

Third Party Auditing with De-Duplication

Shrikant Kale¹, Palash Kondhare², Shreyas Landge³, Santosh Pawar⁴, Asst. Prof. D. V. Jadhav⁵
^{1, 2, 3, 4, 5} Department Of Computer Engineering

^{1, 2, 3, 4, 5} Zeal College of Engineering and Research, Savitribai Phule Pune University, India

Abstract- Identifying bank cheques is one of the important task occurring in cash transactions, as banks receives huge number of different bank cheques every day from various account holders. Identifying bank cheques manually can be done but It is very time consuming. User written information such as date, amount, account number, payee name, cheque number and signature has to be visually verified. Nowadays, the banks have made a compulsion for the use of cheque truncating system (CTS) complaint cheques for saving much time and effort for deposition of cheques. An attempt is made in this paper to reduce the human efforts, time and money for automatic processing of cheques. To protect outsourced data in cloud storage against corruptions, adding fault tolerance to cloud storage together with data integrity checking and failure reparation becomes critical. Recently, regenerating codes have gained popularity due to their lower repair bandwidth while providing fault tolerance. In this paper, we propose a third party auditing with de-duplication scheme for checking de- duplication and regenerating-code-based cloud storage. To solve the regeneration problem of failed authenticators in the absence of data owners, we introduce a web service, which is privileged to regenerate the authenticators, into the traditional third party auditing with de-duplication system model. Moreover, we design a novel third party verifiable authenticator, which is generated by a couple of keys and can be regenerated using partial keys. Thus, our scheme can completely release data owners from online burden.

Keywords- Image Processing, Cloud Computing, Security, De-Duplication, Bank Cheque Clarification processing •

I. INTRODUCTION

Third party auditing with de-duplication is system to stop the flow of the physical cheques issued by a drawer to the drawee branch. The physical instrument is truncated at some point en route to the drawee branch and an electronic image of the cheque is sent to the drawee branch along with the relevant information like the MICR fields, date of presentation, presenting banks etc. This would eliminate the need to move the physical instruments across branches, except in exceptional circumstances, resulting in an effective reduction in the time required for payment of cheques, the associated cost of transit and delays in processing, etc., thus speeding up the process of collection or realization of cheques.

Machine simulation of human reading has become a promising area of research after the arrival of digital computers. The main reason for that is not only the challenge in simulating the human reading but also its utility in developing document processing systems capable of transferring data present on documents like bank cheques, commercial forms, government records and envelopes into machine readable format. Paper cheques still play a big role in the non-cash transaction in the world even after the arrival of credit cards, debit cards and other electronic means of payment. In many developing countries, the present cheque processing procedure requires a bank employee to read and manually enter the information present on a cheque (or its image) and also verify the entries like signature and date. As a large number of cheques have to be processed every day in a bank, an automatic reading system can save much of the work. Even with the success achieved in character recognition over the last few decades, the recognition of handwritten information and the verification of signatures present on bank cheques still remain a challenging problem in document image analysis

Cloud is used heavily these days for data storage but the security is still a question. The uploaded data cannot be guaranteed as original data or unharmed data .So there is need of some third part taking care of this and keeping privacy of the data.

II. LITERATURE SURVEY

Automated cheque processing system has become the most important aspect in all banking sector. Since it involves less time for clearance and comparatively less paper work is required, this form has become a huge success.

- [1] M. Jasmine Pemeena Priyadarsini et al. proposed a system in which Bank Cheques can be classified and can be authenticate using signature. A system which flags the cheques that are to be impounded and also it go through cross pattern verification to acclaim the authoritative cheques. The aim of this system is to make verification of signatures size and angle invariant. Scaling and rotational manipulations on the target image is used to achieve the invariance. This system verifies a cheque by recognizing and analyzing the measure details in cheque, which

includes account holders signature. It falls through image capturing, grey scale image conversation, binarization, Edge detection, segmentation, which is then localized & the signature is compared.

- [2] Mukta Rao et al. implemented an automated system for Authentication of Bank Cheque Signatures using Recurrent Neural Network. Author described an optimization relaxation approach which is based on Hopfield Neural Network (HNN) or recurrent network. In this approach customer's sample signature is cross matched with one supplied on the cheque. Next the various percentages are obtained by calculating the different pixels of both images. The used network is built in such a way that each pixel of different image is a neuron in network. This neuron is differentiated by its states, if particular pixel is changed. The network converges to an wavering condition according to energy function that is calculated in experiments. The HNN enables each node to take of 2 binary state values for each pixel (changed/unchanged). The performance of this system is evaluated based on different binary and grey scale images.
- [3] K.A.Valal et al. proposed a system that performs Off-line Signature Recognition and Verification which is used to recognize and verify handwritten signature of individual's. This system provides authorization in business and financial transaction. So, as the demand for processing of identification of individual faster and accurately, the automatic signature system is required. This paper presents a brief survey for various off-line signature recognition and verification system. In this paper also differentiate between off-line verification and on-line verification. The off-line verification method is performed off-line there the hardware signature of individual are acquired by scanner. The scanned signature is used for verification task. The on-line verification method is performed on-line, where the signatures are acquired by touch screen, stylus and digitizer. The device generates dynamic values- location, signature's speed, pressure of pen, co-ordinate values, signature's time etc.
- [4] "An Evaluation on Offline Signature Verification using Artificial Neural Network Approach": Othman o-khalifa, Md. KhorshedAlam& Aisha Hassan Abdalla, Department of Electrical and Computer Engineering International Islamic University Malaysia, Kuala Lumpur, Malaysia.

The signature verification is the oldest security technique to verify the identification of persons. Recently, the signature recognition schemes are growing in the

world of security technology. It offers two different types of schemes those are offline and online method. The offline technique means to verify a signature written on paper which is scanned to convert it into a digital image, whereas the online system required an online device such as Tablet PC, touch screen monitor by a pressure sensitive pen to verify the signature. This paper discusses a review of offline signature verification schemes which considered as a highly secured technique to recognize the genuine person's identity. It addresses the offline signature verification technique using Artificial Neural Network (ANN) approach. It also explains the fundamental characteristics of offline signature verification processes and highlights the comparison among various offline signature verification approaches and various signature recognition issues.

III. PROPOSED SYSTEM

Image-based Clearing System (ICS) is for faster clearing of cheques. Third party auditing with de-duplication is based on online image-based cheque clearing system where cheque images, digital signature and magnetic ink character recognition (MICR) data are collected at the bank where all data gathered and transmitted electronically.

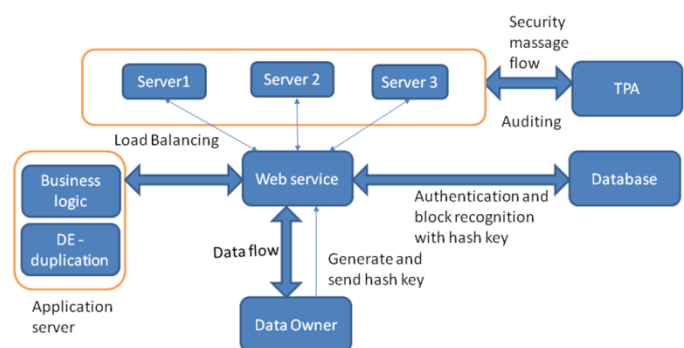


Figure: System Architecture

Cheque clearing by image processing means stopping the flow or drawbacks of the physical cheques issued by a drawer to the drawee branch. The physical instrument is eliminated at some point to en route to the drawee branch and an electronic image of the cheque is sent to the drawee branch along with the information like the MICR fields, date of presentation, presenting banks, digital signature etc. This would eliminate the need of moving the physical instruments across branches, except in exceptional circumstances, it results in an effective reduction in the time required for process of payment of cheques, the associated transit cost and delays in processing, etc., thus enhancing the process of collection or realization of cheques.

1. DYNAMIC WEB APPLICATION:

- a. GUI Design, Servlets, JSP.
- b. Login – User, TPA, Admin
- c. Registration
- d. Upload
- e. Download
- f. Scan System
- g. Alerts

2. DATABASE:

ER Database is the rational data works in background. The database used is nothing but file structure. We have used special database such as MySQL, ER database etc. MySQL, pronounced either "My S-Q-L" or "My Sequel," is an open source relational database management system. It is based on the structure query language (SQL), which is used for adding, removing, and modifying information in the database. Standard SQL commands, such as ADD, DROP, INSERT, and UPDATE can be used with MySQL. MySQL can be used for a variety of applications, but is most commonly found on Web servers. A website that uses MySQL may include Web pages that access information from a database. These pages are often referred to as "dynamic," meaning the content of each page is generated from a database as the page loads. Websites that use dynamic Web pages are often referred to as database-driven websites. Many database-driven websites that use MySQL also use a Web scripting language like PHP to access information from the database. MySQL commands can be incorporated into the PHP code, allowing part or all of a Web page to be generated from database information. Because both MySQL and PHP are both open source (meaning they are free to download and use), the PHP/MySQL combination has become a popular choice for database-driven websites.

3. RSA:

RSA is one of the first practical public-key cryptosystems and is widely used for secure data transmission. In such a cryptosystem, the encryption key is public and differs from the decryption key which is kept secret. In RSA, this asymmetry is based on the practical difficulty of factoring the product of two large prime numbers, the factoring problem.

4. Template Matching Algorithm:

- (i) A basic method of template matching uses a convolution mask (template), tailored to a special feature of the search

image, which we want to detect. This technique can be easily performed on grey images or edge images.

- (ii) The convolution output will be highest at places where the image structure matches the mask structure, where large image values get multiplied by large mask values.
- (iii) This method is normally implemented by first picking out a part of the search image to use as a template:
- (iv) We will call the search image $S(x, y)$, where (x, y) represent the coordinates of each pixel in the search image. We will call the template $T(x, y)$ where (x, y) represent the coordinates of each pixel in the template.
- (v) We then simply move the center (or the origin) of the template $T(x, y)$ over each (x, y) point in the search image and calculate the sum of products between the coefficients in $S(x, y)$ and $T(x, y)$ over the whole area spanned by the template.
- (vi) As all possible positions of the template with respect to the search image are considered, the position with the highest score is the best position. This method is sometimes referred to as Linear Spatial Filtering and the template is called a filter mask.

5. WEB SERVICE:

Web service is a method of communication between two electronic devices over a network. It is a software function provided at a network address over the Web with the service always on as in the concept of utility computing. The W3C defines a Web service generally as, a software system designed to support interoperable machine-to-machine interaction over a network.

6. DUPLICATION CHECKING:

De-duplication is a technique where the server stores only a single copy of each file, regardless of how many clients asked to store that file, such that the disk space of cloud servers as well as network bandwidth are saved.

7. FILE SPLITTING:

File splitting will done many part, as defined by the admin, as replication factor.

V. MATHEMATICAL FUNCTION

Let 'S' be the | Third party auditing with de-duplication in cloud as the final set

$$S = \{D, M, H\}$$

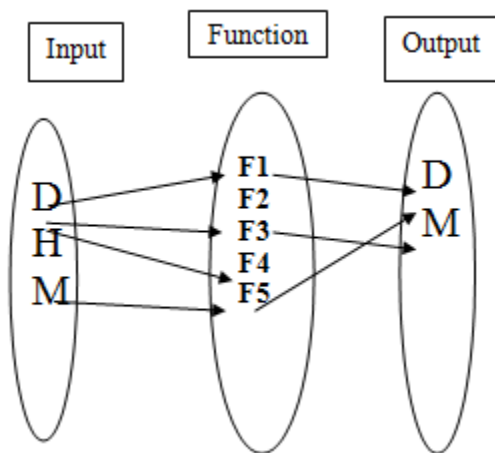
Identify the inputs as D, M

$$S = \{D, M, H\}$$

$$D = \{D1, D2, D3, D4 \dots\} \text{ 'D' given Data files}$$

$M = \{ M1, M2, M3 \dots \mid \text{'M' gives the Security message} \}$
 $H = \{ H1, H2, H3 \dots \mid \text{'H' gives hash key} \}$
 Identify the outputs as O
 $S = \{ D, M \dots \}$
 $D = \{ D1, D2, D3, D4 \dots \mid \text{'D' given Data files} \}$
 $M = \{ M1, M2, M3 \dots \mid \text{'M' gives the Security message} \}$
 Identify the functions as 'F'
 $S = \{ D, M, F \dots \}$
 $F = \{ F1(), F2(), F3(), F4() \}$
 $F1(D) :: \text{Upload data}$
 $F2(D) :: \text{Auditing}$
 $F3(H) :: \text{check for de- duplication}$
 $F4(M) :: \text{Send Audit result}$
 $F5(D) :: \text{Do authentication}$

Hence the functionality can be shown as,



VI. CONCLUSION

Third party authentication is a move in the right direction and will ensure efficiency, transparency, security and faster credit of funds benefiting both you as well as your bank. Hurdles due to jurisdiction and geographical restrictions will be obviated and you can have faster access to your funds. It's a win-win situation for banks as well as you.

ACKNOWLEDGEMENT

We are grateful to present paper on Third Party Auditing with De-Duplication. We acknowledge the efforts taken by the experts who have contribution towards development of different cheque clarification system. We acknowledge the help, suggestions given by our project guide Prof D. V. Jadhav sir, our teachers and team members.

We also acknowledge the support and platform given by the reviewers of the journal for modifications and suggestions to improve quality of paper.

REFERENCES

[1] M. Armbrust et al., "Above the clouds: A Berkeley view of cloud computing," Dept. Elect. Engg. Computer. Sci., Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2009-28, 2009.

[2] G. Ateniese et al., "Provable data possession at untrusted stores," in Proc. 14th ACM Conf. Comput. Commun. Secure. (CCS), New York, NY, USA, 2007, pp. 598–609.

[3] A. Juels and B. S. Kaliski, Jr., "PORs: Proofs of retrievability for large files," in Proc. 14th ACM Conf. Comput. Commun. Secur., 2007, pp. 584–597.

[4] "An Evaluation on Offline Signature Verification using Artificial Neural Network Approach": Othman o-khalifa, Md. Khorshed Alam & Aisha Hassan Abdalla, Department of Electrical and Computer Engineering International Islamic University Malaysia, Kuala Lumpur, Malaysia,

[5] A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment- Mayanka Katyal, Atul Mishra

[6] Dynamic Data Deduplication in Cloud Storage- Leesakul, W. ; Sch. of Comput., Univ. of Leeds, Leeds, UK ; Townend, P. ; Jie Xu

[7] "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment"- Mayanka Katyal, Atul Mishra