ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

Avoiding Anonymous Users in Multiple Social Media Networks (SMN)

[1]Mr.T.A. Mohana Prakash, [2]Manigandan .G, [3]Chidambaram .Y, [4]Rajesh .V

[1]Assistant Professor, Department of CSE Panimalar Institute of Technology, Chennai, India
[2, 3, 4] Department of CSE- IV YEAR, Department of CSE Panimalar Institute of Technology, Chennai, India

Abstract: **The main aim of this project is secure the user login and data sharing among the social networks like Gmail, Facebook and also find anonymous user using this networks. If the original user not available in the networks, but their friends or anonymous user knows their login details means possible to misuse their chats. In this project we have to overcome the anonymous user using the network without original user knowledge. Unauthorized user using the login to chat, share images or videos etc This is the problem to be overcome in this project .That means user first register their details with one secured question and answer. Because the anonymous user can delete their chat or data In this by using the secured questions we have to recover the unauthorized user chat history or sharing details with their IP address or MAC address. So in this project they have found out a way to prevent the anonymous users misuse the original user login details.**

Keywords: **Data mining, Cross-Platform, Social Media Network, Anonymous Identical Users, Friend Relationship, User Identification.**

I. INTRODUCTION

Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets ("bigdata") involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.

The most commonly used techniques in data mining are:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.

Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification And Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k ³ 1). Sometimes called the k-nearest neighbor technique.

Rule induction: The extraction of useful if-then rules from data based on statistical significance.

An Architecture for Data Mining

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. When new insights require operational implementation, integration

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

with the warehouse simplifies the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as promotional campaign management, fraud detection, new product rollout, and so on. Figure 1 illustrates an architecture for advanced analysis in a large data warehouse.

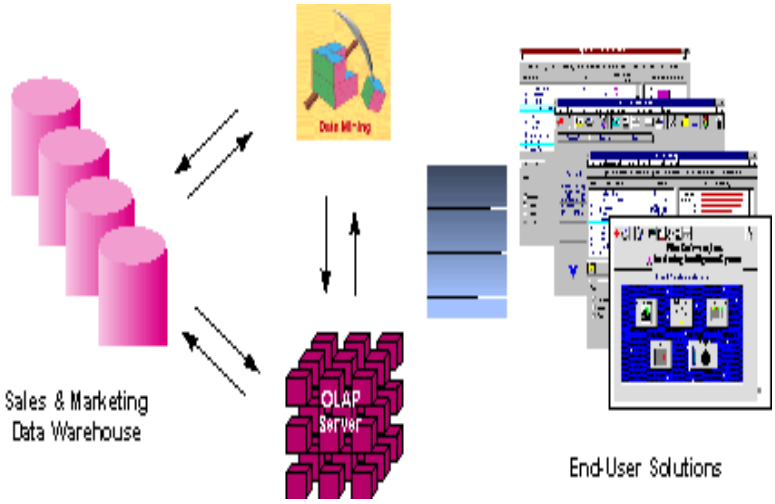

Figure 1.1 - Integrated Data Mining Architecture

The ideal starting point is a data warehouse containing a combination of internal data tracking all customer contact coupled with external market data about competitor activity. Background information on potential customers also provides an excellent basis for prospecting. This warehouse can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and should be optimized for flexible and fast data access.

An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user business model to be applied when navigating the data warehouse. The multidimensional structures allow the user to analyze the data as they want to view their business – summarizing by product line, region, and other key perspectives of their business. The data mining server must be integrated with the data warehouse and the OLAP server to embed ROI-focused business analysis directly into this infrastructure. An advanced, process-centric metadata template defines the data mining objectives for specific business issues like campaign management, prospecting, and promotion optimization. Integration with the data warehouse enables operational decisions to be directly implemented and tracked. As the warehouse grows with new decisions and results, the organization can continually mine the best practices and apply them to future decisions.

This design represents a fundamental shift from conventional decision support systems. Rather than simply delivering data to the end user through query and reporting software, the Advanced Analysis Server applies users' business models directly to the warehouse and returns a proactive analysis of the most relevant information. These results enhance the metadata in the OLAP Server by providing a dynamic metadata layer that represents a distilled view of the data. Reporting, visualization, and other analysis tools can then be applied to plan future actions and confirm the impact of those plans.

Social Media are computer-mediated tools that allow people or companies to create, share, or exchange information, career interests, ideas, and pictures/videos in virtual communities and networks. *Social media* is defined as "a group of Internet-based applications that build on the ideological technological foundations of Web 2.0, and that allow the creation and exchange of user generated content.

Cross-Platform, **Multi-Platform**, or **Platform Independent**, is an attribute conferred to computer software or computing methods and concepts that are implemented and inter-operate on multiple computing platforms. Cross-Platform software may be divided into two types; one requires individual building or compilation for each platform that it supports, and the other one can be directly run on any platform without special preparation, e.g., software written in an interpreted language or pre-compiled portable bytecode for which the interpreters or run-time packages are common or standard components of all platforms.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

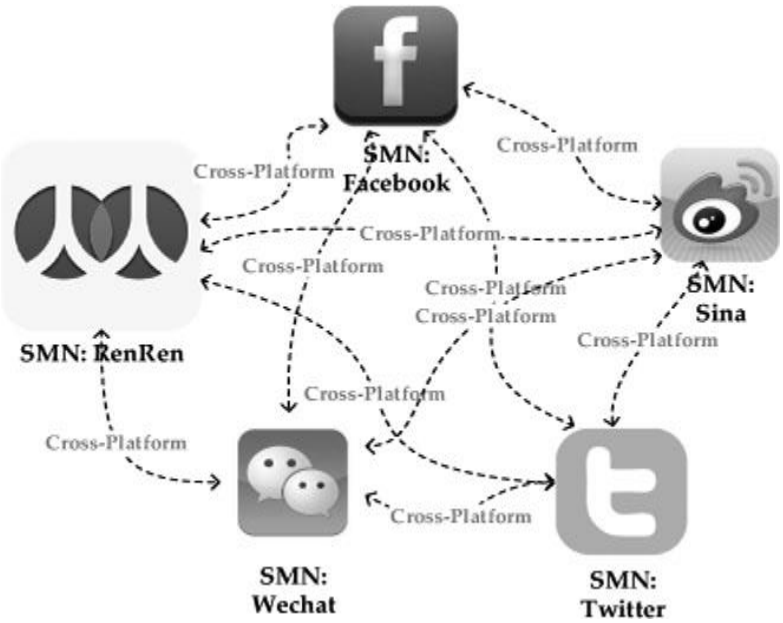

Fig 1.2 Cross Platform Identification

Anonymous is a loosely associated international network of activist and hacktivist entities. A website nominally associated with the group describes it as "an Internet gathering" with "a very loose and decentralized command structure that operates on ideas rather than directives". The group became known for a series of well-publicized publicity stunts and distributed denial-of-service (DDoS) attacks on government, religious, and corporate websites.

In our analysis of cross platform SMNs, we deeply mined **friend relationship** and network structures. In the real world, people tend to have mostly the same friends in different SMNs, or the friend cycle is highly individual. The more matches in two unmapped users' known friends, the higher the probability that they belong to the same individual in the real world. Based on this fact, we proposed the FRUI algorithm. Since FRUI employs a unified friend relationship, it is apt to identify users from a heterogeneous network structure. Unlike existing algorithms. FRUI chooses candidate matching pairs from currently known identical users rather than unmapped ones.

User Identification refers to now a days more and more people have their virtual identities on the Web. It is common that people are users of more than one social network and also their friends may be registered on multiple web sites. A facility to aggregate our online friends into a single integrated environment would enable the user to keep up-to-date with their virtual contacts more easily, as well as to provide improved facility to search for people across different websites. In this paper, we propose a method to identify users based on profile matching. We use data from two popular social networks to study the similarity of profile definition. We evaluate the importance of fields in the web profile and develop a profile comparison tool. We demonstrate the effectiveness and efficiency of our tool in identifying and consolidating duplicated users on different websites.

The **anchor text**, **link label**, **link text**, or **link title** is the visible, clickable text in a hyperlink. The words contained in the anchor text can determine the ranking that the page will receive by search engines. Since 1998, some web browsers have added the ability to show a tooltip for a hyperlink before it is selected. Not all links have anchor texts because it may be obvious where the link will lead due to the context in which it is used. Anchor texts normally remain below 60 characters. Different browsers will display anchor texts differently. Usually, web search engines analyze anchor text from hyperlinks on web pages. Other services apply the basic principles of anchor text analysis as well. For instance, academic search engines may use citation context to classify academic articles, and anchor text from documents linked in mind maps may be used too.

A **digital footprint** is a trail of data you create while using the Internet. In includes the websites you visit, emails you send, and information you submit to online services. A "passive **digital footprint**" is a data trail you unintentionally leave online.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

Digital Footprints are classified into passive and active. A passive digital footprint is created when data is collected without the owner knowing, whereas active digital footprints are created when personal data is released deliberately by a user for the purpose of sharing information about oneself by means of websites or social media. Passive digital footprints can be stored in many ways depending on the situation. In an online environment a footprint may be stored in an online data base as a "hit". This footprint may track the user IP address, when it was created, and where they came from; with the footprint later being analyzed. In an offline environment, a footprint may be stored in files, which can be accessed by administrators to view the actions performed on the machine, without being able to see who performed them.

Crawlers consume resources on the systems they visit and often visit sites without tacit approval. Issues of schedule, load, and "politeness" come into play when large collections of pages are accessed. Mechanisms exist for public sites not wishing to be crawled to make this known to the crawling agent.

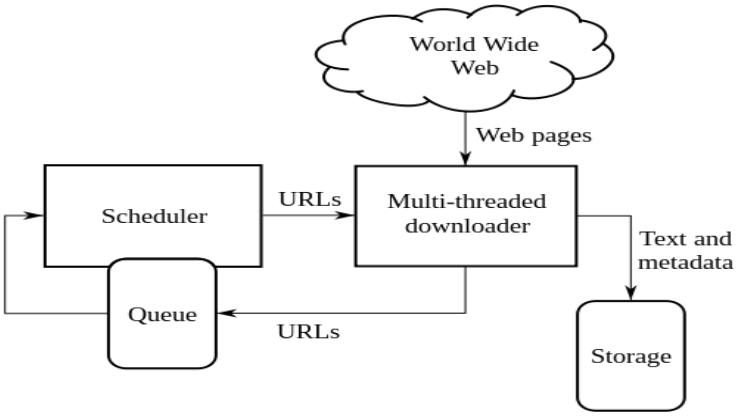

Fig.1.3 Architecture of web crawler

II. RELATED WORK

Bin Zhou Jian Pei [1] proposed preserving privacy in social networks against neighborhood attacks by a technique called anonymization algorithm. The attacks in need to plant a set of deliberative structures before the social network data is anonymized, which is a task hard to achieve in some situations. As shown before, even without planting deliberative structures, the released social network data is still in danger, as neighborhood attacks are still possible. One of the privacy concerned problems is publishing microdata for public use, which has been extensively studied recently. As social network data is much more complicated than relational data, privacy preserving in social networks is much more challenging and needs many serious efforts. Modeling adversarial attacks and developing privacy preservation strategies are critical**. "Preserving Privacy in Social Networks Against Neighborhood Attacks".**

Xiangnan Kong, *et al* [2] proposed inferring anchor links across multiple heterogeneous social networks by a technique called multi-network anchoring. The proposed Multi-Network Anchoring (MNA) method consistently out performs other baseline methods. This result supports the intuition of this paper: Multiple heterogeneous social networks can provide different types of information about the users. The intractability of the problem, existing methods usually rely on practical heuristics to solve the alignment problem. By explicitly consider the users heterogeneous data within the networks. It shows that by incorporating the one-to-one constraint in the inference process can further improve the performance of anchor link prediction**. "Inferring Anchor Links across Multiple Heterogeneous Social Networks"**

Reza Zafarani [3] proposed connecting corresponding identities across communities by a technique called link analysis algorithm. The relationship between usernames selected by a single person in different communities, and on some of the web phenomena regarding usernames and communities The unrevealing nature of the web and the fact that most communities preserve the anonymity of users by allowing them to freely select usernames instead of their real identities and the fact that different websites employ different username and authentication systems Nevertheless, if there exists a mapping between usernames across different communities and the real identities behind them, then connecting communities across the web becomes a straight forward task. **"Connecting Corresponding Identities Across Communities"**

Paridhi Jain, *et al* [4] proposed identifying users across multiple online social networks by a technique called identity search algorithms. we introduce two novel identity search algorithms based on content and network attributes and improve on traditional identity search algorithm based on prole attributes of a user that exploiting multiple identity search methods, surfaces the identities similar to the given identity in different aspects other than the traditional ways (e.g., similar name) and therefore, increases the accuracy of finding correct identities users across social networks. In this work, they attempt to understand if inclusion of search methods based on an identity's content and network attributes, along with search methods based on an identity's prole attributes. "**Identifying Users Across Multiple Online Social Networks".**

E.-P. Lim, *et al* [5] proposed exploring linkablility of community reviewing by a technique called matching algorithm. The techniques that extract frequent pattern write-prints are characterized by one author. Typically reflect one's experience when dealing with a buyer or a seller. Unlike our general-purpose reviews, these comments do not review products, services or places of different categories in addition, although they take advantage of ratings and categories to boost LRs, they need to further explore usage of other non-textual features, such as sub-categories of places, products and services reviewed as well as the length of reviews. In fact, it would be interesting to see how the LR can be improved without resorting to lexical features, since they generally entail heavy processing**. "Exploring Linkablility of Community Reviewing"**

Paridhi Jain, *et al* [6] proposed finding nemo: Searching and resolving identities of users across online social networks using algorithm called profile search. Our knowledge, majority of the approaches proposed exploited either one or two dimensions for an identity search and linking, thereby leaving other hints uninvestigated to leverage available information about the user and create a set of candidate identities for a user on a social network. To adapt to real-time search, limited availability of information and usage of the auxiliary information left unexplored. Researchers have developed a set of approaches which assume that the considered dimension is constituted in a similar fashion by a user across her multiple identities**. "Finding Nemo: Searching and Resolving Identities of Users Across Online Social Networks".**

O. de Vel [7] proposed mining email content for author identification forensics by algorithm called vector machine learning algorithm. Many methods that automatically learn rules have been proposed for text categorisation. No set of significant style markers have been identified as uniquely discriminatory. There does not seem to exist a consensus on a correct methodology, with many of these techniques suffering from problems such as questionable analysis, inconsistencies for the same set of authors, failed replication etc. features may not be valid discriminators. Prescriptive grammar errors, profanities etc. which are not generally considered to be idiosyncratic Just as there is a range of available stylometric features there are many different techniques using these features for author identification**. "Mining E-mail Content for Author Identification Forensics"**

Reza Zafarani, *et al* [8] proposed connecting users across social media sites: A behavioral-modeling approach by an algorithm called learning algorithm. The proposed behavioral modeling approach exploits information redundancy due to these behavioral patterns. An alternative solution addressing the age verification problem by exploiting the nature of social media and its networks The information available on all social media sites (usernames) to derive a large number of features that can be used by supervised learning to connect users across sites. Users often exhibit certain behavioral patterns when selecting usernames. It includes analyzing these possibilities and discovering features indigenous to specific sites, beyond those constricted to usernames, and incorporating them into MOBIUS for future needs. "**Connecting Users Across Social Media Sites: A Behavioral-Modeling Approach".**

Nitish Korula, *et al* [9] proposed an efficient reconciliation algorithm for social networks by using Learning algorithms. A deeper understanding of the characteristics of a user across different networks helps to construct a better portrait of her, which can be used to serve personalized content or advertisements to the best of our knowledge, it has not yet been studied formally and no rigorous results have been proved for it. Even if certain behavior can be observed in several networks, there are still serious problems because there is no systematic way to combine the behavior of a specific user across different social networks and because some social relationships will not appear in any social network. For these reasons, identifying all the accounts belonging to the same individual across different social services is a fundamental step in the study of social science. **"An Efficient Reconciliation Algorithm for Social Networks"**

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

Elie Raad, *et al* [10] proposed user profile matching in social networks by using decision making algorithm. They searched for the total number of possible combinations that refer to the same physical person. Then they calculated the number of combinations of found profiles by our method that also exist in the initial set R. We also calculated the number of profiles combinations that were detected by our approach as being the same physical person. They addressed the issue of providing intersocial network operations and functionalities. In this work, they proposed a framework for user profile matching in social networks. This framework is able to discover the biggest possible number of profiles that refer to the same physical user that existing approaches are unable to detect. They are planning to further explore and propose more interesting inter-social operations and functionalities. "**User Profile Matching in Social Networks"**

III. IMPLEMENTATION

In this paper we use Friend Relationship Based User Identification Algorithm (FRUI) to recover the data which has been deleted or changed by anonymous user. This can be achieved by providing a history option and that will be designed in such a way that it will provide the actions that have been performed till the current session.

In this paper we have developed an application through which we can perform data sharing and chatting. To register to this application we have to give username, password and answer two level of security questions. After giving all these details the user of this application will be successfully registered. At the time of login it will ask username, password and one security question through which we have registered. By logging into this application we can perform chatting and data sharing. And in this additionally history menu is also provided which provide the actions performed till the current session. By viewing this we can come to know whether anonymous user used our login and we can our data if any irrelevant actions have been performed.

Architecture of the system

The basic idea behind this paper is to recover the data which has been deleted or changed by the anonymous user. In the architecture diagram, the user logins into his account and he will perform his actions on his account. And in the same way if anonymous user can come to know the login details of the original user he can also make modifications in that account. To record those actions which have created a history menu in that application.

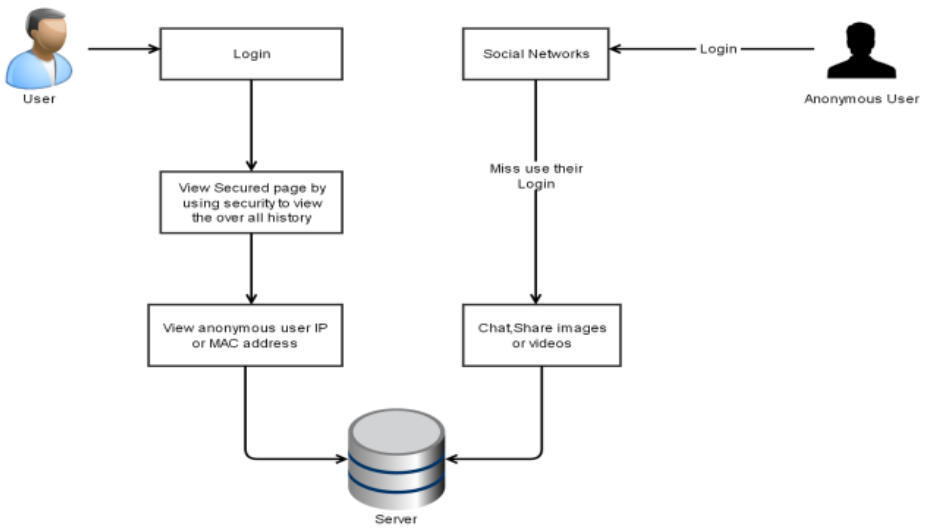

Fig 3.1 Architecture of the system

The user can retrieve his details which has been changed by anonymous user by simply answering the second level of security question which is provided inside the history menu. And once if he answers the question a dialog box will be opened listing the actions which have been performed till the current session. The user can recover his data from that dialog box if any anonymous activities have been performed. And he can also view the anonymous user IP or MAC address.

Phases of the system:

The six phases of the system are:

i. User Matched Pair

ii. Network Structure Based User Identity

iii. User Identification

iv. Friend Relationship Based User Identification (FRUI)Algorithm

v. Recovery misused details

vi. Find IP or MAC

Phase 1: User Matched Pair

In this module first user register their details with security questions that will help to recover the original data. The reason why we choose the Q & A means if other secret password or other values are put in that place, easily hackers can find out the password.

Phase 2: Network Structure Based User Identity

In this module user can easily find their login is misused or not. By sending the notification details like last time out, logout time and IP address we can find out the user identity. The IP address is used to find the logout system where located after that user can change their password.

Phase 3: User Identification

In user identification module we have to find the hacker IP address. We have mentioned that the hacking system IP is used to find the location. If the location is nearby we can easily find the location of original hacker.

Phase 4: Friend Relationship Based User Identification (FRUI) Algorithm

We proposed the Friend Relationship-Based User Identification (FRUI) algorithm. FRUI (Friend Relationship Based User Identification Algorithm) calculates a match degree for all candidate User Matched Pairs (UMPs), and only UMP with top rank share considered as identical users. We also developed two propositions to improve the efficiency of the algorithm.

Phase 5: Recovery Misused Details

In this project it is mentioned that the user create their login with secret question, that question will help to recover the misused details. In this original login page as the original user or the hacker whatever they do that will shown in front page. If the original user wants to know their hacked or misused files or profile that time the secret question will helps to them.

Phase 6: Find IP or MAC

In this module by using the secured questions we have to recover the unauthorized user chat history or sharing details with their IP address or MAC address. So in this project we have to find out the anonymous users misuse the original user login details.

IV. ALGORITHMS

This algorithm is used to provide the details which have been deleted in the chat history. Inside this application we have a separate tab called history and this provides the details which has been deleted in the conversation tab.

Input: When the user logs into the system he has to provide the username, password and select one level of security question and answer it. When he performs all these actions the user successfully logs into the system.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

For a new user of this application the user has to register with all the details such as username, password and selecting one level of security question and answering it and then giving a password to access the history tab. After giving all these details the user gets successfully logs into the system. Once we get login into the application the user has access to six different tabs.

Profile: In this user can perform the actions such as edit his profile by changing the details which he have been at the time of registration.

Friends: In this user can have access to three different tabs such as Friend List, Notifications, New Friends. New Friends provides sending friend request who are registered with this application. Notification tab has the notifications of all the other users who has provided us the friend request. Friend List provides the list with all the friends we have in this application. And in this we also have the facility which is we can unfriend them.

Chat: As the name implies this chat tab is used for chatting with users. This displays the list of friends same as that of Friend List through which we can chat with those users.

Conversation: This conversation tab has the details of all the information which we have chatted. And in this we have the facility to clear those chats.

History: When clicking on this history tab the user will be asked with history password which is given at the time of registration. When it is given correctly it provides the details which have been deleted inside the conversation tab along with the IP and MAC address.

Logout: Once you finished all your actions by clicking on this Logout tab it will take out of this application.

V. EXPERIMENTAL RESULTS

A. Home Page:

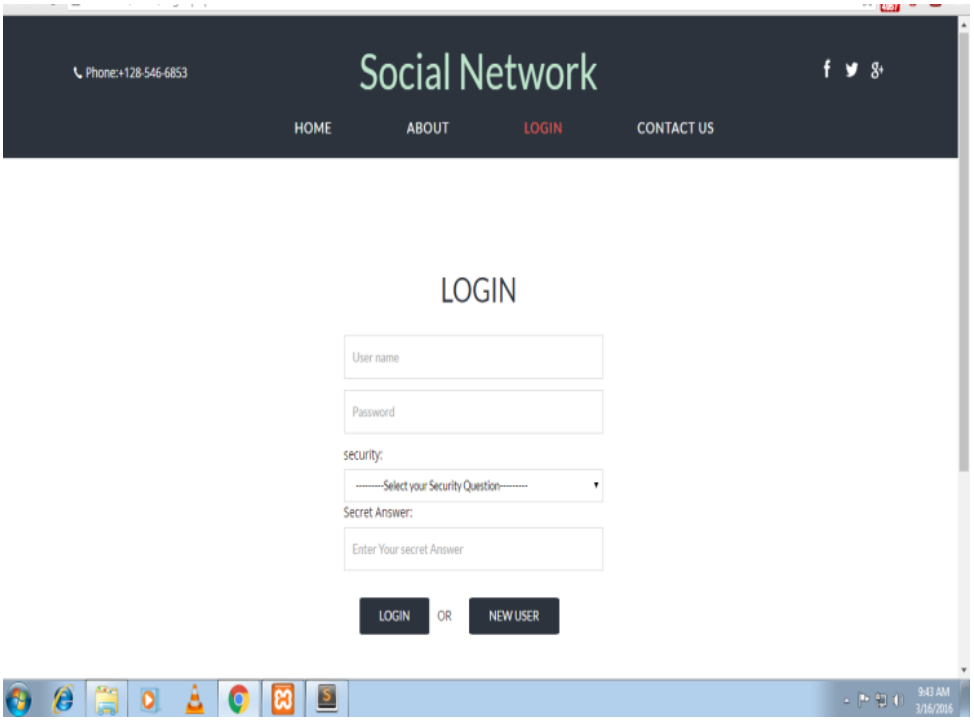

Fig 5.1 Home Page

In this page the user will login into his account if he is an existing user or otherwise he will register his information and login. At the time of login the user has to provide with username, password and select a question in the drop down box and answer it which he has answered at the time of registeration.

ISSN 2350-1022

International Journal of Recent Research in Mathematics Computer Science and Information Technology
Vol. 2, Issue 2, pp: (188-198), Month: October 2015 – March 2016, Available at: www.paperpublications.org

B. Sending Friend Request And Notification Page:

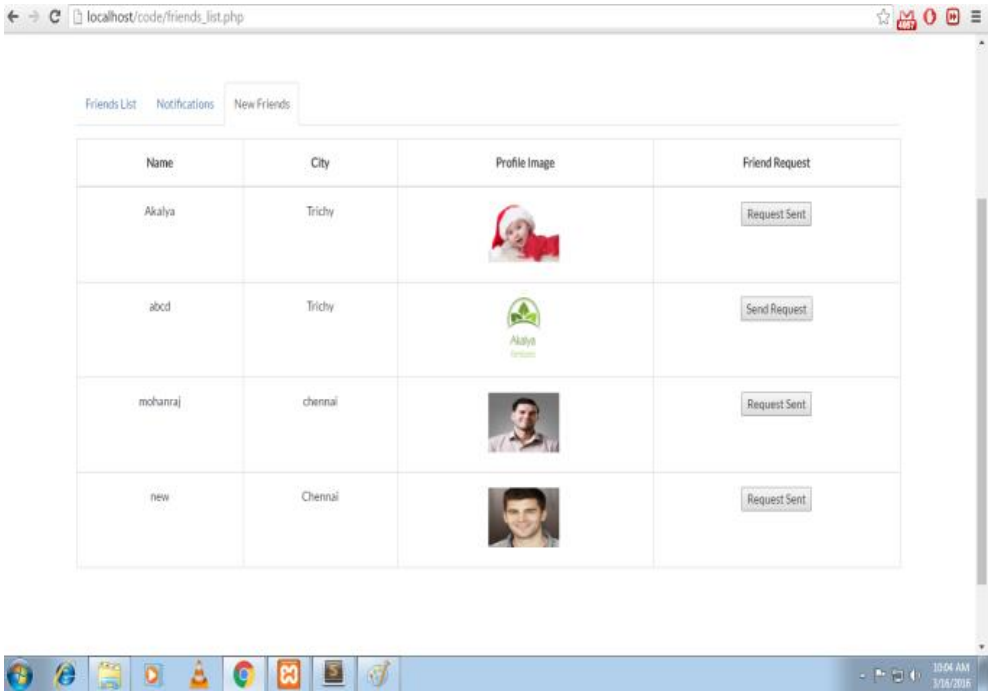

Fig 5.2 Friend Request And Notification Page

Here the user can send the friend request to the person he wish to send and in the notifications tab he can view the persons who has sent him the friend request. And according to his opinion he can accept or deny the request. In the friends list tab it shown him with the persons whom he is friend with and he also has the option to unfriend that person.

C. Clearing the Chat Details:

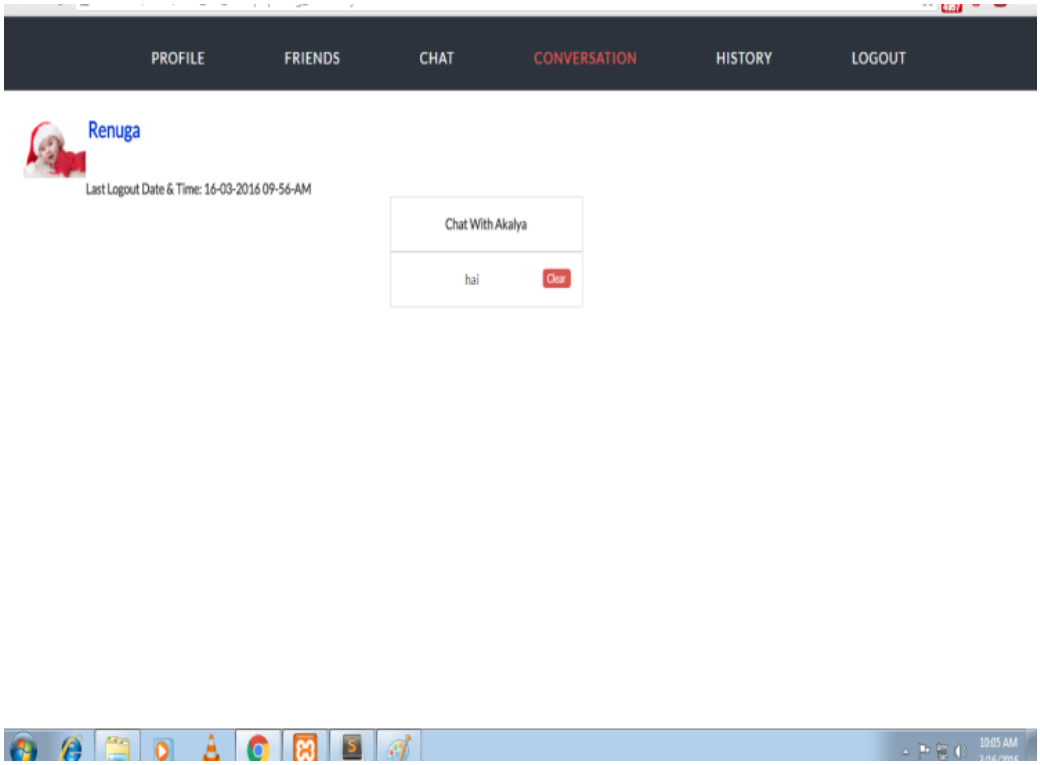

Fig 5.3 Clearing Chat Details

If anonymous user uses our account he may chat with the persons whom we know and he has the option to delete the chat which he has performed. Once performing all his chats in the chat tab he may delete the chats in the conversation tab which is shown in Fig 5.3.

D. Deleted Chat Details:

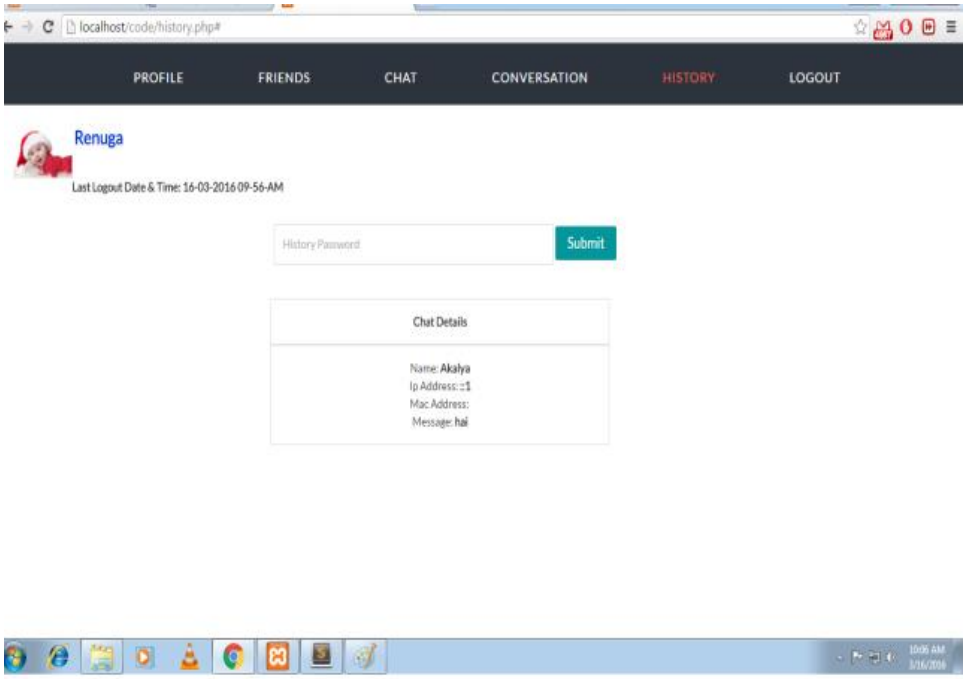

Fig 5.4 Deleted Chat Details

In this history tab which is shown above in Fig 5.4 contains the deleted chat details which is performed inside the conversation tab.

To open this history tab the user has to give the history password which he has given at the time of registeration. Once if he has entered it correctly a dialog box will be opened showing the deleted chat details.

VI. CONCLUSION

Thus the project "Cross Platform Identification of Anonymous Identical Users in Multiple Social Media Networks" provides uniform network structure-based user identification solution. Moreover our project can be easily applied to any SMNs with friendship networks, including Twitter, Face-book and Foursquare. Since only the Adjacent Users are involved in each iteration process our method is scalable and can be easily applied to large data sets and online user identification applications.

REFERENCES

[1] B. Zhou and J. Pei, "Preserving privacy in social networks against neighborhood attacks," Proc. Of the 24th IEEE International Conference on Data Engineering (ICDE'08), pp. 506–515, 2008.

[2] X. Kong, J. Zhang, and P.S. Yu, "Inferring anchor links across multiple heterogeneous social networks," Proc. of the 22nd ACM International Conf. on Information and Knowledge Management (CIKM'13), pp. 179-188, 2013.

[3] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," Proc. of the 3rd International ICWSM Conference, pp. 354-357, 2009.

[4] P. Jain, P. Kumaraguru, and A. Joshi, "@ i seek 'fb. me': identifying users across multiple online social networks," Proc. of the 22nd International Conference on World Wide Web Companion, pp. 1259-1268, 2013.

[5] M. Almishari and G. Tsudik, "Exploring linkability of user reviews," Computer Security–ESORICS 2012 (ESORICS'12), pp. 307324, 2012.

[6] P. Jain and P. Kumaraguru, "Finding Nemo: searching and resolving identities of users across online social networks," arXiv preprint arXiv:1212.6147, 2012.

[7] O. De Vel, A. Anderson, M. Corney, and G. Mohay, "Mining email content for author identification forensics," ACM Sigmod Record, vol. 30, no. 4, pp. 55-64, 2001.

[8] R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach, " Proc. of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13), pp.41-49, 2013.

[9] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," arXiv preprint arXiv:1307.1690, 2013.

[10] E. Raad, R. Chbeir, and A. Dipanda, "User profile matching in social networks," Proc. Of the 13th International Conference on Network-Based Information Systems (NBiS'10), pp.297-304, 2010.