SECUMAIL
[Secure Email System]

Shrihari Ahire
Department of Computer Engineering
AISSMS College of Engineering
Pune, Maharashtra, India
shreehari.ahire@gmail.com

Vishakha Panjabi
Department of Computer Engineering
AISSMS College of Engineering
Pune, Maharashtra, India
vishakha4punjabi@gmail.com

Rahul Jagtap
Department of Computer Engineering
AISSMS College of Engineering
Pune, Maharashtra, India
rahul728482@gmail.com

Madhuri Bagul
Department of Computer Engineering
AISSMS College of Engineering
Pune, Maharashtra, India
bagul.madhuri1@gmail.com

A.S.Deokar
Department of Computer Engineering
AISSMS College of Engineering
Pune, Maharashtra, India
deokar.anu@gmail.com

Abstract— Data Leakage is situation where data is disclosed to personnel who are unauthorised to access that data. It can be done either intentionally or unintentional. Most organizations have to face the problem of data leakage and thus the security practitioners have always deal with data leakage issues that arise from various ways like email and other mediums. Hence, there is a need to filter each and every e-mail to prevent data leakage and prevent the sensitive data by getting disclosed to unauthorised person. This can be done by using an intelligent system "SECUMAIL" which will filter email for organization's sensitive data. Principle used in e- mail filtering is we filter e-mail based on the contents and attachments of email, white list consisting of hash of data which can be attached and black list of words which should not be send through message body is maintained. The hash of email attachments is computed, the hash and contents of email is compared with the white list data and black listed words depending on the match the email is either blocked or forwarded.

Index Terms— Email filtering, Data leak, Data Hash, White list, Word list, Email continuity, Admin, Client.

1 INTRODUCTION

A data leakage may be intentional or unintentional process that explores secure information to an unauthorized environment. A data leak is a security aspect in which sensitive and secured data is copied and transported by authorized person for illegal inadvertent use. Despite using security policies and tools, employees around the world are engaged in risky behaviors that put corporate and sensitive data at risk. Email is the best medium through which large amount of data can be leaked quickly and easily. Considering this threats of data leakage through email, an organization blocks emails of the employees or restrict the employees to access the email hosting sites through the use of firewall.Securmail is an intelligent system which allows the employees to compose emails and forward them by filtering the email contents for organization's sensitive data. The type of data being leaked through e-mail can be in video format, textual format, audio format, graphical format, zip folders or files etc. In the proposed system, the owner of the data is the administrator of organization and the agents are the employees of the organization. The aim of the system is to provide email continuity to employees at the same time prevent the leakage of organization's sensitive data through email.

2 PROPOSED WORK

The aim of the proposed system is to provide email continuity to employees along with the filtering capability to prevent sensitive, secured or confidential data by getting disclosed to unauthorized environment. "SECUMAIL" covers the drawbacks of the existing mechanisms such as use of firewall which restricts access to email hosting sites and other various policies which organizations implement to prevent data leakage. The following diagram depicts the architecture of the proposed system.

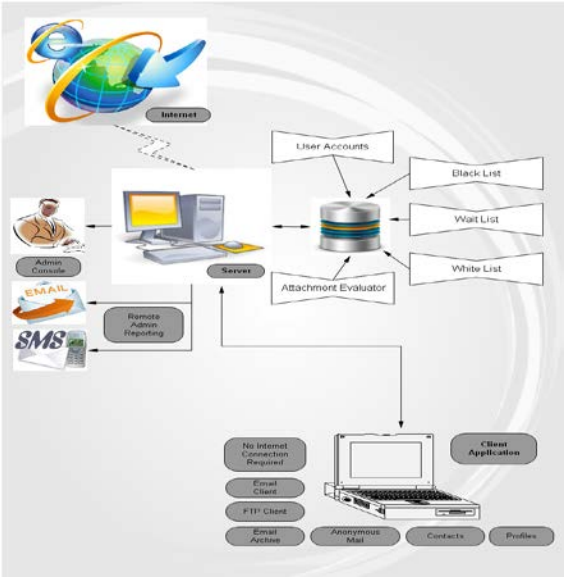


Fig. 1 Proposed System Architecture

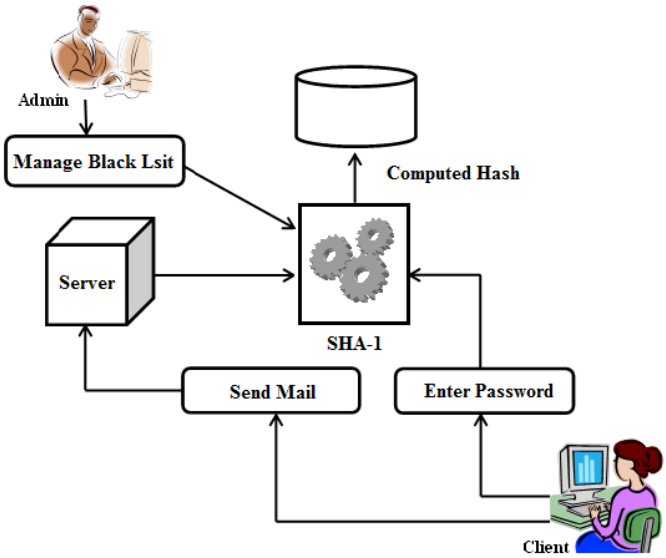


Fig 2: SHA-1 Working

T The proposed system is based on client- server architecture. Employees of the organization are the clients and administrator is at the server side. Only admin is provided with internet connection and email sending at client side is achieved with the help of servlet mechanism which transfers the mail to the server and further it is sent to the intended destination after filtering process. The admin consist of white list, black word list, sent mail list, block mail list, active mail list and white list of employee ids. The client consists of mail inbox, account management, change password, compose mail and manage contacts options.

3 METEHEDOLOGIES USED

A] SECURE HASH ALGORITHM [SHA-1]

SHA1 (Secure Hashing Algorithm) is a standard algorithm that produces 160bit digest of any size of file or data. The other algorithm similar to SHA1 is MD5.Both are hashing algorithms. The SHA1 returns a 160 byte hash whereas MD5 returns a 32 byte hash. The security of the MD5 hash function can be compromised. Also MD5 is not suitable for long data whereas SHA1 is appropriate for large data. So considering these limitations of MD5, SHA1 is used in the proposed system both at the client and server side to obtain hash of data. At client side the hash of the password is obtained using SHA1.At server side SHA1 is used to obtain the hash of the sensitive data and it is stored as a black listed data. The following architecture depicts the use of SHA-1 in the proposed system.

B B] TERM FREQUENCY

Term frequency is a weighting scheme that refers to the assignment of weight to each term in the document that depends on the number of occurrences of the term in that document.

The term frequency is denoted as ({tf}_{t,d}) with the subscripts denoting the term(t) and document(d),based on the weight of term(t) in document(d).

Equation for Term Frequency (tf) is given as:

$W_t = c_t \log (N/f_t)$

Where W_t is the weight of term f_t is the number of times the term in the mail, c_t is number of times the term in the passage, N is the total number of terms in the mail.

In our proposed system, we are using term frequency algorithm in which the data in the mail will be checked with word list. Threshold value for each word will be stored in the word list. If the occurrence of any word crosses the threshold value, the mail will be blocked.

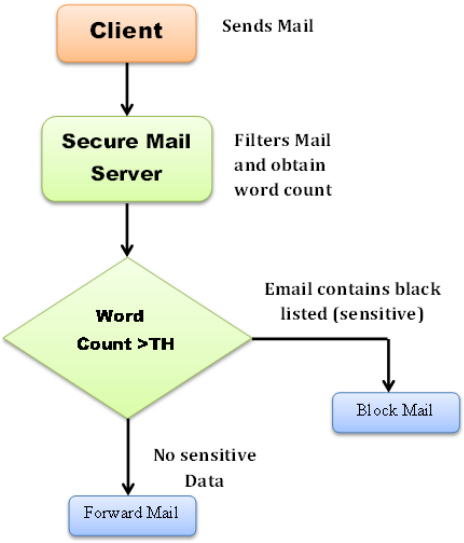
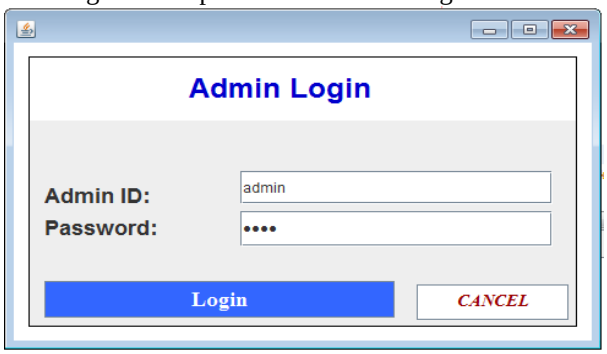


Fig 3: Term Frequency Working

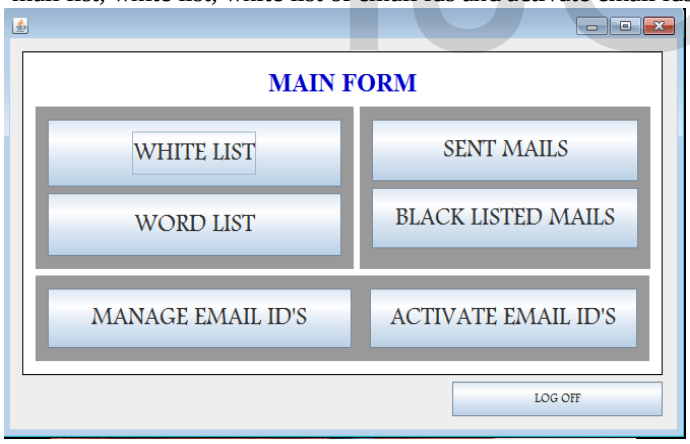
4 IMPLEMENTATION

I. Admin module:-

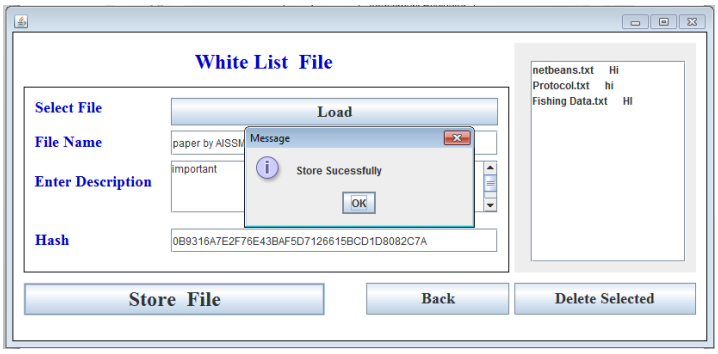
1. Admin login id and password secured using SHA-1.



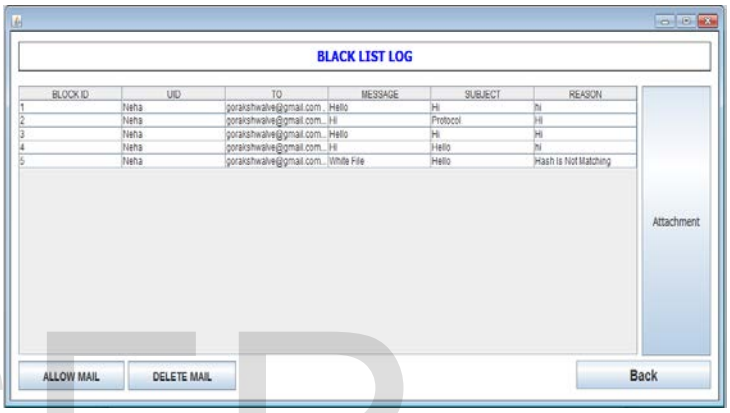
2. The admin manages black list of words, block list of mails, sent mail list, white list, white list of email ids and activate email ids.



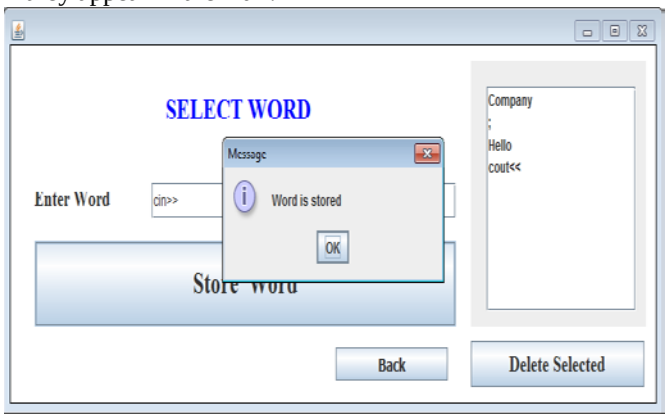
3. The white list contains hash of data which is only allowed in the mail as attachments.



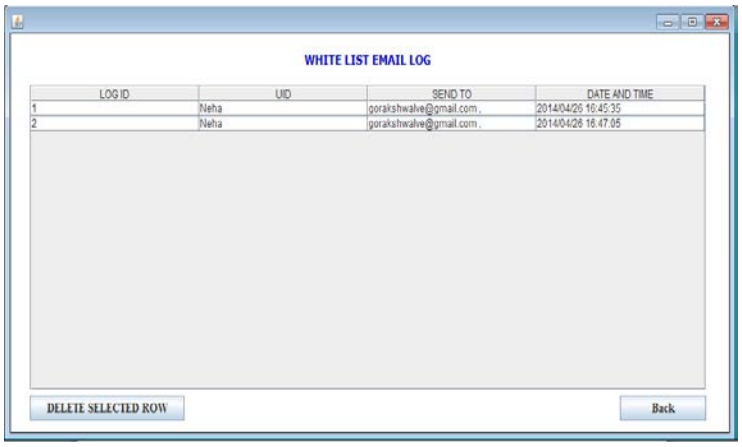
4. The block mail list contains mails which are blocked by the filtering mechanism and admin can allow these mails if he wants under some circumstances.



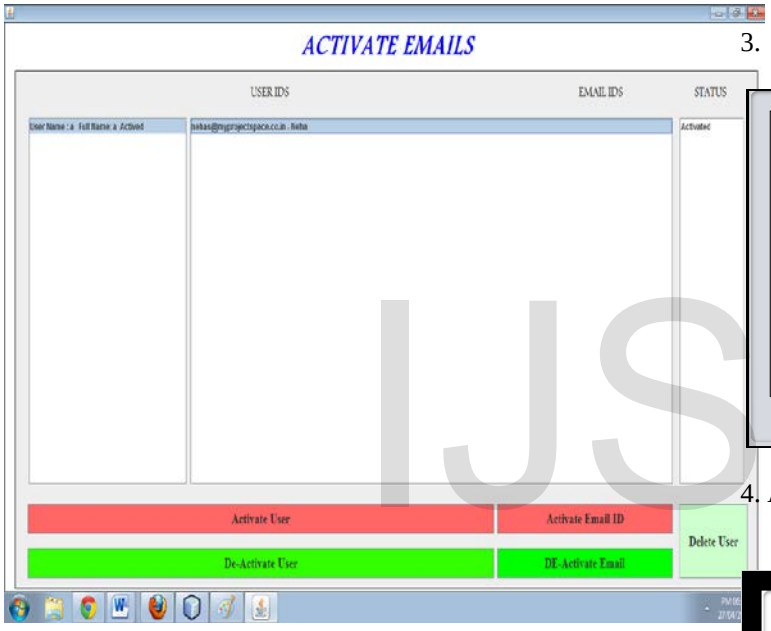
5. 5. The black listed word list contains words which will be blocked if they appear in the mail.

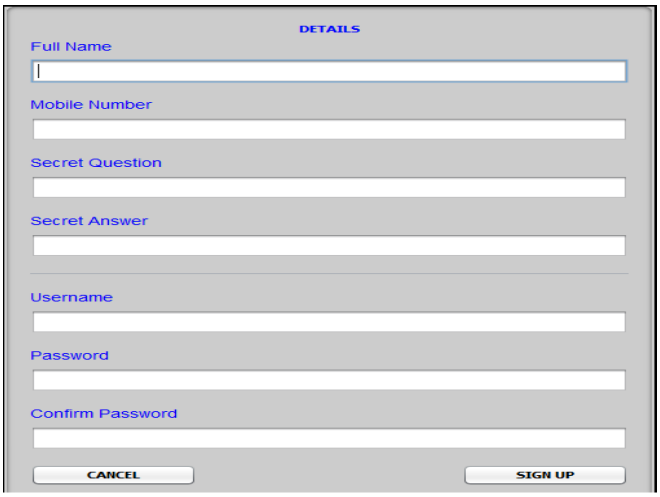


6. 6. The admin can analyse the sent mails from the sent mail list which will contain all the mails which has been sent.

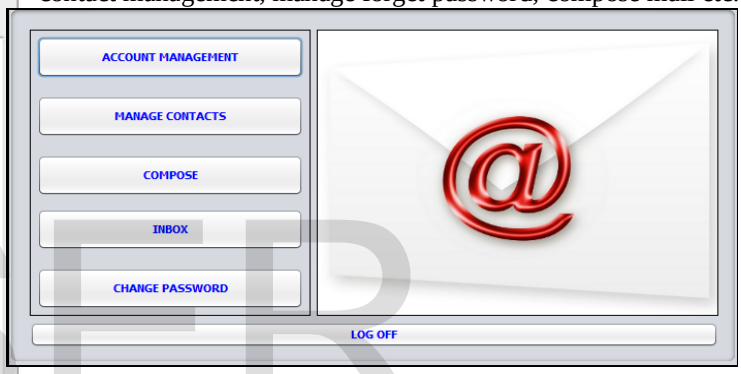


7. 7. Admin activates or deactivates individual user and his email ids.

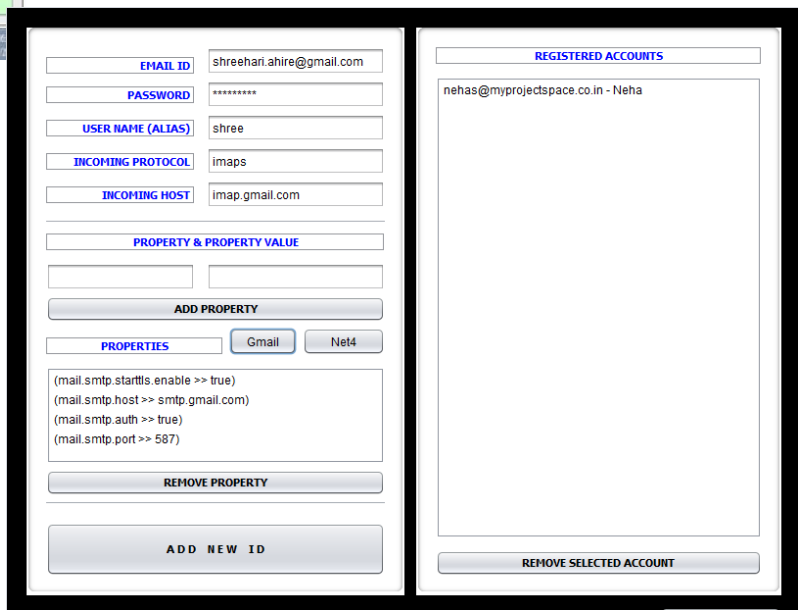




3. Client provided with different options such as manage accounts, contact management, manage forget password, compose mail etc.



4. Account management provides option to the client to register their email ids through which they can send mail.

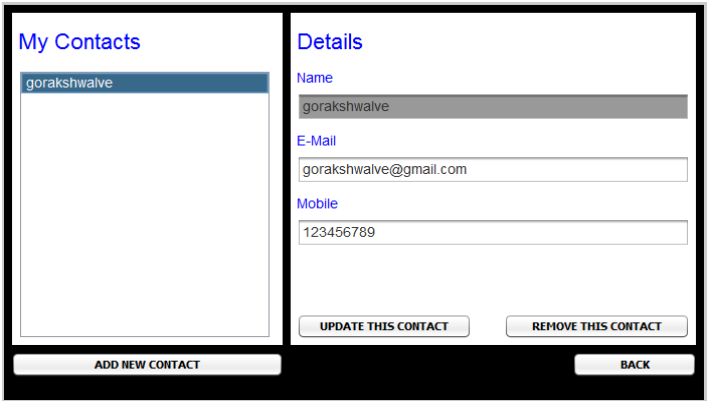


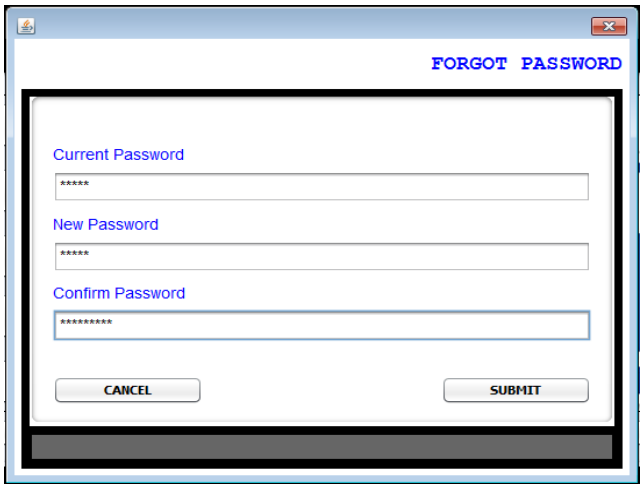
II. Client **module:-**

1. Client login id and password secured using SHA-1.

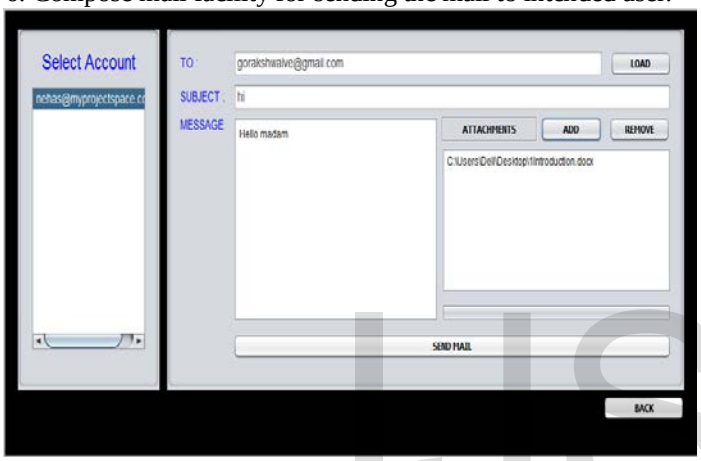
2. Client registration form to create id and password.

5.5. Contact management to store contacts and their email ids.





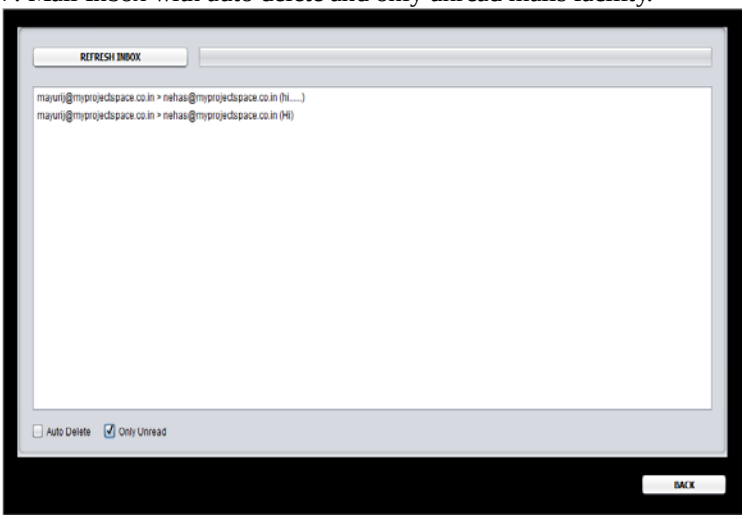
6. 6. Compose mail facility for sending the mail to intended user.



III. **Server and filtering mechanism:-**

When client sends mail, it will come to the server. The server will filter the mail in two phases. In the first phase it will calculate the hash of the mail attachment and compare it with the white list. If the match is successful than only the mail will be forwarded else it will be blocked and added in the block mail list. In the second phase if the mail is not blocked means if the attachments is a white listed data than the server will check the contents of mail and matches it with black listed word list and if the count of particular word crosses the threshold value than the mail is blocked and added to block mail list else the mail will be forwarded and added to the sent mail list. There is one condition in which there will be no filtering on the mail at that is the id sending the mail is a white list mail id.

7. Mail Inbox with auto delete and only unread mails facility.



8. Change and forgot password facility.

.

4 CONCLUSION

Many organizations handle confidential data on daily basis. The technologies that make this data easily available also increase the risk of data leakage. Some mechanisms have been implemented to prevent data leakage such as firewall mechanism which restricts access to email sites which hampers email continuity. Considering these the limitations, we have develop a system which will provide email continuity along with filtering capability for sensitive data leakage without any internet connection at client side. The algorithmic strategy used provides email filtering for any size and type of data.

.

ACKNOWLEDGMENT

We would like to thank all the professors of Computer Engineering Department of AISSMS College Of Engineering, Pune-01.We are indebted to Prof. Mrs. A.S.Deokar our project guide who was very generous in providing us with technical-support,

material and otherwise. Her invaluable suggestion and time have helped in making this project possible.

REFERENCES

[1] Ankit Agarwal, Mayur Gaikwad, Department of Information Technology, University of Pune, Kapil Garg, Vahid Inamdar, Department of Computer science, university of pune, *Robust Data Leakage and Email Filtering System* International Conference on computing Electronics and Electrical Technologies [ICCEET], 2012.

[2] Saadat Nazirova, Institute of Information technology of Azerbaijan National Academy of sciences 9,F.Agayevstreet,Baku,Azerbaijan,*Survey on spam Filtering techniques Communication and Network*, 2011, 3,153-160 doi:10.4236/cn.2011.33019 published Online August 2011(http://www.SciRP.org/ Journal/cn).

[3] Christina V M.Phil Reasearch Scholar P.S.G.R Krishnammal College For Women, Karpagavalli S Senior Lecturer GR Govindarajulu school of Applied Computer Technology, Suganya G M Phill Research Scholar P.S.G.R Krishnammal College for women, *A Study on Email Spam Filtering Techniques*, International Journal Of computer Application (0975-8887) Dec 2010.

[4] Herbertschildt, *Java Complete Reference*, 7, Osbome, 2011.