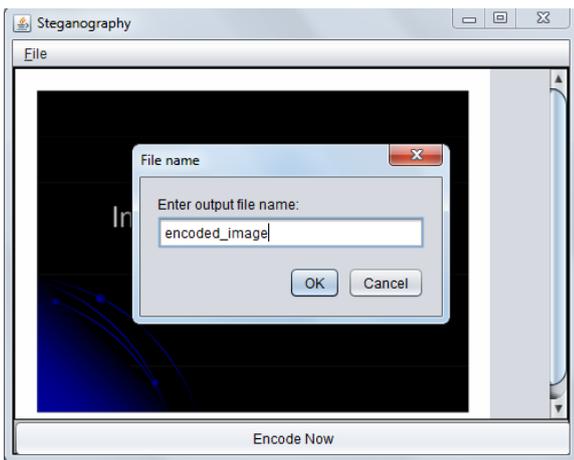


Fig 8:naming stego image

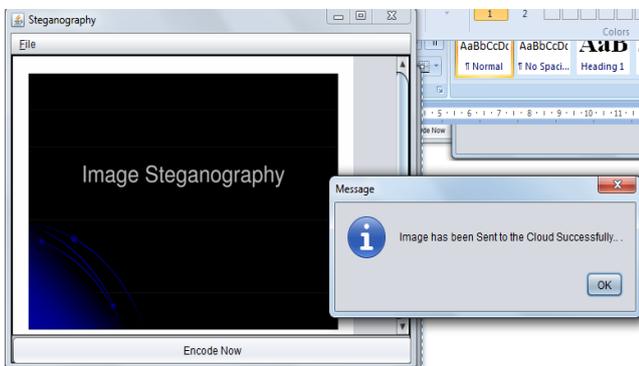


Fig 9:Stego image saved on the cloud

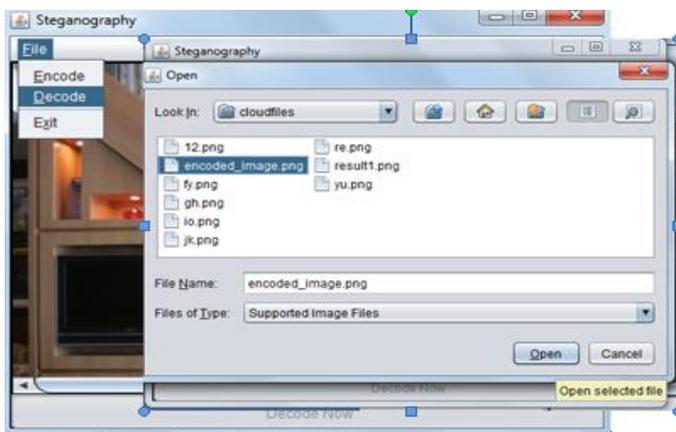


Fig:10 selecting the stego image from cloud folder

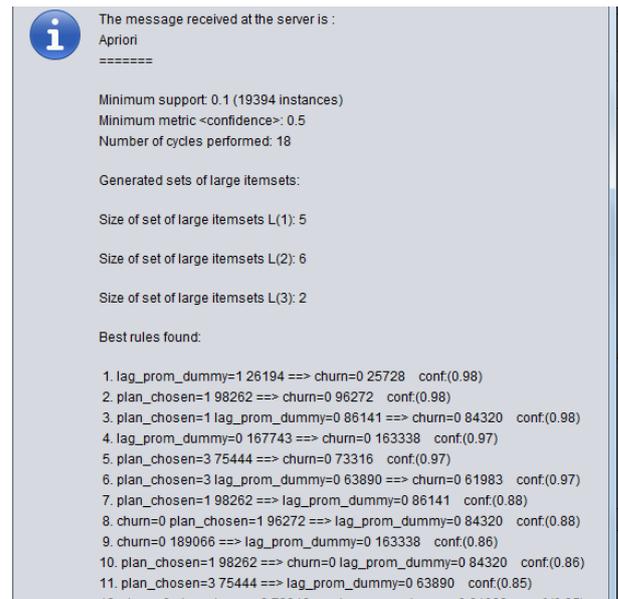


Fig:11 Report

## 5. CONCLUSION

Data mining technologies provided through Cloud computing is important in today's internet and business environment to make knowledge driven decisions, proactive as it helps them have future trends and behaviors predicted. The combination of facts such as the extensive growth of internet, the vast financial possibilities opening up in electronic trade, and the lack of truly secure systems make it an important field of research. So Steganography , the method for secret communication plays a important role. This paper provides an overview of the necessity and utility of data mining in cloud computing with Edge detection based approach for image steganography. Combining the effects of the three technologies provide us with a novel approach in field of research with the main objective of maintaining data integrity with no change in visual quality.

## Histogram

Encrypted Image

Original image

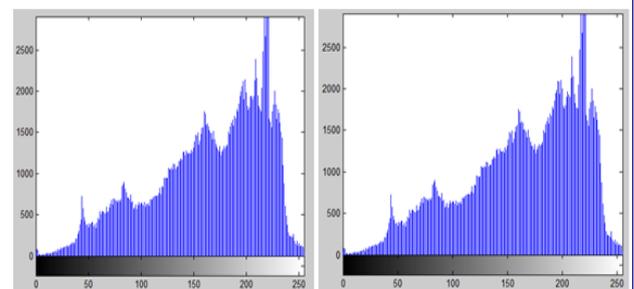


Fig:12 Histogram

## 6. FUTURE SCOPE

Data mining should play an important and increasing role in the business industry due to the large amounts of high quality data available, the competitive nature of the industry and the advances being made in data mining. An efficient and

effective management of the stored data and mining of data to get knowledge from stored data is required. As the knowledge to be extracted from the data is stored over the cloud storage; it is impossible technically and for organizational reasons to get all the data to a centralized storage. Hence there is a need for a data mining algorithm that works over non centralized data storage over distributed environment.

## **7. ACKNOWLEDGMENTS**

Our deep sense of gratitude to Assistant Prof. Sheveta Vashisht, Dept. of Computer Science & Engineering, Lovely Professional University, Jalandhar, India, for her generous guidance, help, continuous encouragement and supervision throughout the course of present work

## **8. REFERENCES**

- [1] Ruxandra-Ştefania PETRE (2012), “Data mining in Cloud Computing” Database Systems Journal vol. III, no. 3/2012
- [2] Sneha Arora and Sanyam Anand, ”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
- [3] Peter Mell, Timothy Grance “The NIST Definition of Cloud Computing”, NIST Special Publication 800-145
- [4] Ahlam Shakeel Ahmed Ansari, Kailas Kisan Devadkar (2012), “Secure Cloud Mining”, IEEE International Conference on Computational Intelligence and Computing Research 2012
- [5] University Vitez, Travnik, Bosnia and Herzegovina, “Data Mining and Cloud Computing”, Journal of Information Technology and Applications JITA 2(2012) 2:75-87
- [6] Michael J.A.Berry,Gordon S.Linoff (2009) “Data Mining Techniques”, Pearson Publications,page(s)347-359,386-393,401-434
- [7] Jain Nitin, Meshram Sachin and Dubey Shikha (2012) “Image Steganography Using LSB and Edge – Detection Technique”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July.
- [8] Agrawal, R., Imielinski, T. and Swami, A.N., Mining Association Rules between Sets of Items in Large Databases. In Proceedings of SIGMOD, 20716, 1993.
- [9] Jayshree Jha and Leena Ragha, Educational Data Mining using Improved Apriori Algorithm International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3, Number 5 (2013).