SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)

EPRA International Journal of Research and Development (IJRD)

Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

FRAUDULENT FILE SHARING PREVENTION USING SECURE COMPUTING

Anandu Viswanath*1, Vasantha S*2

*1*Dept. of MCA, Sir M Visvesvaraya Institute of Technology, Bengaluru, India.*

*2*Dept. of MCA, Faculty of MCA, Sir M Visvesvaraya Institute of Technology, Bengaluru, India.*

ABSTRACT

Data security is very important. The vast majority of companies using desktop applications place more emphasis on the importance of data and how it relates to their products. Most security service systems need to be offline. If it is online, there is a possibility of getting hacked and tampered with. But today everything is done online. Therefore, we need to increase the security of files and data. Many companies create intra-applications to prevent data breaches. Therefore, this document proposes a secure file sharing system to prevent data leakage. Source code is one of the most important and sensitive data for IT companies. Source code breaches can lead to tampering with the products and services they offer, resulting in significant business losses. Therefore, this system aims to provide a secure in-file sharing system that can detect unauthorized sharing of sensitive information such as source code. The system uses Principal Component Analysis for this purpose

KEYWORDS: *File Sharing, Prevention, Principal Component Analysis, Pattern matching, Anomalies*

I. INTRODUCTION

Any organization's primary priority is security. When it comes to information technology, firms are highly concerned. Because most systems are online, there is a good likelihood that an attack to steal data will occur. When sensitive data is released, it can be exploited and interfered with, resulting in a loss of business and a reduction in the company's service quality. Sensitive information frequently needs to be shared with a third party. As a result, creating a secure sharing system is critical. An internal employee can also be a data leaker. As a result, it's critical to keep track of what information employees are sharing. Source code is one example of sensitive data for an IT firm. As a result, only authorized individuals should have access to it. In this paper the aim is to develop one such system which monitors what data being shared internally by an employee. The current approach has the restriction of not being able to determine the contents of the file or data that is being exchanged. The content of the file is detected in our system using **Principal Component Analysis**. The user will be blocked if the detection system detects the sharing of a code file. Even if the user tries to share the code by copying it to a text file, the system analyses the text file's contents and stops the procedure if it detects code. PCA, or principal component analysis, is a statistical process that allows you to summaries the information content of big data tables using a smaller collection of "summary indices" that can be viewed and examined more readily. Measurements defining attributes of production samples, chemical compounds or reactions, process time points in a continuous process, batches from a batch process, biological people, or DOE-protocol trials can all be used as the underlying data. PCA can be used for following scenario.

- When the dimensions of the input features are large (e.g., a large number of variables), PCA can be employed.
- When analyzing data with **multi-collinearity** between features/variables, the PCA technique comes in handy.
- **Denoising** and data reduction

The data set is treated as a file with a high number of lines in our system. When there are fewer lines, it is possible to check for code on each line individually. When the number of lines increases, however, checking each line for code becomes complex and time consuming. As a result, we're utilizing PCA to condense a huge number of lines into a tiny sample containing all of the relevant data. Then it compared the sampling result with a predefined keywords which are present in another data set when a match is found. It is considered as an unauthorized file sharing is taking place. The result of PCA is then compared to a list of predetermined keywords found in another data set. If a match is identified, it is assumed that unlawful file sharing is occurring. The proposed system's accuracy is high and it is efficient than the existing system

II. METHODOLOGY

A. Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised learning approach used in machine learning to reduce dimensionality. With the use of orthogonal transformation, it translates observations of correlated features into a set of linearly uncorrelated data. The Principal Components are the new transformed features. PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.

The construction of relevant features is achieved by linearly transforming correlated variables into a smaller number of uncorrelated variables. This is done by projecting (dot product) the original data into the reduced PCA space using the eigenvectors of the covariance/correlation matrix aka the principal components (PCs).

B. Implementation of PCA

The technique divides the huge dataset into smaller groupings using PCA. A file can have a lot of lines. Checking each line takes time and reduces system efficiency. It is reduced to a smaller set using PCA while maintaining the relevant information in the dataset. The PCA's output is compared to the predetermined keywords. These are reserved programming language keywords. There is no way to write a language without the reserved terms. When these keywords appear in PCA's reduced data, it's assumed that a code file was attempted to be shared with an unauthorized person. This technology can be used to detect the source of a data leak, remove the source from the system, and assure data security.

III. MODELLING AND ANALYSIS

SYSTEM ARCHITECTURE

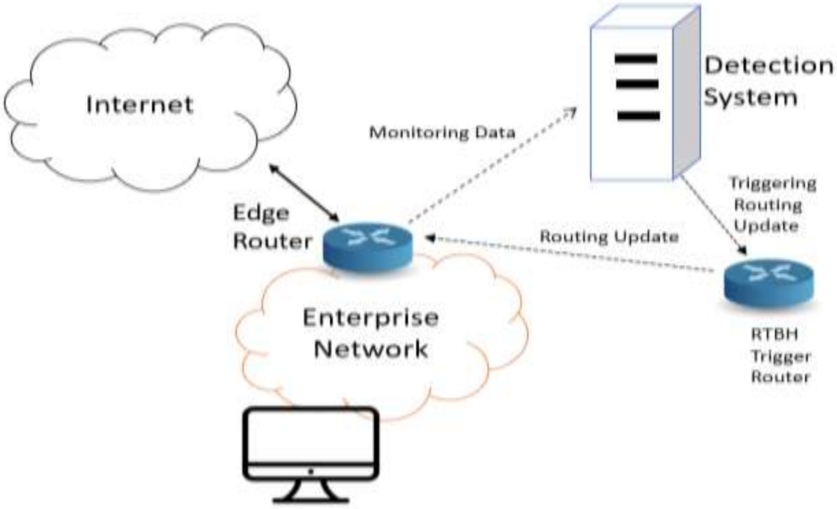


Fig 1: System Architecture

The detection and identification component of the above architecture is the system that monitors and detects any unlawful file sharing. When it discovers any, it informs the system's administrator, who may then ban the user's activity by blocking his IP, effectively prohibiting any kind of sharing from the system.

DATA FLOW

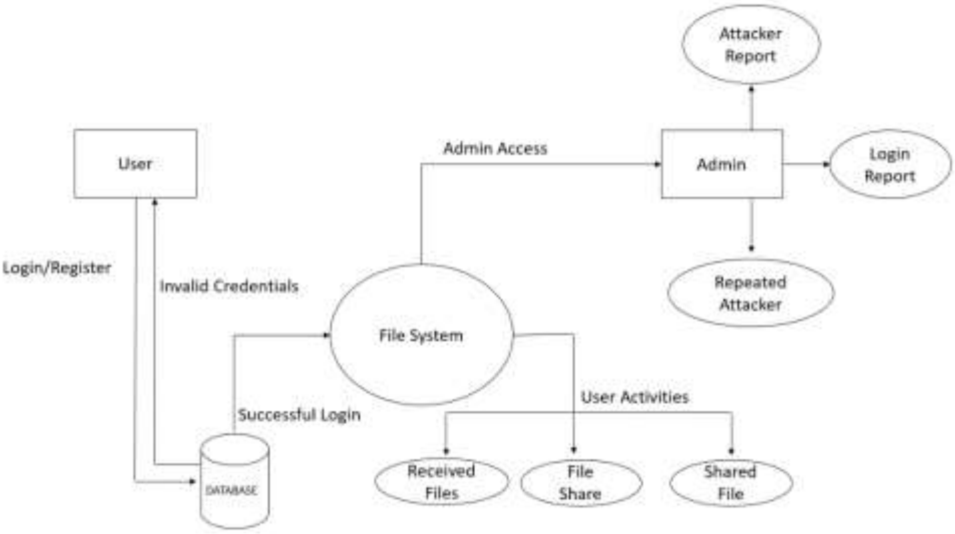


Fig 2. Data Flow Diagram

The proposed system has the following functionalities

To gain access to the system, the user must login with the proper credentials. After logging in successfully, the user can share files over the network with another user. When a file is shared, it passes through a detection mechanism that examines the contents for the existence of sensitive data. If any sensitive data is found the system reports to the administrator.

The administrator has access to information on the users who attempted to disclose sensitive information. The administrator can then ban his IP address to prevent any future sharing from that system.

IV. RESULT

A. User Registration Page

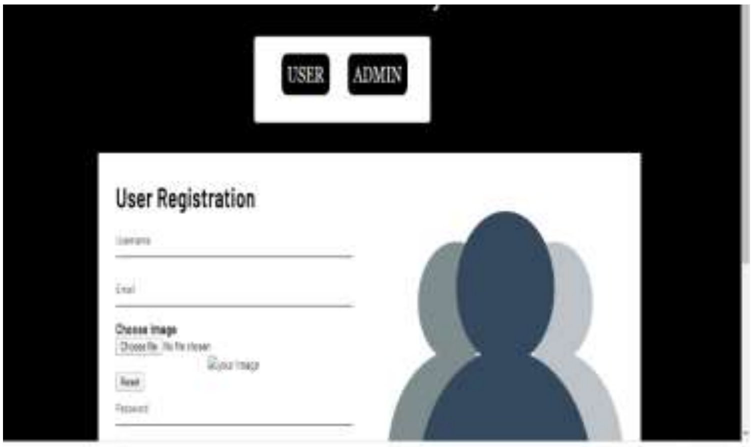


Fig 1: User Registration Page

This page displays the user registration page where the user can get registered themselves

B. User Dashboard

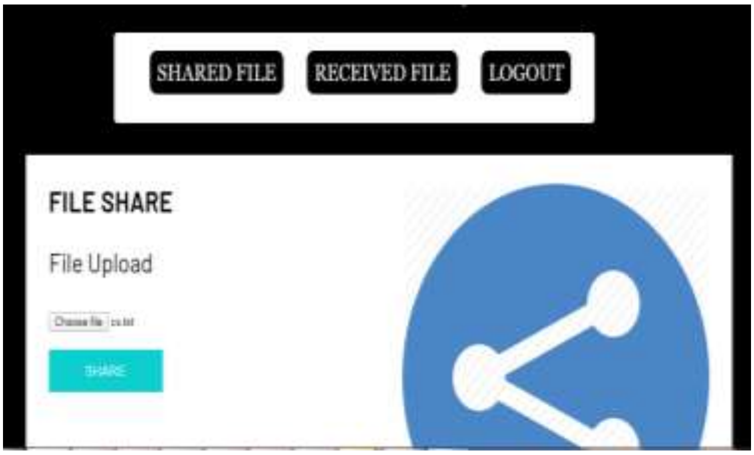
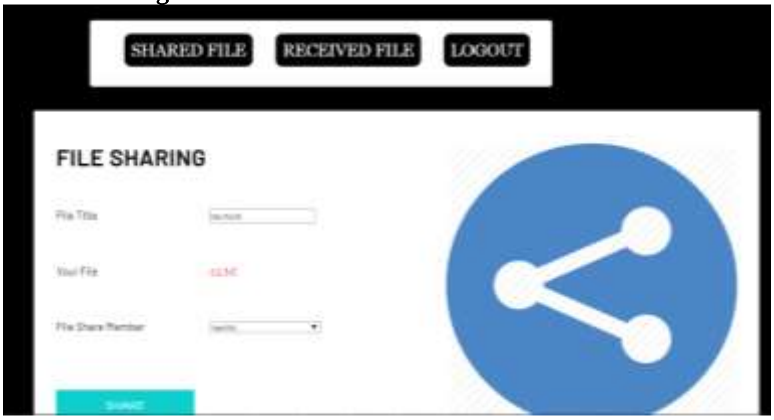


Fig 2. User Dashboard

After the successful login the user will be taken into the user dashboard where the user has the option to share a file to another user and view received files.

C. File Share Page



FILE SHARING

File Title [illegible]

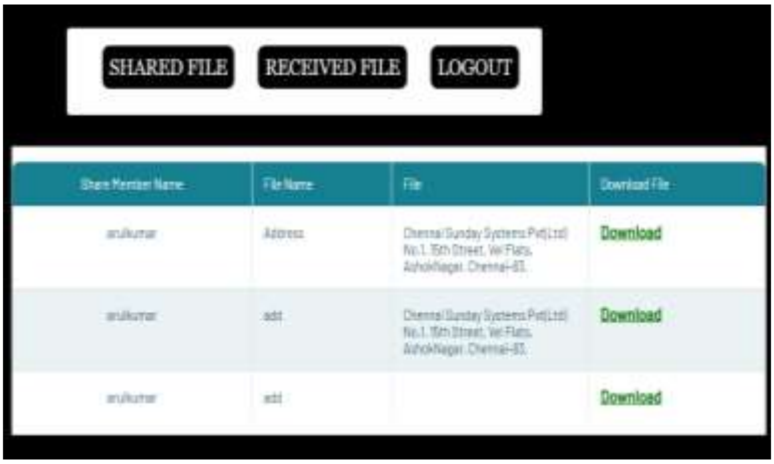
Your File [illegible]

File Share Member [illegible]

Fig 3. File Share

This is the page where the user can select the file to be shared and the recipient to whom the file is sharing.

D. Shared Files

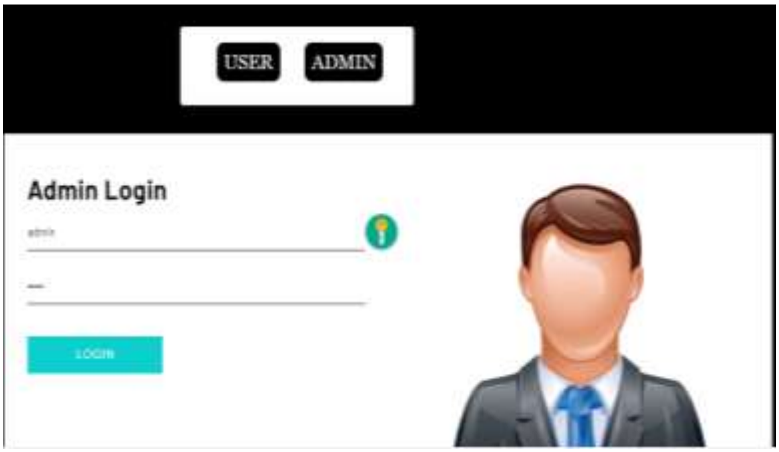


Share Member Name	File Name	File	Download File
arulkumar	Address	Chennai(Sunday Systems Pvt(Ltd) No.1, 15th Street, Vel Flats, AshokNagar, Chennai-83.	Download
arulkumar	adt	Chennai(Sunday Systems Pvt(Ltd) No.1, 15th Street, Vel Flats, AshokNagar, Chennai-83.	Download
arulkumar	adt		Download

Fig 4. List of shared files

This page displays the list of all shared files which are shared by the user. This page includes information such as recipient name file name, file description and option to redownload the shared file

E. Admin Login



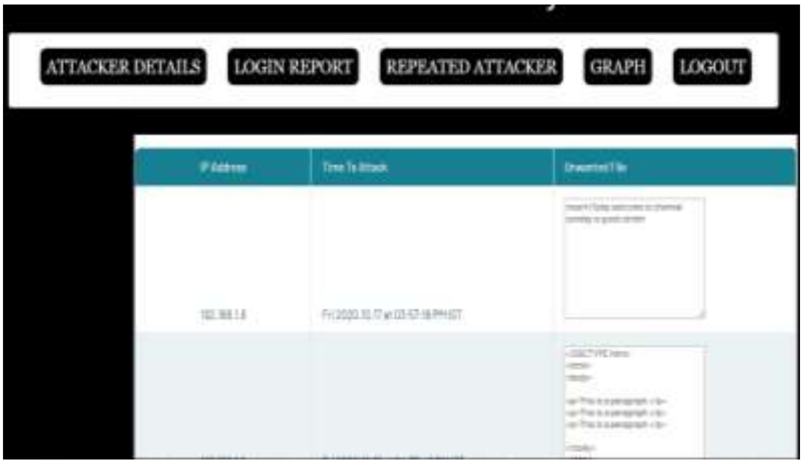
Admin Login

admin

Fig 5. Admin Login Page

This page contains the login page for admin where the admin can login into system using proper credentials

F. Admin Dashboard

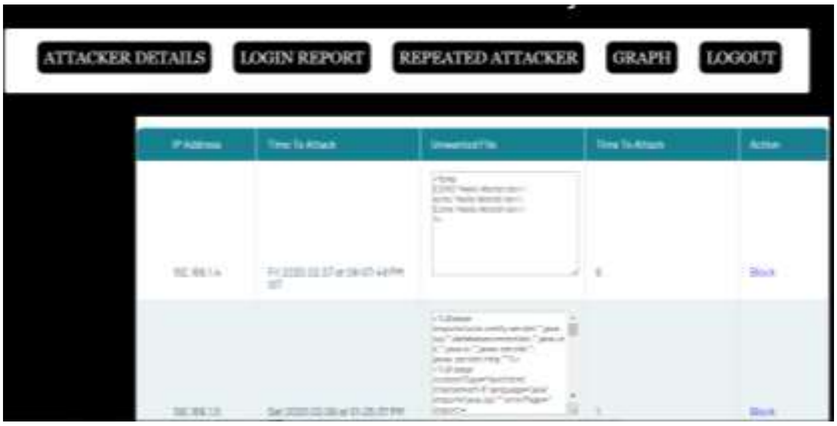


IP Address	Time To Attack	Unwanted File
192.168.1.6	Fri 2020.10.17 at 03:57:18 PM IST	[illegible]
[illegible]	[illegible]	<!DOCTYPE html> <html> <body> <p> This is a paragraph.</p> <p> This is a paragraph.</p> <p> This is a paragraph.</p> </body>

Fig 6. Admin Dashboard

The admin dashboard is displayed on this page. This dashboard has some options. The information regarding the IP address from which a fraudulent file sharing attempt was made is displayed in the attacker details section.

G. Block User

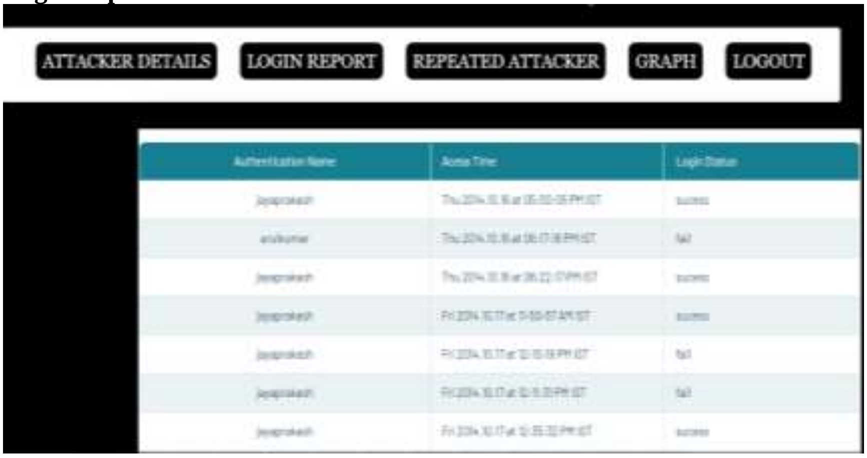


IP Address	Time To Attack	Unwanted File	Time To Attack	Action
192.168.1.4	Fri 2020.10.17 at 04:07:44 PM IST	[illegible]	6	Block
192.168.1.5	Sat 2020.10.18 at [illegible] PM	[illegible]	1	Block

Fig 7. Option to block user

This page displays the details of the user who tried to share sensitive data multiple times and the admin has the option to block the user by blocking his/her IP address

H. Login Reports



Authentication Name	Access Time	Login Status
jayaprakash	Thu [illegible] PM IST	success
anilkumar	Thu [illegible] PM IST	fail
jayaprakash	Thu [illegible] PM IST	success
jayaprakash	Fri [illegible] AM IST	success
jayaprakash	Fri [illegible] PM IST	fail
jayaprakash	Fri [illegible] PM IST	fail
jayaprakash	Fri [illegible] PM IST	success

Fig 7. Displays list of users logged in

This page displays the list of users who logged into the system

CONCLUSION

This paper proposed a mechanism for preventing the exchange of files containing sensitive information. This technology assists in identifying the source of data leakage and removing the source from the system, so preventing any sensitive information from being released or tampered with. Implementing this project in large scale can be beneficial for the organization as it provides a convenient way to secure data from being leaked. Implementing Principal Component Analysis made it easier to reduce the size of large data sets and allowed for faster and more accurate file recognition, as well as a reduction in overall process time.

REFERENCES

1. *G. Xie et al., "Fast low-rank matrix approximation with locality sensitive hashing for quick anomaly detection," in Proc. IEEE INFOCOM, May 2017, pp. 1–9.*
2. *K. Xie et al., "Fast tensor factorization for accurate Internet anomaly detection," IEEE/ACM Trans. Netw., vol. 25, no. 6, pp. 3794–3807, Dec. 2017,*
3. *H. Huang, H. Al-Azzawi, and H. Brani. (Feb. 2014). "Network traffic anomaly detection." [Online]. Available: https://arxiv.org/abs/1402.0856*
4. *D. Jiang, Z. Xu, P. Zhang, and T. Zhu, "A transform domain-based anomaly detection approach to network-wide traffic," J. Netw. Comput. Appl., vol. 40, no. 2, pp. 292–306, Apr. 2014.*
5. *I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.*
6. *A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide trafficanomalies," in Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM), vol. 34. Oct. 2004, pp. 219–230.*
7. *H. Hotelling, "Analysis of a complex of statistical variables into principal components," J. Educ. Psychol., vol. 24, no. 6, p. 417, 1933.*
8. *J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131-137, Jan. 2004.*
9. *L. Huang et al., "In-network PCA and anomaly detection," in Proc. Adv. Neural Inf. Process. Syst., 2006, pp. 617–624*