IMPLEMENTATION OF SECURE HYBRID DATA MINING APPROACH IN CLOUD

Manju Bala
Khalsa College of Engineering & Technology,
Amritsar,(Punjab),India

Pooja
CT College of institute Engineering,Management
&Technology,Jalandhar(Punjab),India

VijayKumar
M.Tech,PTURegionalCenter

Abstract-The continuous growth associated with cloud computing will be providing solution to alot more cloud services, on account of which in turn security associated with cloud services, especially Data privacy protection, gets more critical. This Research work explores principle top features of Data mining techniques in cloud computing and securing the data. The paper explores the current status of the cloud computing security, the potential data privacy issues and security loop holes that the cloud data faces. The main issue that has been given much emphasis is the Implementation of data mining techniques which have encouraged the unauthorized users to extract meaningful hidden predictive information from the confidential data present on cloud. This may result into a big loss to any enter pries or Organization. To address this issue, an image Steganography based data hiding approach is proposed which involves the use of Canny Edge Detection to extracted ges and further uses LSB Embed ding technique for data concealing.

Key words –Data Mining; LSB; SSIS; Steganography;

I. INTRODUCTION

The constant growth regarding information Technology in numerous grounds regarding human Life has triggered the broad volumes regarding Information storage in a variety of formats such As records, documents, images, sound recordings, videos, scientific data, and a lot many new data representations. The information as assembled from diverse applications requires legitimate knowledge /information extraction system to help play a role in superior decision making. Data Mining enables the 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 prospect sin check. Data mining in corporates numerous methods and algorithms to discover and extract patterns of stored data. Association rule mining is an interesting and significantly important research topic of data mining; its task is to find all subsets of items which are found to occur more frequently, along with determining the relationship between them. There are a lot of different association rule algorithms. The paper uses the Apriori algorithm to mine the interesting patterns. Data mining can be quite a prospective danger to cloud security considering the fact that entire data owned by a certain user is stored in a cloud. This can result Into a serious data privacy
problem if the cloud Providers misuses the information or any attacker extracts the valuable knowledge from the Business data. 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 paper involves the use of image steganography attained using LSB embedding approach.

I. RELATED WORK
Rabi Prasad Padhy [5] et. Al examined different models, security issues and research challenges in cloud computing. Information security is a real and critical issue for Cloud Computing. In addition to this, there are a few other security challenges including security parts of system and virtualization. This paper has highlighted all the conceivable issues of cloud computing and thus Emphasizes the need of new security techniques Since the conventional security strategies need to Be fund a mentally changed to be able to work with The cloud sarch it architecture. Rajeev Kumar et.al [18] gave a proposal to secure the business data on cloud. The paper presents the spread Spectrum Image steganography strategy to secure the data on cloud. Spread Spectrum Image Steganography (SSIS) strategy performs the data hiding by storing a message as Gaussian noise in an image. The Image degradation is undetectable by the human Eye at low noise power levels, while the noise Appears as snow or speckles at higher levels.
T.V. Mahendra et.al [20] proposed an algorithm to perform cloud data mining using sector/sphere Frame work with association rules. The paper explains the sectors to rage cloud as a distributed storage system encouraging the customers to utilize and download vast data set from any area with a decent network connection with the frame work. To achieve better reliability and accessibility, sector automatically replicates files.
The Sector system consist sofa master node Responsible for maintaining the file system, slave Nodes which store the data. Unlike it, in a sphere Data base system, user define functions perform The computation as they can be autonomously Applied to individual element in the data set and The final obtained outcome scan be written to Either the local disk or common destination files On other nodes. Sneha Arora et.al [21] proposed a strategy to conceal the information in to the colored image With the aid of edge detection technique. It is hard to distinguish the alterations in the edges so more Data can be concealed under the edges without any compromise in the quality of image. In this technique, edges of an image are detected by scanning using 3x3 windows and then text message is concealed in edges using first component alteration technique. The proposed scheme achieved high embedding capacity and high quality of encoded image.
Mamta Juneja et.al [22] proposes an enhanced LSB based Steganography procedure for images be stowing better data security. It exhibits and Embedding algorithm for hiding ciphered Messages in nonadjacent and irregular pixel areas in edges and smooth regions of images. The edges in the cover-image are detected using improved edge detection filter. The encrypted message bits are then embedded in the least significant byte of randomly selected edge pixels and some specific LSBs of red, green, blue components respectively. Such type of steganography technique ensures least chances of suspicion about message bits hidden in the image and it gets hard to estimate the true message length by standard steganography detection methods. The Proposed approach shows better results in PSNR value and Capacity As compared to other existing techniques.
Saiful Islam et.al [23] proposed a Steganography technique which is found to have Brilliant security against steganalysis is attacks and is an augmentation of edge embedding in color image. The selection of edges for embedding is decided by the factors such as the length of pay load and the image. With the expansion of the pay load size, a weak threshold is disused for the selection of edge so as to include more edges into the selection space to accommodate the increased amount of data. For embedding purpose, the proposed technique uses two-bit LSB substitution approach and consequently diminishes the quantity of pixels to be distorted. Although the alteration of two bits of the selected Pixels results into a significant change in pixel intensity but the change get sun detected due to Significant difference in intensity of edge and non edge pixels.

Youssef Bassilet.al [24] proposes an image steganography technique based on the Canny edge detection algorithm. The proposed technique starts with reading a24-bit uncompressed-type colored input image and then transforming all the pixels into black or white subsequently into a mono chrome version. The black pixels are represented by bit0 while the white pixels are represented by bit 1. Finally, the edges are extracted by applying the Canny edge detection algorithm on the transformed image. The extracted edges are recognized as white lines and curves composed of only white pixels with value1. The secret data bits are concealed in the three LSBs of every pixels that make up the extracted Edges in the original 24-bit colored input image.

Guowei Yanget.al [25] presents a new edge detection operator named as Log _Sobel which processes optical density to deal with the uneven illumination intensity effects. In addition to this, it inherits smooth function of Sobel operator and improves anti jamming of the detection. The operator processes image with logarithm of luminous density. A sit processes fast, it is more suitable for online detection. The experimental results demonstrate that image processed by Log _Sobel operator is better than the other Existing methods.

Simrat Pal Kaur et.al [26] emphasize son using Steganography along with 2k correction method & edge detection strategy. The proposed methodology is observed to be superior to any thing prior systems on account of its ability of conveying more pay loads with better imperceptibility. This can be accomplished by inserting more information in edge zones rather than with smooth regions of the image as human eye can't identify the twisting at edges effort lessly. The proposed algorithm yields better PSNR values When contrasted with past calculations.

Himel Dev et.al [27] presents away to deal with Data mining based attacks on cloud data. The key Thought of the proposed methodology is to split The user data into chunks and as sign the chunks To the best possible cloud suppliers. Hence, the Proposed approach is all about the categorization, Fragmentation and distribution of data. The paper Shed slight on the security risks that cloud data May face from data mining techniques and further Demonstrates the distributed architecture of cloud to mitigate the risks. The proposed system provides an effective approach which successfully protecting privacy from mining attacks but the drawback is the performance over head incurred in case the client demands a frequent data access.

Sonam Saluja et.al [28] presented a detailed Theory of various edge based image segmentation Methods which involves the use of Gradient-based And Laplacian based Edge Detection techniques. The paper gives an overview about Robert, Prewitt, Sobel, LoG
and Canny detection methods. It is understood that different edge detection methods can be opted according to the needs. But a robust and an adaptive edge-detection algorithm is necessary to deal with the varying noise levels. All the techniques have been analyzed with respect to different parameters such as the sensitivity to noise and computational complexity.

Poonam Dhankhar et al [29] presents a survey of different methodologies to achieve image segmentation incorporating the edge detection approaches. The investigation of various Edge recognition strategies and the experimental results proves Canny edge detection to be the best out of others. The paper presents are view on various edge detection techniques including the discontinuity intensity levels. The performance comparison of various techniques has been achieved using a single image in MATLAB. The results reveal that LoG and Canny edge detectors produce nearly same edge map. For a selected image, Canny results are observed to be superior due to the fact that different edge detections work better under different conditions.

Nitin Jain et al [30] presents an edge detection technique for data hiding which is different in a certain way from other existing edge detection techniques. The proposed technique marks a different color to identify the dark area of image. The selected dark areas hide the secret data but rather than putting the data directly in those pixels, it is put into the low bits of each eight bit pixel. The most crucial aspect is the selection of pixel to embed since the LSB insertion modifies the pixels. The modified pixels are found very noticeable in such areas of image where the pixels are some what similar to their neighbors. The LSB method is preferable for its simplicity to embed the bits of the message directly into the LSB plane of cover image.

Sanjima Manocha et al [31] describes the basic fundamentals of data mining techniques in cloud computing and data security using edge detection method. The paper presents a hybrid approach integrating data mining techniques into cloud computing and image processing. The paper explains the utility of data mining in cloud computing with Edge detection based approach for image steganography.

III. RELATIONSHIP BETWEEN DATA MINING AND CLOUD COMPUTING

Cloud computing is considered one of the today’s most appealing technological innovations due to its cost-efficiency and flexibility. It is a continuously evolving paradigm providing the users a reliable to scalable, customized to dynamic computing environment for which user has to pay only as per use. Cloud computing is a marketing term referred for the technologies providing various services like computation, software, data access, and storage facilities with an abstraction of the physical location and configuration of the source system. 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.

Despite being a potential cause of high storage and reasonable cost service, cloud computing has not lived up to its reputation due to several significant security issues like assurance and confidentiality of user data in cloud. The potential threat to the cloud data security may be data mining which aims to discover patterns or fabricate models using particular algorithms from various scientific disciplines including artificial intelligence, machine learning, data base systems and statistics. Since the attacker can extract sensitive information out of the highly
confidential data using data mining algorithms, hence data mining and cloud computing both are considered closely related to each other. This relationship offers its own set of benefits and drawbacks. The cloud providers may use the data mining algorithms to provide better results to the clients hence offering benefits. While the darker side associated with it is that the attackers who do not have any authorized access to the cloud data also get away to see that confidential information. There exist a number of data mining algorithms to Extract valuable information from the gigantic Volumes of data. The paper involves the use of A priori algorithm for the mining purpose. The most general, simple and widely used association rule mining algorithm is A priori algorithm which is designed with the motive to operate on transactional data bases. The algorithm functions in breadth-first-search fashion and a tree structure issued to count candidate item sets efficiently. It uses a layer search iterative method to generate (k+1)item sets from k item sets. A k item set is considered to be frequent only if all of its sub-item sets are frequent. This process continues to iterate until no more frequent k-item set scan be generated for some k. The basic rule parameter son which the A priori algorithm depends are–support, confidence and number of cycles used. In A priori Algorithm, the number of best rules generated is independent of the number of instances and attributes but are dependent on the value minimum support taken.

Fig.1KDD

IV.SECURINGCLOUDDATAFROM MININGATTACKS

Security can be a critical barrier towards the much wider use of cloud computing. Even though cloud research guarantees reduce expenses, fast scaling, easier maintenance, as well as support Availability at any place, whenever, a vital concern Is the best way to assure as well as build self-confidence that the user data is handled securely By the cloud. To make the cloud computing be adopted by users and enterprise, the security concerns of users should be rectified first to make cloud environment trust worthy. The three aspects of information security which needs to applied to data stored at any place and at cloud storage are: confidentiality, integrity and availability. 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. Steganography is the Procedure of concealing any data
into different Sources of information such as text, picture or audio file, with the goal that it is not noticeable to the natural perspective. The security module which issued 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.

The image steganography is attained using Least Significant Bit approach which hides the Secret data into a digital image. The data is concealed into the pixels that define the edges/ Boundaries of the objects

![Image Steganography Process](image)

Fig.2 Basic image steganography process

It is not preferable to conceal the secret data into all pixels because it is very common fashion and is very easy to recover the secret data out of the carrier file. Hence, the secret data should be concealed using certain specified pattern which cannot be interpreted by anyone else other than these under and the intended receiver.

The edge pixels are detected using Canny Edge Detection Algorithm. The algorithm consists of three parameters- the size of the Gaussian Filter, The maximum and the minimum threshold values. The parameterization of the algorithm supports the diverse availability of different career pixels for the same input image and same data.

Canny Edge Detection algorithm works in five stages:
1) It starts with Smoothing in which image Is blurred to eliminate noise. This step uses the Gaussian filter.
2) Further, the edge strengths in the Image are found by taking gradient with large magnitudes.
3) The local maxima are marked as edges And then on-maximum are suppressed
4) Thresholds are computed to ensure the possible edges
5) Final edges are found by discarding all Edges that are not connected to strong edges.

V. PROPOSED METHODOLOGY

The paper involves the use of LSB approach to embed the secret data into the detected edge Pixels which are indicated by the canny edge detection algorithm. This method modifies the lowest order bit of each pixel to embed the secret message.

LSB steganography can be described as if the LSB of the pixel value \( I(i, j) \) is equal to the Message bit \( m \) to be embedded, \( I(I, j) \) remain unchanged; if not, set the LSB of \( I(i,j) \) to \( m \).

The message embedding procedure can be formulated as:

\[
\text{Is} (i, j) = \begin{cases} 
\text{I} (i, j)-1 \text{ LSB} (\text{I} (i, j)) = 1 \text{ and } m = 0 \\
\text{I} (i, j) \text{ LSB} (\text{I} (i, j)) = m \\
\text{I} (i, j) + 1 \text{ LSB} (\text{I} (i, j)) \neq 0 \text{ and } m = 1 
\end{cases}
\]

The proposed methodology is as:
1) The procedure starts with the collection of data followed by the pre-processing of Data for mining with generating the Mining report.
2) Next step is to choose an image and Apply Canny edge detection method to find its edges which are used as pixel key pattern and the edges hide the data mining report in it.
3) Randomize the pixels and generate the Sequence of pattern positions to hide the report.
4) Convert the report into bits and replace them with the pattern positions.
5) Send this encrypted image to cloud storage.
6) At user end, the image is decoded, and the desired secure mining report is generated. 

VI. EXPERIMENTAL OBSERVATION

The proposed methodology is simulated in CloudSim simulator and Java Eclipse.

Fig.3 Comparison of A priori and Predictive A priori Algorithm

The relevant patterns are mined from the extensive business data using A priori and Predictive A priori algorithm. The graph shows that The A priori algorithm shows better performance in Terms of different parameters.

Fig.4 Encoding/Decoding Time Analysis

The graph shows the message Encoding/Decoding time comparison and clearly reveals that the decoding takes more time. It reveals that the embedded secret message is not easy to recover from the carrier image.

Fig.5 Histogram of original and encrypted image

The histogram of original image and the Encrypted image are found to be same. Hence, it becomes nearly impossible for someone to suspect the presence of any critical data in the image. Hence, the proposed algorithm is found capable of achieving the real objective in terms of data hiding.

REFERENCES


loud Computing”, DOI: 978-1-4673-0009-4/12, IEEE 2012

[15] Jiawei Han, Hong Cheng, Dong Xin, “Xifeng Yan Frequent Pattern mining: current status and future”, Data Min Knowl Disc 15:55–86, DOI: 10.1007/s10618-006-0059-1, 2007
[16] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, “From Data Mining to Knowledge Discovery in Databases”, American Association for Artificial Intelligence, 0738-4602-1996


[27] Himel Dev, Tanmoy Sen, Madhusudan Basak and Mohammed Eunus Ali, “An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks”,


