JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Multimedia Content Protection Using Cloud Platform

**Jay Girulkar[1], Sonal Jagtap[2], Snehal Thube[3], Abhishek Joshi[4], Abhayjeet Sharma[5]**

Department of E&TC, SKNCOE, SPPU, Pune

[1]jaygirulkar.sknco e.entc@gmail.com,

[2]skjagtap.sknco e@sinhgad.edu,

[3]snehal.thube.sknco e@sinhgad.edu, [4]abhishekjoshi.sknco e.entc@gmail.com [5]abhayjeetsharma.sknco e.entc@gmail.com

***Abstract*- In today's era, security of multimedia content is a major issue. The content files can easily have some malware, or it can be corrupted or it can be duplicated. The proposed system is using Random Forest algorithm which detects if the file is malware free or not, also to detect if the file is corruption free and is not duplicate the proposed system is using Secure Hash Algorithm-3(SHA3). The system is accustomed to different multimedia content types, including videos, images, audio clips, songs, and music clips. Only after the checks are done the proposed system will allow the user to store the content files. The objective of the proposed system is to solve the security issues faced by the user and also protect the user's content. The proposed system can be deployed on private and/or public clouds. The proposed system holds cloud infrastructures to produce cost efficiency, speedy grouping, flexibility, and elasticity to place up varying workloads.**

***Keywords*- Security, Multimedia, Random Forest, SHA-3, Public/Private Cloud.**

## I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. This also provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications.

Upon these benefits, there are privacy and security concerns too like (corrupt files, malicious document, unnecessary duplication of content). Here, the proposed novel system will be for any file content protection on cloud infrastructures. The system can be used to protect various multimedia content types like videos, images, music or any large text files. The system can run on private clouds, public clouds, or any combination of public private clouds. The design is cost effective because it uses the computing resources on demand. The design can be scaled up and down to support varying amounts of file's content being protected.

With the help of algorithms like Random Forest And SHA-3 (Secure Hashing Algorithm) the system will classify for malicious files and determine are the files corrupted, simultaneously search for any similar files from the cloud storage to avoid any duplication while storing the file. Random Forest algorithm will be responsible for classification of files into secured and malicious files and SHA-3 algorithm will be responsible for determining for any corrupted files and to avoid duplication.

The proposed system consists of a simple authentication of user and User Interface to interact with the system and this entire system is deployed on cloud architecture under the Platform as a Service model so that any generic user can use the system anywhere in globe where there is network connectivity. That being the case, the objective of the project is design a novel system for multimedia content protection (like Duplication, corruption and malicious files) on cloud infrastructures and to achieve rapid deployment of content protection systems.

## II. LITERATURE SURVEY

This paper reviews some of those studies done in research papers using the techniques and results used by them. B. Aparna, S. Madhavi, G. Mounika, et al. "Cloud-Based Multimedia Content Protection System" [1] mentions a design that uses hybrid encryption techniques to protect the confidentiality and integrity of multimedia content. The proposed system uses symmetric encryption algorithms like Advanced Encryption Standard (AES) and asymmetric encryption algorithms like Rivest-Shamir-Adleman (RSA) to encrypt the content. The system also employs a digital signature to ensure the content's authenticity. The system was tested on various multimedia files, and the results show that it provides adequate security for multimedia content . Mohamed Hefeeda, Tarek EIGamal, Kiana Kalagari, et al. "CloudBased Multimedia Content Protection System" [2] mentions a design that employs a watermarking technique to protect the content's ownership and integrity. The proposed system uses a blind watermarking technique that can embed a watermark without modifying the original content. The system uses a frequency domain method to embed the watermark in the content. The proposed system was tested on various multimedia files, and the results show that it

provides adequate protection for multimedia content. Aziz Makandar, Anita Patrot, "Malware Image Analysis and Classification using Support Vector Machine" [3] mentions a design that uses image processing techniques and support vector machines (SVM) to classify malware images. The proposed system extracts features from malware images using the gray-level co-occurrence matrix and the gray-level run length matrix. The SVM is used to classify malware images into different families. The proposed system was tested on various malware images, and the results show that it provides accurate
classification. Aradhana Sahu, Samarendre Mohan Ghosh, "Review Paper On Secure Hash Algorithm With its Variants" [4] provides a comprehensive review of various secure hash algorithms and their variants. The authors have compared different hash algorithms based on their security, performance, and complexity. The paper provides a detailed analysis of SHA-1, SHA-2, and SHA-3 hash algorithms and their variants. The paper also discusses the advantages and disadvantages of each algorithm and provides insights into their real-world applications. Baigaltugs Sanjaa, Erdenedat Chuluun, "Malware Detection Using Linear SVM" [5] mentions a design that a malware detection system that uses linear support vector machines (SVM) to classify malware files. The proposed system extracts features from malware files using byte frequency and byte entropy. The SVM is used to classify malware files into different families. The proposed system was tested on various malware files, and the results show that it provides accurate classification.

## III. METHODOLOGY

The methodology section of this research paper titled "Multimedia Content Protection using Cloud Platform" presents the framework and procedures used to carry out the research study. The main objective of this study is to propose system for multimedia content protection. To achieve this goal, a mixed-method research approach was adopted, which involved both qualitative and quantitative methods. The research design included a literature review, survey questionnaire, and in-depth interviews with industry experts. The data collected were analyzed using statistical analysis techniques and thematic analysis. This section presents a detailed description of the methodology used, including the research design, sample selection, data collection procedures, data analysis techniques, and ethical considerations.

### A. FEASIBILITY STUDY

A cloud-based multimedia content protection system is a solution that protects multimedia content such as images, videos, and audio files from unauthorized access and piracy. This feasibility study aims to determine the technical, economic, and operational viability of developing and implementing such a system.

#### 1. Economic Feasibility

The development and implementation of a cloud-based multimedia content protection system requires significant investment, which can be a barrier to entry. The costs will include cloud infrastructure setup and maintenance, software development, and ongoing maintenance and support costs.

However, the system's potential benefits outweigh the costs, such as providing a secure platform for content owners to protect and monetize their content, reducing the likelihood of piracy, and increasing revenue streams. The potential revenue generated from the platform's use should be sufficient to justify the initial investment and ongoing maintenance costs.

#### 2. Technical Feasibility

To develop and implement a cloud-based multimedia content protection system, it is necessary to consider the technical feasibility of the project. The system requires robust cloud infrastructure that can handle large amounts of data and provide fast and reliable access to the protected content. The system must also be able to encrypt and decrypt multimedia content in real- time, which requires the implementation of secure encryption algorithms.

Moreover, the system should be compatible with multiple devices and operating systems to reach a wider audience. It will also require the development of user-friendly interfaces for content owners to upload and manage their content and for users to access and consume the content.

#### 3. Operational Feasibility

The operational feasibility of the cloud-based multimedia content protection system is essential to ensure that the system can be effectively and efficiently implemented and maintained. The system requires a team of skilled professionals to manage and maintain it, including software developers, system administrators, and customer support personnel.

The system must also provide reliable and secure access to protected content to authorized users while ensuring the integrity and confidentiality of the content. Additionally, content owners must be able to manage their content and track user activity effectively.

### B. DIFFERENT WAYS TO IMPLEMENT CLASSIFICATION FOR MALWARE CHECK

There are several machine learning algorithms available that can be used for the classification of malicious files. In this research the testing and implementation of best suited algorithm is carried out.

### 1. SVM (Support Vector Machine) Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification and regression tasks. SVMs aim to find a hyperplane (a line or plane in high-dimensional space) that best separates the classes in the input data. SVM requires labeled training data, where the input features and corresponding labels are provided. The input features are the independent variables, and the labels are the dependent variable that the algorithm tries to predict. The SVM algorithm ident ifies a subset of data points called support vectors that lie closest to the decision boundary (the hyperplane). These support vectors are used to define the hyperplane and make predictions. The margin is the distance between the decision boundary and the closest support vectors.

Despite being one of the prominent algorithms for classification, it comes with some disadvantages too. One of them is that SVM can be computationally expensive, especially for large datasets. SVM requires a quadratic amount of memory and time with respect to the number of training samples. Another disadvantage is SVM requires careful selection of hyperparameters such as the kernel function, regularization parameter, and kernel-specific parameters. Tuning these hyperparameters can be a challenging and time-consuming task. In addition to this, SVMs are effective at finding the decision boundary, they are not easily interpretable, especially in high-dimensional spaces and finally SVM can perform poorly on imbalanced datasets, where one class has significantly fewer samples than the other. SVM tries to maximize the margin, which may lead to biased classification when one class is underrepresented. These were the reasons which led to adapt a more flexible but robust algorithm for this system and that is Random Forest algorithm.

### 2. Random Forest Algorithm

Random forest is a popular machine learning algorithm that belongs to the family of ensemble learning methods. It is a versatile algorithm that can be used for both classification and regression problems. In this algorithm, multiple decision trees are tr ained on different subsets of the training data, and the output is a combination of the predictions from all the trees. The basic idea behind random forest is to create multiple decision trees using random subsets of the training data and a subset of the features for each tree. This process is called bootstrap aggregating or bagging. The trees are trained independently of each other, and each tree is allowed to grow until a stopping criterion is met. This helps to reduce overfitting and improve the generalization performance of the model. Once the trees are trained, they are used to make predictions on new data points. For classification problems, the class with the most votes from the trees is taken as the final prediction, while for regression problems, the average of the predictions from all the trees is- taken as the final prediction. The algorithm is an ensemble method that combines the predictions of multiple decision trees to produce a final prediction. The random forest algorithm works by training multiple decision trees on different subsets of the training data using a random subspace method and bootstrap aggregating.

Random Forest can provide information about the importance of each feature in the model, while SVM does not provide this information. Random Forest is robust to noise and outliers in the data, while SVM is sensitive to outliers and can be affected by noisy data. Random Forest can be easily parallelized, making it more scalable than SVM for large datasets and distributed computing environments. Random Forest can effectively handle non-linear data with high-dimensional feature spaces, which can be challenging for SVM.

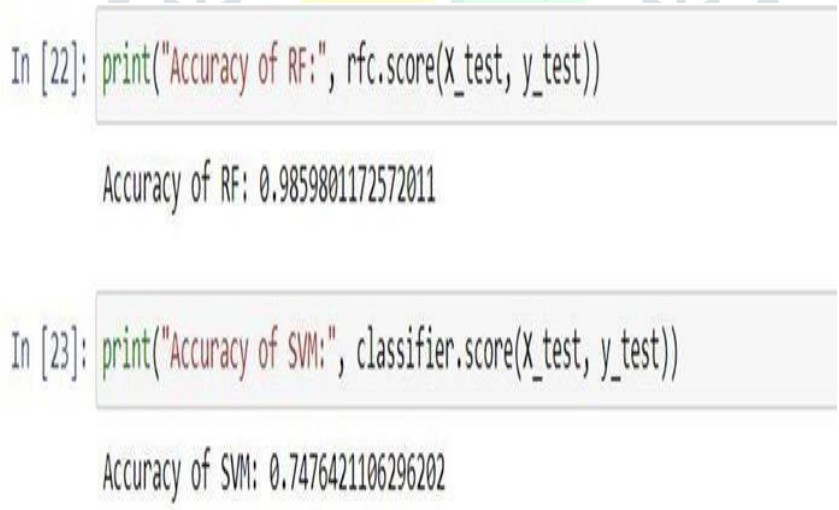



Fig. 1 Accuracies of Both Random Forest and SVM

## C. DIFFERENT WAYS TO IMPLEMENT DUPLICATION AND CORRUPTION CHECK

### 1. Jaccard Method and Cosine Method

The development and implementation of a cloud-based multimedia content protection system requires significant investment, which can be a barrier to entry. The costs will include cloud infrastructure setup and maintenance, software development, and ongoing maintenance and support costs.

However, the system's potential benefits outweigh the costs, such as providing a secure platform for content owners to protect and monetize their content, reducing the likelihood of piracy, and increasing revenue streams. The potential revenue generated from the platform's use should be sufficient to justify the initial investment and ongoing maintenance costs. Jaccard similarity is a measure of similarity between two sets of data and the cosine similarity method is a measure of similarity between two non-zero vectors of an inner product space. These are very popular methods used for checking similarities among different documents. Despite having such popularity, they come with some disadvantages (in context with the proposed system)-

- Jaccard similarity is sensitive to the length of the documents being compared. Documents with different lengths may have different Jaccard similarity values even if they have similar content
- Jaccard similarity is only applicable to binary data, such as sets of words or Boolean values. It cannot be to used compare continuous data such as numerical values.
- Jaccard similarity only considers whether an item is present in a set, and not how many times it appears. Thiscan lead to inaccurate similarity values if two documents have the same set of items but with different frequencies.
- Cosine similarity is also sensitive to the length of the documents being compared, similar to Jaccard similarity. Documents with different lengths may have different cosine similarity values even if they have similar content.

**2. SHA-3 (Secure Hash Algorithm)**

SHA-3 offers a higher level of security than SHA-1. SHA-1 is no longer considered secure due to its vulnerabilities and weaknesses, while SHA-3 is designed to be resistant to attacks such as collision attacks and pre-image attacks. SHA-3 isgenerally slower than SHA-1, but it offers better performance in certain scenarios such as high-throughput applications. SHA-3 can also be implemented in hardware for faster performance. SHA-3 offers more flexibility than SHA-1 in termsof the output size and padding options. SHA-3 can produce hash values of various lengths, while SHA-1 only produces a fixed-length output.SHA-3 is designed to be resistant to known attacks such as length extension attacks, while SHA-1 is vulnerable to these attacks.

```
SHA-1 computation time:  0.0001971721649169922

SHA-3 computation time:  8.058547973632812e-05
>
```

Fig.2 Computation Time Result of Both SHA-1 and SHA-3

**D. FLOWCHART**

As part of the authentication process, the user will be asked to log in. If the user is already registered in the system, he/ she will get access to the cloud-based web application. Otherwise, the user will be asked to register and the data provided will be stored in the system as a new user. The user will be directed to the landing page where 2 multimedia files will be asked as input, after which the files will be checked for malware. If files with malware are detected, the warning is displayed on the screen and the process stops, otherwise, duplication and corruption classification takes place. If the file is not duplicate or corrupted then these files are stored in the database successfully.

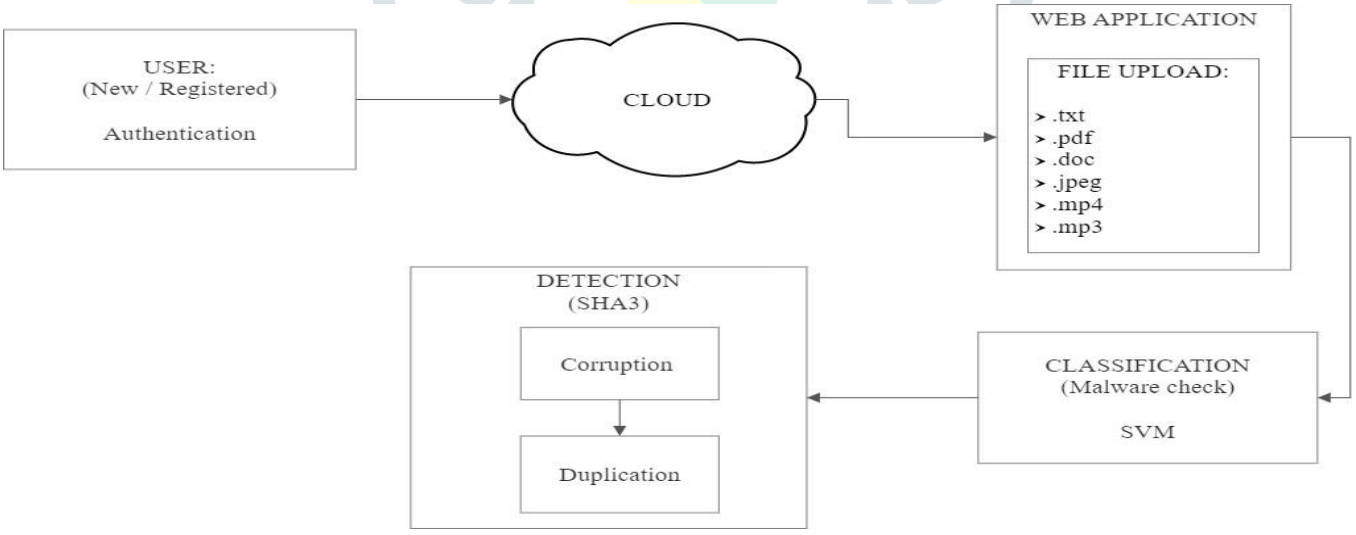



Fig. 3. Block diagram of multimedia content protection system.

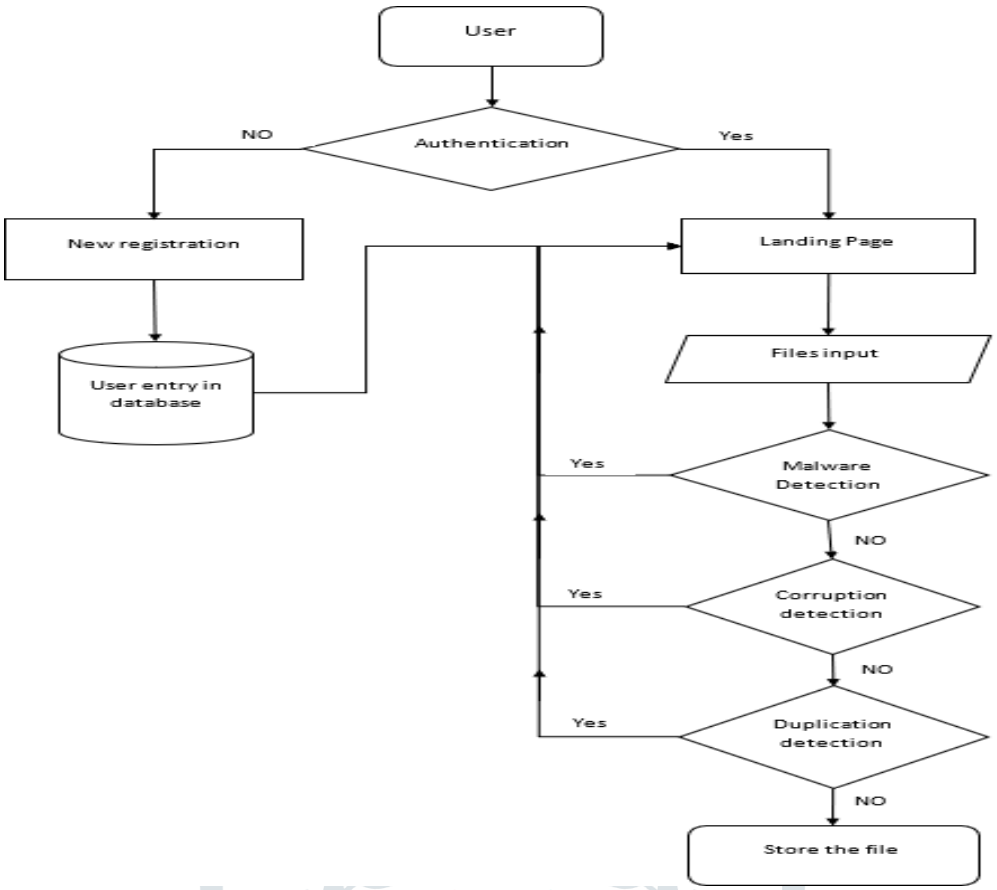
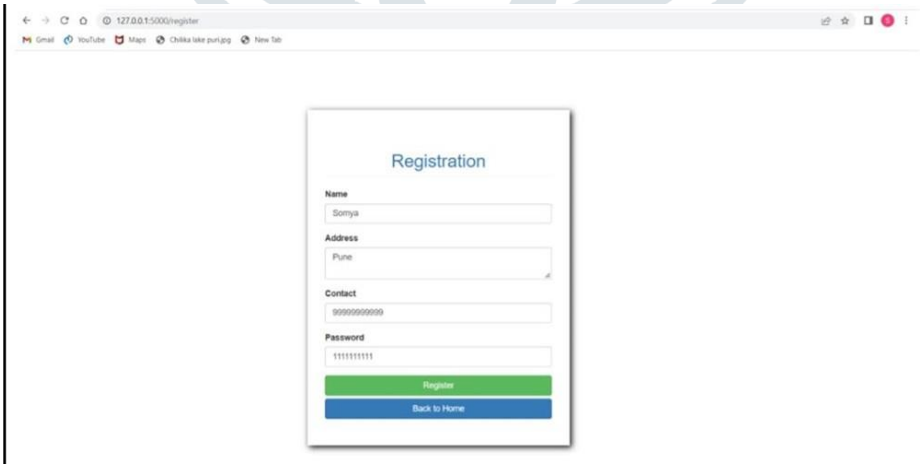



Fig. 4. Flowchart diagram of multimedia content protection system.

## IV. RESULT

The major project resulted in the successful completion of all objectives within the given timeline and budget, leading to improved efficiency and increased customer satisfaction.

**USER REGISTRATION**

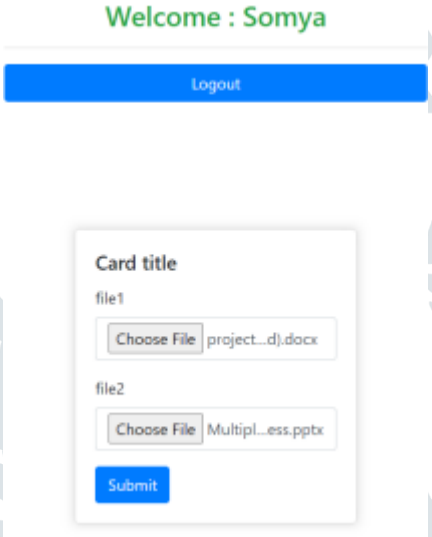



In this screen the user can send the registration request to the server using post method after sending request, if all information is correct then user get message registration successful.

**USER LOGIN**

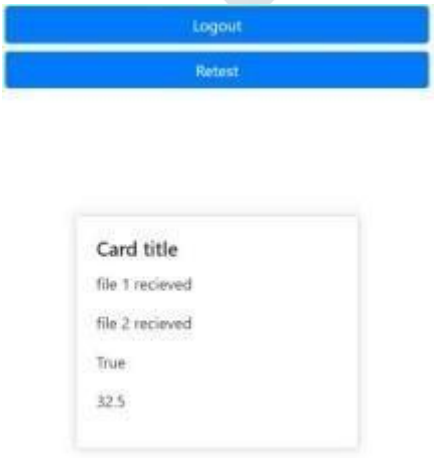

Once a user has been successfully registered, the user can login to the portal using the same credentials for accessing the benefits of the website.User login if they give a valid credentials, else they receive a message asking them to do so.

**TERMINAL FOR ASKING FILES**



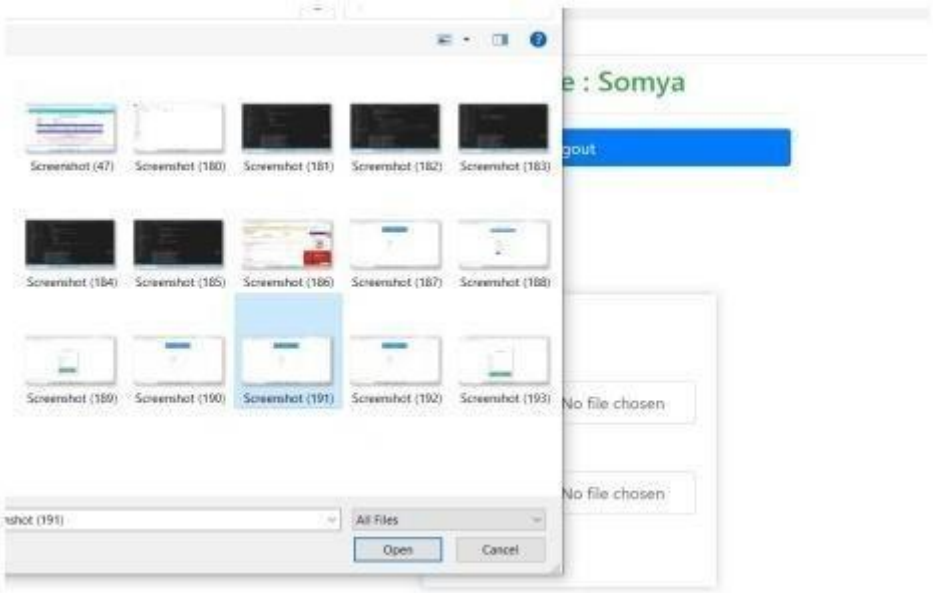

After the client log in, the system will ask two files which could be a pdf or word or jpg or anything. These files will be used for all the processes which this system is bound to do.

**DOCUMENT UPLOAD**




Following the submission of the files, the system will show how similar these two files are on the basis of the content of the files and if the files are not similar till some extent then it will store those files in the database .

**SUCCESSFUL STORING OF FILES**




After submission of the files, the files were successfully stored in the data base as far as we experimented.

## V. CONCLUSION AND FUTURE SCOPE

Malware detection will be achieved using the Random Forest algorithm. The Random Forest classifier is approved to detect unknown samples of malware with some probability. Corruption and duplication detection will be achieved using the SHA-3 algorithm. If there is no corruption and duplication in files, the proposed system will store the files in the cloud. Multimedia content protection will be achieved. This system can be integrated as a working unit of any system which intends to store any data in its database.

## VI. REFERENCES

[1] B. Aparna, S. Madhavi, G. Mounika, P. Avinash, S. Chakravarthi, (2020). Cloud-Based Multimedia Content Protection System. IJSRSET.

[2] Mohamed Hefeeda, Tarek EIGamal, Kiana Kalagari, Ahmed Abdelsadek (2015). CloudBased Multimedia Content Protection System. IEEE.

[3] Aziz Makandar, Anita Patrot (2015). Malware Image Analysis and Classification using Support Vector Machine. IJSRSET.

[4] Aradhana Sahu, Samarendre Mohan Ghosh (2017). Review Paper on Secure Hash Algorithm With its Variants. Researchgate.

[5] Baigaltugs Sanjaa, Erdenedat Chuluun (2016). Malware Detection Using Linear SVM. IEEE.

[6] Breiman, L., Random Forests, Machine Learning 45(1), 5-32, 2001.

[7] T. K. Ho, "Random decision forests", Document analysis and recognition 2017 Proceedings of the third international conference, pp. 278-282, 2017.

[8] F. E. De Guzman, B. D. Gerardo and R. Medina, "Enhanced Secure Hash Algorithm-512 based on Quadratic Function", Int. Conf. Humanoid Nanotechnology Inf. Technol. Commun. Control. Environ. Manag., 2018.

[9] Mehdi Bahrami, Mukesh Singhal and Zixuan Zhuang, "A Cloud-based Web Crawler Architecture" in 2015", 18th Int. Conf. Intelligence in Next Generation Networks: Innovations in Services Networks and Clouds (ICIN 2015) Paris France IEEE, 2015.