AUTHENTICATION SCHEME FOR PREVENTING DATA DUPLICATION IN CLOUD DATA CENTER

[1]Hammed, M. & [2]Akinbola, S. M.
[1&2]Computer Science Department, Federal Polytechnic, Ilaro, Ogun State
[1]mudasiru.hammed@federalpolyilaro.edu.ng & [2]serifat.akinbola@federalpolyilaro.edu.ng

Abstract

Cloud computing is an emerging information technology field, which allows storage, access of data, programs, and their execution over the internet with offering a variety of information and other related services. Cloud information services is an essential for data and information to be stored without duplication. Literature revealed that cloud data storage has suffered from issues related to data duplication which has attracted the attention of researchers. Despite the fact that different researchers focus on numerous techniques to eliminate data duplication especially in the cloud datacenter but some did not consider the authentication scheme as one of the tools to eliminate data duplication. Those who have proposed authentication used biometric. This study used authentication scheme using Luhn's algorithm which attained high degree of accuracy to prevent data duplication in the cloud datacenter. This method enhanced cloud performance when datacenter are free from duplication.

Keywords: *Cloud Computing, Data Storage Deduplication, Authentication Scheme, Luhns Algorithm Apriori Algorithm*

Introduction

Cloud computing is one of the emerging technology, which helped several organizations to save money and time adding convenience to the end users. Thus the scope of cloud storage is vast because the organizations can virtually store their data's without bothering the entire mechanism. Cloud Computing provides key advantage to the end users like cost savings, Able to access the data irrespective of location (Shobana, *et.al.,* 2016).Cloud computing provides virtualized information technology (IT) resources to ensure the workflow desired by user at any time and location; it allows users to borrow computing resources such as software, server and storage (Hakjun, *et. al.,* 2021). Internet spread, and usage, data availability in social media. These technologies make wider influences on big data in our day-to-day activity. Many organizations like Amazon, Flipkart and Netflix perform data collection, mining, and analysis from various sources. Sharing large volumes of data over the network has been made easily accessible through cloud storage (Vinoth, *et. al.,* 2021). Storage such as database play an important role in cloud computing and many industries and systems depend on the accuracy of databases to carry out operations. Therefore, the quality of the information (or the lack thereof) stored in the databases can have significant cost implications to a system that relies on information to function and conduct business (Ahmed, *et. al.,* 2007). Large amounts of storage in cloud infrastructure are occupied by duplicate data records and the cost of cloud storage can be reduced by deduplication process which avoids storing of same data at multiple times in real time (Vinoth, *et. al.,* 2021).However, aiming to address this challenge, researchers have focused on techniques for data de-duplication using biometric de-duplication with user authentication. This study, used the Luhns algorithm to authenticate every data that will be stored in the cloud storage to ensure that such data has not already occurred in the storage (datacenter). This authentication method used by this study eliminates set of duplicates data and keeps only unique and essential data, thus, it is significantly clearing storage space.

3.0 Related Works

Al-Assam, & Zeadally, (2019), surveyed various biometric-based authentication methods in cloud environments. The traditional password-based authentication lacks security when it comes to cloud data. To this end, multifactor authentication is suggested which allows two or more authentication parameters along with password-based security. This review focuses on the various available biometric authentication models. This study proposed an authentication scheme using Luhn's algorithm which free datacenter from duplication.

Dulari, & Bhushan, (2019), proposed a authentication method for user in cloud computing based distributed environment. In this process user's biometric information are stored as template in cloud server. Further user verification is done with several participants. Here users feature vector query is compared with template saved in cloud server. In this method, homomorphic based encryption is used for matching the protocol. The homomorphic encryption used by the author lacks practical implementation, but the practical implementation of Luhn's algorithm is more feasible compare to homomorphic encryption for authentication scheme.

4.0 Method

This research work focuses on authentication/ using Luhn's algorithm augmented with to verify any data that will be stored in the cloud storage (database) whether the data is already occurred in the base or not in order to avoid data duplication. The system automatically generates set of digits for every data that must be stored in the database while Luhn's algorithm determines the check digit for the set of digits generated by the system. Thereafter, the algorithm performs modulo10 on the set of digit and the result of modulo10 will be matched with the check digit. If the result of modulo10 matches the check digit, the data is free from duplication and it is valid to be stored in the database but error message will occur if the result of modulo10 does not match the check digit. This shows that the data want to be stored has already occurred in the database. The Luhn"s algorithm enhanced authentication for reducing data duplication.

Algorithm 1: Luhn's Algorithm

Step 1: Starting with the second to the last digit and moving to the left, double the value of all alternating digits. If the product obtained from this step is greater than 9, then subtract 9 from the product.

Step 2: Add the digits of the products together with the digits from the original number. Exclude the check digit.

Step 3: Divide the sum by 10 and check on whether the remainder is 0. If so, then that is the check digit. However, if the number is not equal to 0, then subtract the remainder from 10. The resultant number is the check digit. .

Assuming this set of digits (5342135411422512) are being generated for an incoming data (i.e. data to be stored) and can be demonstrated as follows:

Step 1: Starting with the second to the last digit and moving left, double the value of all alternating digits. If product of this doubling operation is greater than 9, then subtract 9 from the product.

Result of Step 1 of Luhn's Algorithm

Digits generated	5	3	4	2	1	3	5	4	1	1	4	2	2	5	1	2
Digits generated Doubled	10		8		2		10		2		8		4		2	
Result	1		8		2		1		2		8		4		2	

Step 2: Add the digits of the products together with the digits from the original number. Exclude the check digit (digits in brackets are the products from Step 1).

(2) + 5 + (4) + 2 + (8) + 1 + (2) + 4 + (1) + 3 + (2) + 2 + (8) + 3 + (1) = 48

Step 3: Divide the sum by 10 and verify whether the remainder is equal to 0. If the remainder is 0, then that is the check digit. If the number is not equal to 0, then subtract the remainder from 10. The resultant number is the check digit.

48 mod 10 = 8

10 – 8 = 2

The result (2) matches the check digit (2), which shows that such data has not occurred in the base. But, if the result did not match the check digit it shows that the data occurred in the base. The system flowchart in figure 1 depicts the flow of information in the system.

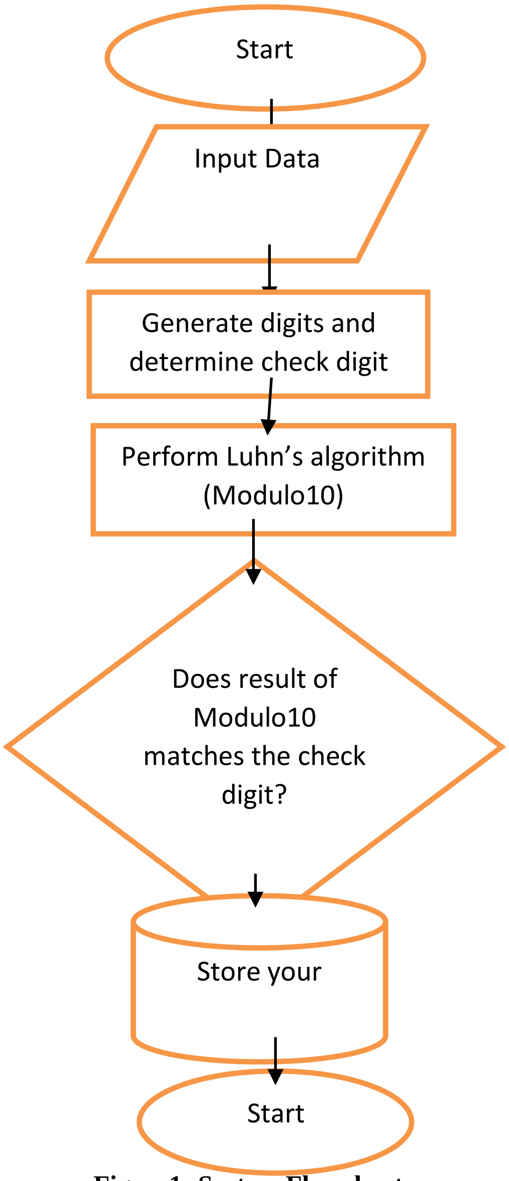
Start
Input Data
Generate digits and determine check digit
Perform Luhn's algorithm (Modulo10)
Does result of Modulo10 matches the check digit?
Store your
Start

Figure1: System Flowchart

5.0 Implementation and Result

The system was implemented with virtual datacenter which allow multiple data to be stored in the database. Different data were uploaded to the datacenter, all the data that were the same was detected by Luhn's algorithm when the data were authenticated. The Luhn's algorithm attained a very high accuracy for authenticating every incoming data when it was tested.

Conclusion

Authentication scheme using Luhn"s algorithm is a powerful tool to eliminate data duplication when data are to be stored on the cloud datacenter. This method enhanced the cloud performances when the cloud storage are being free from data duplicate.

References

Ahmed, K. E., Panagiotis G. I. & Vassilios, S. V. (2007). Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering, 19(1), Pp 1-16*

Al-Assam, H., Hassan, W,, & Zeadally, S. (2019). Automated biometric authentication with cloud computing. In: Obaidat M, Traore I, Woungang I, eds. *Biometric-based physical and cybersecurity systems. Cham: Springer, 455–475 DOI 10.1007/978-3-319-98734-7_18.*

Dulari P, & Bhushan B. (2019). A novel approach for cloud data security enhancement through cryptography and biometric in the government cloud environment. *International Journal of Computer Science and Mobile Computing 8(12):59–63.*

Hakjun, L. , Dongwoo K. , Youngsook L. , & Dongho W., (2012). Secure Three-Factor Anonymous User Authentication Scheme for Cloud Computing Environment, *Hindawi Wireless Communications and Mobile Computing Volume 2021, Article ID 2098530, 20 pages*

Shobana, R. , Shantha, K., Shalini, S., Leelavathy & .Sridevi, V., (2016).. De-Duplication of Data *In Cloud. Int. J. Chem. Sci.: 14(4), pp 2933-2938*

Vinoth K. M. , Venkatachalam,K., Prabu, P., Abdulwahab, A. & Mohamed, A.(2021). Secure biometric authentication with de-duplication on distributed cloud storage, *Peer J Computer Science, DOI 10.7717/peerj-cs.569*