International Journal of Operations Research and Information Systems
Volume 12 • Issue 3 • July-September 2021

An Enhanced Security Measure for Multimedia Images Using Hadoop Cluster

Prakash Mohan, Karpagam College of Engineering, India
https://orcid.org/0000-0002-9476-3142

Balasaravanan Kuppuraj, M.P.Nachimuthu M.Jaganathan Engineering College, India

Saravanakumar Chellai, St. Joseph's Institute of Technology, India

ABSTRACT

Information is generated over the internet every second. This information is not fully secured. To increase the security of this information, there are two methods. Cryptography and steganography are combined to encrypt the data using RSA algorithm as well as to hide the data in multimedia image in Hadoop cluster. Features of the resultant image such as color are extracted and stored separately in Hadoop cluster to enhance security. The features of the stenographic image for secret image retrieval, which has been split into image and secret information, are combined. At last, decrypting the secret information, the authors retrieve the actual information. Application of this system in Hadoop will increase the speed of execution of the process.

KEYWORDS

Big Data, Clustering, HDFS, MapReduce, Multimedia, Security

1. INTRODUCTION

Now a days, huge amount of information are generated through mobile phones, satellite, digital cameras and other sources. These massive information are not fully secured. With the growth of the technology, several algorithms are developed in order to provide security. Cryptography and Steganography (Choras, R. S. 2007). methods are used to encrypt the data as well as to hide the data in multimedia image. Cryptography is used to encrypt information that is converting information into cipher text. The human and the computer are unable to read this encrypted (Patil, S. M. 2012). information without the proper cipher to decrypt it. Strong ciphers should be difficult to break, that is inability on the part of the code breaker to break it in less time or by using limited computer resources. Steganography is a technique that is used to hide the information within another information. The information can be hidden in multimedia image also using LSB algorithm (Prakash, M., Sayeed, R. F., Princey, S., & Priyanka, S. 2015) in which we first do the scanning process in which it scan the image row by row and encode it in binary. Then encode the secret message in binary and check the size of the image and the size of the secret message. For each of the pixels of the image, choose one pixel of the image randomly and divide the image into four parts (Alpha, Red, Green and Blue parts) and hide two by two bits of the secret message in each part of the pixel in the two least significant bits. The encrypted data is embedded in the image so that it cannot be seen. It establishes a communication

DOI: 10.4018/IJORIS.20210701.oa4

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

that is not directly visible by hiding information in an image (Annamalai, R., Srikanth, J. 2015). It provides a situation that an unauthorized person is unaware about the presence of hidden information and so they will be unable to access that original information.

The features can be extracted from that image on a distributed computing platform by using Apache Hadoop framework. With the rapid growth of technology and social media, lots of information in the form of text documents and multimedia is generating. To process such a huge amount of information is a big data (Sagiroglu, S., & Sinanc, D. 2013) problem in the present domain. As a solution to this big data problem Hadoop came up that almost solved this problem. By this approach memory is increased and execution time will be faster due the parallel execution of Hadoop cluster. Hadoop takes care of single point of failure, the system is highly error sophisticated and less susceptible to node failures. Hadoop is an open-source implementation of the Map Reduce platform and distributed file system (Katal, A., Wazid, M., & Goudar, R. H. 2013). The Hadoop system is written in Java. Hadoop Streaming is a utility that comes with the Hadoop distribution and that can be used to invoke streaming programs that are not written in Java (such as Ruby, Perl, Python, PHP, R, or C++). Using this utility, we can execute database programs written in the Ruby programming language. The utility also allows the user to create and run Map and Reduce jobs with any executable programs or scripts as the mapper and the reducer. An execution of a program when using Hadoop Streaming (Yamamoto, M., & Kaneko, K. 2012) and a description of the Map and Reduce functions in the Ruby programming language. Key-value pairs can be specified to depend on the input–output formats.

The Hadoop framework is one of the common frameworks that processes distributed big data by applying the MapReduce method. The Hadoop framework uses a distributed, scalable, and portable Hadoop Distributed File System (HDFS) (Liao, H., Han, J., & Fang, J. 2010), which is written in the Java programming language in order to perform these jobs (Patel, A. B., Birla, M., & Nair, U. 2012). HDFS (Zikopoulos, P., & Eaton, C. 2011) allows big data to form blocks, distributed to nodes, processed and retrieved successfully (Wu, Y., Ye, F., Chen, K., & Zheng, W. 2014).

Hadoop (Rao, B. T., 2012) consists of two components namely Map Reduce and the Hadoop Distributed File System (HDFS), performing distributed processing by master and slave servers. Map-Reduce is the programming model that works on the large datasets with parallel and distributing algorithm on cluster. JobTracker schedules jobs to task tracker and monitors different components. Tasks are accepted by task tracker and then it executes these tasks to return the result. Name node manages and executes file system name space operations. The mapping of blocks to DataNodes is also determined by the name node. DataNode manages the storage and performs block operations in response to the direction from NameNode. Secondary NameNode is a helper to the NameNode which is responsible for supporting periodic checkpoints of the HDFS (Kaushik, R. T., Bhandarkar, M., & Nahrstedt, K. 2010) metadata. The combination of cryptography and steganography efficiently hides the encrypted information within an image that is supported by feature extraction. The combination of cryptography and steganography efficiently hides the encrypted information within an image (Prakash, M., & Ravichandran, T. 2012). The human and the computer are unable to read this encrypted information without the proper cipher to decrypt it. Strong ciphers should be difficult to break, that is inability on the part of the code breaker to break it in less time or by using limited computer resources. The additional approach of feature extraction of the image added high level security to the information. By this approach memory is increased and execution time will be faster due the parallel execution of Hadoop cluster.

HDFS is designed to consistently to store huge volume of files across device in large cluster. It is inspired by the Google File System. HDFS is composed of NameNode and DataNode (Yamamoto, M., & Kaneko, K. 2012). HDFS stores each file as a sequence of blocks (currently 64 MB by default) with all blocks in a file the same size except for the last block. Blocks belonging to a file are replicated for fault tolerance. The block size and replication factor are configurable per file. Files in HDFS are write-once and can have only one writer at any given time.

The remainder of this paper is organized as follows. In section 2, we have discussed about the various survey of Hadoop HDFS and MapReduce. The proposed system of multimedia cluster is discussed in section 3. In section 4, a framework model of multimedia cluster with architecture diagram is discussed. In section 5, we describe the implementation module and section 6 describes about the conclusion.

2. LITERATURE REVIEW

Visual Cryptography Schemes (VCS) is a technique of image encryption to hide the secret information in images. In this established VCS, the secret image is encrypted into n number of shares arbitrarily and extend to the n number of participants. The secret image can be recovered basically by stacking the shares lacking any complex calculation concerned. With the use a novel encoding scheme, the technique has a unique flexibility that enables a single encryption of a color image but enables three types of decryptions on the same cipher text (Zhang, X., Zhou, Z., Jiao, Y., Niu, Y., & Wang, Y. 2018). The three different types of decryptions enable the recovery of the image of varying qualities. The physical transparency stacking type of decryption enables the recovery of the traditional visual cryptography quality image. An enhanced stacking technique enables the decryption into a halftone quality image. Finally, a computation-based decryption scheme makes the perfect recovery of the original image possible. However this approach suffers a safety, pixel extension and noise trouble. The major drawback of this scheme is that it suffers from high transmission risk as the shares are like noise which causes the attackers attention and the shares can be intercepted.

In Steganography, some information is hidden within other information. Nivedhitha, R., Meyyappan, D. T., & Phil, M. (2012) in this paper it is clearly shown that how the information are hidden within the digital images. This paper introduces two new methods where in the combination of cryptography and steganography are used to encrypt the data as well as to hide that data within the image.

In Cryptography, the secret information are encrypted and decrypted in order to provide security. Bharti, P., & Soni, R. (2012) paper the science of using mathematics to encrypt and decrypt data has been shown. This paper shows how the information is converted into cipher data which is not understandable to the people. This method is combined with the Steganography to hide the information within an image. The quality of the image that we get is better than that of other methods.

The hadoop is a new technology that acts as a solution to the big data. Yamamoto, M., & Kaneko, K. (2012). in Parallel Image Database Processing with MapRedue and Performance Evaluation in Pseudo Distributed Mode says how to perform parallel distributed processing of a video database by using the computational resource. At present, the Apache Hadoop (Zhang, J., Wu, G., Hu, X., & Wu, X. 2012) redistribution for open-source cloud computing is available from MapReduce.

According to Shirale (2015), the abundance of online photo storage and social media from websites such as Facebook, WhatsApp, Instagram, Twitter and Picasa, and the volume of image data available is larger than ever before and growing more rapidly every day. This alone provides an incredible database of image and video that can scale up to billions of data.

One of the most widely used methods for processing, analyzing, storing, and retrieving big data is the MapReduce method by Slagter, K., Hsu, C. H., Chung, Y. C., & Zhang, D. (2013). The MapReduce method, developed by Google in 2004, has become an accepted standard in the field of big data analysis discussed in Dean, J., & Ghemawat, S. (2008).

3. PROPOSED SYSTEM

We use two methods where in Cryptography and Steganography are combined to encrypt the data using RSA algorithm as well as to hide the data in multimedia image in Hadoop Cluster. The use

of RSA algorithm for encryption adds advantage to maximum level. RSA is an algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm.

Asymmetric means that there are two different keys. RSA involves a public key and private key. The public key can be known to everyone that can be used to encrypt messages. Messages encrypted using the public key can only be decrypted with the private key. LSB algorithm is used to hide information within an image. It embeds the secret message in an image based on its binary coding.

Steganography is not the same as cryptography data hiding techniques have been widely used to broadcast of hiding secret message for long time. Assuring data security is a big dispute for computer users. Business men, professionals, and home users all have some important data that they want to secure from others. Even though both methods provide protection, to add multiple layers of security it is always a good practice to use Cryptography and Steganography together. By combining, the data encryption can be done by a software and then embed the cipher text in an image or any other media with the help of steganographic key. The combination of these two methods will enhance the security of the data embedded.

Using the LSB algorithm we first do the scanning process in which it scan the image row by row and encode it in binary. Then encode the secret message in binary and check the size of the image and the size of the secret message. For each of the pixels of the image, choose one pixel of the image randomly and divide the image into four parts (Alpha, Red, Green and Blue parts) and hide two by two bits of the secret message in each part of the pixel in the two least significant bits. After that set the image with the new values and finally set the image with the new values and save it. Features of the resultant image such as color are extracted and stored separately in Hadoop cluster. Then combining features of the steganographic image for secret image retrieval, by this image has been split into image and secret information to enhance security. At last, decrypting the secret information to retrieve the actual information.

4. SYSTEM DESIGN

Hadoop is an open source usage of the MapReduce parallel preparing structure. Hadoop conceals the points of interest of parallel handling, including appropriating information to preparing hubs, restarting subtasks after a disappointment, and gathering the aftereffects of the calculation. This structure enables designers to compose moderately straightforward projects that emphasis on their calculation issue, as opposed to on the stray pieces of parallelization.

In order to provide security to the information we first use RSA algorithm to encrypt the information and convert it into cipher data. The use of RSA algorithm for this purpose adds advantage to maximum level. RSA is an algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two different keys. RSA involves a public key and private key. This cipher data is not understandable to the people. Then we use Steganography technique using LSB algorithm to hide this cipher data into an image to generate a secret image. LSB algorithm is used to hide information within an image. It embeds the secret message in an image based on its binary coding. Using the LSB algorithm we first do the scanning process in which it scan the image row by row and encode it in binary. Then encode the secret message in binary and check the size of the image and the size of the secret message. For each of the pixels of the image, choose one pixel of the image randomly and divide the image into four parts (Alpha, Red, Green and Blue parts) and hide two by two bits of the secret message in each part of the pixel in the two least significant bits. After that set the image with the new values and finally set the image with the new values and save it. The secret information is embedded within an image is more secured than any other technique. The quality of the image is not lost even after embedding information within that. Then color features like red, green and blue are extracted and stored separately in order to enhance security. As the part of retrieval of the secret information, the secret image is first

retrieved from the color features stored in the database. Then we split the image and the cipher data from which we can retrieve the secret information after decrypting the cipher data as shown in Fig 1.

The above figure shows the architecture diagram. Initially the secret information is encrypted using RSA algorithm which generates a cipher data that is not understandable to the people. The use of RSA algorithm for this purpose adds advantage to maximum level. RSA is an algorithm used by modern computers to encrypt and decrypt messages. Strong ciphers should be difficult to break, that is inability on the part of the code breaker to break it in less time or by using limited computer resources. A strong cipher should take very long time to break it. This encrypted information is then embedded within an image using the method known as steganography based on LSB algorithm. It will generate a secret image which contains the secret information. Using the LSB algorithm we first do the scanning process in which it scan the image row by row and encode it in binary. Then encode the secret message in binary and check the size of the image and the size of the secret message. For each of the pixels of the image, choose one pixel of the image randomly and divide the image into four parts (Alpha, Red, Green and Blue parts) and hide two by two bits of the secret message in each part of the pixel in the two least significant bits. After that set the image with the new values and finally set the image with the new values and save it. The color features like Alpha, Red, Green and Blue are extracted from the image and stored separately in HDFS. Then combining features of the Steganographic image for secret image retrieval, which has been then split into image and encrypted information. At last, decrypting the encrypted information, we retrieve the actual information. Application of this system in Hadoop will increase the speed of execution of the process.

5. RESULTS AND DISCUSSION

The multimedia images are used as an input and Java code is used for the HDFS. The secret information to embed in the multimedia image is given via the console window of HDFS clustering as shown in Fig 2. The sample image as the input is shown in Fig 3. The given secret information is embedded into the original image, there wont be any difference in the images of alpha, beta and gamma. The various colors of RGB images are represented in Fig 4 to Fig 6. The hidden secret information cannot be extracted in the human visualization and the images are provided with the security, so its difficulty to extract the information from brute force method. In the social media various images are shared, while sharing the information in this Hadoop clustering the secret information is also embed into it.

6. CONCLUSION

This paper has described a method to enhance security for the information that is processed within the Hadoop cluster. The combination of cryptography and steganography efficiently hides the encrypted information within an image. The additional approach of feature extraction of the image added high level security to the information. In future, the enhancement of the image security is also provided to the video.

REFERENCES

Annamalai, R., Srikanth, J., & Prakash, M. (2015). Accessing the Data Efficiently using Prediction of Dynamic Data Algorithm. *International Journal of Computers and Applications*, *116*(22).

Bharti, P., & Soni, R. (2012). A new approach of data hiding in images using cryptography and steganography. *International Journal of Computers and Applications*, *58*(18).

Choras, R. S. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *International Journal of Biology and Biomedical Engineering, 1*(1), 6-16.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107–113. doi:10.1145/1327452.1327492

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: issues, challenges, tools and good practices. In *Contemporary Computing (IC3), 2013 Sixth International Conference on* (pp. 404-409). IEEE. doi:10.1109/IC3.2013.6612229

Kaushik, R. T., Bhandarkar, M., & Nahrstedt, K. (2010). Evaluation and analysis of greenhdfs: A self-adaptive, energy-conserving variant of the hadoop distributed file system. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on* (pp. 274-287). IEEE. doi:10.1109/CloudCom.2010.109

Liao, H., Han, J., & Fang, J. (2010). Multi-dimensional index on hadoop distributed file system. In *Networking, architecture and storage (NAS), 2010 IEEE fifth international conference on* (pp. 240-249). IEEE. doi:10.1109/NAS.2010.44

Nivedhitha, R., Meyyappan, D. T., & Phil, M. (2012). Image security using steganography and cryptographic techniques. *International Journal of Engineering Trends and Technology*, *3*(3), 366–371.

Patel, A. B., Birla, M., & Nair, U. (2012, December). Addressing big data problem using Hadoop and Map Reduce. In *Engineering (NUiCONE), 2012 Nirma University International Conference on* (pp. 1-5). IEEE. doi:10.1109/NUICONE.2012.6493198

Patil, S. M. (2012). Content based image retrieval using color, texture & shape. *International Journal of Computer Science and Engineering Technology*, *3*(9).

Prakash, M., & Ravichandran, T. (2012). An Efficient Resource Selection and Binding Model for Job Scheduling in Grid. *European Journal of Scientific Research*, *81*(4), 450–458.

Prakash, M., Sayeed, R. F., Princey, S., & Priyanka, S. (2015). Deployment of MultiCloud Environment with Avoidance of DDOS Attack and Secured Data Privacy. *International Journal of Applied Engineering Research: IJAER*, *10*(9), 8121–8124.

Rao, B. T., Sridevi, N. V., Reddy, V. K., & Reddy, L. S. S. (2012). *Performance issues of heterogeneous hadoop clusters in cloud computing*. arXiv preprint arXiv:1207.0894.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE. doi:10.1109/CTS.2013.6567202

Shirale, S., Patmas, M., Gunjal, P., & Rane, D. (2015). A online multimedia data processing on cloud and hadoop platform. *Int J Comput Technol Electron Eng*, *5*(2).

Slagter, K., Hsu, C. H., Chung, Y. C., & Zhang, D. (2013). An improved partitioning mechanism for optimizing massive data analysis using MapReduce. *The Journal of Supercomputing*, *66*(1), 539–555. doi:10.1007/s11227-013-0924-9

Wu, Y., Ye, F., Chen, K., & Zheng, W. (2014). Modeling of distributed file systems for practical performance analysis. *IEEE Transactions on Parallel and Distributed Systems*, *25*(1), 156–166. doi:10.1109/TPDS.2013.19

Yamamoto, M., & Kaneko, K. (2012). Parallel image database processing with mapreduce and performance evaluation in pseudo distributed mode. *International Journal of Electronic Commerce Studies*, *3*(2), 211–228. doi:10.7903/ijecs.1092

Zhang, J., Wu, G., Hu, X., & Wu, X. (2012). A distributed cache for hadoop distributed file system in real-time cloud services. In *Grid Computing (GRID), 2012 ACM/IEEE 13th International Conference on* (pp. 12-21). IEEE. doi:10.1109/Grid.2012.17

Zhang, X., Zhou, Z., Jiao, Y., Niu, Y., & Wang, Y. (2018). A Visual Cryptography Scheme-Based DNA Microarrays. *International Journal of Performability Engineering*, *14*(2), 334.

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

M. Prakash received B.E degree from the University of Madras, Chennai, India and M.E. degree from Sathyabama University, Chennai, India in 2001 and 2007 respectively, and Doctorate of Philosophy in Faculty of Computer Science and Engineering from Jawaharlal Nehru Technological University Hyderabad India in 2014. He is currently working as Professor in the Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, India. His research interest includes Big Data, Cloud Computing, Networks, Machine Learning and Fuzzy systems. He has published more than 30 papers in the refereed journal and life member in CSI and ISTE.

K. Balasaravanan received M.E. degree from Anna University, Chennai, Tamilnadu, India in 2008 and Ph.D. Degree from Anna University, Chennai, Tamilnadu, India in 2014. He is currently working as Associate Professor in the Department of Computer Science and Engineering, M.P.Nachimuthu M.Jaganathan Engineering College, Erode, Tamilnadu, India. His research includes in Data mining, Spatial Database, Opinion Mining, Theory of Computation and Compilers. He has published more than ten papers in the referred journals and he is the life member in CSI, ISTE.

Saravanakumar Chelliah received B.E degree in Information Technology from Bharathidhasan University, Trichy, India, in 2003, and the M.E degree in Computer Science and Engineering from, Vinayaga Mission's University, Salem, India, in 2007. Currently he is working as an Associate Professor, Department of Information Technology, St. Joseph's Institute of Technology, Chennai, Tamilnadu, India. He is a life member of Computer Society of India (CSI) and published many national and international research articles. His research interests are Cloud Computing, Software reliability and Distributed Systems.