Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Email Correlation And Analyzer Using Pattern Matching Technique

Prof K.S. Wagh[1], Shraddha Bhagwat[2], Ashwini Kawade[3], Pratima Khambad[4], Shrirang Potale[5]

[1]*Department of Computer Engineering, AISSMS Institute of Information technology, Pune,*
[2]*Department of Computer Engineering, AISSMS Institute of Information technology, Pune,*
[3]*Department of Computer Engineering, AISSMS Institute of Information technology, Pune,*
[4]*Department of Computer Engineering, AISSMS Institute of Information technology, Pune,*
[5]*Department of Computer Engineering, AISSMS Institute of Information technology, Pune.*

Abstract —*.Email or Electronic mail is been widely used to exchange messages between people. Electronic mail or an Email can be personal, commercial, business related messages. These messages can be used to determine association between mail holders. A report or summary can be made by retrieving mails and grouping them to get a count of total mails that has been exchanged from a particular mail id. This will help to find out an association between the two contacts and reduces the burden of reading and going through each and every mail which is being send or received. The generated report is useful in means of cyber forensics and in means of data mining, where emails are studied to find out a person, company or an organization with whom the user has been communicating using emails and how many times they have exchanged emails. Proposed system generates a report on number of mail's being exchanged with different user and provides a graphical representation of the same. In addition, to the above feature, an advanced features is added in the system which helps to monitor the workers or group members and gets feedback on progress of work assigned to them by using emails*

Keywords; *Email, Cyber Forensic , Email Correlation, Data Mining , Pattern Matching, AES, Email Analyzer,KMP algorithm*

I. INTRODUCTION

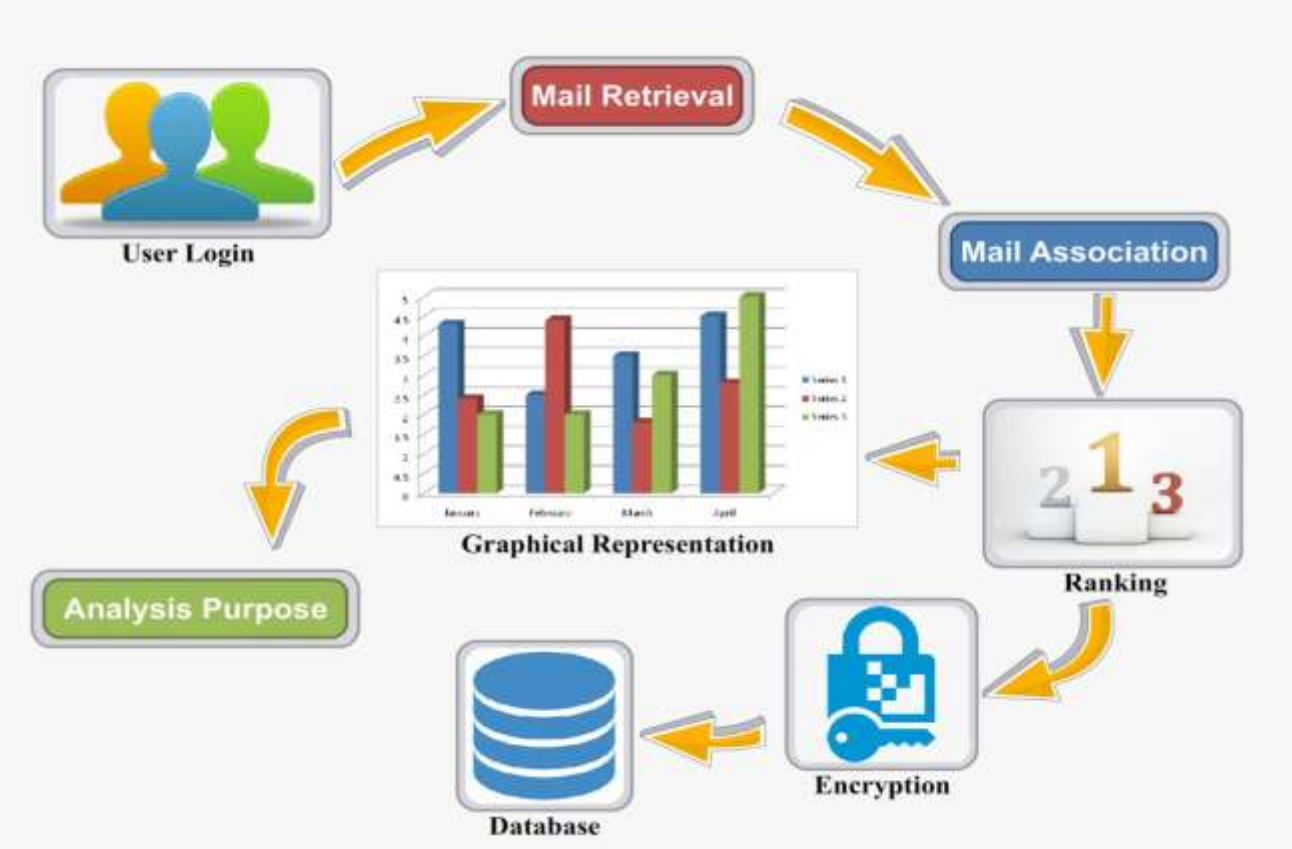
Email has become a common mean of communication. People widely use email to share personal official and even commercial messages. Over the period of time email has become a convenient way for exchanging messages and for communication. An average person has hundreds or thousands of mail's on his or her email account, the number of mail's increase exponentially if we talk about mail id of a company or organization. Situation may occur in which there maybe need of finding email contacts which have been communicating with an email id belonging to a person or even company. These instances can be from the frame of cyber forensics, in which email id of a convict or a victim is to be monitored or these instances can be frame of data mining of a company wanting to monitor email datasets of its own or other company.

Project Idea

The proposed system retrieves sender mail id, email subject, body and other information. Arranges them according to time and shows association between contacts which includes total number of mails sent or received from a particular mail address or user name. This association is shown in form of report as well as it is represented graphically for better documentation and understanding. The proposed system also has an extra feature of email client in which important mails can be monitored by the sender and receiver can update the progress of that mail to the sender without composing a new mail

II . RELATED WORK

This paper presents the foxmail client system which extracts the information from email evidence file by using MD5 hash algorithm [1]. Email forensic analysis is recently developed software by using machine learning techniques with social networking techniques which helps to detect email related crimes [2]. Further for email collection classification keyword search technique i.e. Knuth Morris Pratt (KMP) algorithm is used and forensic results are visualized using related layout techniques such as radial tree layout and spring force layout for better understandings [3]. For the confidentiality of an information, here encryption and decryption of the information is stated by using standardized algorithm called Advanced Encryption Standard (AES) [4]. E-mail communication and how an E-mail can become digital evidence for forensic investigators is given here. Statistical analysis is used for graphical representation. This will further be further used for forensic analysis. And for the classification decision tree classifier J48 is used]5].



III . PROPOSED ARCHITECTURE

The system proposed in this paper provides a summary or a report showing all email ids or username with which a particular email account has interacted. Report also shows number of mails exchanged with interacted mail accounts. Another important feature of proposed system is to make monitoring of assigned work easy. If a group leader sends an important mail to his group members and want the mail to be processed within a period of time, then this important mail will reside on the top of inbox until the time period assigned by the group in-charge expires. Group members can also mark percentage of task completed which would be visible to group in-charge

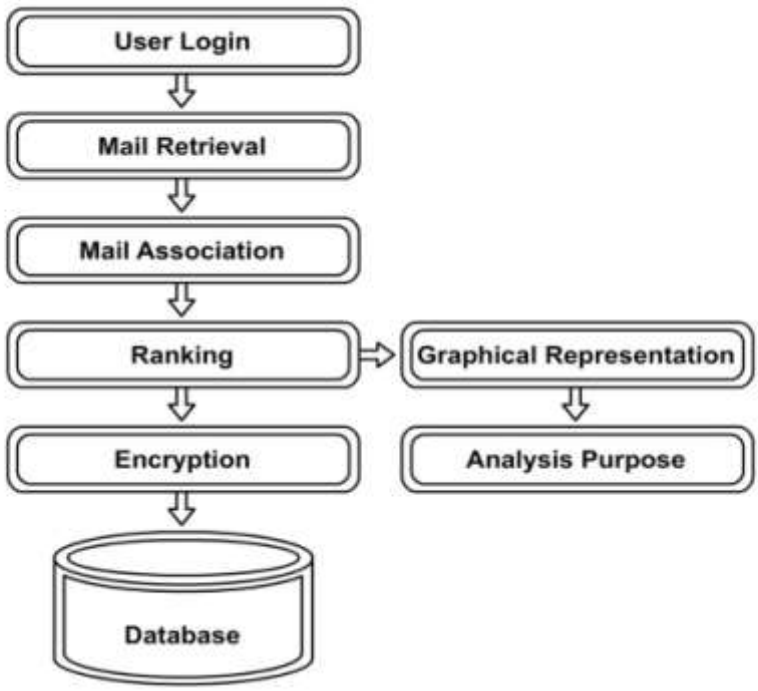


Fig. 2 Architecture View.

IV . ALGORITHM

```
algorithm KMP(P[1, . . . , m], T[1, . . . , n])
input: pattern P of length m and text T of length n
preconditions: 1 ≤ m ≤ n
output: list of all numbers s, such that P occurs with shift s in T
q ← 0;
i ← 0;
while (i < n) /* P[1, . . . , q] == T[i − q + 1, . . . , i]
{
if (P[q + 1] == T[i + 1])
{
q ← q + 1;
i ← i + 1;
if (q == m)
{
output i − q;
q ← π(q); /*slide the pattern to the right
}
}
else /* a mismatch occurred
{
if (q == 0) { i ← i + 1 }
else { q ← π(q) }
}
}
```

S = { I, O, F, S, Ff }
$I = \sum^{2}_{n=1} I_n$
I_1 = Username, Password
I_2 = Mails_extracted
Pre-requisite - live mail account
$O = \sum^{4}_{n=1} O_n$
O_1 = Mail Association
O_2 = Ranking
O_3 = Graphical Representation
O_4 = Encrypted Data

$F = \sum^{5}_{n=1} F_n$
F_1= Login, Logout
F_2 = Send mail, Receive mail, Add attachment
F_3 = Associate mails, Ranking
F_4 = Graphical Representation
F_5 = Encryption, Database storage, Decryption

$S = \sum^{5}_{n=1} S_n$
S_1 = Successful Login
S_2 = Mails are sent and received successfully
S_3 = Mails are associated properly
S_4 = Data is encrypted and stored
S_5 = Successful Logout

$Ff = \sum^{5}_{n=1} Ff_1$
Ff_1 = Unsuccessful Login
Ff_2 = Mails are not sent and received successfully
Ff_3 = Mails are not associated properly
Ff_4 = Data is not encrypted and stored
Ff_5 = Unsuccessful Logout

V . CONCLUSION AND FUTURE SCOPE

The proposed system would display the association between email contacts. This information is important by means of both cyber forensic and email data mining. the proposed system uses open sources software and can be installed on

computers with low configuration, which makes the system more feasible. Proposed system also provides a feature of email client which would help users to monitor important mails send to colleagues, and progress on these important mail can be seen to by the user once updated by his or her colleague which reduces the effort of composing a new mail for sending updates to user.

System can be made feasible and efficient if we are able to trace and count those emails which are deleted from server's mail account and making application of the proposed system which would not be limited to desktop and can be used in multiple machine any time any where can provide a bright future to cyber forensics and email data mining.

VI . REFERENCES

[1] Zhenya Chen, Ying Yang, Lijuan Chen, Liqiang Wen, Jizhi Wang, Guang Yang, Meng Guo," Email Visualization Correlation Analysis Forensics Research" 2017

[2] Rachid Hadjidj, Mourad Dcbbabi, Hakim Lounis, Farkhund Iqbal, et al, "Towards an integrated e-mail forensic analysis framework,"Digital Investigation ,vol. 5,Mar.2009, pp.124-137, doi:10.1016/j.diin.2009.01.004.

[3] Fanlin Meng, Shunxiang Wu, Junbin Yang, and Genzhen Yu, "Research of an E-mail Forensic and Analysis System Based on Visualization," 2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications(PACIIA 2009),Nov.2009,pp.281-284, doi:10.1109/PACIIA.2009.5406437.

[4] Daniel F. Garcia, "Performance Evaluation of Advanced Encryption Standard Algorithm", 2015 Department of Informatics, University of Oviedo Gijion, Spain.

[5] Sobiya R Khan, Smita M Nirkhi, R V Dharaskar, "E-mail Data Analysis for Application to Cyber Forensic Investigation Using Data Mining",2013, International Journal of Applied Information Systems(IJAIS).

[6] Fanlin Meng, Shunxiang Wu, Junbin Yang, Genzhen Yu, "Research of an E-mail Forensic and Analysis System Based on Visualization", 2009, Department of Automation Xiamen University, China

[7]Fajian XU, "Application of Associated Model in Police Information System" JOURNAL OF FUJIAN POLICE COLLEGE,2010,pp.25-28.

[8]Appavu, R. Rajaram, M. Muthupandian , G. Athiappan, and K.S.Kashmcera, "Data mining based intelligent analysis of threatening email,"Knowledge-Based Systems, vol. 22, July. 2009,pp. 392-393.

