iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** IV **Month of publication:** April 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Privacy-Preserving using Profile Matching for Medical Report Analysis

Pratik Soni[1], Kiran Mehtani[2], Samiksha Poojari[3], P.S. Dhotre[4]
[1,2,3,4]STES's Sinhgad Institute of Technology and Science, Narhe, Pune, MH, India

Abstract— Wireless and mobile communication has brought up major features to individuals. Mobile Social Networks (MSNs) is not so far behind considering its effective use. Utilizing these services, individual has to pay its personal information. In existing systems, the sensitive and personal information should be revealed on their website to utilize the services. However, in many applications, user doesn't want to disclose the user's personal profiles containing sensitive information publicly. In this paper, we would like to introduce personal information privacy mechanism based on profile matching in medical field. The mentioned system allows individuals/ patients to control their medical information or reports for disclosing and sharing. We have described about the safety protocol that we have implemented for identification of users with similar interests/ attributes. We extend the privacy in such a way that abstracted information regarding the user profile is shared/ exchanged.

Keywords—Privacy, Mobile Social Networks (MSN), Wireless Communication, Abstraction.

I. INTRODUCTION

One necessary perform provided by social network is friend discovery. Social networking is the grouping of individuals into specific groups, like small rural communities or a neighbourhood subdivision. The matter of finding individuals of the same attribute/ interest/ community has long been studied in the context of social network. As an example, profile-based friend discovery will advocate those that have similar attributes/ interests; topology-based friend discovery will recommend individuals from constant community. When it involves on-line social networking, websites are usually used. Once you're granted access to a social networking web site you'll be able to begin to socialize. This socialization could embrace reading the profile pages of alternative members and presumably even contacting them. As mentioned, social networking involves grouping specific individual and organization along, whereas there are variety of internet sites concentrate on explicit interests which suggests anybody will become member, notwithstanding what their hobbies or interest are, once you're within the community you'll be able to build friends of common interest and may eliminate those friends. To deal with user identification, it's essential to disclose least and necessary personal information.

One special demand of algorithms operative on social network is that it should be privacy-preserving. As an example, social network nodes could also be willing to share their attributes/ interests with individuals having similar profile or they may be willing to share their raw connections with people within the same community.

However, it's unfavourable to leak that personal information to arbitrary strangers. Towards this end, the friend discovery routine ought to solely expose minimal necessary info to concerned parties.

II. MOTIVATION

However, such systems additionally raise variety of privacy considerations. Let us examine a noteworthy situation. In a hospital, patients could embrace their illness symptoms and medications in their personal profiles so as to seek out similar patients, for physical or mental support. During this situation, associate initiating user (initiator) might want to seek out the patient having the utmost range of identical symptoms along with his/ her, whereas being reluctant to disclose his/ her sensitive illness information to the remaining users, and also the same for the users being matched with. If user's non-public profiles area unit directly changed with one another, it'll facilitate user identification wherever that information are often simply collected by a close-by user, either in a lively environment or passive way; and user's information could also be exploited in unauthorized ways. For instance, a salesperson from a pharmacy could submit malicious matching queries to get statistics on patient's medications for selling his products.

III. RELATED WORK

In this section, we review related work focusing on social networking services and applications. A considerable body of work has contributed to this area. Leveraging networked portable devices such as smart phones and PDAs as platforms, MSN not only enables people to use their existing online social networks (OSNs) at anywhere and anytime, but also introduces a myriad

of mobility-oriented applications.

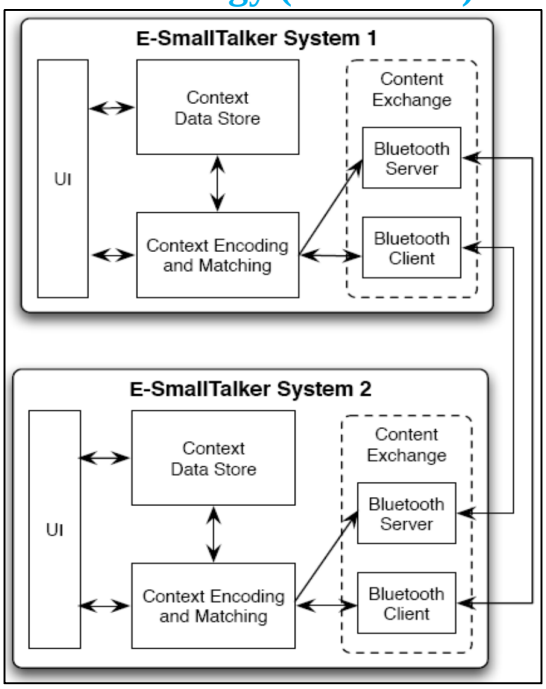


Fig. 1: E-SmallTalker Architecture [3]

For example, MagnetU and E-SmallTalker [3] are MSN applications that match one with nearby people for dating or friend-making based on common interests. Fig. 1 describes the architecture of E-SmallTalker. In such an application, a user only needs to input some (query) attributes in her profile, and the system would automatically find the persons around with similar profiles. The scopes of these applications are very broad, since people can input anything as they want, such as hobbies, phone contacts and places they have been to. The latter can even be used to find "lost connections" and "familiar strangers".

Also, WeChat and Whatsapp are social networking applications where anyone can be discovered through his/ her personal contact number. These social networking applications leak personal information to arbitrary strangers.

To overcome these problems we propose following solutions for privacy preservation using profile matching techniques.

IV. RESEARCH CHALLENGES

A. *System Model*

Our system consists of N users (parties) denoted as P1 … PN, each possessing some specific set of attributes. We denote the initiating party (initiator) as P1. P1 launches the matching process and its goal is to find one party that best "matches" with it, from the rest of the parties P2, ... PN, which are called candidates. Each party Pi's profile consists of a set of attributes Si, which can be strings up to a certain length. P1 defines a matching query to be a subset of S1, and in the following we use S1 to denote the query set unless specified. Also, we denote n = |S1| and m = |Si|, i > 1, assuming each candidate has the same set size for simplicity. Note that, we assume that the system adopts some standard way to describe every attribute, so that two attributes are exactly the same if they are the same semantically. There could be various definitions of "match". In this paper, we consider a popular similarity criterion, namely the intersection set size |S1 ∩ Si| The larger the intersection set size, the higher the similarity between two user's profiles. User P1 can first find out her similarity with each other users via our protocols, and then will decide whether to connect with a best matching user based on their actual common attributes.

B. *Design Goals*

Since the users may have different privacy requirements and it takes different amount of efforts to achieve them, we hereby (informally) define two levels of privacy where the higher level leaks less information to the adversaries.

Definition 1 (Privacy level 1 (PL-1)): When the protocol ends, $P1$ and each candidate Pi, $2 \leq i \leq N$ mutually learn the intersection set between them: $I1,i=S1 \cap Si$. An adversary A (whose capability is defined in Sec. II-B) should learn nothing beyond what can be derived from the above outputs and its private inputs.

Definition 2 (Privacy level 2 (PL-2)): When the protocol ends, $P1$ and each candidate Pi, $2 \leq i \leq N$ mutually learn the size of

their intersection set: $m_{1,i}=|S_1 \cap S_i|$. The adversary A should learn nothing beyond what can be derived from the above outputs and its private inputs.

In PL-2, except when $m_{1,i} = |S_1|$ or $|S_i|$, P_1 and each P_i both will not learn exactly which attributes are in $I_{1,i}$. The adversary needs to run the protocol multiple times to obtain the same amount of information with what he can obtain under PL-1 when he assumes the role of P_1 [1].

For profile matching, it is desirable to involve as few human interactions as possible. In this paper, a user only need to explicitly participate in the end of the protocol run, e.g., decide whom to connect to, based on the common interests. In addition, the system design should be lightweight and practical, i.e., being enough efficient in computation and communication. Finally, different users (especially the candidates) shall have the option to flexibly personalize their privacy levels.

C. Our Contributions

In this paper, we implement a set of privacy-preserving profile matching schemes. We have outlined many privacy levels [1] for secure profile matching. However, it's difficult to seek out the matching users privately while expeditiously.

In this paper, we describe the flaws of the present system and tried subsequent main contributions in our project:-

(1) We formulate the privacy preservation drawback of profile matching. Two levels of privacy area unit defined alongside their threat models, wherever the higher privacy level offers the abstracted data concerning the profile of the initiating user to the opponent than the lower level.

(2) We propose two totally distributed privacy-preserving profile matching schemes, one amongst them being Private Set Intersection [2] (PSI) protocol and the other is Private Cardinality of Set Intersection (PCSI) protocol. However, solutions supported existing PSI schemes area unit off from efficient.

(3) We propose Shamir's secret sharing scheme [1] for distributing a secret amongst a group of participants, each of which is allocated a share of the secret.

D. System issues

The issues that may affect the operations capabilities are:

1) Finding the best matched profile
2) Proper initialization of profile matching algorithms
3) Selection of proper attributes from datasets
4) Optimal matching rate

E. Assumptions and Constraints

Following are the assumptions for the implementation are:

1) The data is based on various heterogeneous attributes.
2) The dataset (profiles) must be in large number to be classified and matched, before sharing an attribute/ interest.

V. SYSTEM DESIGN

In this system, we formulate the privacy preservation problem of profile matching for Medical Analysis. Two levels of privacy are defined along with their threat models, where the higher privacy level leaks less profile information to the adversary than the lower level. We propose two fully distributed privacy-preserving profile matching schemes, one of them being a Private Set Intersection Protocol (PSIP), PSIP used to calculate intersection of matching attributes; and the other is a Private Cardinality of Set-Intersection Protocol (PCSIP), PCSIP used to calculate number/ count of total attributes matching. However, solutions based on existing PSI schemes are far from efficient. We leverage secure multi-party computation based on polynomial secret sharing, and propose several key enhancements to improve the computation and communication efficiency. Shamir secret sharing to compute the intersection between user profiles privately, and due to the smaller computational domain of secret sharing, our protocols achieve higher performance and lower energy consumption.

The systems aim is to use data mining tools such as profile matching, secret sharing and cluster analysis in order to:

1) Preserve Privacy using profile matching percentage
2) Consult Doctors
3) Provide Treatments information
4) Give feedback
5) Suggest about NGOs/ Extra Help
6) Socialize with Patients having same Disease

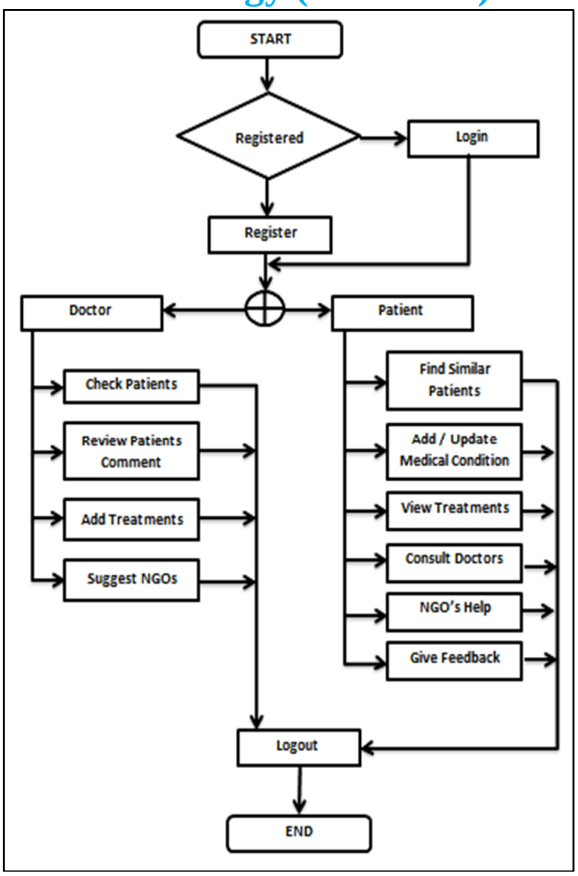


Fig. 2: System Flow Diagram

VI. RESEARCH METHODOLOGIES

A. *Shamir's Secret Sharing Scheme*

A (*t, w*)-SS [1] scheme shares secret *s* among *w* parties by giving each party *Pi* the value $[s]_i^{t,w}$, and if any at most *t* parties collude they cannot gain any information about *s*. Thus their protocol realizes randomization and degree-reduction in one round by letting each Pi pick a random t-degree polynomial and re- share $[\alpha]_i^{t,w}$ $[\beta]_i^{t,w}$ to others:

Round 1: Each party Pi shares the value $[\alpha]_i^{t,w}$ $[\beta]_i^{t,w}$ by choosing a t-degree random polynomial hi(x), s. t. $Hi(0)$ =$[\alpha]_i^{t,w}$ $[\beta]_i^{t,w}$. He sends the value $H(j)$ to party Pj, $1 \leq j \leq w$.

Round 2: Every party Pj computes his share of αβ, i.e., the value $H(j) = [\alpha + \beta]_i^{t,w}$ under a t-degree random polynomial H, by locally computing the linear combination H(j) =$\sum_{i=1}^{w} \lambda_i h_i(j)$, where λ1, ..., λw are known constants. An additive homomorphic encryption scheme E allows one to compute E(m1+ m2) given E(m1) and E(m2), without knowing the plain texts. This is used in our protocol for PL-2.

Notations:

N, t: Number of Parties, maximum number of colluders

$[s]_i^{t,w}$: Party P_i's secret share of s

S_1 , S_i: P_1's query attribute set, P_i's profile attribute set

$x_j, 1 \leq j \leq n$: P_1's query set elements, n = $|S_1|$

$y_{ij}, 1 \leq j \leq m$: P_i's profile set elements, m = $|S_i|$

$I_{1,i}$: Intersection set between P_1 and P_i; $m_{1,I} = |I_{1,i}|$

Δ_1: SS-Add: $[\alpha + \beta]_i^{t,w} = [\alpha]_i^{t,w} + [\beta]_i^{t,w}$

Δ_2: SS-Mul: $H(j) = \sum_{i=1}^{w} \gamma_i h_i(j)$

B. *Private Set Intersection Protocol*

Increasing dependence on anytime-anywhere availability of data and the commensurately increasing fear of losing privacy motivate the need for privacy-preserving techniques. One interesting and common problem occurs when two parties need to privately compute an intersection of their respective sets of data. In doing so, one or both parties must obtain the intersection (if one exists), while neither should learn anything about other set. Although prior work has yielded a number of effective and elegant Private Set Intersection (PSI) [2] techniques, the quest for efficiency is still underway.

Consider the following examples:

1. A government agency needs to make sure that employees of its industrial contractor have no criminal records. Neither the agency nor the contractors are willing to disclose their respective data-sets (list of convicted felons and employees, respectively) but both would like to know the intersection, if any.
2. Two national law enforcement bodies (e.g., USA's FBI and UK's MI5) want to compare their respective databases of terrorist suspects. National privacy laws prevent them from revealing bulk data, however, by treaty; they are allowed to share information on suspects of common interest.
3. Two real estate companies would like to identify customers (e.g., homeowners) who are double-dealing, i.e., have signed exclusive contracts with both companies to assist them in selling their properties.
4. Federal tax authority wants to learn whether any suspected tax evaders have accounts with a certain foreign bank and, if so, obtain their account records. The bank's domicile forbids wholesale disclosure of account holders and the tax authority clearly cannot reveal its list of suspects.
5. Department of homeland security (DHS) wants to check its list of terrorist suspects against the passenger manifest of a flight operated by a foreign airline. Neither party is willing to reveal its information; however, if there is a (non-empty) intersection, DHS will not give the flight permission to land.

These example motivate the need for privacy-preserving set operations, in particular, set intersection protocols. Such protocols are especially useful whenever one or both parties (who do not fully trust each other) must compute an intersection of their respective sets (or some function thereof).

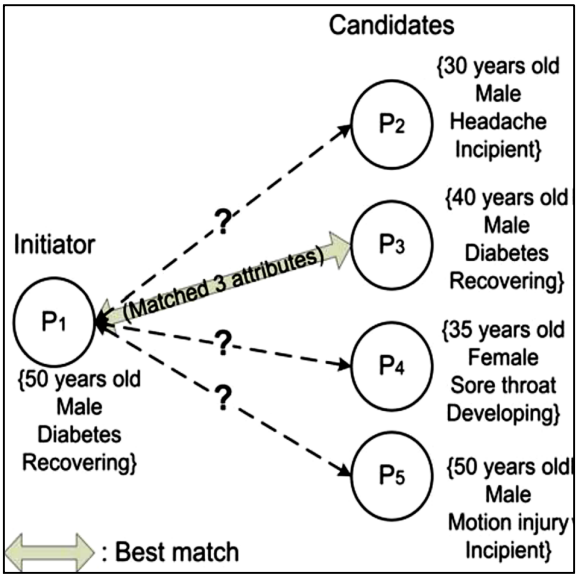


Fig. 3: Private Profile Matching [4]

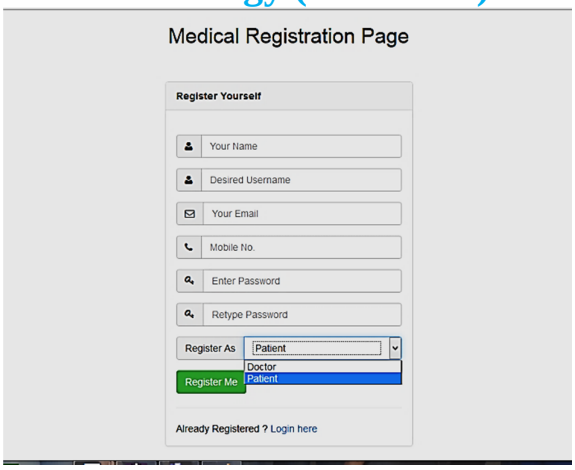
Private Set Intersection (PSI) is a cryptographic protocol that involves two players, each with a private set. Their goal is to compute the intersection of their respective sets, such that minimal information is revealed in the process. In other words, they should learn the elements (if any) common to both sets and nothing (or as little as possible) else. This can be a mutual process where, ideally, neither party has any advantage over the other.

VII. RESULTS AND DISCUSSIONS

In this section, we discuss the results/ final outcome of our website.

The flow of our website is as follows:-

Volume 3 Issue IV, April 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



a. Registration Page

Any user must register on this website before using any services of the website. Users can register as a patient if he wants to interact with people with same medical conditions. If user is doctor by profession and is willing to help patient through their medical anomalies, he can register as doctor. After registration, users can login with their registered username and password

b. Patient's Home Screen

The Patient's Home Screen consists of various services as follows:-

a. Home Screen displays your profile (1) that is the attributes that are public to other users.

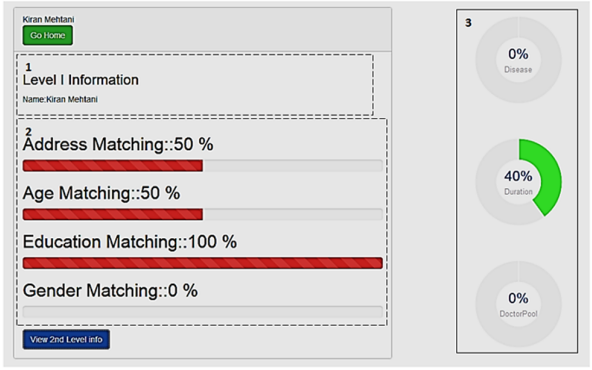
b. On Right hand side of the page, it displays people with matching attributes (2) that are people with same medical condition or within same localities, etc.

c. Update your profile (3) option helps you with updating your personal info or Medical condition.

d. Settings (4) option is used to define visibility (i.e. Private, Public and Shared) of each attribute. By default, visibility is set to share.

e. Friend request (5) tab displays the pending request from other patients.

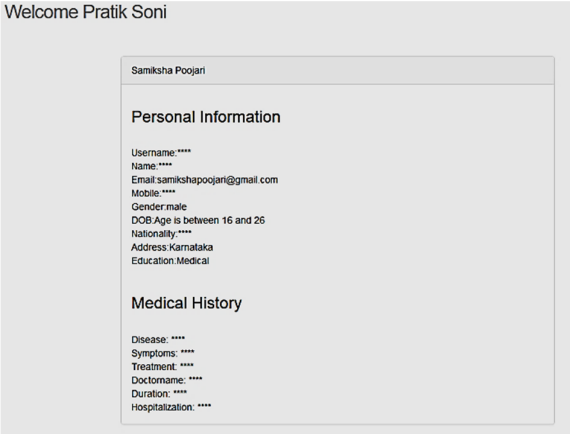
f. Friends (6) tab displays the list of users, who have accepted your request for further interaction and are ready to disclose their shared attributes.



c. Profile Matching Percentage

Search profile displays various attribute matching percentage in graphical form (Bar graph and Donut Graph). Page view is divided into following parts:-

a. Displays Personal attributes (1) of the user that are public to others.
b. Displays bar graph (2) of the personal matching attributes (Level 1 Information).
c. Displays donut graph (3) of medical matching attributes (Level 2 Information).



d. Abstracted Profile view

This page displays the Profile of the patient/ user (within friend list). It displays the attributes of the patients maintaining its visibility and privacy level.

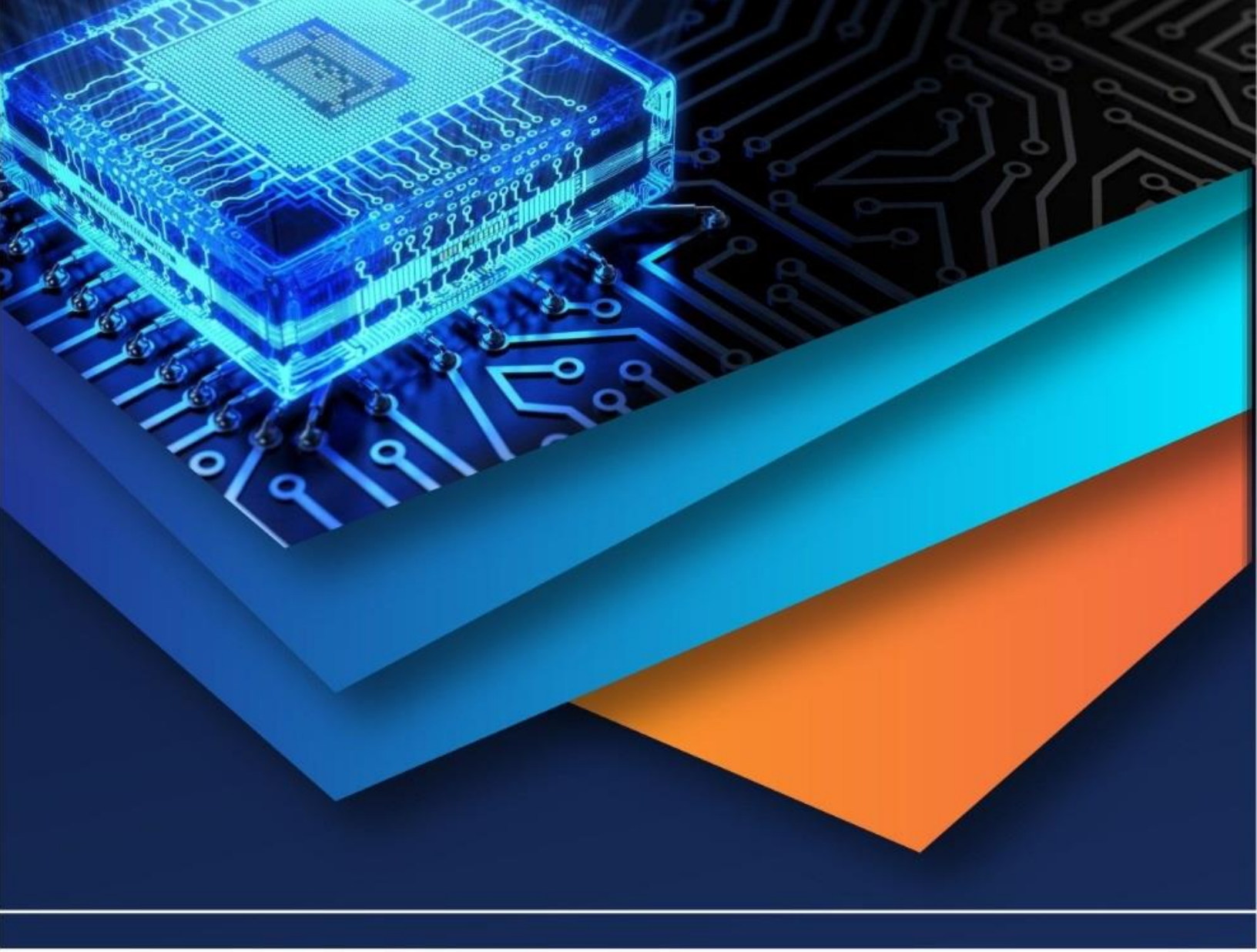
If an attribute is Private, it is not visible to any user. If an attribute is Public, it is visible to all the users (even though they are not in friend list). If an attribute is Share, it is visible to user within their friend list with a specific level of abstraction.

VIII. CONCLUSION

In this paper we have surveyed different Profile Matching Techniques. We compared different technique based on their performance as we have studied in the papers. By surveying we have seen that the privacy of the profile of users is the major issue in mobile social network, we implement the best technique which is less prone to attacks and requires less communication cost and computation cost. This paper describes group matching techniques which formalize private multiusers, private input sets, private attributes and stranger's privacy without gaining private information. We overcome the problem in privacy concern and implement secure and efficient group matching in social network. More secure and privacy preserving profile matching uses attribute based verification in social networks.

REFERENCES

[1] "Privacy-Preserving Distributed Profile Matching in Proximity-based Mobile Social Networks", Ming Li, Shucheng Yu, Ning Cao, and Wenjing Lou, 2013 5th IEEE Transaction on Wireless Communication
[2] "Practical Private Set Intersection Protocols with Linear Computational and Bandwidth Complexity", Emiliano De Cristofaro and Gene Tsudik
[3] "E-SmallTalker: A Distributed Mobile System for Social Networking in Physical Proximity", Zhimin Yang, Boying Zhang, Jiangpeng Dai, Adam C. Champion, Dong Xuan and Du Li
[4] M. Li, N. Cao, S. Yu, and W. Lou, "FindU: Private-Preserving Personal Profile Matching in Mobile Social Networks," in Proc. IEEE INFOCOM, 2011, pp. 2435 – 2443
[5] R. L. Rivest, A. Shamir, and Y. Tauman, "How to Leak a Secret," in Proc. ASIACRYPT. Springer-Verlag, 2001, pp. 552–565



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY