ISSN 2320 – 2602
Volume 3, No.11, November 2014
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2014/ijacst043112014.pdf

E-Mail Fusion With Graph Based Approach Stipulating OPass Security

Shankar Kalwal[1], Kawaljeet Singh Bagga[2], Rameshwari Pohekar[3], Priyanka Morey[4]
[1]PGMozeCoE, Affiliated to Savitribai Phule Pune University, India, shankarkalwal@gmail.com
[2] PGMozeCoE, Affiliated to Savitribai Phule Pune University, India, kawaljeetsbagga@gmail.com
[3] PGMozeCoE, Affiliated to Savitribai Phule Pune University, India, rameshwaripohekar@gmail.com
[4] PGMozeCoE, Affiliated to Savitribai Phule Pune University, India, priyankamorey25@gmail.com

ABSTRACT

Most users are satiated with separately logging in into their multiple accounts, and hence end up spending a large amount of time in accessing them. It becomes a tedious task to remember all email id's and passwords for all their accounts. This paper presents a single framework in which we club all the user's accounts into one. It also presents a framework for multi-folder email classification using graph mining as the underlying technique. The user authentication protocol proposes the OPass enhancement to protect user identity. OPass only requires each participating website possesses a unique mobile number, and involves a telecommunication service provider(TSP) in registration and recovery phases for the creation of one-time password. OPass is efficient and inexpensive compared with the conventional web authentication mechanisms. Therefore OTP mechanism that has enhanced security using private key infrastructure is used to prevent integrity problem due to phishing attack and key-loggers.

Key words : Clustering, Email, Graph Based, OPass.

1. INTRODUCTION

Electronic mail is a fast, efficient, affordable and one of the most preferred way of communication among a large group of people[1]. Nowadays there are a number of Email services which are provided by many firms such as Google, Yahoo, Rediffmail, Hotmail, etc. Every firm has its own authentication process for logging in into their system. User has to go through this process in order to use the Email services. Furthermore, maintaining these services and remembering all their id's and passwords is a hectic task, For example, if a user wants to search a mail but is unable to recollect that on which domain he received that mail, then he will have to access all his Email accounts which may consume a lot of time in authentication processes. To ease this difficult task, this paper presents a single framework in which all the user accounts are clubbed together which can be accessed through a single login.

Some users prefer different login id's and passwords on their different accounts for security reasons, but memorizing so many id's and passwords can become a challenging job for any user. That is why, most users prefer using same id's and passwords for their multiple accounts in order to eliminate the need of memorizing their multiple id's and passwords. If the user tends to use the same password across different websites then possibilities of getting hacked are more. This paper presents Opass mechanism to overcome this problem. To use Opass mechanism, user must have unique phone number and involve telecommunication service provider for registration and recovery phase of one time password. Opass is efficient and affordable compared with the conventional web authentication mechanisms[2]. Opass generates a 6-digit number.

2. BACKGROUND

2.1 Database

The database used in this system is MySQL. The database accepts the user details which he inputs at the time of registration phase. The credentials(ID, Password) of all the users are stored in encrypted form in the database. At the time of login phase, these credentials are compared with the login credentials entered by the user by the database.

2.2 One Time Password

The one-time passwords in OPass are generated by a secure one-way hash function(H). With a given input, the set of one-time passwords is established by a hash chain through multiple hashing. Assuming we wish to prepare one-time N passwords[3], the first of these passwords is produced by performing hashes on input c [2].

$$c = \mathcal{H}(P_u||ID_s||\phi)$$

2.3 SMS Channel

SMS is a text-based communication service of telecommunication systems. SMS channel is used by Opass to provide a secure authentication protocol for the user against password stealing attacks. SMS represents the most successful data transmission of telecom systems; hence, it is the most widely used mobile service across the globe. We chose SMS channel because of its security benefits. The SMS network is a closed platform as compared with TCP/IP network. hence, it raises the difficulty of internal attacks, e.g., manipulating and tampering attacks. Therefore, Short Messaging Service(SMS) is an out-of-band channel that

protects the exchange of messages between users and servers[3].

3. LITERATURE SURVEY

In the existing system, if the user has to access mails from his multiple accounts, he has to separately log in into all those accounts and perform his respective task which consumes a lot of his time. Also, it becomes vital for the user to memorize all his user id's and passwords for all his accounts.

Concepts used in this system are elaborated below:

3.1 Advanced Encryption Standard (AES)

All of the cryptographic algorithms we have looked at so far have some problem. The earlier ciphers can easily be broken with ease on modern computation systems. The DES algorithm was broken in 1998. It was far too slow in software as it was developed for mid-1970's hardware and does not produce efficient software code. Triple DES on the other hand, has three times as many rounds as DES and is correspondingly slower. As well as this, the size of 64 bit block of triple DES and DES is not very efficient and is questionable when it comes to security.

A brand new encryption algorithm was required which would be resistant to all known attacks. Like DES, AES is a symmetric block cipher. This means that the same key for both encryption and decryption. However, AES is quite different from DES in a number of ways. The Rijndael algorithm allows for a variety of key and block and key sizes and not just the 64 and 56 bits of DES' block and key size. The key and block not to be chosen dependently from 128, 160, 192, 224, 256 bits and need not be the same. However, the Advance Encryption Standard states that the algorithm can only accept a block size of 128 bits and a choice of three keys - 128, 192, 256 bits. The standard name is modified to AES-128, AES-192 or AES-256 respectively, depending on which version is used,. As well as these AES differences differs from DES in that it is not a feistel structure and half of the data block is used to modify the other half of the data block and then the halves are swapped. In this case, processing of the entire data block is in parallel during each round using substitutions and permutations.

A number of AES parameters depend on the key length. For example, if 128 key size is used then the number of rounds is 10 whereas it is 12 and 14 for 192 and 256 bits respectively. At present 128 bit key is the most common key size likely used. This description of the AES algorithm therefore describes this particular implementation.

To fulfill following characteristics Rijndael algorithm was designed:

- Resistance against all known attacks.
- Speed and Code optimization on a wide range of platforms.
- Design Simplicity.

3.2 Opass

One form of attack on networked computing systems is eavesdropping on network connections to obtain authentication information such as the login IDs and passwords of legitimate users. This information, after capturing, can be used at a later time to gain access to the system. One-time password systems are depicted to counter this type of attack, called a "replay attack".

To generate a sequence of one-time passwords, the authentication system described in this document uses a secret pass-phrase. With this system, the secret pass-phrase of the user never needs to cross the network at any time such as during authentication or during pass-phrase changes. Thus, it is not exposed to replay attacks. The property provides added security that no secret information need to be stored on any system also including the server being protected. The One Time Password system protects against external passive attacks against the authentication subsystem. It does not prevent a network hacker from gaining access to private information and does not provide protection against either "social engineering" or active attacks.

There are two entities in the operation of the one-time password system. The generator must outcome the appropriate one-time password from the user's secret pass-phrase and from information provided in the challenge from the server. The server must send a challenge that consist of the appropriate generation parameters to the generator, must verify the OTP received, must store the last valid OTP it received, and must store the corresponding OTP sequence number. The server must also provide the changing of the user's secret pass-phrase in a secure manner. The OTP system generator sends the user's secret pass-phrase, including a seed received from the server as part of the challenge, through multiple repetitions of a secure hash function to produce a OTP. After each successful authentication, the number of secure hash function repetitions is reduced by one and also a unique sequence of passwords is generated. The server verifies the OTP received from the generator by computing the secure hash function once and comparing the result with the previously accepted one-time password. Leslie Lamport first suggested this technique.

3.3 Clustering

A graph representation becomes a natural choice to represent complex relationships as the data visualization process is relatively simple as compared to a transactional representation. Representation of data in the form of graph saves the structural information of the data which can be lost if it is represented in another form.

Classification of document that considers the document as a set of words(bag) which has no particular structure is a form which is different from our approach to utilize the structure of a document which is unique. Instead of just a set of words, we believe that designing a classification system that represents an email as a *graph* has a better scope for achieving higher accuracy as compared to some of the other conventional

approaches. Hence, for addressing the problem of email classification, we adopt graph-based data mining.

The importance of the this approach is in providing graph mining techniques for multi-folder email classification. This approach performs an extensive evaluation of graph mining techniques and their effect on various parameters used for classification and determining the values of these parameters both analytically and experimentally. Extensive experimental evaluation has been performed to establish the soundness of this approach[1].

4. PROPOSED FRAMEWORK

As discussed earlier, commemorating the credentials is a hassle task. If the user is not able to reminiscence the credentials, then he will be impuissant to access the E-mail services. The Corporate users who are mostly dependent on E-mail services, for them, this situation can prove to be critical. To overcome this situation, system provides a single framework in which E-mail fusion occurs. E-mail Fusion is a method in which all the user accounts are bought together that reside in a single frame which can be accessed using the same credentials provided by the user at the time of registration.

4.1. Registration Phase

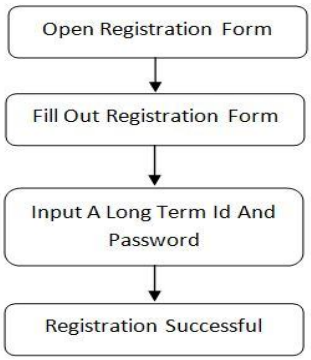

Figure 1: Registration Phase

In registration phase, user id and password are the mandatory fields which are required to be filled by the user. These registration details along with the optional fields are accepted during the initial stage by the system which are later stored in the database. After this, the registration phase is completed as shown in Figure 1.

4.2. Login Phase

In login phase, authentication process is carried out. This phase requires credentials entered by the user at the time of login, to be cross checked with the database for validity. If the result of cross checking fails, the system notifies the user about invalidity of the credentials.

If the system claims the validity of the credentials, then the system generate a One Time Password(OTP) which is routed to the user's cell phone. This OTP must be entered by the user to ensure that the user is authenticated and is a human as shown in Figure 2 which shows login system.

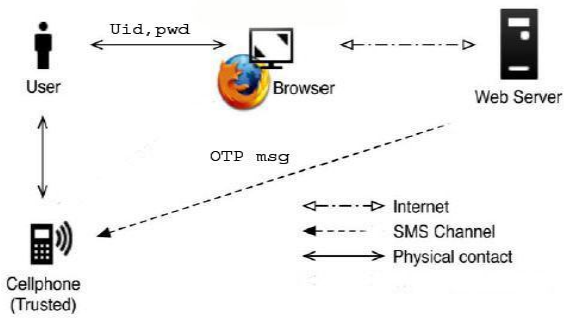

Figure 2: Login Phase

It can also happen that the OTP is wrongly entered by the user. In either case, OTP is cross verified by the system for validity. If it is found to be incorrect, the system notifies the user, and it re-generates the OTP, if requested by the user.

4.3. Graph Based Clustering

The organization of documents is a task that we face as computer users daily. This is particularly true for email management. Email documents are organized in directory structures, which reflect the users outlook with respect to his daily communication needs. Since users activities are constantly changing this may render email classifications increasingly inaccurate and manual maintenance is a painstaking task.

In this framework, we represent email as a graph. The suggested graph scheme naturally models an email corpus in the sense that it forms a direct layout of the information included within the corpus. The graph entities correspond to objects of particular types, including documents and terms, as well as email addresses, persons and dates. The graph edges are directed and correspond to relations like sent-by, sent-on-date etc. We use this framework to derive a similarity metric between email entities. This framework integrates content, timeline and social networks in a structural graph as shown in Figure 3.

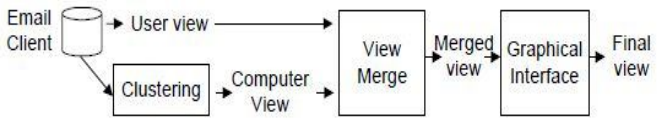

Figure 3: Clustering Email

5. CONCLUSION

In this paper, clubbing of Email domains with graph based approach by providing o-pass security is proposed, which

consists of clustering of emails, generation of OTP. It also generates a graph and a list view of searched content and at the same time it is saved into the database. In the registration phase, a new user will register into the application. The details, given at the time of registration will be used for login process and OTP will be generated. After logging in into the system, user can access his multiple email domains with no separate login for each domain. Clustering allows filtered mails based on priority. After accessing the email domains, if user wants to generate a graph to view the percentage of communication with other users, he can generate a graph by entering a text into the search bar, and also list will be generated by the same way. Both graph and list will be saved into the database for future use. As system is having only one user name and password, O-pass security is provided to secure the system. After clustering, list or graph will be displayed to user's search query and will be saved into the database.

REFERENCES

[1]. S. Chakravarthy, A. Venkatachalam and A. Telang. **A Graph-Based Approach for Multi-Folder Email Classification**, IEEE International Conference on Data Mining, 2010, pp. 78-87.

[2]. Hung-Min Sun, Yao-Hsin Chen and Yue-Hsun Lin. **oPass: A User Authentication Protocol Resistant to Password Stealing and Password Reuse Attacks**, IEEE Transactions On Information Forensics And Security, vol. 7, no. 2, pp. 651-662, April 2012.

[3]. Ms. R.R.Karthiga and Mr.K.Aravindhan. **Enhancing Performance of User Authentication Protocol with Resist to Password Reuse Attacks**, International Journal Of Computational Engineering Research (ijceronline.com), vol. 2 issue. 8, pp. 106-115.

[4]. Q Song, Jingjie Ni and G Wang. **A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data**, vol. 25, no. 1, 2013.

[5]. S F Shazmeen and J Gyani, **A Novel Approach for Clustering E-mail Users Using Pattern Matching**, 2011, pp. 205-209.

[6]. M Aery and S Chakravarthy. **eMailSift Email Classification Based on Structure and Content**, vol. 59, no. 6, April 2013.

[7]. Md R Islam, W Zhou, M Gua and Y Xiang. **An innovative analyser for multi-classifier e-mail classification based on grey list analysis**, pp. 357-366, February 2008.

[8]. M Sappelli, S verberne and W kraaij. **E-mail categorization using partially related training examples**.

[9]. E Minkov and W Cohen. **An Email and Meeting Assistant using Graph Walks.**

[10]. E Giacoletto and K Aberer. **Automatic Expansion of Manual Email Classifications Based on Text Analysis**, 2003 pp. 785-802.