Privacy-Enhanced Image De-Duplication Technique for Internet of Vehicles

Dr.CHANDRA MOHAN

MTech student, Department of Computer Science and Engineering

Bangalore Institute of Technology, Bangalore, Karnataka, India

Email-bhoomikasuresh33@gmail.com

Dr.ECCLESTON

Assistant Professor, Department of Computer Science and Engineering

Bangalore Institute of Technology, Bangalore, Karnataka, India

Email-nmanirr@gmail.com

Abstract:
Internet of Vehicle is playing very important role in the present growing technology. This technology is changing from physical world to cyber world, where the users are concerned more regarding their privacy leakage, delay problem and security for data collection of Cyber Intelligence Technology (CIT) [1]. the problem is that if the user collected data or images is uploaded directly to the cloud this may bring the huge load balance on the cloud storage along with this there will be a presser on network which will lead to the malicious attacker to attack and steal the user data. To end this, the proposed schema consists of Map Reduce technique, Hash code generation technique and redundant data elimination technique using block generation using MD5 (Message Digest version 5) algorithm.

Keyword: Privacy Leakage, Delay, Cloud storage, Map Reduce technique, Hash code generation, Redundant data and Block generation.

I. Introduction

As a part of integrated application of network technology, smart city is playing very important role which is directing in development of growing generation of network technology. Internet of Vehicle is a very important scenario and application for smart city. Internet of Vehicles performs as a wireless communication which helps in exchanging information only with the specific and particular communication protocol along with data interaction standard [2]. However, while providing services to vehicle network applications using cloud platform it faced few problems such as, privacy, delay and load balancing on cloud. Along with these unauthorized or malicious attackers who access the vehicles data which leads to the loss of user privacy.

To the end, the design contains a data collection and preprocessing scheme, which adopts Map Reduce technique, Splitting and merging of chunks or blocks using MD5 algorithm, hashing technique and finally upload and download process where users can upload and download the images or files which were collected by vehicles these images may be road images, etc.,

On one side, human brain can automatically detect and forget unnecessary or irrelevant data in data analysis and interpretation machines, where it can also perform data processing, quality inspection and also reduce robustness of the system. on other side, edge computing of distributed network in certain paradigm will reduce redundant network.

Therefore, it is very important to design a new preprocessing scheme to filter a large amount of redundant data. Similarly, Internet of Vehicle is the part of this present growing technology but in recent years there were many cases where the user's data's were stolen where we can say as there were malicious took place and an unauthorized users created fake details and stole the data entered by the user.

The attackers stole the data when the data was entered into the cloud, during the process of time the data took to reach the cloud storage was high this made the attackers to commit the crime and before the data reach the cloud the attackers made changes on the data such as deleting the data or editing the data and even storing the fake or false information or data into the cloud. This made the criminals who committed crime to escape easily. To avoid this and type of attacks the present project is implemented to provide user privacy and avoid the leakage of data by reducing the time it takes the data to reach cloud storage along with reducing the redundant data into the cloud.

Privacy leakage and information delay have become the key problems that restrict network intelligence technology. Therefore, it is veryimportant to design a new preprocessing scheme to filter a large amount of redundant

data. Paper propose an image duplication avoidance system using image processing techniques, which improves the efficient memory storage.

In the existing system, the old parallel programming techniques, like passing of messages as well as shared-memory threads both are too oppressive for many developers. They need from the programmer to domesticate concurrency through making threads as well as synchronizing those using messages. Also, they need manually management of data locality. So, this job is very difficult and writing correct as well as scalable parallel codes is very difficult too.

The main contribution is summarized as:

- This paper is using Chunk splitting and Chunk merging process, whatever the image file is uploaded that will be divided into number of chunks based on the packet size. In the chuck merging process, the blocks are downloaded from database and merged and finally client will get back original image.
- In Hashing technique, the hash code will be generated using MD5 (Message Digest 5) algorithm for divided chunks or blocks.
- In Map Reduce phase, image files will be divided into blocks and for those each and every blocks hash code will be generated, where different block will have different hash code and verify whether the hash code is already uploaded into the database, if it is already uploaded then it will map to the existing block.
- Finally, upload and download process, here blocks will be uploaded into cloud storage and sent to database. While downloading the file, image files will get downloaded based on LBA (Logical Block Addressing) of the image file, blocks will be downloaded to the server and merged, later the original file will be sent to the client system.

II. Literature Survey

A. Duplicate image detection:

Employee learning technique was proposed to detect or identify an unauthorized copied of images [3]. Classification learning was used to identify the boundary between copied and irrigated images with more accuracy. To identify the copied images EFS framework with the help of prior simulated attacks. The proposed approach was built using GMM, SVM and Multivariate Gaussian method. With the help of these methods it was able to identify the existing copy images.

B. Data Collection:

The data was collected from only trusted node to reduce invalid data and also to improve the life cycle of Internet of Things system [4]. Cloud was used only to process the data which was generated by the sensor networks and which was able to provide service to the user request. MST (Minimum Spanning Tree) was built by considering trusted cluster of nodes.

C. Drive HQ:

Cloud Drive HQ is a private cloud, it provides IT cloud security, it is also one stop shop for many cloud IT services. It provides easy access to the cloud storage [5]. Here the images are stored in the form of blocks or chunks, so that the storage space on cloud will be less and the load pressure will also be reduced.

III. Privacy-Enhanced Image Data Collection and Processing Scheme

As shown in Figure 1, defines architectural design of Privacy-Enhanced Image Data Collection and Processing Schema. It is a step by step process, which defines how the files will be uploaded and processed as:

Upload Image File: File will contain all the image file which are uploaded from vehicle through system and those images can be road images, vehicle images etc.,

Image Block Generation: In image block generator, image blocks will be generated for the input or uploaded images and they will be sent to cloud storage, Image de-duplication check and hashing technique.

Hashing Technique: In hashing technique, for each and every chunks or blocks in hashing technique hash code will be generated for all the unique chunks. Later those blocks will be sent to the image de-duplication check and Logical Block Addressing (LBA).

Cloud Storage: All image blocks will be stored in the cloud storage for the first image file but for the later uploaded image files, the system will compare with each and every block in the cloud and it will neglect the similar blocks or chunks and it will only store the unique chunks or blocks into the cloud storage.

Image De-Duplication Check: Here the system will compare and check if there are any similar blocks of images and consider those similar blocks as duplicate blocks and neglects those similar blocks and doesn't all system to upload those blocks into the database.

Logical Block Addressing (LBA): It will identify the block for the file.

File Request: If the user wants to download any file then the user should request for the file and the request will be forwarded to the Image File Block Identification.

Logical Block Addressing will address the selected image file based on image file ID.

Retrieval of Blocks: Retrieval of blocks will collect the blocks which are stored in the cloud and send those blocks for merging.

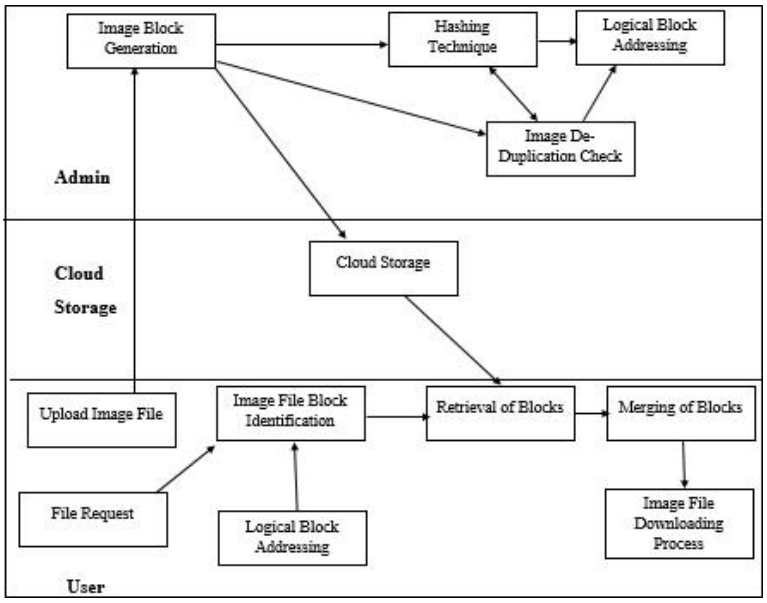

Figure 1: Architectural design for Privacy-Enhanced Image Data Collection and Processing Scheme

Merging of Blocks: it will merge the blocks which were spited will be merged again and the original image file finally will be sent to the client system without data lose.
Image File Downloading Process: in this phase the user can finally download the image file which ever he/she requested to download.

IV. Algorithm

Message Digest algorithm 5 (MD5), it is a 5th version of message digest algorithm, where this algorithm will produce 128 bits of message digest. This algorithm considers any text of 512 bits and converts it into 128 bits' plain text, 512 bit blocks will be divided into 16 blocks where each block in those 16 blocks will be 32 bits and producers 128 bits of message digits, the main purpose of developing this algorithm was for security [6]. It is useful and easy to store small hashes instead of large length. Hence, MD5 algorithm is used in this project to split the image into (5x5) blocks i.e., 25 blocks and for each block this algorithm will produce hash code. The hash code will be different from another hash code, hash code will be generated for unique blocks but it won't generate hash code for the similar blocks and those blocks hash code will be stored in database.

V. Experimental Process and Result

Paper introduce a data preprocessing scheme based on cloud computing. Cloud server directly collects data from the user terminal, only collect the latest model training results on the cloud computing. When a user generates new data, the data is used to train the edge model, and the model gradient update is uploaded to the cloud after the training is completed. The data center is used to update the shared model of the cloud. This learning mode avoids user privacy leakage and thus protects the user's private information [7][8]. This system model, efficient map reducing technique is implementing, while uploading the IMAGE file to the vehicle to cloud, file will be divided into number of chunks for each chunks hash code will be generated, for all the unique hash codes, unique id will be assigned and create <key, value> pair whereas key is the unique hash code id, and value is the respective hash code (which resembles the chunk) that are maintaining in the hash table. The unique chunks will be stored in the cloud. If the file blocks are already existing in the cloud, we are mapping to the existing chunk and will not upload the duplicate blocks to the cloud. File blocks details will be maintained as the Logical Block Addressing (set of keys) [9]. While downloading the file, based on the Logical Block Addressing we get the keys and with the keys value will be retrieved, where value is the hash code present in the hash table. Based on the hash codes chunks will be downloaded to the server and will be merged and the original file will be sent to the client system without data lose. The algorithm used in this paper is MD5 algorithm.

In this system have two actors and there are admin and user. Admin has to login using his/her username and password. After admin login he can able to create the user details and also he can see the existing user details and each and every user transaction admin can able to check it based on user ID. When the user is uploaded image and how many blocks are divided in that image. User has to login using user id and password. User has to select the image to upload so the selected image has to get divided into small blocks (500 bytes each block) [10].

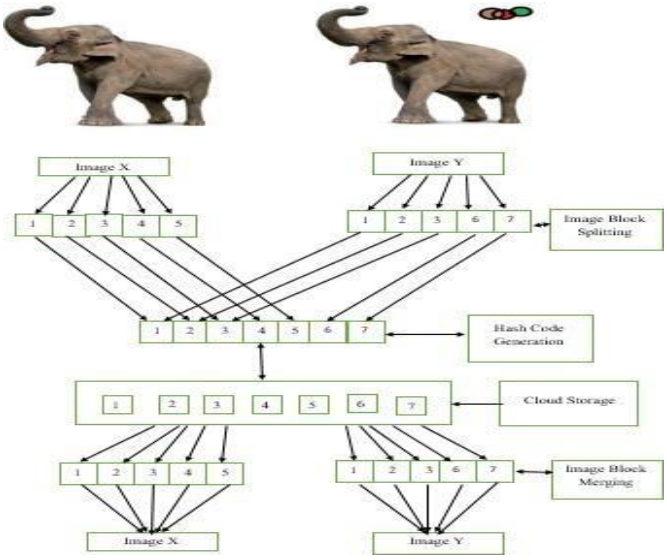

Figure 2: Image Splitting and Merging Segmentation for input image

Generate hash tag for all block is generated. Compare generated hash block with existing hash tag from database if hash tag is matched in that case we will not upload that block into cloud, we will increase number of instance of that block in database table. If hash tag is not matched in that case, we will add that block hash details in database and upload that block in cloud. LBA (Logical Block Addressing) technique is used to identify what and all blocks are present in a file. User has to download the image, firstly it has to select LBA address of selected image based on image file id, it has to download from cloud then have to merge those blocks and display the image.

As shown in Figure 2, divide an input Image X & Y into 5x5 blocks, Compute for each block, and compare it with the same block in the other images. Image X (1,1) with image Y (1,1). Compare the similarity between the two input images (should be transform invariant). Similarly, in merging all the spited blocks which are stored in the cloud storage and database should get merged. Split and merging segmentation is an image processing technique used to segment an image. The image will successively split into quadrants based on a homogeneity criterion and similar regions are merged to create the segmented result and sent to client system.

Figure 3: Admin and User page

Figure 3 shows the home page for both the actors user and admin. Where both the actors have to enter their ID and password to login and perform the tasks as shown in Figure 4 (a) and 4 (b).

Figure 4 (a): Admin Login Page

Figure 4 (b): User Login Page

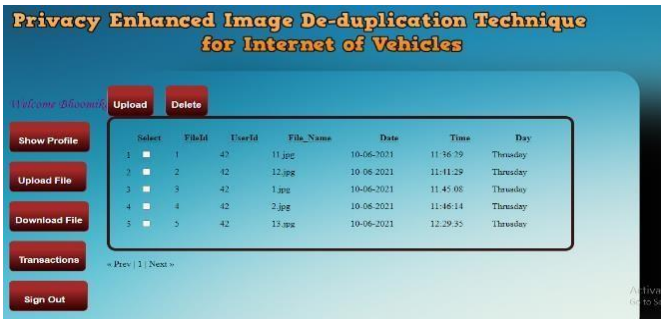

Figure 5: User Upload File

Figure 5 shows the page when the user login into his account and uploads the file or images to the server, the file will contain file ID, User ID, File Name, Date, Time and Day when the file was uploaded.

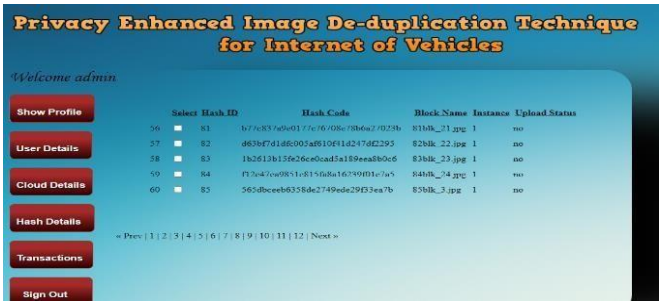

Figure 6: Admin Page where the Hash Code of Image file is Present

The admin can be able to check the hash code which is generated for the images once after the input image is splitted into blocks as shown in Figure 6.

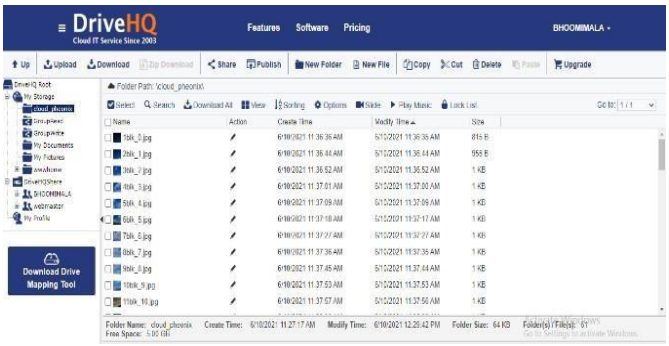

Figure 7: Image Blocks stored in Cloud DriveHQ

The input images are stored in the form of blocks into the Cloud DriveHQ as shown in the Figure 7.

Figure 8: User Download File

If the user wants to download the file, then he/she should select the file and give download later the blocks or chunks which were splitted will get merged and provide original file to the user or client according to their request as shown in the Figure 8.

VI. Conclusion

In the current technology, Internet of Vehicle is facing various problems such as privacy, delay and load balancing on cloud. Along with these unauthorized or malicious attackers who access the vehicles data which leads to the loss of user privacy. The paper proposes a Privacy-Enhance Inage De-Duplication technique as a solve for these problems, MD5 algorithm was proposed to split the image into blocks and generate Hash Code for the chunks or blocks. The images were uploaded by user, those images were split into blocks and the block files were stored in Cloud Drive HQ. Similar blocks were not stored in cloud. User will be able to upload and download the image files and, this will provide user privacy by logging in into his own profile by using his own login ID and password, with the help of this user can have his own privacy. Load balance is reduced on cloud by storing the images in the form of blocks or chunks where the similar data will be neglected and it will only store the unique chunks or blocks into the cloud storage. In future enhancement, instead of using images as input file in future the user can consider using video. MD5 algorithm can be replaced by SHA (Secure Hash Algorithm) for even faster performance. Whateverthe file is deleted by the user, the admin should be able to download that image or file.

VII. References

[1] Yingmao Li and Nicholas R. Gans, "Predictive-RANSAC: An Effective Data Fitting and Tracking Method with Application in Curve Tracking in Video," 2016 American Control Conference (ACC) Boston Marriott Copley Place, July 6-8, 2016. Boston, MA, USA.

[2] Zi Wang, Teng Jiang, Lin Zhang, Yu Liu," Mobility and QoS oriented 802.11p MAC scheme for Vehicle-to-Infrastructure communications", 978-1-4577-0101 ©2018 IEEE

[3] Jen-Hao Hsiao, Chu-Song Chen, Lee-Feng Chien, and Ming-Syan Chen, "A New Approach to Image Copy Detection Based on Extended Feature Sets", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 8, AUGUST 2007.

[4] T. Wang, L. Qiu, G. Xu, A. K. Sangaiah, and A. Liu, "Energy-efficient and trustworthy data collection protocol based on mobile fog computing in internet of things," IEEE Transactions on Industrial Informatics, 2019,10.1109/TII.2019.2920277.

[5] https://www.drivehq.com/secure/Account.aspx

[6] https://www.educba.com/md5-alogrithm/

[7] K. Yan, W. Shen, Q. Jin, and H. Lu, "Emerging privacy issues and solutions in cyber-enabled sharing services: From multiple perspectives," IEEE Access, vol. 7, pp. 26 031–26 059, 2019.

[8] Xiaojie Wang, Xiang Wei, Lei Wang, "A Deep Learning Based Energy-Efficient Computational Offloading Method in Internet of Vehicles", IEEE, 16(3), 81-91, 2019.

[9] Tian Wang, Zhihan Cao, Shuo Wang, Jianhuang Wang, Lianyong Qi, Anfeng Liu, Mande Xie, and Xiaolong Li, "Data Collection Based on Deep Learning for Internet of Vehicles", 1551-3203 (c) 2019 IEEE.

[10] Nupur Thakur, Deepa Maheshwari, "A review of image classification techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 04 Issue: 11 | Nov -2017.