World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

An improved data leakage detection system in a cloud computing environment

Prisca I. Okochi [1, 2, *], Stanley A. Okolie [1] and Juliet N. Odii [1]

[1] *Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, PMB 1526 Owerri, Imo State Nigeria.*

[2] *Department of Computer Science, College of Physical and Applied Sciences, Michael Okpara University of Agriculture Umudike Umuahia Abia State Nigeria.*

World Journal of Advanced Research and Reviews, 2021, 11(02), 321–328

Publication history: Received on 13 July 2021; revised on 22 August 2021; accepted on 24 August 2021

Article DOI: https://doi.org/10.30574/wjarr.2021.11.2.0385

Abstract

An Improved Data Leakage Detection System is designed to mitigate the leakage of crucial and sensitive data in a cloud computing environment. Generally, leakage of data in computing system has caused a lot of irreparable damage or catastrophe to various institutions or organizations worldwide. Therefore, this research aims at detecting and preventing any intentional or non-intentional data leakages using dynamic password or key for data decryption security mechanisms. To achieve this the OOADM methodology was adopted. The new system was implemented using ASP.net MVC and Microsoft SQL Server Management Studio as the backend. And by incorporating an Audit trail/Transaction log mechanism, the new system monitors the activities within and outside the computing environment with date and time stamp. Hence, the system can be applied in any environment for the prevention and detection of any data leakage.

Keywords: Data leakage; Cloud computing; Audit trail; Transaction log

1. Introduction

Data leakage is the unauthorized transfer of classified information from a computer or datacenter to an unintended outsider. Data leakage can be simply accomplished by mental means, or by physical removal of disks /reports or by subtle means such as data hiding. There has been a shift from battling to be free from intrusions, viruses, or spam, to wrestling to be free from data leakage. Data leakage can be intentional or accidental. It can be exposured of legally protected personal information, intellectual property or trade secrets. These days, the exposure of confidential information has become the number one threat for several enterprises. Data leakage occurs always when confidential business information like customer or patient data, ASCII text file or design specifications, tariffs, property and trade secrets, and forecasts and budgets in spreadsheets are leaked out. When these are leaked out it leaves the company unprotected and goes outside the jurisdiction of the organization. This uncontrolled data leakage puts business in a vulnerable position. Once this data is not any longer within the domain, then the company is at serious risk. Today data leakages can impact hundreds of thousands often millions of individual consumers, and even more individual records, all from a single attack on one company. Many previous researchers have deployed several enabling security technologies such as firewalls, encryption mechanisms, access control mechanisms, identity management, and machine learning/context-based detectors to offer protection against data leakage threat. Though a good number of researchers such as [5], developed a three-tier application that makes data access secure through a safe channel monitoring system using watermarking to ensure data do not leak out. But this technique implements a unique code in every copy of data. If later, a distributed copy of data is found at some unauthorized location, the guilty agent can be detected very easily

* Corresponding author: Prisca I Okochi
Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, PMB 1526 Owerri, Imo State Nigeria.

Copyright © 2021 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution Liscense 4.0.

[2]. Hence despite the efforts of previous researcher to prevent data from leaking out, data leakage is still a big challenge especially to many organizations and institutions.

2. Literature Review

Data leakage didn't begin when companies began storing their protected data digitally. In fact, data leakages have existed for as long as individuals and companies have maintained records and stored private information. Before computing became commonplace, a data leakage could be something as simple as viewing an individual's medical file without official permission or viewing sensitive documents that weren't properly disposed of. Still, publicly-disclosed data leakages greatly increased in frequency in the 1980s. And in the 1990s and early 2000s, public awareness of the potential for data leakages began to rise. Most information on data leakages focuses on the time period from 2005 till date. This is largely because of the advancement of technology and rapid increase of electronic data throughout the world, making data leakages a top concern for both enterprises and individuals. Sensitive data of companies and organizations includes Intellectual Property (IP), financial information, patient information, personal credit-card data etc. Furthermore, in many cases, sensitive data is shared among various stakeholders like employees performing from outside the organizational premises (e.g., on laptops), business partners and customers [7]. Data leakage is defined as the accidental or unintentional distribution of personal or sensitive data to unauthorized entity [7]. Data leakages have gained widespread attention as businesses of all sizes become increasingly reliant on digital data, cloud computing, and workforce mobility. Sensitive business data are stored on local machines, on enterprise databases, and on cloud servers, leaking a company's data has become as simple or as complex as gaining access to restricted networks. Data leakage also is a situation whereby crucial information is illegally transferred to the outside world [16]. Traditionally, leakage detection is handled by watermarking, e.g., a singular code is embedded in each distributed copy. In case a replicate is found later within the hands of a third or an unauthorized party, the leaker can be identified. Watermarks are often very useful in some cases, but again, involve some modification of the first data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious. E.g. A hospital may give patient records to researchers who will devise new treatments.

3. Methodology

The methodology adopted for this research work is the Object-Oriented Analysis and Design methodology (OOAD). The existing system from the perspective of objects and similar objects are grouped as classes and their characteristics are handled as properties while their behaviors are treated as the actions or methods within the same bundle of object. This methodology was chosen because it is best used to handle a system where different objects exist. It is a structured approach that is used in analyzing and designing a system. It applies object-oriented concepts and develop a set of graphical system model during the development lifecycle of the software.

3.1. Proposed System Architecture

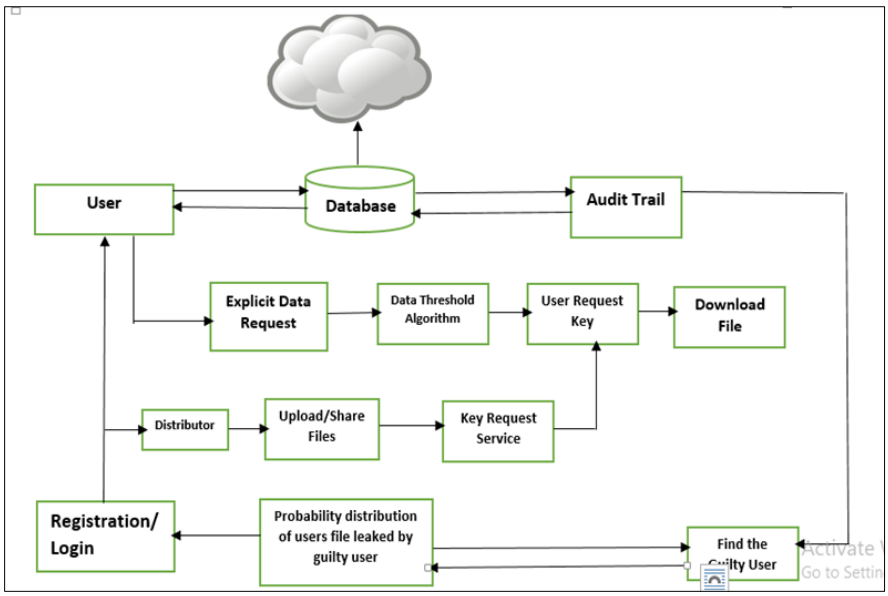

Figure 1 Architecture of the proposed system

The New application system has in it an audit trail/transaction log system which checks also monitors user transactions and activities to computing resource at each stage of the process. The system administrator registers each user and assigns roles to them. The roles include an administrator, a distributor and a user. The distributor uploads new files into the system, approves, sends key request to the user, also shares file to each user. Above depicts the architecture of the proposed system:

4. Results

4.1. Input Interface

In this system, data is supplied by the user during registration which in turn is used to create login details and the secret key of the user. The interface makes use of Graphical User Interface components such as radio buttons, text boxes to accept input from the users into the system.

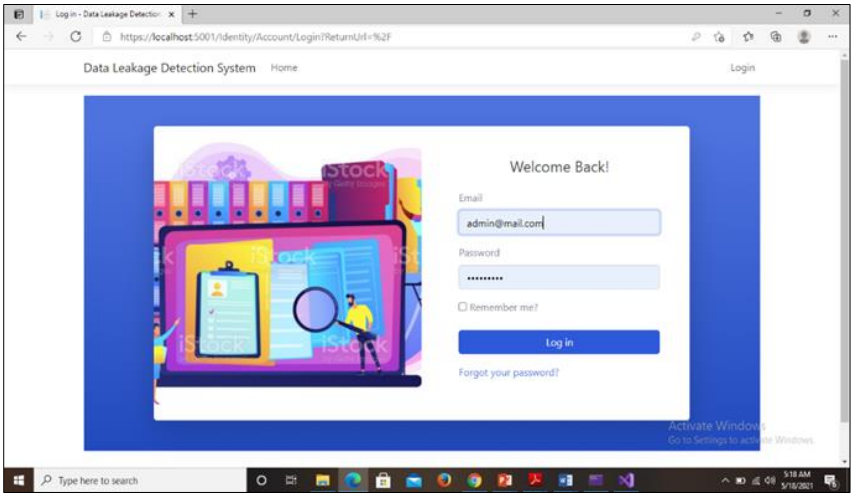

Figure 2 Admin sign in Module

Figure 2 shows the module where the admin logs in to the system in order to create new users and assign role to the users.

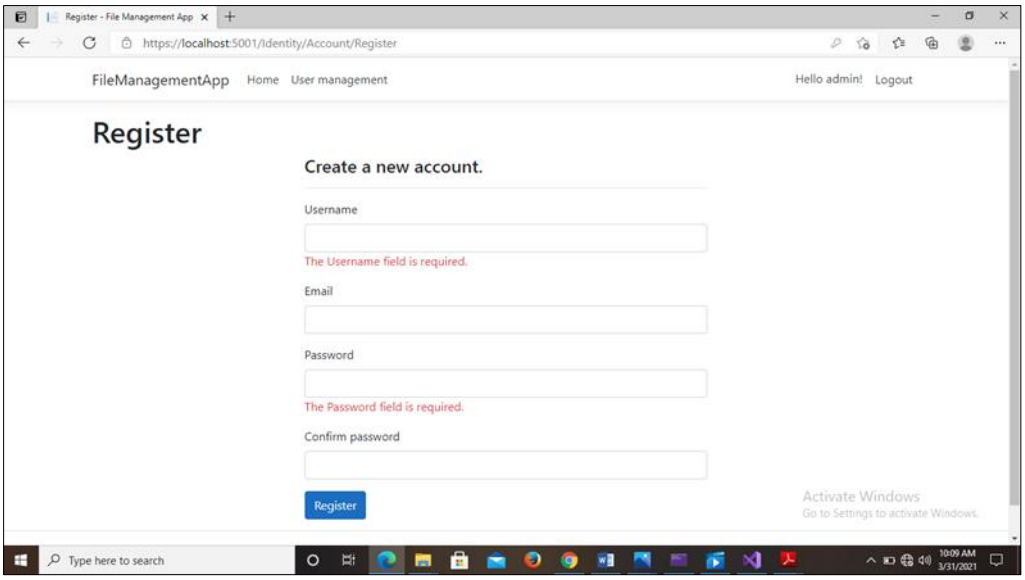

Figure 3 Admin registers a new user

Figure 3 illustrates where the administrator creates a new user. The “Register” button is clicked after entering the fields.

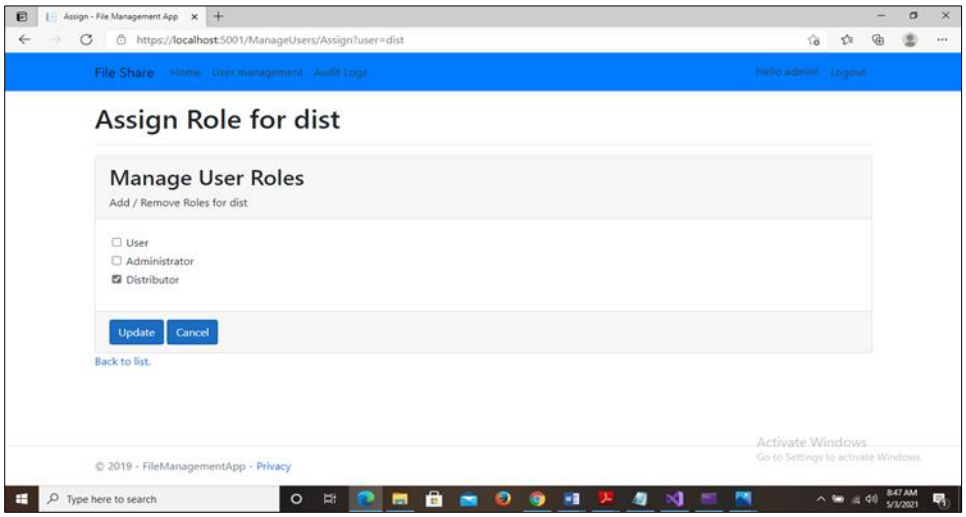

Figure 4 Admin assigns roles

Figure 4 shows where the administrator assigns role to the users, one user can be a distributor of files, or another administrator or only a user.

4.2. Output Interface/ Test Results

Firstly, the administrator logs into the system and has the privileges of registering users, viewing the list of all registered users, assign roles to members, and checking the audit trail to identify a leaker.

In Figure 5 below, the admin views the set of all users that have registered and also can assign roles to them.

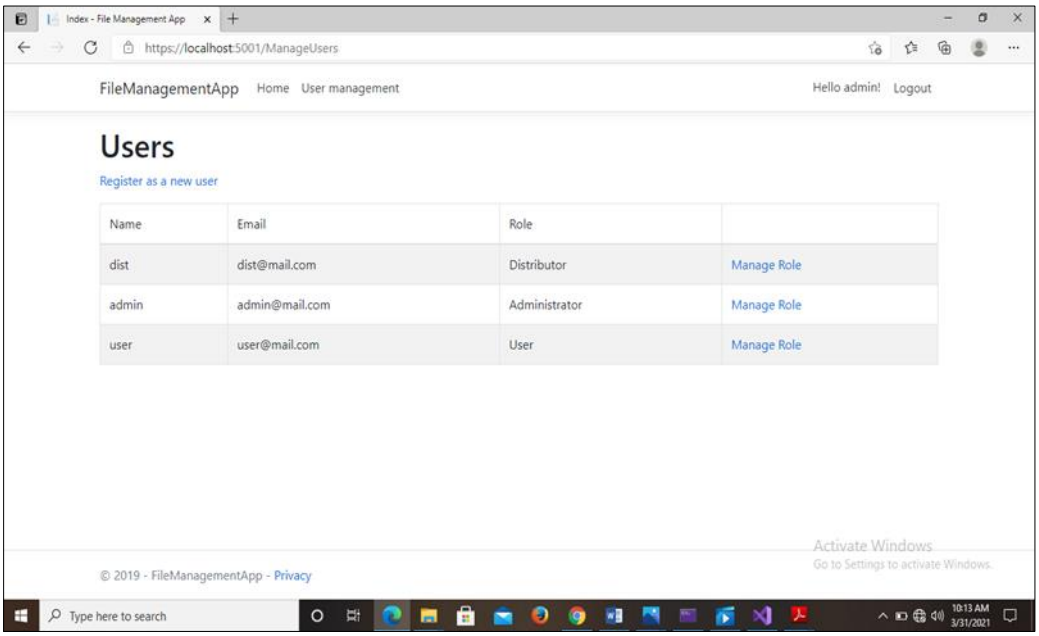

Figure 5 Admin dashboard

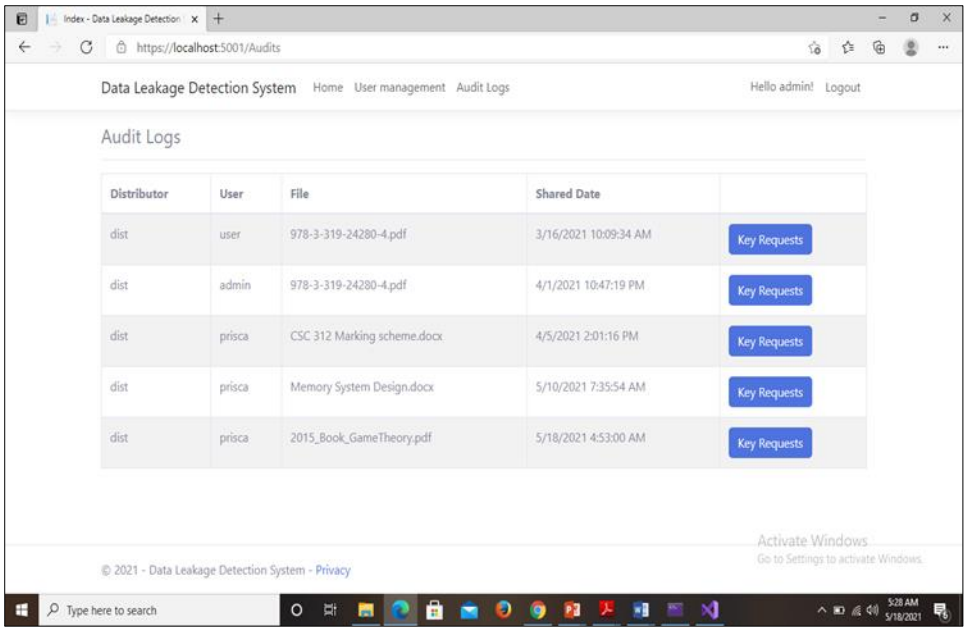

Figure 6 Audit Trail/Transaction log module

Figure 6 shows the audit log or trail that tracks user activities in the system. It shows the various activities performed by the user (admin) including the date and time and actual activity performed at each stage of the application process. This audit trail table has no link or dependency to the other tables in the program making it safe from manipulations from other tables.

It is intended to be activated at every login stage for user identification and verification purposes. It has been tied to application accessibility, which means a user must be successfully verified before access is granted into any software application. This will ensure that users who are not biometrically verified through this system have no access to any software program. It is just another level of the user identification and authorization within the cloud computing architecture (a middleware application).

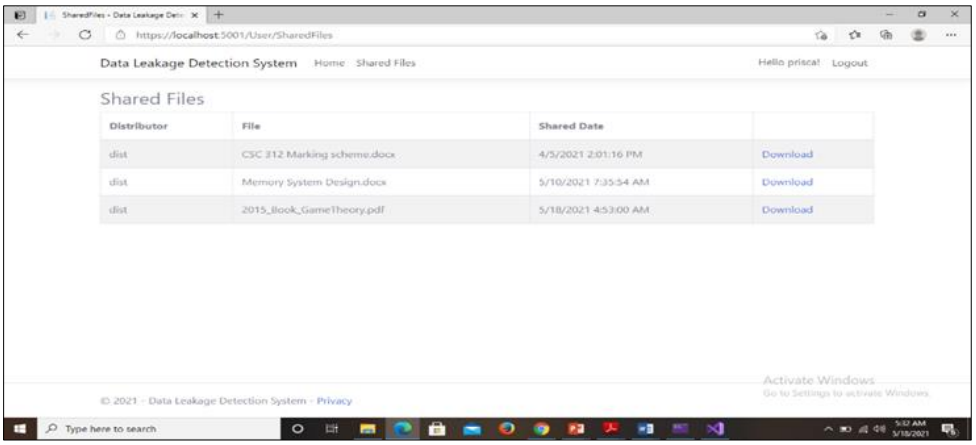

Figure 7 Shared file module

Figure 7 shows the files shared to a particular user and can be downloaded through the approval of key request by the distributor or administrator.

Figure 8 Graph of Entries without Transaction Log/Audit Trail

Figure 8 depict graphically, 10 clients entered in to the without transaction log/Audit trail system. The graph shows that for any member of clients entered there is no any mechanism to profile user activity in the system

Figure 9 therefore shows that for each client entered there is a corresponding transaction log/Audit Trail System that profile user activity in the system.

Table 1 Table of Entries without Transaction Log/Audit Trail

No. of Users	0	1	2	3	4	5	6	7	8	9	10
Audit Trail	0	0	0	0	0	0	0	0	0	0	0

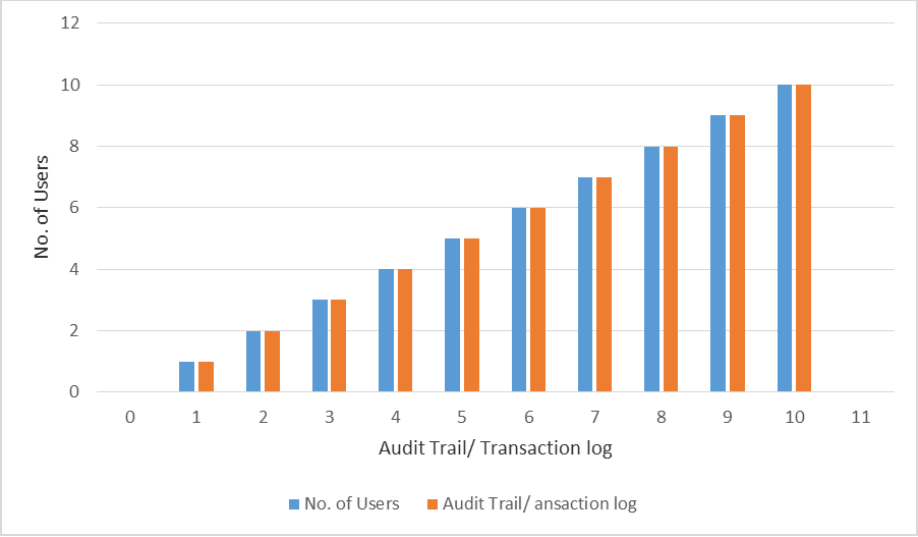

Figure 9 Graph of Entries with Transaction Log/ Audit Trail

Table 2 Table of Entries with Transaction Log/Audit Trail

No. of Users	0	1	2	3	4	5	6	7	8	9	10
Audit Trail	0	1	2	3	4	5	6	7	8	9	10

5. Conclusion

Data security is not an easy task even top leading organizations suffer the fear of data leakage. This research work discussed some of the latest work carried out in the area of detection of data leakage and prevention algorithms, tools and technologies supported. And the researchers were able to develop a system that provides security to overcome issues related to cloud environment. This helps to protect the data from leakage by tokenization of files before distributing and setting the time bound for each and every file that particular user needs to download from the cloud storage. The system protects the data leaked from guilty agent who act as a third party and security is provided using dynamic key generation which is an auto generated random unique number for every file when user or an employee makes attempt to view the content of file. Incorporated in it also is an Audit Trail/ Transaction log which profiles user activities in the system as against the old system that does not have an Audit trail. The audit trail will monitor when a user sends out organization's information with date and time stamp.

Compliance with ethical standards

Acknowledgments

We acknowledge the services of Charles Osedeke and Beloved Obinnaya for ensuring that all necessary grammatical corrections were made. We also appreciate the MOUAU librarian who granted us access to materials used in the work.

Disclosure of conflict of interest

No conflict of interest. As this is the contribution of the aforementioned authors.

References

[1] Archana Vaidya, Prakash Lahange, Kiran More, Shefali Kachroo, Nivedita Pandey. Data Leakage Detection, International Journal of Advances in Engineering & Technology. 2012; ISSN: 2231.

[2] Bhatt C, Sharma R. Data Leakage Detection, International Journal of Computer Science and Information Technologies. 2014; 5(2): 2556-2558.

[3] Buneman P, WC Tan. Provenance in Databases, in SIGMOD ACM, Bejing, China. 2007; 1171-1173.

[4] Malsoru V, Naresh Bollam. Review on Data Leakage Detection, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com. 2016; 1(3): 1088-1091 1088.

[5] Monali U Pawar, Shraddha A Mankar, Snehal S Mandhare, Siddhi N More, Rashmi R Patil. Enhancement of Data Leakage Detection Using Encryption Technique. 2019.

[6] Priyanka Barge, Pratibha Dhawale, Namrata Kolashetti. Ass. Prof, A Novel Data Leakage Detection", Department of Computer Engineering, NIRMALA CHOUHAN International Journal of Modern Engineering Research (IJMER). 2013; 3(1): 2249-6645.

[7] Sandip A Kale, Prof SV Kulkarni. Data Leakage Detection, Department Of CSE, MIT College of Engg, Aurangabad, Dr. B. A. M. University, Aurangabad (M.S), India International Journal of Advanced Research in Computer and Communication Engineering. November 2012; 1(9).

[8] Sion R, M Atallah, S Prabhakar. •\Rights Protection for Relational Data, Proc. ACM SIGMOD. 2003; 98-109.

[9] Patil Rashmi, SM Sangve. Public Auditing System: Improved Remote Data Possession Checking Protocol for Secure Cloud Storage. 2015.

[10] Peneti S, BP Rani. Data Leakage Pevention System with Time Stamp, in International Conference on Information Communication and Embedded System (ICICES). 2016; 1-3.

[11] Plummer D, Clearly D, Smith D. Cloud Computing – Confusion leads to Opportunity. 2008; 10-20.

[12] Pon Periyasamy AR, E Thenmozhi. DataLeakage Detection and Data Prevention Using Algorithm International Journal of Advanced Research in Computer Science and Software Engineering. 2017.

[13] Prashant Khobragade, Ashish Golghate. Data Leakage Detection and Security in cloud computing. 2016.

[14] Praveen S Kumar, Y Srinivas, D Suba Rao, Ashish Kumar. A Novel Model for Data Leakage Detection and Prevention in Distributed Environment, International Journal of Engineering and Technical Research (IJETR). 2016.

[15] Rajat Verma, Vipin Gautam, Chandra Prakash Yadav, Ishu Gupta, Ashutosh Kumar Singh. A Survey on Data Leakage Detection and Prevention. International Conference on Data Analytics and Management (ICDAM 2020).

[16] Davis Ziff. PC Magazine Encyclopedia, Definition of Cloud. 2009.