Volume 4, Issue 1, January 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Secure Email System for SOHO (Small Office Home Office)

Shrihari Ahire, Vishakha Panjabi, Rahul Jagtap, Madhuri Bagul, A.S.Deokar

Department of Computer Engineering

AISSMS College of Engineering

Pune, Maharashtra, India

Abstract— Data Leakage is scenario where the data is disclosed to unauthorized personnel either by malicious intent or inadvertent mistake. Most of the organizations face the problem of data leakage. Security practitioners have always had to deal with data leakage issues that arise from various ways like email and other internet channels. Hence, there is a need to filter these e-mails. This can be done by using an intelligent system which will filter email for organization's sensitive data. Principle used in e- mail filtering is we classify e-mail based on the message bodies, black lists consisting of hash of organization's data are maintained. The hash of email contents is computed and compared with the black list data and depending on the match the email is either blocked or forwarded.

Keywords— Email filtering, Data leak, Data Hash, Black list, Word list, Email continuity.

I. INTRODUCTION

A data leak may be intentional or unintentional process that releases secure information to an illegal environment. A data leak is a security incident in which sensitive, secured or confidential data is copied, transported, seen, moved secretly or used by an unauthorized person to do so. Despite the security policies, procedures, and tools currently in place, employees around the world are engaging in risky behaviors that put corporate and personal data at risk. Email is the medium through which sensitive data of an organization can be leaked. Considering the 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.Securemail 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 and the trusted agents are the employees of the organization. The aim of the system is to prevent the leakage of organization's sensitive data through email.

II. RELATED WORK

Presently most of the organizations use firewall to prevent data leakage through email by configuring the firewall to restrict access to email hosting sites like Gmail, Yahoo etc. So the employees of the organization cannot send email from within an organization. The problem with firewall approach is that it cannot operate on individual files and email, as firewall only works on packet level. Due to this limitation firewall cannot block emails by filtering their contents. Another issue with firewall is that SOHO (Small Office Home Office) has limited resources, so the firewall product they implement must be relatively easy to use and maintain and cost-effective. A system previously developed to prevent data leakage through emails used the concept of fake object [1]. Fake objects are added to the databases which appear exactly as the original data. No one is aware of the fake objects. The algorithm used for filtering the emails is the K-nearest algorithm. The problem with fake objects concept is that it is only suitable for the fixed data. It is not appropriate when the data is variable. Organizations used Microsoft Outlook express to handle their emails. Outlook provides a heterogeneous environment but outlook has no provision for email filtering. Organizations mostly use firewall mechanism along with outlook for security mechanism. Also while using outlook internet connection is needed for every node in the network. Considering the limitations of firewall which restricts the email continuity which is very important part of organization's work, the problem of internet connection for every node in the network and the cost effectiveness the proposed system will cover these limitations and will be an asset for SOHO.

III. PROPOSED SYSTEM

Due to number of drawbacks of existing systems it is not possible to check the contents of email and also the employees of the organization are restricted from accessing the email sites and send email from within an organization which affects email continuity. Firewall implementation is not suitable for SOHO (Small Office Home Office) and not cost efficient. The fake object mechanism is suitable only for fixed data and requires internet connection for every node in the network .The aim of the system is to cover the above limitations and provide email continuity but at the same time preventing the leakage of organization's sensitive data through email. The following diagram depicts the architecture of the proposed system.

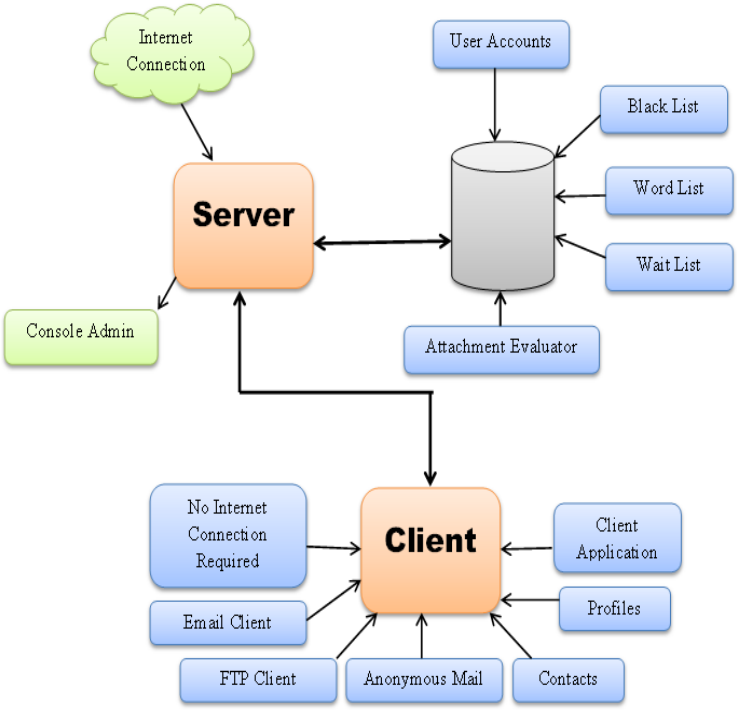

Fig. 3 Proposed System Architecture

The proposed system is based on client- server architecture. Employees of the organization are at the client side and administrator is at the server side. Internet connection is provided only at server side and clients communicate with server through servlet mechanism. The server consists of black list, waitlist, and word list and attachment evaluator. The black list contains hash of organization's sensitive data which is maintained by the administrator. The admin will allow or block the mails from the wait list. When client sends mail, it first comes to the server. The server will calculate hash of the mail's data and attachment which will be compared with the black list data and depending on the match, the email will be either forwarded or blocked. If the server fails to detect whether the email contents are black listed or not, email will be sent to the wait list. The system will classify e-mail based on the message bodies, the white and black lists consisting of hash of organization's data are maintained. The hash of email contents is computed and compared with the white and black list data and depending on the match the email is either blocked or forwarded. The following diagram shows how and when the mail will be blocked.

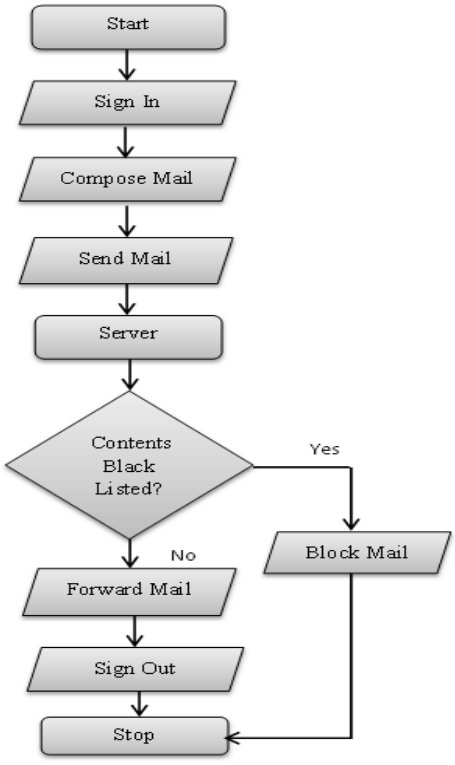

Fig 2: Flow Diagram

IV. METHODOLOGIES 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.

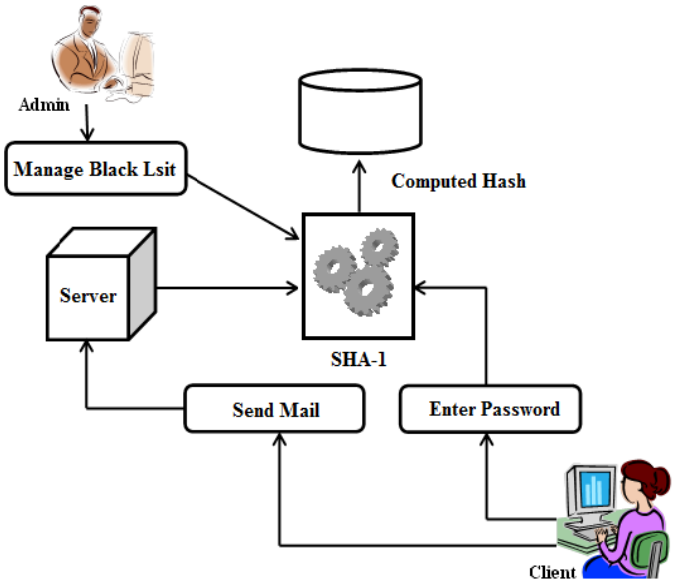

Fig 3: SHA-1 Working

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.

Advantage:

As the term frequency algorithm does not require information regarding the structure or grammar of the natural language. Therefore the algorithm may be used in many natural languages.

V. CONCLUSIONS

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, filtering email using fake object mechanism which is only suitable for fixed database. Considering these the limitations, we have shown that it is possible to 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.

