

USER-CENTRIC SECURE CROSS-SITE INTERACTION FRAMEWORK FOR
ONLINE SOCIAL NETWORKING SERVICES

by

Moo Nam Ko

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Information Technology

Charlotte

2011

Approved by:

Dr. Mohamed Shehab

Dr. Gail-Joon Ahn

Dr. Cem Saydam

Dr. Richard Hartshorne

©2011
Moo Nam Ko
ALL RIGHTS RESERVED

ABSTRACT

MOO NAM KO. User-centric secure cross-site interaction framework for online social networking services.

(Under the direction of DR. MOHAMED SHEHAB)

Social networking service is one of major technological phenomena on Web 2.0. Hundreds of millions of users are posting message, photos, and videos on their profiles and interacting with other users, but the sharing and interaction are limited within the same social networking site. Although users can share some content on a social networking site with people outside of the social networking sites using a public references to their content, appropriate access control mechanisms are not supported. In this dissertation, we outline a cross-site interaction framework and identity mapping approaches that enable social network users to share their content across social networking sites. We propose a cross-site interaction framework *x-mngr*, allowing users to interact with others on other social networking sites, with a cross-site access control policy. We also propose identity-mapping approaches that map user's identities across social networking sites. The partial mapping approach based on a supervised learning mechanism which provides user's identity mapping based on a training set composed of a small subset of the profile mappings. We provide mechanisms to enable users to fuse identity-mapping decisions that are provided by their friends or others on the social network. Furthermore, we propose a Game With A Purpose (GWAP) approach that provides identity-mappings using a social network game. The proposed framework and game are implemented on real social networking sites such as Facebook and MySpace. The experiments are performed to evaluate the feasibility of our approaches. A user study is also performed and the result is included as part of our

evaluation efforts for the proposed framework.

ACKNOWLEDGEMENTS

Throughout my time at UNC Charlotte, I have had the great fortune to work with outstanding colleagues. Therefore, I dedicate this dissertation to colleagues who have directly and indirectly contributed to this dissertation.

First and foremost, I would like to express my sincere gratitude to my advisor Professor Mohamed Shehab. His expert guidance and support have made this work possible and were the foundation of a great graduate research experience. I also give great gratitude to Professor Gail-Joon Ahn who as my former advisor guided the user-centric identity management research and gave me numerous supports in research.

Many Thanks to the numerous individuals who worked with me on collaborative papers, and on other topics related to my research over the last few years, including Dr. Seok Won Lee, Dr. Jing Jin, Dr. Wenjuan Xu, Dr. Napoleon Paxton, Hongxin Hu, Said Marouf, Hakim Touati, Gorrell Cheek, Sherif Fawzy, and Mitesh Doshi. Their feedback and criticism enriched my work and helped me to align it with other research projects.

I also wish to express my deep appreciation to my Ph.D. committee members, Professor Mohamed Shehab, Professor Gail-Joon Ahn, Professor Cem Saydam and Professor Richard Hartshorne. The assistance and guidance they provided in the preparation of this manuscript have been invaluable.

Finally, I would like to thank my wife Hyun Jung Lee. Her support, encouragement, patience and unconditional love enabled me to surpass hardships and complete my Ph.D. study successfully. I also thank my parent Kyung Ryul Ko and Ok Hee Sim, my sisters Sung Hee Ko and Jung Hee Ko(with brother-in-law Goon Se Lee), and my parent-in-law

Ki Wan Lee and Young Ja Lee. They were always supporting me and encouraging me with their best wishes. Finally, I would like to thank my two princesses Ellie Yumi Ko and Marie Yuji Ko. They gave me the motivation and courage to get through this work.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Statement of the Hypothesis and Approaches	4
1.2 Summary of Contributions and Dissertation Organization	5
CHAPTER 2: BACKGROUND INFORMATION	7
2.1 Definition of Digital Identity	7
2.2 Digital Identity and Privacy on the Web	8
2.3 Digital Identity and Privacy on Social Networking Services	11
CHAPTER 3: PROBLEM DOMAIN ANALYSIS	19
3.1 Sharing Mechanisms on Social Networking Sites	21
3.2 User Experience Survey for Social Networking Sites	23
3.2.1 Survey Results	24
CHAPTER 4: CROSS-SITE INTERACTION FRAMEWORK	30
4.1 User Profile on Social Networking Site	30
4.2 Cross-Site Interaction Framework	31
4.2.1 Architecture	32
4.2.2 Local Policy	32
4.2.3 Cross-Site Policy	33
4.2.4 Secure Interoperation	35
4.2.5 Policy Levels	37
4.3 Implementation of <i>x-mngr</i>	39
4.4 User Study	40

	viii
4.4.1 User Study Results	42
CHAPTER 5: PARTIAL IDENTITY MAPPING	45
5.1 Supervised Learning	46
5.1.1 Training Set Generation	49
5.1.2 Attribute and Network Distances	50
5.1.3 Classifier Selection and Fusion	52
5.2 Implementation of experimental site	55
5.3 Experimental Results	56
CHAPTER 6: IDENTITY MAPPING USING GAMES	63
6.1 Game with a Purpose	64
6.2 Definition of Profile Mapping	64
6.3 General Game Description	66
6.4 Recommendation Generation	68
6.5 Game Theoretic Analysis	69
6.6 Implementation Details	72
6.6.1 Collusion and Irrational Behavior	73
6.7 Experiments	74
6.7.1 Evaluation of Mapping Results	76
6.7.2 Evaluation of Irrational Player Detection	82
CHAPTER 7: CONCLUDING REMARKS	87
7.1 Summary	87
7.2 Future Work	88
7.2.1 Assertion Based Cross-Site Interaction Framework	88

	ix
7.2.2 Portable Social Graph with Policy	90
BIBLIOGRAPHY	92
APPENDIX A: SURVEY RESPONSE	98

CHAPTER 1: INTRODUCTION

With the evolution of the Web, various identity management models and privacy technologies have been introduced to solve the identity and privacy issues on the Web. In Web 1.0, most identity management system models such as silo model, centralized model, and federated model are designed from organization's perspective. In this environment, user's privacy concern is focused on how much user's information is stored by service providers and how much user's information is shared with other parties. Moreover, it is difficult for users to obtain information about actual data practices. In other words, privacy concern is raised by storing the user's information in service providers. To reduce this privacy concern, various privacy technologies such as P3P [20], APPEL [43] and PREP [4] have been introduced. These technologies describe the service provider's privacy policy and user's privacy preference in a machine-readable form, and provide comparison mechanisms to help users to be aware of the service provider's privacy policy practice.

With the introduction of Web 2.0, the digital identity industry recognized that existing identity management models are designed without consideration of user experience, which lead the proposal of the user-centric identity management model that allows users to control their own digital identities in the middle of the transaction between identity providers and service providers. Therefore, users have more rights and control over their identities. The users are able to decide which identity attributes they want to share with other service

providers in the middle of transaction. However, the rapid growth of online social networking services in the Web 2.0 changes the user's privacy game. Hundreds of millions of users have accounts on social networking sites. Users build social connections with families, friends, and coworkers by sharing various contents via their profile pages. Updating the user profile pages with attractive content is a form of self-expression that increases the interactions between friends within the social networking sites. The posted content on the user's profile pages is shared with friends or others in public, but the users are often not aware of the size of the viewers accessing the content on their profile. The posted content can be re-distributed by the viewers, and eventually the content can be shared with unintended users who were not explicitly allowed to view that content. Such open sharing availability of social networking sites exposes the users to a number of privacy risk [64]. Therefore, how to control the sharing of content with friends on the social networking sites becomes critical to protect the user's privacy [22, 47, 52].

Social networking sites provide different sets of services. For example, Facebook and MySpace provide services that help users to connect with people and share contents (messages, photos, and videos). On the other hand, LinkedIn provides services that help users exchange information and opportunities with a broader network of professionals. Depending on context (i.e. age, gender, location, and interest) and purpose, users select different social networking services. For instance, major users of MySpace are teenagers, and 61 % of Facebook users are 35 old or older [59]. From a location perspective, Facebook is the most popular service in North America and Europe, where Orkut is more common in India and Brazil [46]. In order to enjoy these different services, users need to create accounts on different sites and manage their scattered profiles and friends on different social network-

ing sites. For example, 64 % of Facebook users have MySpace accounts [56]. However, sharing contents with scattered friends on different social networking sites is a bothersome task to the users. Current social networking architecture provides limited content sharing mechanisms across multiple networks. Thus, to be able to share content with friends that are on different sites, the users have to upload duplicate content and set up their policies on each site. Moreover, scattered friends do not generally migrate to other social networking sites from their favorite social networking sites to access a shared content.

Given these environments, we need to address several challenging questions:

- How to build a cross-site interaction framework that enables users to share the content with scattered friends on different social networking sites?
- How to provide the cross-site interaction securely?
- How to control the friend's access across social networking sites to protect the user's privacy?
- How to map friend's identities across social networking sites effectively?

These are critical questions to be answered to assure the secure content sharing across social networking sites. Some approaches have been proposed to address the content sharing issues on Web 2.0 [15, 30, 45, 71, 72, 74]. However, as these approaches do not handle the content sharing issue between social networking sites, our study clearly indicates that there is a need to design a secure cross-site interaction framework that is general and flexible enough to cope with the specific access control requirements as well as identity mapping issues associated with the environment. In this dissertation work, we would make one step towards this direction.

1.1 Statement of the Hypothesis and Approaches

Therefore, this research hypothesizes that:

Effective identity management and access control are key to enabling cross-site interaction framework across the online social networking services.

We first explore the current content sharing mechanisms of social networking sites and conduct an online survey to understand the users' content sharing experience. From these investigation results, we formulated a set of core requirements for the cross-site interaction framework. These requirements are reflected and addressed in our proposed cross-site interaction framework *x-mngr* that manages content sharing, identity mapping, and access control across social networking sites. We present a cross-site policy which enables users to setup policies that allow/deny access to their shared contents across different social networking sites with different policy levels. To enable secure cross-site sharing, we design the *x-mngr* to support the principles of secure interoperation. We also propose three policy levels to provide different policy enforcement. In order to evaluate the feasibility and usability of the *x-mngr*, we implement a proof-of-concept application *MyCrossAlbum* and conduct a user study.

We also propose identity-mapping approaches that map user's identities across social networking sites. The partial mapping approach based on a supervised learning mechanism provides user's identity mapping refer to a small subset of the profile mappings. We provide mechanisms to enable users to fuse identity-mapping decisions that are provided by their friends or others on the social network. Furthermore, we propose a Game With A Purpose (GWAP) approach that provides identity-mappings using a social network game.

The proposed identity mapping approaches are implemented on real social networking sites such as Facebook, MySpace, and Twitter. The experiments are performed to evaluate the feasibility of our approaches.

1.2 Summary of Contributions and Dissertation Organization

The contributions of our cross-site interaction framework and identity mapping approaches are summarized as follows:

- We conduct a survey to investigate users' social networking site experience, privacy setting, and content sharing experience.
- We formulate a set of core requirements of cross-site interaction framework.
- We propose secure cross-site interaction framework with the cross-site policy and policy level.
- We evaluate the secure cross-site interaction framework through performing the user study on our proof-of-concept application.
- We propose the partial mapping approach and mechanisms that fuse identity-mapping decisions.
- We propose a Game With A Purpose approach for solving the profile mapping problem as a game supported by social verification.
- We prove the equilibrium of the game scoring mechanism using game theory to ensure that rational players will provide accurate profile mappings while playing the game.

- We implement our game as an online social networking game in Facebook, MySpace and Twitter.

The remainder of this dissertation is organized as follows. Chapter 2 reviews digital identity and privacy management from Web 1.0 to Web 2.0 and discusses the social network connect services. In Chapter 3, we explore the current content sharing mechanisms of social networking sites and discuss the survey results. Chapter 4 proposes a cross-site interaction framework *x-mngr*, introduce a prototype implementation, and discuss the user study results. Chapter 5 elaborates our partial identity mapping approach. A Game With A Purpose approach is explained in Chapter 6. Finally, Chapter 7 summarizes this dissertation and presents some directions for future work.

CHAPTER 2: BACKGROUND INFORMATION

2.1 Definition of Digital Identity

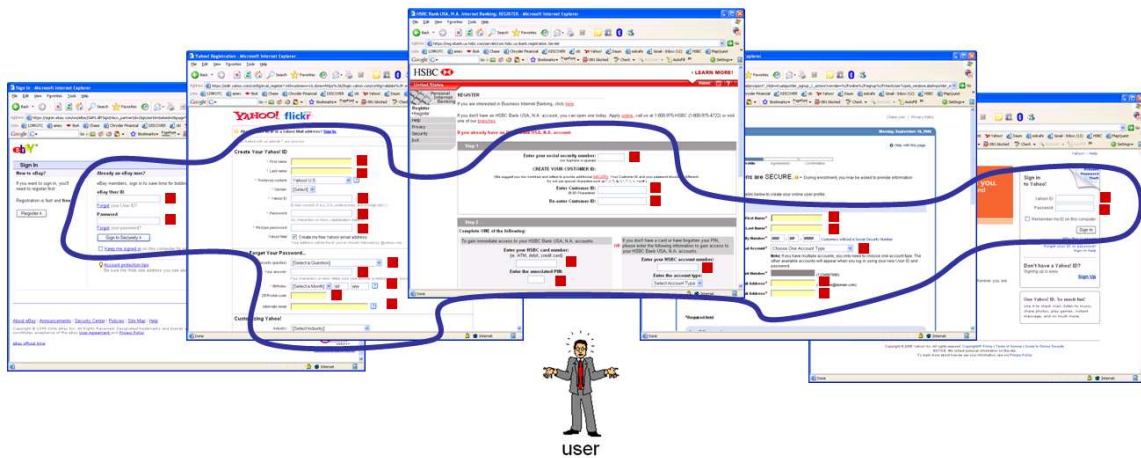


Figure 2.1: Digital Identity: Global Set of Attributes of a User

There are various definitions of digital identity. Depending on organizations, systems, and contexts, the diverse definitions of digital identity have been created and used. From our perspective, we define a user's digital identity as the global set of attributes that make up an online representation of who and what an entity is. It can include access credentials, personal attributes and personal references. Over the Internet, a user has numerous access credentials that are issued from different sites and different or duplicated personal attributes and references on each site. We believe all of these attributes should be considered as the user's digital identity as shown in Figure 2.1. In each site, a user can be represented by subsets of these attributes. Depending on the situation and the context, different subsets of attributes are used to represent the same user's identity on the Internet. For example, in an

auction site, a subset of a user's attributes such as username, password, shopping history, and reputation record represent the user's identity in this site, while a subset of the user's attributes such as a student ID number, class record, and GPA may represent the user's identity in a university site.

2.2 Digital Identity and Privacy on the Web

The rapid changes in the Internet environment have demanded the development of new identity management models with privacy technologies that support the new demands of the continuously evolving Internet environment. In the early stages of the Internet, silo identity management models were commonly used, where each organization forms its own identity management domain and has its own way of maintaining user identities that include employees, customers, and partners. In this environment, it is difficult for users to obtain information about actual data practices, which leads to online privacy concerns. Although some organizations post their human-readable privacy policies on their web sites to help build user confidence and trust in the process of personal information disclosure, it is not enough to solve the privacy concerns since the user has a lack of knowledge and the privacy policies can be complex. Moreover, the users must take additional time and effort to understand the content of the privacy policies to check whether the web site conforms to their personal privacy preferences. To reduce these efforts, privacy technologies such as Platform for Privacy Preferences (P3P) [20] and P3P Preference Exchange Language (APPEL) [43] were developed. P3P allows privacy policies to be encoded in the machine-readable form and APPEL provides a machine-readable rule set for the user's privacy preferences. P3P user agents such as web browsers and AT&T Privacy Bird [21] shows the conformance

of the service provider's privacy policy with the user's privacy preferences. These privacy agents help users to be aware of the web site's privacy policy practice.

With the evolution of the Internet, centralized identity management and federated identity management models were introduced as the next step of identity management approaches among organizations. The centralized identity management model has a single identity provider that brokers trust to other participating members or service providers in a circle of trust. A single identity provider has a centralized control over the identity management task, providing easy access to all service provider domains with simplicity of management and control. Hence, this model can reduce the maintenance cost of identity management systems. The drawback of this model is a single point of failure. If the single identity provider fails to provide authentication service, the entire systems in the circle of trust will be affected. User convenience can be also achieved partially in a case where the single sign-on for users is only effective within service providers in the same circle of trust. Microsoft Passport is a well-known centralized identity management model. Federated identity management has multiple identity providers that securely share confidential user identities with trusted organizations within or across the circle of trust. Every member agrees to trust user identities vouched for by other members of the federation. It also facilitates single sign-on and trust, thereby allowing businesses to share the identity management cost with its partners. Liberty Alliance is based on the federated identity management model. Since identity federation is likely to facilitate the voluminous exchange of sensitive user information, privacy concerns associated with such exchanges are key issues in federated identity management which have been addressed by several research projects [4, 6, 58, 69]. The multi-leveled policy approach [60] is a simplified mechanism

for handling privacy preference within Liberty Alliance framework using the standardized policy levels. It allows the reply parties to represent their intended usage for users' attributes by indicating one of the standardized policy levels. It also allows users to represent their privacy preferences for their attributes by indicating one of the standardized policy levels. This approach simplifies policy comparison and conflict resolution. Ahn et al. [2, 4] proposed a privacy preference expression language called PREP for storing the user's privacy preferences with Liberty Alliance enabled attribute providers. The PREP language enables users to tag their attributes with privacy labels and it facilitates privacy-enhanced attribute exchange.

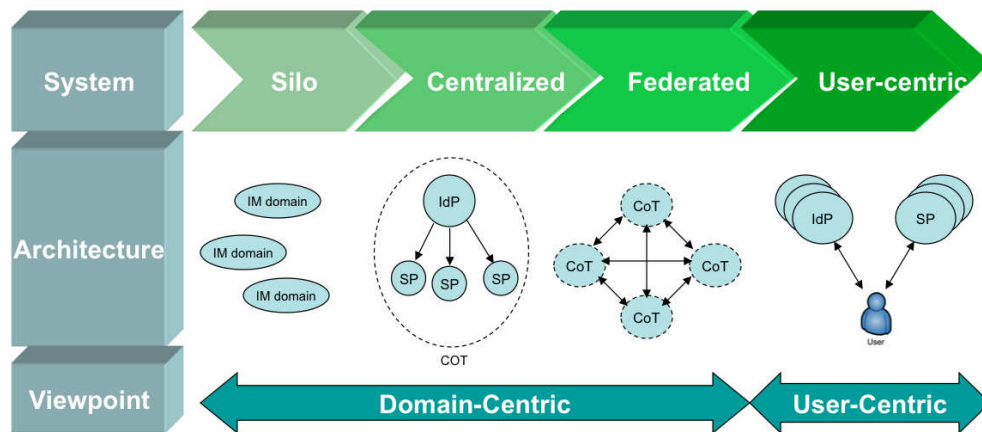


Figure 2.2: Trend of Digital Identity Management

In the beginning of Web 2.0, the digital identity industry recognized that existing identity management systems are designed without consideration of user experience and the non-interoperability between identity management systems. Silo, centralized and federated identity management models are designed from organization's perspective. Users were not considered carefully in the design stage. It leads users to be the weakest link in digital identity management systems. The user-centric identity management shifts the control of

the user digital identities from organizations to users by putting the user into the middle of transaction between identity providers and relying parties. It allows users to decide which identity attributes share with other trusted parties under what circumstance. Thereby better protection of the user's private information is enabled by user. As the users have more rights and responsibilities over their identity information, user can actively control their identities. Well-known user-centric identity management systems are OpenID [63] and MicroSoft CardSpace [10]. In the transaction, the user's understanding of the privacy conflict between relying party's privacy policy and user's privacy preference is also important to help users make a clear decision for the transaction. Ahn [3] proposed two privacy preference management approaches, category-based privacy preference and claim-based privacy preference, to the user-centric identity management model. It helps the user's understanding for the privacy conflict of the requested claims by using the different color of icons on the user interface.

2.3 Digital Identity and Privacy on Social Networking Services

The trend of social networking services began from the need of reconnecting with lost classmate. Through the social networking sites, people build their own social graph with families, friends, and coworkers and share their favorite contents such as videos, photos, and messages. Generally, social networking systems provide a profile page for each user to represent themselves. The profile page includes user's details, friends, groups, photos, videos, updates, messages, installed applications, and so on. Decorating of profile page with attractive contents is a form of self-expression which increases the interaction between friends on social networking sites. This made the social networking sites to be popular

immensely. Various content generated by users and interactions between users increase the security and privacy risks.

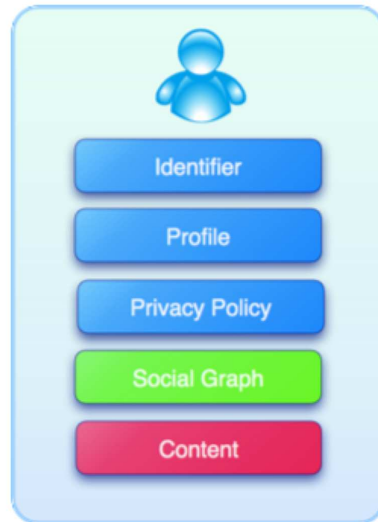


Figure 2.3: User Data on Social Network Service

Figure 2.3 describes the user data of social network services that is composed of three categories represented as: Identity data, social graph data, and content data. The identity data presents “who I am”, which includes the user’s identifier, profile data, and privacy policy. The social graph data presents “who I know”, which includes friendship connections and their descriptions such as family, coworker, and friend. The content data presents “what I have”, which includes the user’s messages, photos, and all other data objects created by user through the social networking activities. In Web 1.0, the user’s privacy concerns mainly focus on their identity data to protect sensitive user data such as birthday, address, and social security number. However, this trend of privacy concerns has been changed in Web 2.0 since all users can easily post content such as message, photo, and video and share it with other users. Especially, the user should be careful to share the content with friends

to protect their privacy.

Most social networking sites provide an access control features to protect user data on social network services. A simple solution is to make profiles either public or private. Public profile can be viewed by anyone, while only an individual's friends can access the private profile. More complex solution is relationship based access control model since it strikes the best balance between ease-of-use and flexibility. Profile owners can define access control policy in privacy menu in their profile page. If a profile owner assigns an appropriate relationship to a content or service, users who have the assigned relationship are allowed to access the content or service. For example, if Alice assigns the friend relationship to her photo album, her friends who have the friend relationship can access the Alice's photo album on her profile page.

Some researchers have proposed different access control scheme for social networking services. Kiran et al. [30] presented a social-networking based access control scheme suitable for online sharing of personal media. The authors consider the user identities as key pair and social relationship on the basis of social attestations. Access control lists are employed to define the access lists of users. Barbara et al. [15] proposed a more sophisticated rule-based access control model for social networks. It enforced complex policies expressed as constraints on the type, depth, and trust level of existing relationships. The authors also proposed using certificates for granting relationships authenticity, and the client-side enforcement of access control according to a rule-based approach, where a subject requesting to access an object must demonstrate that it has the rights of doing that. These papers focused on access control within a single social networking site and did not consider the access control for sharing content with outside of social networking sites.



Figure 2.4: Privacy Setup on Facebook

The architecture of social network service was changed dramatically after launching the developer's APIs and application platforms (containers). By allowing third party developers interact with the social networking sites through exposing web services in the form of APIs, social networking sites, third party sites, and users are possible to enjoy the benefit of APIs. The social network application platforms allow third party developers create applications that run on the social networking sites. The third party applications usually provide new services using the social network data and their own data. It helps users to decorate their profile page with rich contents that encourage the interaction between friends. A well-known social networking platform is the Facebook Application Platform. Although the Facebook Application Platform is powerful, other social networking sites could not use it since this application platform only supports the applications based on Facebook core technologies such as FBML, XFBML, FQL, FBJS and API. Google and other social networking sites introduced OpenSocial that defines a common API for social applications across multiple websites. It allows one OpenSocial application to be executed on multiple social networking sites. Unlike Facebook application, OpenSocial application uses the standard technology such as HTML, XML and Javascript. Many social networking sites including Orkut, MySpace, Hi5, LinkedIn, Netlog, Ning and Yahoo support the OpenSo-

cial.

In the social application model, the data flow looks like this:

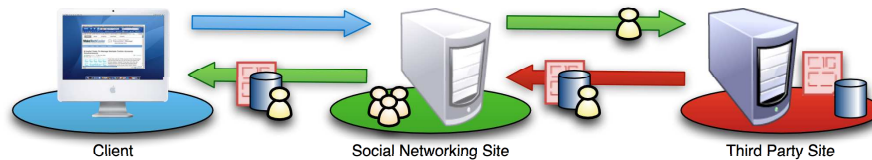


Figure 2.5: Data flow of Social Application

1. A user invokes a social application in a social networking site.
2. The social networking site provides user data such as profiles, social graphs, and their contents to a third party site.
3. The third party site renders a page for the user and sends it the social networking site.
4. The social networking site send the created page to the client machine and then it is showed to the user.

Social application is a big step in the evaluation of social network services in order to move from the walled garden to open environment. It helps the social networking sites to provide various application services to their users. It also helps the third party sites to distribute their services rapidly via social networking sites, and keep in touch with their users via social networking sites again. Moreover, the users can enjoy various applications with contents on their profile page on social networking sites. For example, Facebook users can share music with friends, create playlists and get concert alerts on their profile page by installing the iLike's music application.

These days, major social networking services have launched a new service such as Facebook Connect, Google Friend Connect, and MySpaceID. Here, we will call these new services “Social Network Connect Services”. The Google Friend Connect and MySpaceID use OpenID technology and Facebook Connect use their own technologies. The Social Network Connect Service enables any web site to extend its services to accommodate social services without having to either host or build up its own social network. This allows users to use their social features in other web sites without creating a username, password, filling out a profile, and re-connecting friends. The user’s Internet activities also can be shared with friends on social networking sites. Closed social networking sites lock their users inside of their sites and do not share user’s social web data with others outside the social networking sites. However, the Social Network Connect Service allows users to interact with their friends regardless of where they are and where their friends are and take the advantages of implicit social features in any place. By providing seamless Social Network Connect Service across the web sites, the social networking sites become identity providers in Web 2.0.

Figure 2.6 shows the change of new user’s identity management selection in the registration process in TypePad during three months [38]. They provide various identity providers in login and registration page and give a choice to their users to select their identity providers. At the beginning, most new users selected the traditional silo identity management system that is TypePad’s own identity management system. However, this identity management selection trend was changed from silo to user-centric while the Social Network Connect Service was widely spreading in the web. Three months later, many new users selected the user-centric identity management systems in the registration pro-

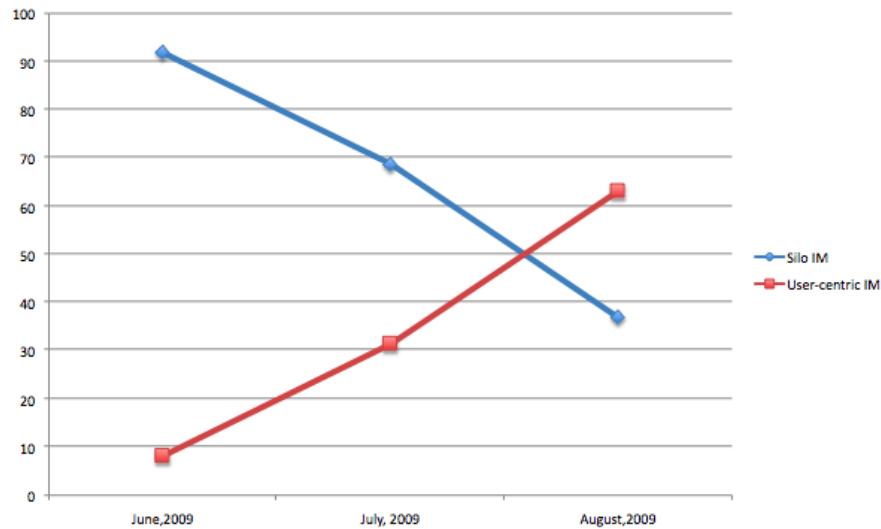


Figure 2.6: Change of Identity Management Selection by Month

cess. From this trend change, we figure out many Internet users select the major social networking sites, Facebook, Twitter, Yahoo and MySpace, as their identity providers.

In the Social Networks Connect Service, the data flow looks like this:



Figure 2.7: Data flow of Friend Connect

1. A user invokes a service in a third party site. (Assumption: the user is already registered with the third party site using Social Networks Connect Service).
2. Third party sites request the user's data to the social networking site.
3. The social networking site provides requested user's data to the third party site.
4. The third party site renders a page for the user using the user data.

5. The third party site sends the created page to the client machine and then it is showed to the user.

CHAPTER 3: PROBLEM DOMAIN ANALYSIS

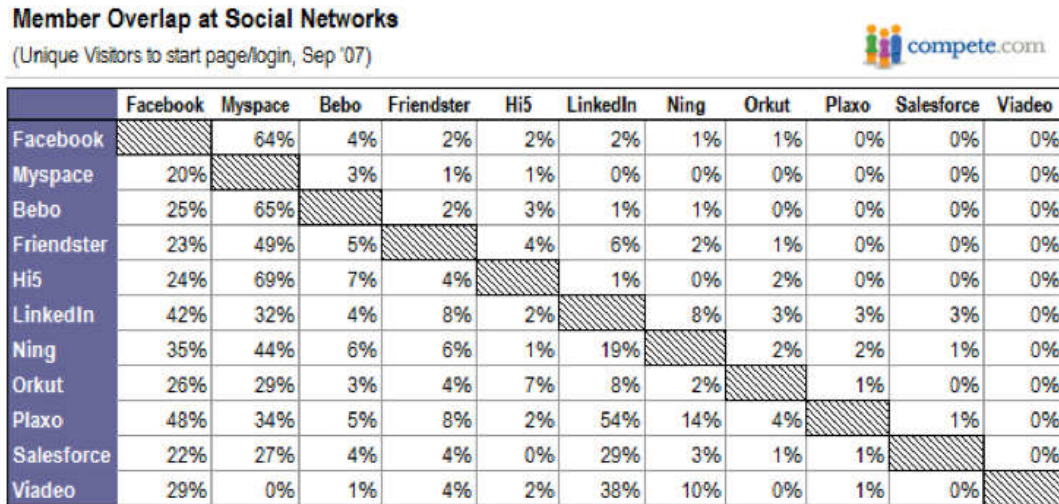


Figure 3.1: Member Overlap between Social Networking Services (image source: [56])

Social networking sites, including Tagged, Xanga, Friendster, LiveJournal, MySpace, Facebook, and LinkedIn have developed on the Internet over the past several years. The popularity of social networking sites on the Internet introduces the use of mediated-communication into the relationship development process. According to ComScore Media Metrix, more teens visit MySpace than Yahoo, MSN, or Electronic Arts gaming site. In addition, more than half or 55 % of all online teens use social networking sites [39]. Currently, a new type of communication behavior is emerging among young Internet users as they explore their identities, experience with behavioral norms, and build friendships. Social networking sites play a key role in youth culture in cyberspace [34]. Different social networking sites

provide users with different sets of services, for example, second-Life provides a virtual 3D environment for users to build their own virtual spaces and interact with other users. Other sites, such as MySpace, do not provide a virtual 3D environment, but at the same time it hosts the highest number of users. To enjoy these services, users end-up creating accounts on different sites. Figure 3.1 shows the member overlap at the social networks. For example, 64 % of Facebook users have Myspace accounts and 69 % of Hi5 users have Myspace accounts [56].

Current social networking architectures do not provide appropriate interaction mechanisms between users on different social networking sites. For instance, a user i who has a friend relationship with the user a in SN_A can not directly access the user a 's photo album in SN_B like Figure 3.2. To share a photo with the user i , the user a have to upload the same photo album in both sites or the user i have to create an account in SN_B and have an appropriate friend relationship with user a .

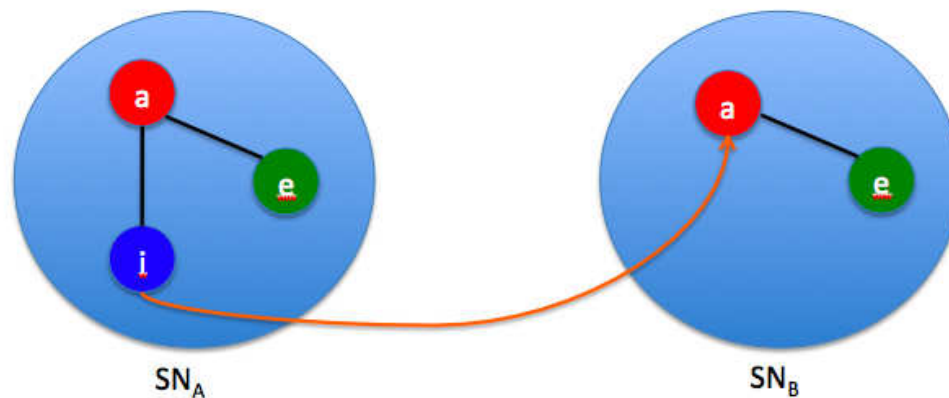


Figure 3.2: Limitation of Interaction between Social Networking Sites

Enabling cross-site interaction beyond social networking site boundaries is a challeng-

ing task that is related to both the semantics and the policies of the involved sites. Moreover, user privacy protection should be considered carefully since inappropriate disclosure of content on social networking sites has led various privacy issues. For instance, people have been denied or lost jobs because of their comments on social networks [5, 62]. Students were suspended after making derogatory comments on Facebook [16]. Therefore, when a user shares content across the social networking sites, a cross-site interaction model must prevent unintentional disclosure of content to an inappropriate user on different social networking sites.

In order to understand the challenge of cross-site interaction more deeply, we explore the current content sharing mechanisms of social networking sites and perform an online survey for users' social networking experience, privacy preference, and content sharing experience.

3.1 Sharing Mechanisms on Social Networking Sites

Content sharing on social networking sites can be classified into the internal sharing and external sharing. Most social networking sites provide similar sharing mechanisms for the internal sharing. Based on users' privacy policies, only authenticated and authorized users are allowed to access shared contents within a social networking site. In Figure 3.3, Alice is the owner of the content (public photo, private photo) and has friend relationship with Bob and Carol. If Alice specifies her privacy policy to allow users who have the friend relationship with her to access the private photo, only Carol and Bob are able to access Alice's private photo album. Ted is not able to access since he does not have the friend relationship with Alice.

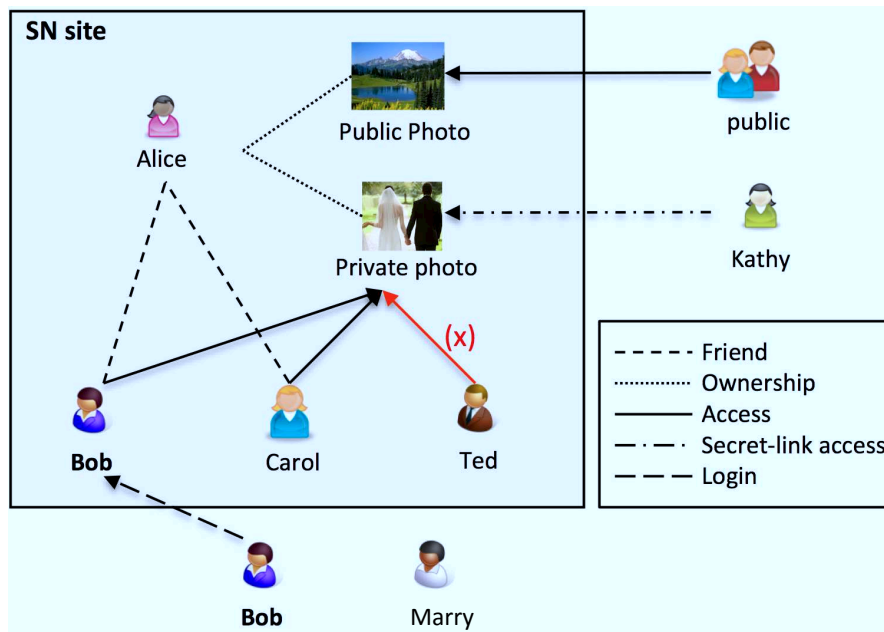


Figure 3.3: Content Sharing Mechanisms on Social Networking Sites

On the other hand, current social networking sites provide immature sharing mechanisms for the external sharing. One solution is to make the content publicly so that everyone can access it. Another solution is to send a secret-link of content to friends outside of a social networking site using email or messenger. As Figure 3.3, Kathy is able to access the private photo from external of the social networking site since she knows the secret-link. Another newly introduced solution is to connect social networking accounts between different social networking sites. For example, if users connect Facebook and Twitter accounts for the status update service, their updated statuses are shared between the two sites. However, these all external sharing mechanisms are not sufficient to users who want to share content from a favorite social networking site to other social networking sites in a controlled manner. Making content in public is inadequate to protect a private content. A secret-link has a usability issue with security concern since users have to send a secret-link

manually to specific friends via email. Moreover, the secret-link can be forwarded to others who are unauthorized users. In order to reduce this security concern, some sites request login to their sites to access the shared content, but it is inconvenient to users who do not have accounts. Connecting social networking accounts only provide opt-in/out options. Strictly, connecting social networking accounts is not a sharing mechanism since it posts the same content across social networking sites. We believe a better cross-site sharing mechanism is necessary between social networking sites.

3.2 User Experience Survey for Social Networking Sites

We conducted a survey to understand user's content sharing experience on social networking sites and identify the necessity of cross-site content sharing¹. The survey investigated users' social networking experience, privacy setting, and content sharing experience. On users' social networking experience, we inquired whether participants have accounts in multiple social networking sites, for what reasons are participants using multiple social networking sites, and how often participants access social networking sites. On user's privacy setting, we inquired what kinds of relationships are on social networking sites, whether participants organize customized group to apply different privacy setting, and whether participants block or except any friends to protect their privacies. On content sharing experience, we inquired whether participants have experienced the cross-site content sharing problem, whether participants like to share a content with a specific list of friends on other social networking sites, and what content sharing services prefer to use. The survey consisted of a mix of multiple choice questions, yes/no questions, and Likert scale questions. We posted

¹IRB Protocol No: 09-03-16, Title: Cross-site Interaction between Social Networks

advertisements to social networking sites, micro blog sites, and campus to attract our desired demographic of the general class of Internet users. We recruited 403 participants to start the survey, of which 306 participants completed the online survey and 97 participants did not complete the survey. The participants received entry into a drawing for 4 iTunes gift cards (\$15), 30 Osfoora for iPhone APP, and 30 OsfooraHD for iPad APP. We investigated the survey results of 306 participants who completed the survey.

3.2.1 Survey Results

- Most participants have accounts on multiple social networking sites.

87.6 % of participants reported they have accounts on multiple social networking sites such as Facebook, MySpace, Orkut, LinkedIn, Twitter, YouTube, and Flickr. They selected Facebook (82.1 %) and Twitter (81.3 %) as mainly used social networking sites. The participants also reported the reason of using multiple social networking sites is to get different services or to meet the scattered friends on different social networking sites (Table 3.1). “Twitter for fun, Facebook to keep in touch with family” one participant noted. In the question about spending time of social networking services, 40.8 % of participants said they access their social networking sites a few times a day and 42.8 % of participants said they access the social networking sites constantly (Table 3.2). It shows most participants’ daily life is connected with their favorite social networking sites. We also asked about their experience of social applications and social connection services such as Facebook Connect, MySpaceID, and TwitterID. 67.0 % of participants have used social applications such as Lockyou and Farmville and 65.0 % of participants have used the social connection services to register or login to other 3rd party sites.

Table 3.1: To get different service and meet scattered friends are main reasons to have accounts in multiple social networking sites. (N=268, multiple responses and manual input allowed)

For what reasons are you using multiple social networking sites?

To get different services	76.5 %
(Facebook: fun, LinkedIn: professional)	
Friends are scattered	51.1 %
(College friends: facebook, Indian friends: orkut)	
To meet others who have similar hobbies	24.3 %
For curiosity	26.5 %
Other	6.0 %

Table 3.2: 91 % of participants access their social networking sites at least one time a day (N=306)

How often do you access social networking sites?

Constantly	42.8 %
A few times a day	40.8 %
One a day	7.5 %
Once or twice a week	6.9 %
Once a month or less	1.0 %
No answer	1.0 %

- Some participants organized friends using the Friend List to apply different privacy settings.

Participants described their friendship mainly consisted of family, school friends, co-workers, and acquaintance. 92.2 % of participants who had accounts on multiple social networking sites reported they used similar privacy setting between social networking sites. 35 % of participants stated they organized their friends using the Friend List to apply different privacy settings. Their average number of Friend List is 4-6 Friend Lists ($\sigma = 0.78$). They generally categorized their friends based on friends' affiliation and friendship (Table 3.4). 90.7 % of them agreed that the Friend List is helpful to protect privacy on social networking sites. We also asked to the participants who were not using the Friend List about the reason why they were not using the Friend List. The participants mentioned various reasons. 30.8 % of participants said they did not know about the Friend List. 30.8 % of participants said they did not have many friends to use the Friend List. 22.7 % of participants said they wanted to use it, but they were lazy. Other participants mentioned that they did not need it. These results showed participants who concerned about their privacies on social networking sites organized friends into different Friend Lists and applied different privacy setting. In addition, 67.3 % of participants reported they had blocked someone on social networking sites to protect their privacy.

- Social networking services become major content sharing tool.

For sharing content such as photo, video, and others, we found social networking services to be the most common route (Table 3.5). 46.1 % of participants reported they preferred to use social networking sites for sharing content. It showed social networking services were closely connected the daily life of participants and they preferred to use it as a sharing tool

Table 3.3: Most participants have similar relationships with friends (N=306, multiple responses and manual input allowed)

What kind of relationships are between you and your friends
on social networking sites?

Family	86.9 %
School friends	84.9 %
Co-workers	78.1 %
Acquaintance	68.0 %
Neighbor	23.9 %
Other	11.1 %

Table 3.4: Most participants categorize their friends based on affiliation and friendship (N=108, multiple responses allowed)

How do you categorize your friends into friend lists?

Based on friends' affiliation (same school or same company)	60.2 %
Based on friendship (best friends or just friends (acquaintance))	63.0 %
Based on location or nationality	6.5 %
Based on common interest	33.3 %
Based on common features (gender, religious, or relationship status)	11.1 %

Table 3.5: Most participants prefer to use social networking service for sharing content (N=306)

When you share a content such as photo, video, and others with friends, what service do you prefer to use?

Social network services such as Facebook and MySpace	46.1 %
Content sharing services such as Flickr and Youtube	16.3 %
Email Services such as Hotmail and gmail	9.1 %
Micro blog service such Twitter	26.8 %
Personal blog services such as LiveJournal and Blogger	1.6 %

than other sharing methods.

- A content sharing service between social networking sites is necessary.

44.4 % of participants had the same experience that they posted same content different social networking sites to share it with scatted friends. We asked participants' opinion about the necessity of a content sharing service between social networking sites. It was measured on a Likert scale (5 point rating scale, where 1 = Strongly Disagree and 5 = Strongly Agree). Participants took positive attitudes toward the content sharing service between social networking sties was necessary ($M = 3.58$, $SD = 1.03$). We also asked participant's opinion about the sharing content with a specific Friend List on other social networking sites using the same Liker scale. Participants also took positive attitudes toward the Friend List ($M= 3.55$, $SD = 0.98$). These results suggested us to share content from one social networking site with specific Friend Lists or friends on other social networking sites.

To illustrate our challenge, we will use the following scenario throughout this disserta-

tion.

“Alice’s high school friends and music club friends are mainly using the social networking site B (SN_B), and her college friends and coworkers are using the social networking site A (SN_A). To maintain online friendship with them Alice has accounts on SN_A and SN_B . Her friends on two social networking sites don’t want to migrate or access other social networking site, so Alice has uploaded same content to both sites whenever she would like to share some content with them. One day, Alice wants to share her wedding album with high school friends in SN_B and college friends in SN_A . However, she does not want to share the wedding album with her ex-boyfriend Bob.”

CHAPTER 4: CROSS-SITE INTERACTION FRAMEWORK

4.1 User Profile on Social Networking Site

Users and relationships between users are the core components of social networks. Each user manages an online profile, which usually includes information such as the user's name, birth date, address, contact information, emails, education, interests, photos, music, videos, blogs, and many other items. Each user $u_i \in V$ maintains a profile, which is composed of N profile attributes, $\{A_1^i, \dots, A_N^i\}$. Each attribute is a name-value pair (an, av) , where an and av represent name and value respectively. For example, a Facebook user profile includes attributes such as birthday, location, gender, religion, etc. Users are also able to post objects such as photos, videos, and statuses to their profiles to share with other users.

Users are connected to a set of friends, using this notion a social network can be modeled as an undirected graph $G(V, E)$, where the set of vertices V is the set of users, and the set of edges E is the set of friendship relationships between users. The edge $(u_i, u_j) \in E$ implies that users u_i and u_j are friends. Using the graph-based model for social networks, we leverage the node network structural properties to provide additional user attributes. These attributes include several small world network metrics such as node degree centrality, betweenness, hit rate, eigen values [12, 50]. Each metric provides a different indicator about the user, for example the degree shows how popular is a user, Short et al [68] used the centrality measures of degree and betweenness to analyse relationships between street gangs

members. For a user u_i , we are able to compute M network metrics $B_i = \{B_1^i, \dots, B_M^i\}$. Each metric provides a different indicator about users in a given social network [51, 50, 36]. Each user u_i in a social networking site maintains a collection of user profile attributes and a set of user friendships of which social network metrics are computed, $P_i = \{A_i, B_i\}$. The neighborhood of user u is the subgraph $\mathcal{N}_u = (V_u, E_u)$, where $V_u = \{v | v \in V, (u, v) \in E\} \cup \{u\}$, $E_u = \{(x, y) | x, y \in V_u, (x, y) \in E\}$.

4.2 Cross-Site Interaction Framework

We propose the *x-mngr* framework for managing content sharing and access control between social networking sites. The *x-mngr* puts the content owners in the middle of the content sharing process between social networking sites. It gives the content owners the right to select a policy level for sharing content to enforce the different levels of policy. The *x-mngr* operates under the principles of secure interoperation. The details of the *x-mngr* framework are discussed in subsequent sections.

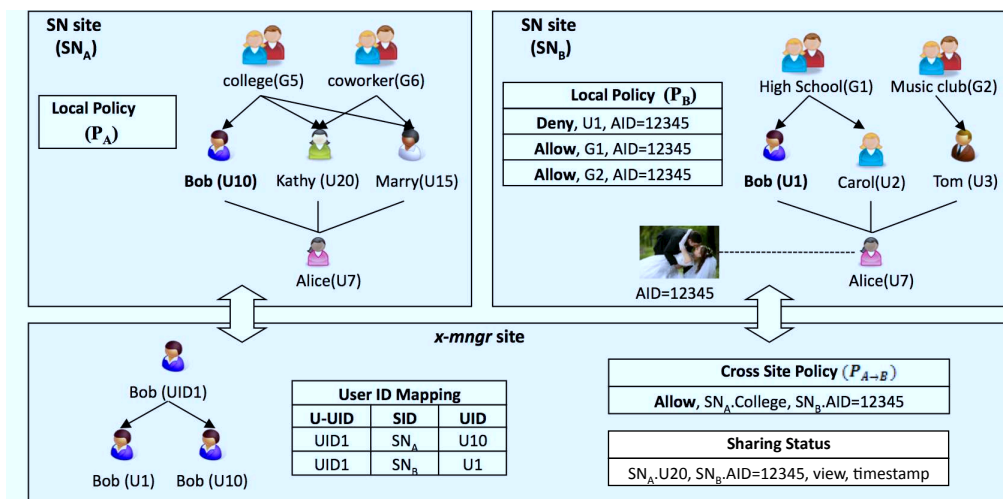


Figure 4.1: Architecture of *x-mngr* Framework

4.2.1 Architecture

As shown in Figure 4.1, the *x-mngr* is located between social networking sites. The social networking sites manage their own users, contents, and local policies. We refer to the SN_B that hosts the shared content as the *target* site, and the SN_A that accesses the shared content as the *viewer* site. We assume that a trusted party operates the *x-mngr* between the social networking sites. The *x-mngr* manages user's identity mapping across the social networking sites. The user identity mapping includes the content owner's identity mapping and friends' identity mapping. Each user's identities are mapped by a unique identifier in the *x-mngr*. For example, Bob's user id is ($U1$) on SN_A and ($U10$) on SN_B , and it is mapped by user id ($UID1$) on the *x-mngr*. The *x-mngr* also manages the cross-site policies that control access across social networking sites and the policy levels for the shared content. All cross-site sharing states are recorded in the Sharing State. It is used to help the content owners to maintain awareness about what they have shared and with who, which policy level is enforced, and who has accessed it.

4.2.2 Local Policy

In each social networking site, a user posting an object (content) O on her profile is allowed to setup an access control policy to specify which friends are allowed/denied access to the posted object within each social networking site. We named this access control policy *LocalPolicy*. The local policy is managed and stored by each social networking site. We define the local policy as:

Definition 1 (*Local Policy*) Given a social networking site, the local policy P of an object O is defined using two access control lists, namely the allow list ACL^+ and the exception

list ACL^- , which are sets of the allowed and the denied users or groups respectively. Access control follows the closed world assumption, where if access is not explicitly specified it is assumed to be not accessible. For an object O given ACL^+ and ACL^- , a user u is given access to O with allowed operation OP iff $u \in ACL^+$ and $u \notin ACL^-$, or in compact form $u \in (ACL^+ \setminus ACL^-)$. The full format of local policy is (O, ACL^+, ACL^-, OP) .

The role of ACL^+ is to enumerate the friends or groups who are allowed to access the content with allowed operation. On the other hand, the role of ACL^- is to enumerate the friends or groups who are not allowed to access the content even if they hold an appropriate group relationship to access the content. For example, in SN_B the user Alice would like all her high school friends (Group G_1) and music club friends (Group G_2) to be able to access her posted wedding photo album except her friend Bob (User $U1$) refer to Figure 4.1 . Accordingly, for this photo album $ACL^+ = \{G_1, G_2\}$, $ACL^- = \{U1\}$, and access is only given to users in $\{G_1, G_2\} \setminus \{U1\}$. This exception based group approach (ACL^+, ACL^-) is commonly adopted by the state of the art of social networking sites such as Facebook.

4.2.3 Cross-Site Policy

The x -mgr manages access control policy for shared content across social networking sites. We named this policy *Cross-Site Policy*. The cross-site policy is defined as follows:

Definition 2 (*Cross-Site Policy*). Given a viewer site SN_A and a target site SN_B , the cross-site policy $P_{A \rightarrow B}$ specifies the access control list (O, ACL^+, ACL^-, OP) w.r.t subjects from the viewer site SN_A and objects from the target site SN_B .

For instance, Alice posted her wedding album in the site SN_B . Alice would like to share the wedding album in site SN_B with her college friends (Group G_5) in the site SN_A . The cor-

responding cross-site policy for Alice's wedding album is $P_{A \rightarrow B} = (ACL^+ = \{SN_A.G_5\}, ACL^- = \{\})$. Figure 4.2 shows that the sites SN_A and SN_B manage the local policy P_A and P_B respectively and the $x-mngr$ manages the cross-site policy $P_{A \rightarrow B}$.

```

<crosssitepolicy>
  <content>
    <aid>12345</aid>
    <type>album</type>
    <url>http://photo.SNB.com/...</url>
    <owner><uid>U7</uid></owner>
    <site><sid>SNB</sid></site>
    <policylevel>strict</policylevel>
  </content>
  <acl>
    <allow>
      <site>
        <sid>SNA</sid>
        <user></user>
        <group><gid>G5</gid></group>
      </site>
    </allow>
    <deny>
      <site>
        <sid>SNA</sid>
        <user></user>
        <group></group>
      </site>
    </deny>
  </acl>
  <permission>read</permission>
</crosssitepolicy>

```

Figure 4.2: The XML format of *Cross-Site Policy*

The cross-site policy includes content information, access control list, and permission elements as described in the XML representation in Figure 4.2. The content information includes details of shared object such as content id, type, url, owner, site, and policy level. The access control list information includes the users and groups who are assigned to ACL^+ and ACL^- respectively. The permission includes the allowed permissions. The default permission is read.

4.2.4 Secure Interoperation

From the background investigation in Chapter 3, we found the current content sharing mechanisms on social networking sites are not enough to meet the desire of the current social networking users who have accounts on multiple social networking sites. The users have the willingness to share their contents from a favorite social networking site with scattered friends on other social networking sites in a controlled manner. We also found the Friend List is useful to organize friends and apply different policies. From those results, we formulate a set of core requirements of *x-mngr* as follows:

- **R1.** The content owner should be able to share contents from his/her favorite social networking site to other social networking sites.
- **R2.** Friends on the viewer sites do not need to create accounts on the target site. Friends are able to access the shared content from their favorite viewer sites.
- **R3.** The content owner should be able to set up access control policies for sharing contents across sites. The content owner should be able to specify the policies using Friend Lists or friends.
- **R4.** Shared content should not be accessed by unintended friends across sites. For example, if Alice blocks Bob to access her wedding album on one social networking site, Bob must be blocked for accessing the shared wedding album from other social networking sites.
- **R5.** The content owner should be able to know which content is being access by who, when, and where, and be able to revoke an authorization at any time if necessary.

- **R6.** The content owner should be able to select a different policy enforcement for sharing content based on her privacy concern for sharing content.

In addition to the core requirements, the *x-mngr* should maintain both the autonomy and security principles of secure interoperation [31, 32, 66]. The autonomy principle requires that any access permitted within an individual site must also be permitted in the same site under secure interoperation. The security principle requires that any access not permitted within an individual site must also be denied under secure interoperation.

Definition 3 (*Safe*). *The x-mngr is safe if it does not deny legal requests or permit illegal requests from a viewer site to a target site.*

The *x-mngr* has no control on enforcing the local policy on local sites. For example, the local policy P_B is controlled and enforced by site SN_B irrespective of the *x-mngr* decisions. It implies that the *autonomy principle* is obeyed. The challenge is to enforce the *security principle* as it requires the *x-mngr* to deny access to objects that would have been denied by the target site's local policy P_B . For an object $O \in SN_B$ with a local policy P_B defined as $P_B.ACL^+$ and $P_B.ACL^-$, and a cross-site policy $P_{A \rightarrow B}$ defined by $P_{A \rightarrow B}.ACL^+$ and $P_{A \rightarrow B}.ACL^-$, a user u from the viewer site SN_A is given access to object $O \in SN_B$ if all the below conditions are satisfied:

- **C1.** $u \in P_{A \rightarrow B}(ACL^+ \setminus ACL^-)$
- **C2.** $M_{A \rightarrow B}(u) \notin P_B(ACL^-)$

The condition (C1) ensures that the requesting user $u \in SN_A$ from the viewer site is permitted access via the cross-site policy $P_{A \rightarrow B}$. The condition (C2) involves the user identity

mapping function $M_{A \rightarrow B} : u \rightarrow v$, where $u \in SN_A$ and $v \in SN_B$, which maps a user u from a viewer site to a corresponding user v from the target site. The mapped user $v = M_{A \rightarrow B}(u)$ is checked against $P_B(ACL^-)$ to ensure that this user is not explicitly denied access by being in the exception access list in the *target* site SN_B . The condition (C2) ensures that the exception list of the *target* site is respected, and it is not violated when requests are made through the *x-mngr* framework.

4.2.5 Policy Levels

In subsection 4.2.4, the condition C1 and C2 are only applied when a viewer has accounts on both SN_A and SN_B and the content owner has specified the $P_B.ACL^-$. Depending on the state of viewer's accounts and the privacy sensitivity of the shared content, the condition C1 and C2 vary. For example, some viewers only have accounts on the viewer site, whereas other have accounts in both sites. Some contents are very private, so it might be shared with a specific friend group on the viewer sites. To support various cross-site sharing cases, we formulate three different policy levels as Figure 4.3 describes.

- *StrictLevel*: The first condition (C1) ensures that a viewer $u \in SN_A$ from the viewer site is permitted access via the cross-site policy $P_{A \rightarrow B}$. The second condition (C2) ensures that the mapped user $M_{A \rightarrow B}(u)$ is not explicitly denied by the local policy $P_B(ACL^-)$. If the viewer u has accounts on both the viewer site and the target site, the viewer must satisfy the condition (C2), but if the viewer u only has an account on the viewer site, the condition (C2) is not enforced. The strict policy level fits to the content owners who want to share a private content, and have blocked friends on the target site.

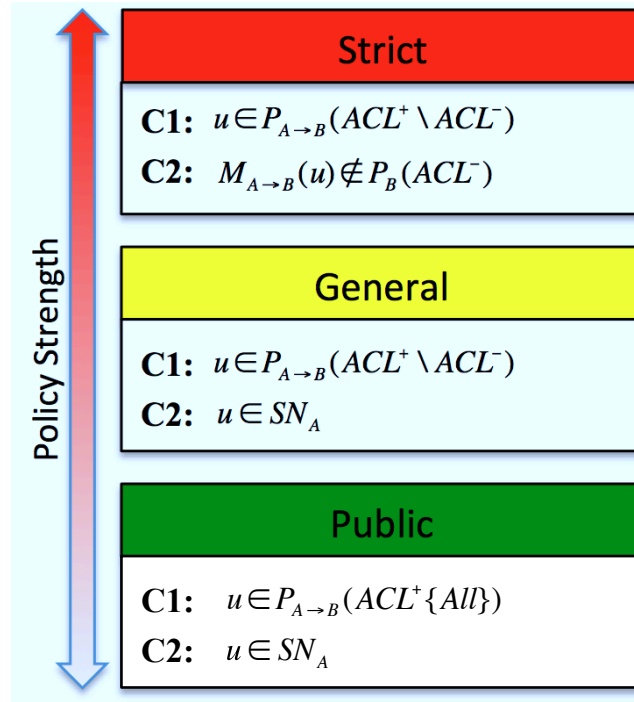


Figure 4.3: Policy Levels

- *GeneralLevel*: it focuses on the cross-site policy. The first condition (C1) ensures that the viewer $u \in SN_A$ from the viewer site is permitted access via the cross-site policy $P_{A \rightarrow B}$. The second condition (C2) is not considered since the viewer u only has account on the viewer site. This policy level is suitable for general content sharing between social networking sites.
- *PublicLevel*: The cross-site policy is setup as $P_{A \rightarrow B}(ACL^+ \{ALL\})$. By adding ALL value to ACL^+ , any viewer $u \in SN_A$ from the viewer site can access the shared content. The second condition (C2) is not considered since the viewer only has an account on the viewer site.

Depending on sensitivity of sharing content and local policy, the content owners are able to assign three different policy levels to the sharing content. It will give the content

owners more flexibility in policy enforcement for sharing content.

4.3 Implementation of *x-mngr*

As a proof of concept, we provided an implementation of *x-mngr* framework between Facebook and MySpace. We developed a social application named *MyCrossAlbum* that enables users to share photos with their friends between Facebook and MySpace. The *MyCrossAlbum* application is built by Adobe Flex 4.0 to provide a rich user experience. By using the Facebook Connect and MySpaceID, the users are able to connect their accounts on the *MyCrossAlbum* application. The Facebook Graph API and MySpace RESTful API were used to access the owner's profiles, Friend Lists, and friends' profiles in both Facebook and MySpace. The *x-mngr* site is developed using PHP and MySQL technologies. It manages user's identity mapping, cross-site policies, and shared state. It provides APIs to the *MyCrossAlbum* application in order to exchange these data. We assumed the *x-mngr* site is a trusted party between involved social networking sites, and the local policy is stored in each social networking site ¹.

Our prototype version of *MyCrossAlbum* has several menus such as *about*, *Sharing Photos*, *My Photos*, *Friends' Photos*, *Friend Mapping*, and *Sharing State*. The *about* menu provide a brief introduction of the application, and Facebook connect and MySpaceID to help the owner to connect their accounts. The *Sharing Photos* menu enables the owner to add a photo to *MyCrossAlbum* and specify the cross-site policy with policy level. The *MyCrossAlbum* sends the sharing photo and policy information the *x-mngr* via API call. The *My Photos* menu displays owner's shared photos. *Friends' Photos* menu helps the

¹Since current social networking sites do not allow 3rd party sites to access user's local policy

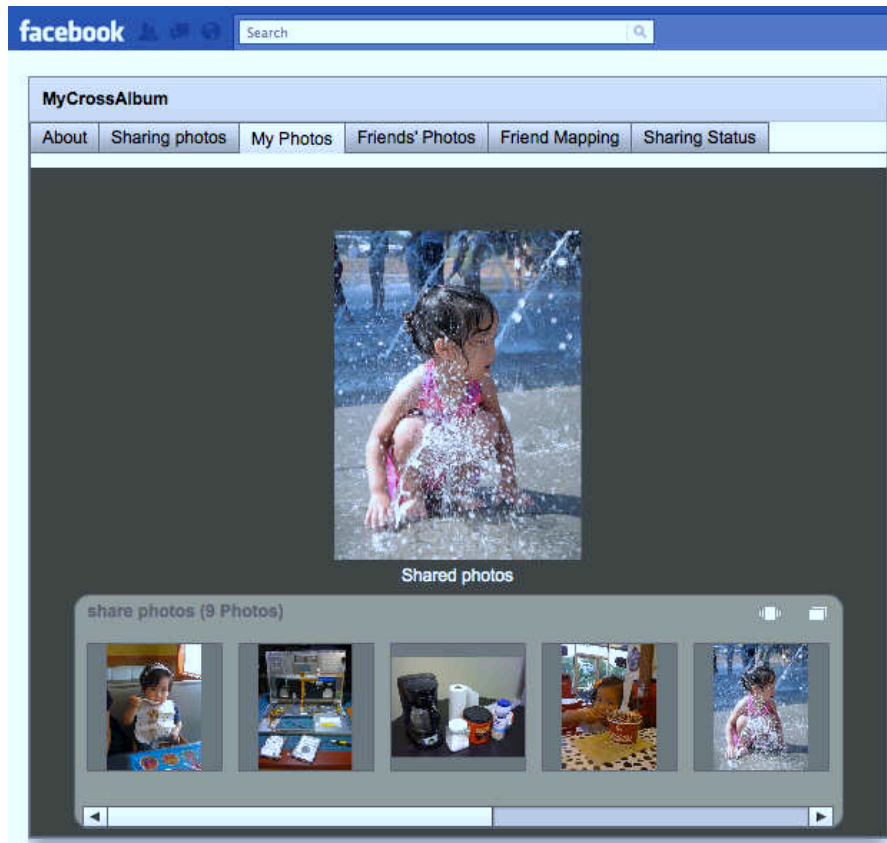


Figure 4.4: *MyCrossAlbum* Application on Facebook

content owner to explore friend's shared photo. The *Friend Mapping* helps users to manage friends' identity mapping based on the recommendation mapping and manual mapping. The last menu *Sharing State* helps user to manage the cross-site policy and know who have accessed photos and from which site. It helps the content owner to get a clear understanding of content sharing state.

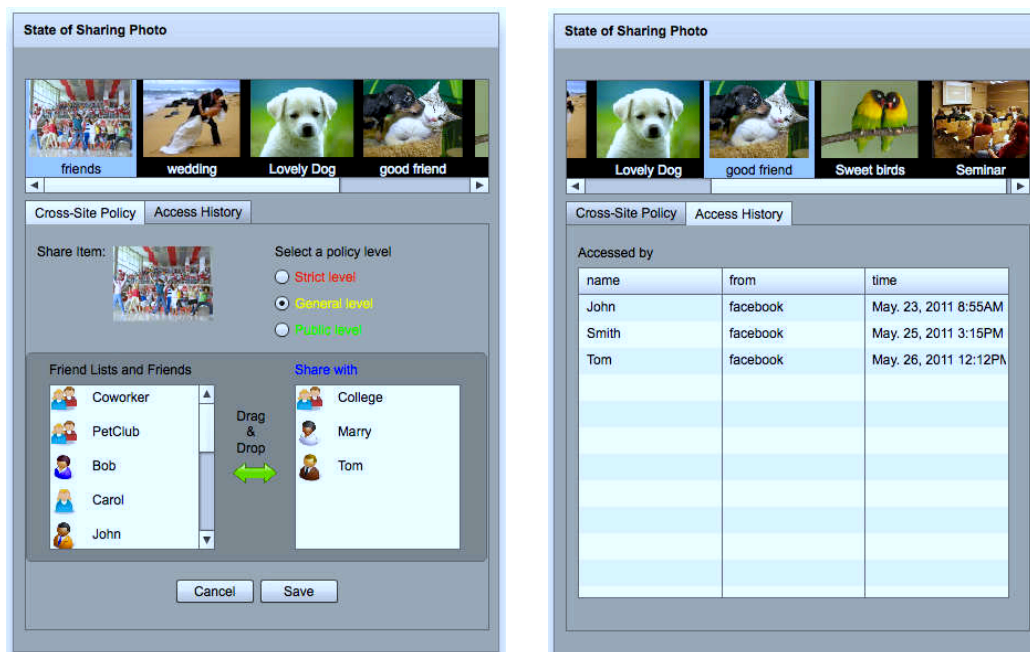
4.4 User Study

To evaluate the approach, we processed a usability user study for the *MyCrossAlbum* application that is a prototype implementation of the *x-mngr* framework.

We conducted a user study to assess participants understanding of the *x-mngr* frame-

work and usability of the *MyCrossAlbum* interfaces. We recruited 13 participants from a university community. At first we conducted a similar survey that we used in the section 3.2 to understand the participants social networking experience, privacy concern, and content sharing experience. After the survey, we showed an animated video describing the concept of *x-mngr* and process of content sharing between two social networking sites. The animated video took approximately 2 minutes to view. After watching the video, the participants answered questions about the *x-mngr* framework. Then, we gave a description of the policy levels that explains the purpose of policy levels and how it works. The participants also answered questions about the policy levels. After that, we conducted usability test for the *MyCrossAlbum* interfaces. The purpose of this usability test is to measure how well the participants specify the cross-site policy on the cross-site policy interface, and to assess the awareness of the participants on the state of the sharing content via the state of sharing photo interface. In the usability test of cross-site policy interface, we gave a mission on the top of the interface and asked the participants to complete the mission using the cross-site policy interface. The mission included assigning a policy level, adding a Friend List with excepting one of members, and adding a friend to the policy. We set up the interface to show a mock Friend Lists, friends and a photo based on the scenario in the Chapter 3. We also recorded the participant's interactions on our server. In the usability test of the state of sharing photo interface, we setup a mock sharing state that shows several sharing photos with cross-site policy and policy level, and access history as Figure 4.5 shows. We gave several questions to the participants and asked them to figure out the answers using the state of sharing photo interface. The survey and questions consisted of a mix of multiple-choice questions, yes/no questions, true/false quizzes and Likert scales. The participants received

entry into a drawing for 5 iTunes gift cards (\$10).



(a) The Cross-Site Policy Panel

(b) The Access History Panel

Figure 4.5: The State of Sharing Photo Interfaces

4.4.1 User Study Results

We asked two true/false quizzes about the *x-mngr* framework after watching the video. One quiz is “The *x-mngr* framework helps the user to share content with friends between different social networking sites”. All participants answered correctly. We also measured the usefulness and preference of the *x-mngr* framework. It was measured on a Likert scale (5 point rating scale, where 1 = Strongly Disagree and 5 = Strongly Agree). The participants indicated that the content sharing across social networking sites using the *x-mngr* framework is useful ($M = 4.34$, $SD = 0.87$), and they have the willingness to use it ($M = 3.92$, $SD = 1.19$). An independent-samples t-test was conducted to compare the willingness between the participants ($N=9$), who have accounts on multiple social networking sites and

the participants (N=4) who have an account on a single social networking site. There was no difference for the willingness between the participants who have accounts on multiple social networking sites (M=4.11, SD=1.36) and the participants who have accounts on a single social networking site (M=3.5, SD=0.57) conditions: $t(11)=0.85$, $p=0.42$. The difficulty of understanding and usefulness for the privacy level was measured on the Likert scale. The question is "The concept of policy level is difficult to understand". The participants disagreed on the question (M=2.46, SD=1.05) and agreed the policy level is a useful way to control the policy enforcement for sharing content across social networking sites (M=4.00, SD=0.58). These results showed the participants understood the *x-mngr* framework and policy level before we conducted the usability test. In addition, the participants had a positive attitude for the *x-mngr* framework and the policy level.

In the usability test of the cross-site policy interface, we assigned the participants the following task: "Alice wants to share her wedding photo with the college Friend List except John. Alice also wants to share the photo with Mary. Alice wants to enforce local policy and cross-site policy together". The task measures whether the participants are able to select right policy level, add right Friend List and except John, and add Mary on the cross-site policy interface. The participants spent average 63 seconds to read and complete the task. Six participants completed the task correctly. Three participants made a mistake on the policy level. Four participants failed the task. After testing the interface, the participants indicated that the cross-site policy interface is easy to use (M=3.69, SD=0.95). However, we would like to know the reason why the participants failed or made mistakes. After finishing all user study, we had an interview with the participants who failed or made mistakes on the task. The participants who failed the task said they did not read the task and

they just tested the interface. The participants who made a mistake on the policy level said they did not recognize the policy level, since they focused to specify the cross-site policy using the drag and drop. From these results, we found out the participants who read the task carefully completed the task easily, and few participants forget to change the policy level. In order to reduce the mistake in the policy level change, one possible solution is to ask the content owner to select a policy level after specifying the cross-site policy.

In the usability test of the state of sharing photo interface, we asked four questions that measure whether the participants can figure out the allowed Friend List and friends, the excepted friend on a Friend List, the accessed friends, and current policy level. One example question is “Who can not access the friends photo in the college Friend List”. All participants reported the correct answers for all questions. The participants indicated that the state of sharing photo interface is easy to use to figure out the current sharing state of photos ($M=4.31$, $SD=0.63$), and the state of sharing photo interface is useful to trace access history, and modify the cross-site policy ($M=4.38$, $SD=0.65$). These results show the state of sharing photo interface is useful to know the current sharing state of photo and modify the cross-site policy.

CHAPTER 5: PARTIAL IDENTITY MAPPING

To support the condition (C2) in the secure interoperation, the *x-mngr* needs a *complete* identity mapping mechanism that provides a complete set of user friends' identity mappings between the viewer site and target site. Requiring a complete set of user identity mappings is not realistic as it will require all users explicitly and truthfully to specify all their accounts in different sites. We explored several identity mapping solutions. One solution is to ask the content owners to indicate all their friends' identity mappings between the target site and viewer site. It might work if the content owners have a small number of friends and have a willingness to provide friend's identity mappings correctly. However, if the friend size is big, it would end up being a very tedious and time-consuming task. Another solution is to compare profiles of all friends between the viewer site and target site. The quality of profile attributes might be one issue for this solution due to deception, errors, or missing attribute. According to [73], 94.9 % of Facebook users and 62 % of MySpace users use their real name on their profiles. It means the quality of user attributes is varying depending on each social networking site. The other solution is to compare email hash values. This solution is only possible when the social networking sites provide user's email hash values to the *x-mngr*. It might generate high accuracy of mapping results when the users use the same email across social networking sites. If users use different email addresses across sites, the email hash based mapping can not map them. The last solu-

tion is to use identity management systems such as federated identity management systems (Liberty [41], Shibboleth [67], or SAML [65]) or user-centric identity management system (OpenID [53]). In order to use federate identity management system, social networking sites need to agree to use a federated identity manage system between them, but it is not realistic. Moreover, many Internet users still do not know what is OpenID even though they already have an OpenID ¹.

5.1 Supervised Learning

In machine learning literature, a learning model is a function f that takes as an input a set of attributes and returns a label or classification. For example, a function that takes the user's age, sex, credit rating and job status and generates a recommendation to either grant a loan or no. A supervised learning mechanism uses previous cases or training data Θ to learn the function f , which we refer to as f_{Θ} .

Taking a simple user centric approach to address the profile-matching problem would require that each focus user (content owner) manually provides mappings between all similar profiles of his friends on different social networking sites. Usually, this is a tedious task, and the user will end up ignoring this task. Furthermore, while users can limit access of their profiles via privacy setting, user's perceptions of visibility do not always match with reality [13], let alone managing cross-site policies. Instead, the approach we adopt is an adapted user centric approach, where the focus user requires only to provide a small subset (α) of the profile mappings. These example mappings are used to compose a training set Θ for the supervised learning algorithm. Basically, we attempt to learn the mapping function

¹Major Internet sites such as Google, Yahoo, and MySpace provide an openID to their users [53]

$f_{\Theta} : \mathcal{X} \rightarrow \mathcal{Y}$, where:

1. \mathcal{X} is a set of attributes describing the profile difference vector (discussed in the next subsection).
2. \mathcal{Y} is a set of labels $\{y_0, \dots, y_m\}$, in our case it is $\{match, no - match\}$, representing match or no-match respectively.
3. Θ is the training set, which is a set of composed of example matching and no-matching friends' profile pairs.

Our goal is to learn the function f_{Θ} based on the provided dataset Θ . Once f_{Θ} is learned, we can automatically decide if a given pair of user profiles $P_i \in SN_A$ and $P_j \in SN_B$ are owned by the same user or no. This learning mechanism is a supervised learning [42] as it requires an example dataset to train and guide the generation of the mapping function f_{Θ} . Given a pair of friends u_p and u_q belonging to the social network SN_A and SN_B respectively, the classifier f_{Θ_i} for user u_i assigns the label y_l to this user pair (u_p, u_q) provided that this label maximizes the classifier's confidence or probability measure $P((u_p, u_q) \rightarrow y_l | \Theta_i)$ based on the training set Θ_i . For more information about supervised learning algorithms the interested reader is referred to [42, 77].

The steps involved in the learning based profile matching process are described in Figure 5.1. The step 1 is a data collection stage in which the *x-mngr* retrieves the focus user friends' profile and network attributes from sites SN_A and SN_B . In the step 2, the *x-mngr* presents the focus user with her friends from SN_A and SN_B , and requests the user to indicate at least α users in both sites. A mapping between user $u_p \in SN_A$ and user $u_q \in SN_B$ is the pair (u_p, u_q) , indicating that user u_p and u_q belong on the same user. A training set is gener-

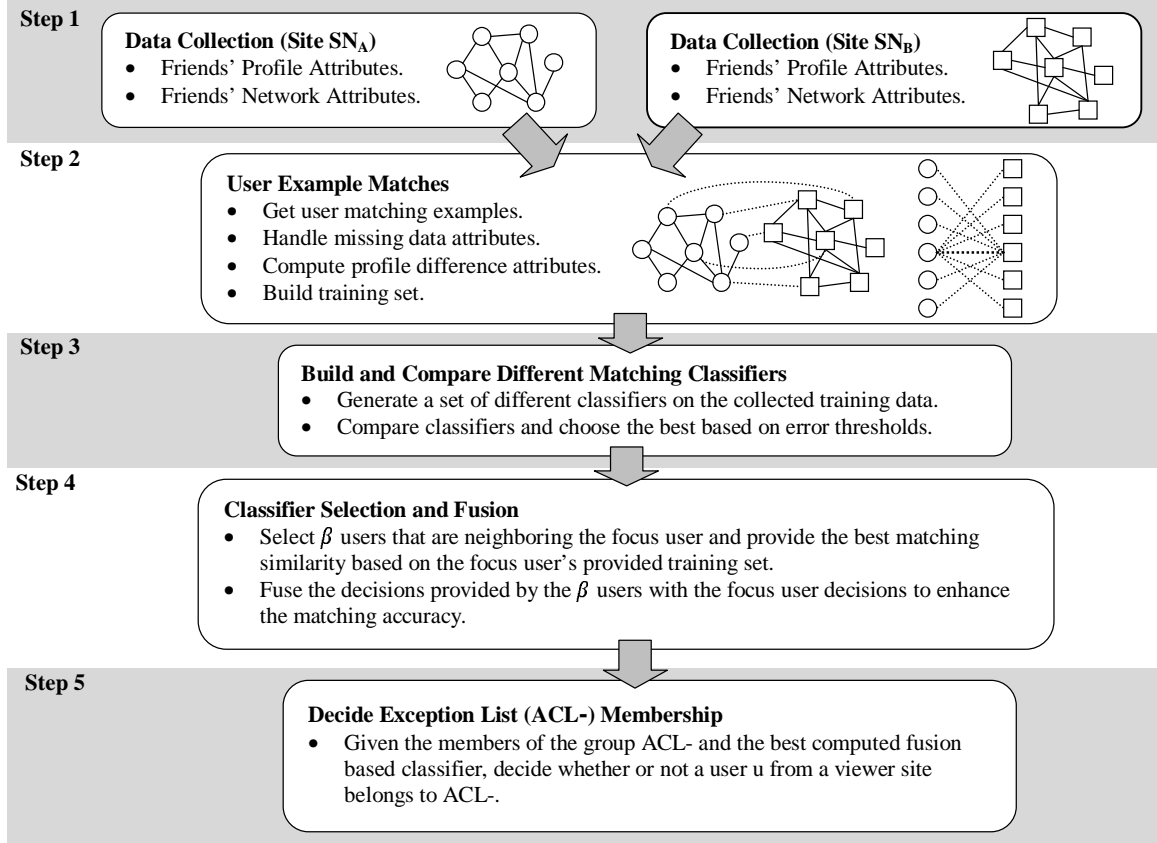


Figure 5.1: Steps in Generating the User-centric Match Classifier.

ated using all the α mapping pairs (u_p, u_q) . In the step 3, the generated training set Θ can be used directly to train a classifier. However, there are several classifiers algorithms and it is crucial to select classifier that is most suited for this specific user instance. The mechanism we adopt is to train and tune several classifiers, then compare their performance based on standard cross validation methods such as n-fold cross validation [77]. Given m classifiers $\{f_{\Theta}^1, \dots, f_{\Theta}^k\}$, the classifier with the lowest error rate is selected, which is denoted as f_{Θ}^* .

In the step 4, the knowledge accumulated by other users in the social network can be utilized further to enhance the classifier accuracy. It is important in this step to seek classification advice from other users who are able to map users similar to the focus user. This is referred to as the selection process where β other user classifiers are selected based

on their accuracy in labeling the focus user's training set. The decisions of the selected β classifiers are fused with the focus user's classifier to generate the focus user's mapping function $\tilde{M}_{A \rightarrow B}$.

Finally, in the step 5, the selected mapping function $\tilde{M}_{A \rightarrow B}$ is used to decide if a user from site SN_A maps to a user in the target site SN_B local policy exception list $P_B.ACL^-$. The details of this approach are discussed in the following section.

5.1.1 Training Set Generation

Given two users $u_i \in SN_A$ and $u_j \in SN_B$, with profile attributes and network metrics $\{A_i, B_i\}$ and $\{A_j, B_j\}$ respectively, we define the distance vector as follows:

$$\begin{aligned} D(i, j) &= [d(A_i, A_j), d(B_i, B_j)] \\ &= [d(a_i^1, a_j^1), \dots, d(a_i^N, a_j^N), d(b_i^1, b_j^1), \dots, d(b_i^M, b_j^M)] \end{aligned}$$

The distance function $d(\cdot, \cdot) \in \mathbb{R}^+$ is dependent on the data attribute domain, where $d(a, a) = 0$. The distance value of each profile attribute and network attribute is considered together in the classification process to decide the matched profiles. Figure 5.2 describes the distance computation. Assume the focus user has R and S friends in SN_A and SN_B respectively,

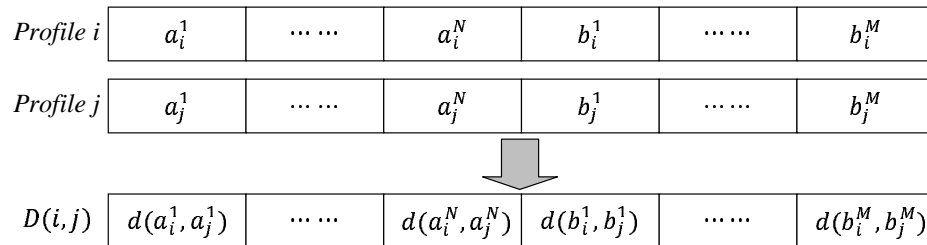


Figure 5.2: Distance Measure of Two Profiles

with a user-mapping (i, j) this provides $R + S - 1$ classified mappings namely:

“Match” (u_i, u_j)

“No-Match” $\forall(u_i, u_s)$ where $u_s \in SN_B \wedge u_s \neq u_j$

“No-Match” $\forall(u_r, u_j)$ where $u_r \in SN_A \wedge u_r \neq u_i$

By explicitly indicating the match (u_i, u_j) , the focus user is implicitly indicating that user u_i is not same to all other friends in SN_B and similarly user u_j is not same to all other users in SN_A . The distance vector is computed for both the explicit match and implicit no-matches, then used as the training set Θ .

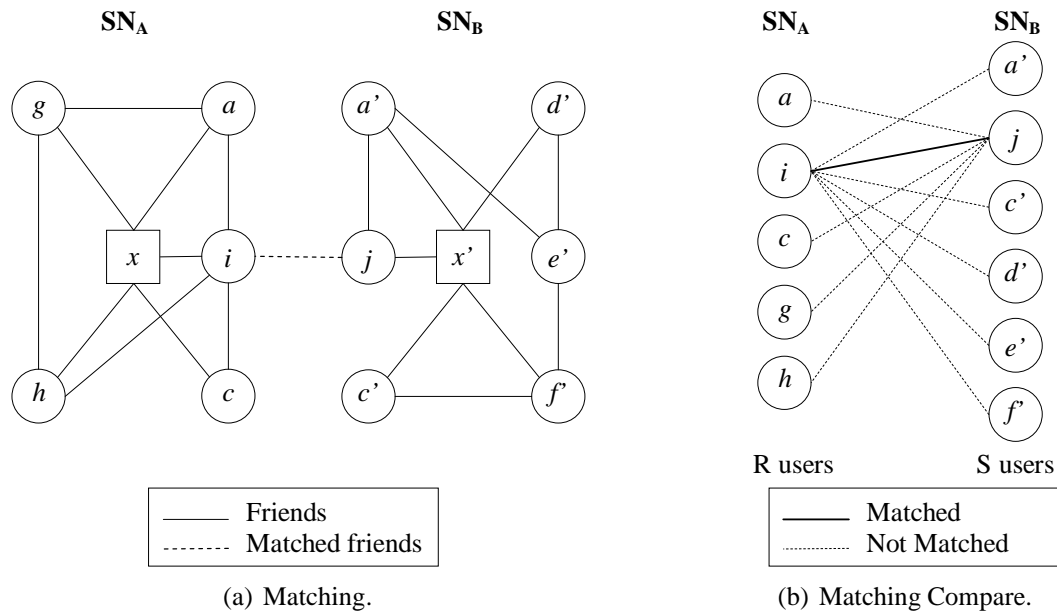


Figure 5.3: Training Set Generation.

5.1.2 Attribute and Network Distances

In order to measure the similarity value of each attribute pair, we consider different similarity methods for different attribute types. In case of string attributes such as school name,

last name and first name, these attributes are tokenized and normalized before computing the distances. Through the tokenization process, the string attribute value is divided into tokens by converting a sequence of characters into a sequence of tokens. For example, if the attribute value of school name is “UNC Charlotte”, it generates (“UNC”, “Charlotte”) as tokens. Then the normalization process, the process of canonicalizing token, matches the semantically equivalent token despite superficial differences in the character sequences. For instance, “UNC” and “University of North Carolina” should be considered as the matched token. In case of attribute value of first name, “joe” and “Joseph” is also considered as the matched token via normalization process. This normalization process increases the accuracy of similarity score for the different format of string attributes. Figure 5.4 describes both the tokenization and normalization process.

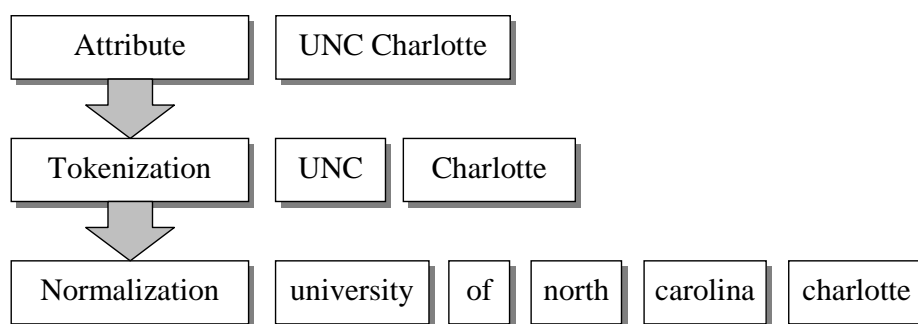


Figure 5.4: Tokenization and Normalization Process

After the tokenization and normalization, we apply the Levenshtein Distance [40] that is a metric for measuring the amount of difference between two strings attributes. For numeric attributes such as age, we use the square Euclidian distance. For address attributes, we first perform the geocoding process of converting the addresses into their geographic coordinates represented as latitude and longitude, then we compute the distance between the two geocoded addresses [33]. The table 5.1 below describes the possible distance mea-

asures for the different attribute domains.

Table 5.1: Distance Measures

Attribute	Distance $d(.,.)$
age	$\ age_i - age_j\ ^2$
address	$\ geo(address_i) - geo(address_j)\ ^2$
name	$lev(name_i, name_j)$
degree	$\ norm(degree_i) - norm(degree_j)\ ^2$

The network metrics are numeric attributes. When comparing metrics computed from different graphs the varying size of the graphs presents a challenge. For example, a user in Facebook might have 300 friends while in MySpace could have only 100 friends, due to the different network sizes the metrics computed will differ considerably. To enable the comparison of metrics computed from different graphs we adopt the approach presented by D. Koschützki et al. [14], which normalizes each network metric based on a specific normalization factor. Then the Euclidian distance is used to compute the distance between the normalized metrics from different social networks.

5.1.3 Classifier Selection and Fusion

The inherent advantage of social networks is the ease of sharing of news, photos, videos and several other data objects among users. We extend this sharing to include the accommodation of user experiences by leveraging their trained match classifiers, where user u_j is

able to share his/her matching function f_{Θ_j} with other users. Assume a user u_i would like to leverage the experience of other users in the social network to improve their matching function f_{Θ_i} . In this section, we use f_{Θ_k} to refer to the best match classifier $f_{\Theta_k}^*$ for user u_k . Given a user u_i and a set of users $S = \{u_1, \dots, u_n\}$, the set S can be chosen from the neighboring trusted friends or other experienced users in the social network. Each user u_k in the set S is willing to share their matching function f_{Θ_k} to improve the matching function of user u_i . As indicated in Figure 5.5, this translates into two sub-steps: (1) The selection of β users from the set S that are best fit to help user u_i in computing an improved matching function, (2) The fusion of the different f_{Θ_k} functions provided by the β users with the focus user's function f_{Θ_i} .

Definition 4 (*Selection*) *Given a user u_i , a set of user trained classifier functions $f_S = \{f_{\Theta_1}, \dots, f_{\Theta_n}\}$, the training set Θ_i for user u_i , and a classifier fitness function $\Phi: f_{\Theta_k} \times \Theta_i \rightarrow \mathfrak{R}$, select the best β classifiers based on the fitness function.*

The selection process is based on the fitness function as defined in Def. 4. The fitness function is a mechanism to rank the classifiers in f_S based on their similarity to the decisions taken by the classifier of user u_i . The fitness function tests each classifier f_{Θ_k} by labeling the tuples in the training set Θ_i and computing the vector $[TP, TN, FP, FN]^T$, where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative. The fitness of f_{Θ_k} is based on the classifier accuracy of recall and precision [8, 57]. The β classifiers with the highest fitness are selected and are denoted by the set $S_\beta = \{f_{\Theta_1}, \dots, f_{\Theta_\beta}\}$.

Given the β classifiers, the next step involves fusing the decisions of these classifiers and the decisions generated by the focus user's classifier (f_{Θ_i}) to improve the classification

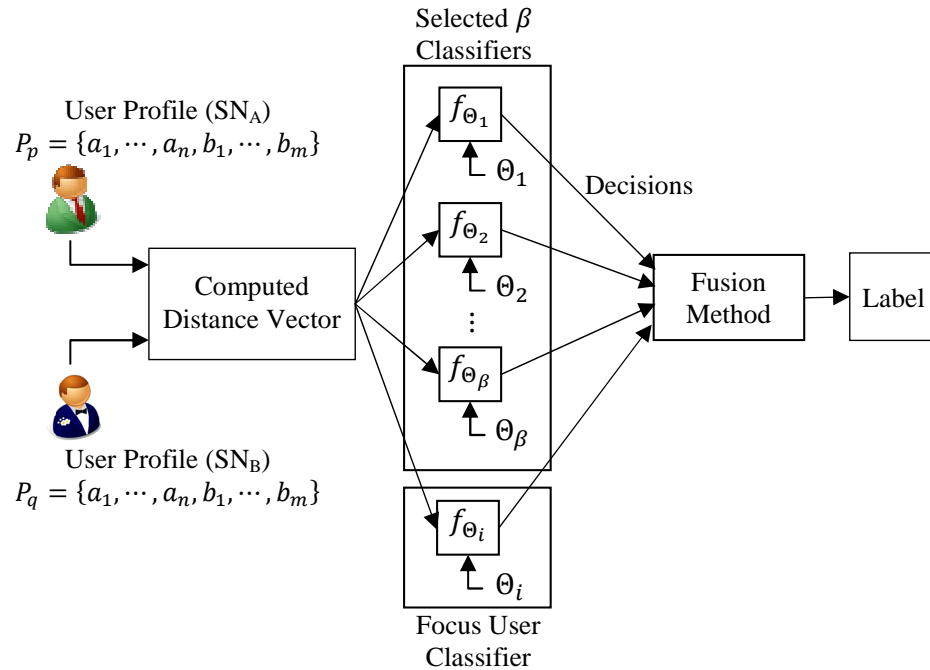


Figure 5.5: Classifier Selection and Fusion

result. We adopt the most relevant classifier fusion algorithms [35]: *group voting*, *group confidence product* and *most confident*. The group voting mechanism is based on selecting the label (e.g., match or not-match) which receives the largest number of votes by the β classifiers. Given a user pair (u_p, u_q) , where $u_p \in SN_A$ and $u_q \in SN_B$, the label w_l is assigned to this user pair if w_l receives votes as follows:

$$\sum_{k=1}^{\beta} \delta_k^l(u_p, u_q) = \max_{r=1, \dots, m} \sum_{k=1}^{\beta} \delta_k^r(u_p, u_q)$$

where

$$\delta_k^r(u_p, u_q) = \begin{cases} 1 & \text{if } f_{\Theta_k}(u_p, u_q) = w_r \\ 0 & \text{otherwise} \end{cases}$$

The group confidence product mechanism is based on selecting the label that maximizes the product of the confidence of all the β classifiers. For a user pair (u_p, u_q) , the label w_l is

selected if group confidence product of w_l is as follows:

$$\begin{aligned} \prod_{k=1}^{\beta} P((u_p, u_q) \rightarrow w_l | \Theta_k) = \\ \max_{r=1, \dots, m} \prod_{k=1}^{\beta} P((u_p, u_q) \rightarrow w_r | \Theta_k) \end{aligned}$$

The most confident mechanism is based on selecting the class that gets the highest confidence from any of the β classifiers. This approach fuses the different classifier confidence and adopts only the label provided by the most confident classifier. For a user pair (u_p, u_q) , the label w_l is selected if the confidence of w_l is as follows:

$$\begin{aligned} \max_{k=1, \dots, \beta} P((u_p, u_q) \rightarrow w_l | \Theta_k) = \\ \max_{k=1, \dots, \beta} \max_{r=1, \dots, m} P((u_p, u_q) \rightarrow w_r | \Theta_k) \end{aligned}$$

After β classifiers with the highest fitness are selected, an appropriate fusion algorithm (of the three listed above) is chosen to fuse the results of the f_{Θ_k} functions producing a predicted label, i.e., match or no-match. This final fused classifier represents the identity mapping function $\tilde{M}_{A \rightarrow B}$ between users in site SN_A and SN_B .

5.2 Implementation of experimental site

We implemented an experimental site namely *ProfileMapping*. We designed the *ProfileMapping* site to collect identity mapping data from the social networking users who have accounts on Facebook and MySpace. We retrieved user data and friend data using the Social Network Connect Service [37].

The data collection was processed in 7 steps. The step 1 and 2 are the login process

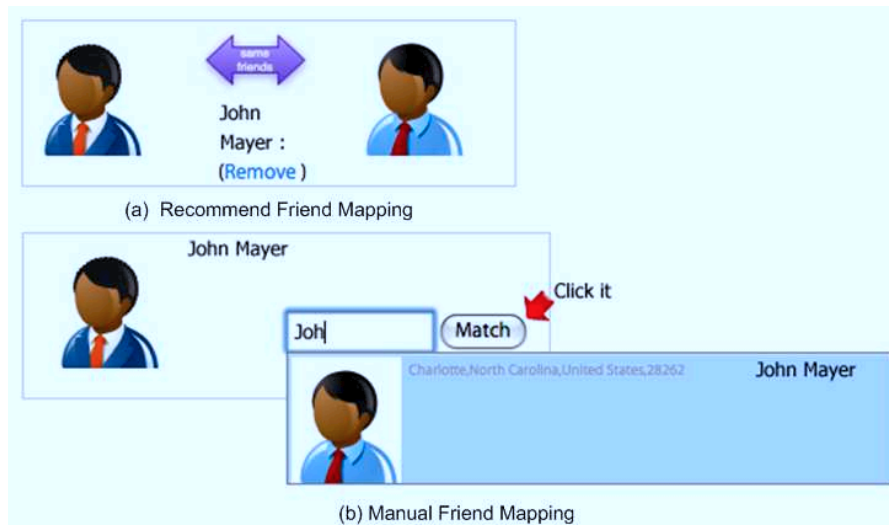


Figure 5.6: Friend Mapping Process

of MySpace and Facebook. The step 3 provides a profile comparing page of participant. The step 4 provides the recommended friend mapping. The participants are able to review the recommended mapping result and make a decision for each mapping result. The step 5 is the manual friend mapping, where the participants search a friend who has accounts on Facebook and MySpace. This was implemented using the jQuery package to enable the user to easily map users by typing a few characters of the friend's name in a text box placed beside each friend profile photo, (See Figure 5.6(b)). The step 6 is the confirmation process, where the participants review all mapping results before submit them. Finally, the step 7 is the last process, where we provide friends' location on the Google map and finished experiment.

5.3 Experimental Results

In order to investigate the effectiveness of the proposed partial mapping approach, we performed an extensive experimental evaluation on the collected data. Especially, we assessed

how different classifiers perform on our data, training set, classifier fusion mechanisms influence to classification results. In order to collect data for this study, we invited 5000 users who have accounts in both Facebook and MySpace to map their friends on the *ProfileMapping* site. 100 users completed the registration and mapping process successfully. The users' profiles, friend's list, friend of friend's lists and profiles were collected. The users were required to provide mappings between their friends between Facebook and MySpace. We collected 5695 profiles in Facebook, 9274 profiles in MySpace, and 960 profiles mappings from the participants. For each user, we accumulated the profile attributes and computed the network metrics. The following profile attributes that were obtained were: First Name, Last Name, Gender, Location, Date of Birth, and Education. In addition, each user's social graph was built and a series of network metrics were computed which include, degree, HUBS, authority, betweenness, closeness, PageRank, Eigenvector, and number of common friends.

The collected data was used to train 7 classifiers namely, AD Tree, BayesNet, Naive-Bayes, NBTree, RandomForest, RBFNetwork, and Ridor. The true positive, true negative, false positive, and false negatives for each classifier were recorded. Figure 5.7 (a-b) shows the accuracy and precision results generated by *x-mngr* for a training set of $\alpha = 20\%$, for the 7 different classifiers and 10 friends selected for fusion ($\beta = 10$). Furthermore, Figure 5.7 (a-b) shows the results obtained by the different classifiers for the fusion mechanisms, and from the figure it is evident that our fusion approach improves the classification result with the voting based approach leading. Our approach consistently provides a classification accuracy of 99 % using any of the fusion approaches. Using the RandomForest Classifier, we are able to generate classifications with a high precision as 98 %, which implies that

our classifier is able correctly to locate the positive matching profiles in our dataset. This implies that relative to our propose framework, the proposed mapping approach presents a high accuracy for mapping the focus user’s friends in both social networking sites.

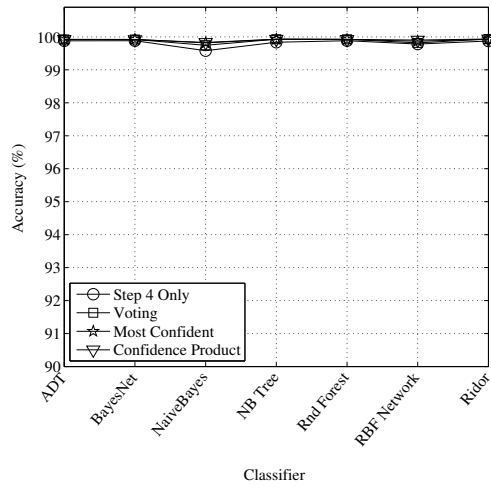
Figure 5.7 (c-d) presents and experiment conducted using the RandomForest classifier while varying the training set α from 10 % up to 60 %. The accuracy and precision were computed and as indicated in the figures our fusion-based approach leads the no fusion approach in all metrics. As expected, the classifier accuracy and precision increase as more user-mappings are provided for training. Note that, our fusion based approach is able to maintain a matching accuracy of 99.8 % and a precision of 85 % at a training size of only 10 %, which means we require the focus user to label only 10 % of his matching friends. To investigate the effect of the size of selected fusion classifiers (β) we conducted experiments holding all parameters constant (RandomForest classifier, $\alpha = 20\%$) while varying β . Figure 5.7 (e-f) depicts the accuracy and precision of the fused classifiers and the best classifier of the focus user (no fusion) for the different β values (10-40). Note that as we increase β the accuracy and precision remain within acceptable bounds. For example, using fusion our approach maintains an accuracy of 99 % and precision around 95 %.

The effectiveness of *x-mngr* depends on whether users are given the right access permissions in the cross-site policy, and whether they are correctly identified in the target social site *local* policy exception list. In order to investigate the effect of the exception list ACL^- on the mapping process, we randomly generated different ACL^- sets and tested different mapping functions based on the trained classifiers. The experiments were repeated multiple times and averaged over all runs. Figure 5.8 (a-b) shows the accuracy and precision

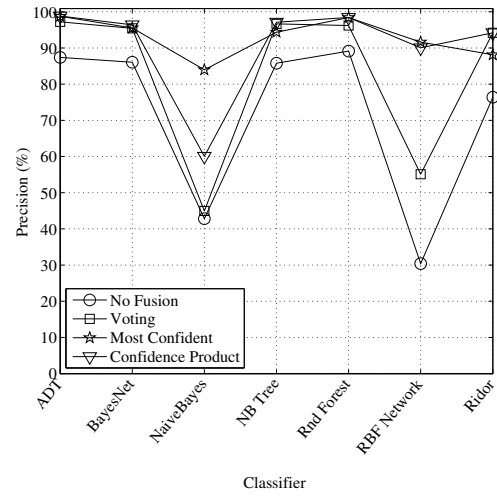
results obtained while using all the 7 classifiers, fixing the training set to $\alpha = 20\%$, and $|ACL^-| = 2$ to represent small exception lists. As depicted in Figure 5.8 (a-b), the results in the fusion based approach maintain an accuracy between 96-98 % and a precision of about 98 %. With Random Forest classifier, our approach was able to identify all the viewer users in the ACL^- with an accuracy of 97.4 % and precision of 99 %. Figure 5.8 (c-d) presents the accuracy and precision for the Random Forest classifier while varying the training set size α from 10 % to 60 %, and fixing $|ACL^-|$ to 2. The results show that the fusion approach perform better than the non fusion approach and the performance increase as we increase the training size, but even with a training set of 20 % we still get a reasonable result of 97.4 % accuracy and 99 % precision. To investigate the effect of the exception list size $|ACL^-|$ on the accuracy and precision of the approach, we conducted experiments holding all parameters constant (RandomForest classifier, $\alpha = 20\%$) while varying $|ACL^-|$. Figure 5.8 (e-f) depicts the accuracy and precision of the fused classifiers and the best classifier of the focus user (no fusion) for the different $|ACL^-|$ values (2-50). Note that as we increase ACL^- the accuracy and precision drop, this is because as the size of ACL^- increases there is a higher probability of false matches which affects both the accuracy and precision. Note that even though the accuracy and precision drops as $|ACL^-|$ increases the fusion based classifier still consistently performs better than the no fusion classifier, and maintains a less steeper decent in accuracy and precision. Furthermore, our fusion based approach still maintains an accuracy of 98.5 % and precision of 98 % for an ACL^- of size 50.

Through the presented experimental results we demonstrated the high accuracy and precision attained by our supervised base approach in user-mapping. Our supervised learning approach shows an accuracy of 99.86 % and a precision of 98 % at the use of 20 % as train-

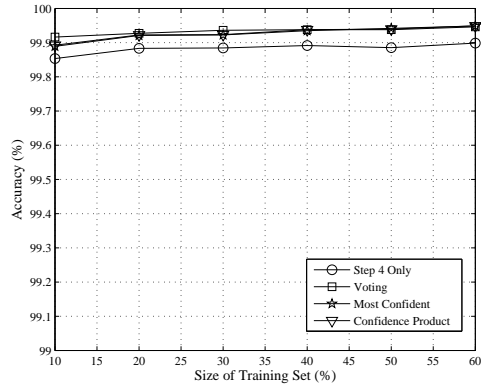
ing set from the mapped friends. Thus, demonstrates the applicability and suitability of our cross-site framework *x-mngr* for enabling secure cross-site interaction between different social networks.



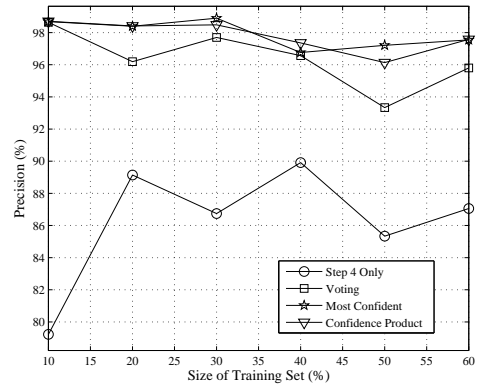
(a) Classifier Type vs. Accuracy



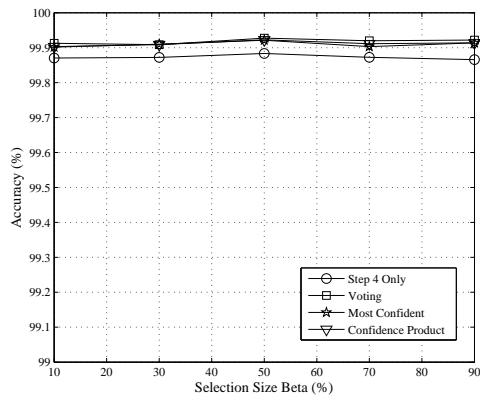
(b) Classifier Type vs. Precision



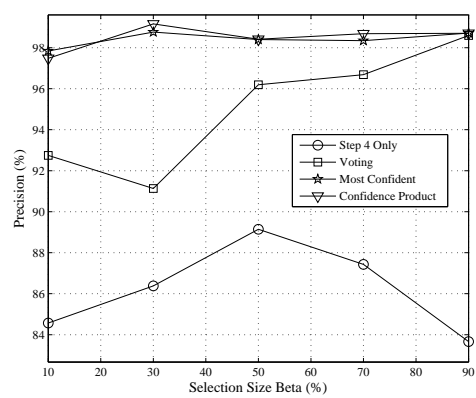
(c) Training Set vs. Accuracy



(d) Training Set vs. Precision

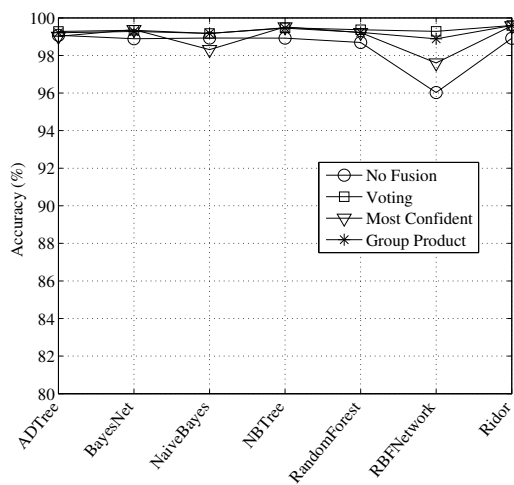


(e) Selected β vs. Accuracy

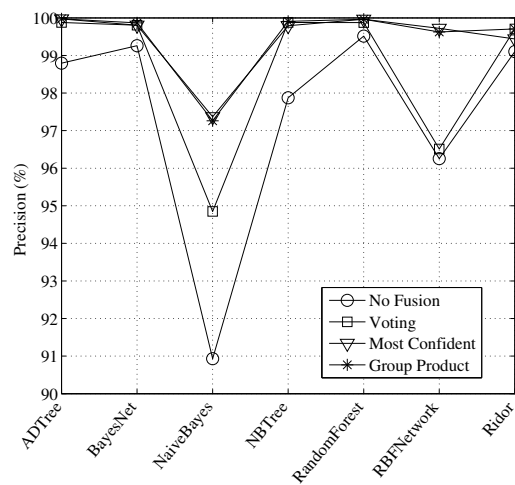


(f) Selected β vs. Precision

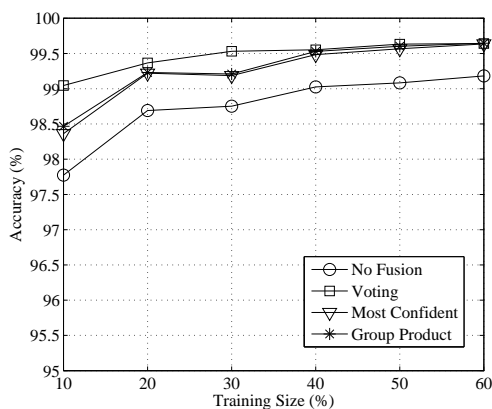
Figure 5.7: User-Mapping Experimental Results



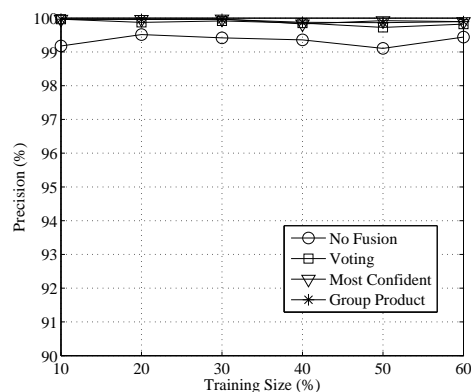
(a) Classifier Type vs. Accuracy



(b) Classifier Type vs. Precision



(c) Training Set vs. Accuracy



(d) Training Set vs. Precision

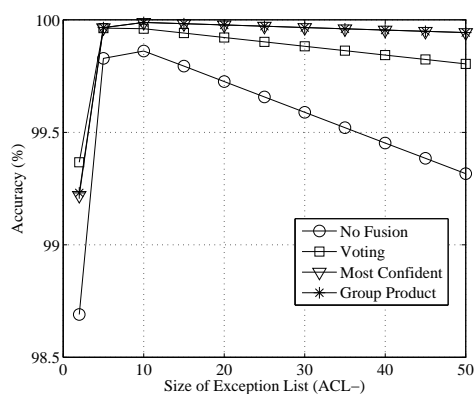
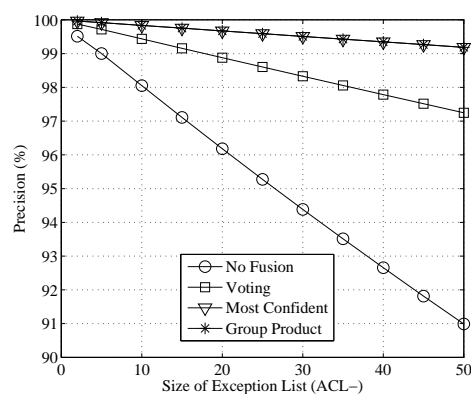
(e) Size of Exception List ACL^- vs. Accuracy(f) Size of Exception List ACL^- vs. Precision

Figure 5.8: Exception List Experimental Results

CHAPTER 6: IDENTITY MAPPING USING GAMES

In this chapter, we propose a Game With A Purpose approach to solve the identity mapping problem in a new way. The proposed approach leverages the game appeal and social community to generate the identity mappings. We designed and implemented an online social networking game (*GameMapping*), the game is fun and is based on human verification. *GameMapping* takes advantage of people's existing perceptual abilities and desire to be entertained. The game will present the player with a user from one social network, and a set of friends from another social network, which represent the *set of mapping recommendations*. The friend's information is summarized in a profile card, which includes the profile photo, name, age, location, etc. The player gets a small number of points for choosing one of the provided mappings, this reinforces a sense of *incremental individual success* in the game. The game also rewards *social success* by awarding the player a large number of bonus points when other users or friends agree to the player's provided mappings. This proposed mechanism is similar to social buying, where buyers are offered discounts discount deals (bonus) if they sign up for a deal in large masses [55]. Users will be allowed to invite their friends to play the game in the hope of gaining the large bonus points. Similar games with a purpose have been successfully proposed to aid in labeling and tagging images over the web [75]. We describe details in the rest of sessions.

6.1 Game with a Purpose

Games with a Purpose (GWAP) is a form of human computation [75, 76], which gets humans to play enjoyable games that are also productive tools. These games are used in tasks that are hard for computers but easy for humans. For example, the ESP game [75] is a two-player game used for labeling and tagging images over the web, the game is setup to reward players providing the same labels by giving them bonus points if their tags match. Our goal is to design a GWAP to solve the profile mapping problem between social networks, by asking players to map their friends in the different social networks. One of the main challenges is the design of a points system that rewards correctly identified profile mappings and to maximize the reward for truthful rational players, and minimize the reward of irrational players. Gaming on social network platforms is becoming very popular with games such as FarmVille in Facebook [27] hosting over 62 million monthly active users. Our proposed game can easily be deployed on social networking sites as an online game, and if it is popular we estimate that most of the account mappings can be properly discovered in a matter of weeks.

6.2 Definition of Profile Mapping

The global profile mapping is defined as follows:

Definition 5 (*Profile Mapping Problem*). *Given social networks SN_A and SN_B , with social graphs $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$ respectively, find the set of profile mappings M of the form $(u_i, u_j) \in M$ where $u_i \in V_A$ and $u_j \in V_B$ belonging to the same user in both social graphs G_A and G_B .*

The problem of mapping data concepts between different sites or platforms has been applied to multiple areas, such as: database schema matching [44, 61], web search [11, 24], ontology mapping [23] and visualization [26, 78]. The graph isomorphism is an NP-Complete problem which involves finding one to one mappings between vertices and edges of a pair of graphs [9, 29]. The subgraph isomorphism graph matching problems has been proven to be NP-complete [28]. Furthermore, the inexact graph matching problem, where $|V_A| \leq |V_B|$, the complexity is proved in [1] to be NP-complete. Several attribute, model, object recognition, and network based techniques were proposed to provide heuristic approaches to solving graph matching problems [7, 18, 19], these approaches are computationally expensive, and require the knowledge of the complete graphs G_A and G_B . In this dissertation, we propose solving the profile mapping problem by using human computation in the form of an online game. This approach has been used in [75, 76] to map tags to images effectively. The main assumption is that with the correct set of incentives, users would enjoy playing a game and at the same time contribute to mapping profiles between users in different networks.

Definition 6 (*Local Profile Mapping Problem*) *Given a user u who has identities u_i and u_j on social network SN_A and SN_B respectively, and user's local neighborhoods $\mathcal{N}_{u_i}^A$, $\mathcal{N}_{u_j}^B$ find the set of mappings $M_u \subseteq M$ mappings between profiles in \mathcal{N}_{u_i} and \mathcal{N}_{u_j} .*

Our proposed approach will leverage the individual and social knowledge of social network users to provide mappings, and to provide mapping verifications which can be then used to solve the local profile mapping problem. The local profile mapping problem does not require knowledge of the whole social network graph, instead it only requires knowledge

of the neighborhood network. Providing incentives to ensure the wide spread adoption of the game would allow solving a large number of local profile mappings, which enables the mapping of all similar profiles in large social networks. In fact, this is equivalent to the generalization of the sub-graph isomorphism mappings of local networks to the maximum number of common sub-graph problem in the global networks [79].

6.3 General Game Description

Our proposed game is called *GameMapping*. The basic idea is that players gain points by providing mappings of their friends' profiles on different social networks. *GameMapping* allows players to map Facebook and MySpace profiles, or Facebook and Twitter profiles.

In order to play the game, the player needs to complete an authentication stage that involves two social networking sites. We implement Facebook Connect, MySpaceID, and TwitterID to enable users to authenticate into the corresponding social networking sites, and to authorize the GameMapping site to access their profiles and friends list. It enables the GameMapping site to retrieve the user's profile and neighborhood social graph data which includes last name, first name, gender, age, country, profile picture, friends list and mutual friendships. These data enable our system to compute the local neighborhood for the current player $(\mathcal{N}_u^A, \mathcal{N}_u^B)$. A user profile referred to as the focus user u_f is picked from smaller neighborhood. Without loss of generality assume the focus user profile u_f is selected from neighborhood \mathcal{N}_u^A , the game then computes the recommended mappings profiles R from neighborhood \mathcal{N}_u^B based on attribute and network distance metric. The focus user and the computed recommendations are then presented to the player. Figure 6.1, shows a screen shot of the game, where the focus user is in the center surrounded by his

possible best recommended mappings displayed in a random order. The users' profile pictures are shown along with their profile information which include, age, gender, and location. Information about the recommended mappings is presented to the user when the mouse is moved over the photo. The player should decide either to map the focus user to one of the recommended profiles or to skip if no map is present. The player is given 40 seconds to make a decision about the presented game dataset, then a new game dataset is presented. The game also presents top 10 players ordered by the points earned. To

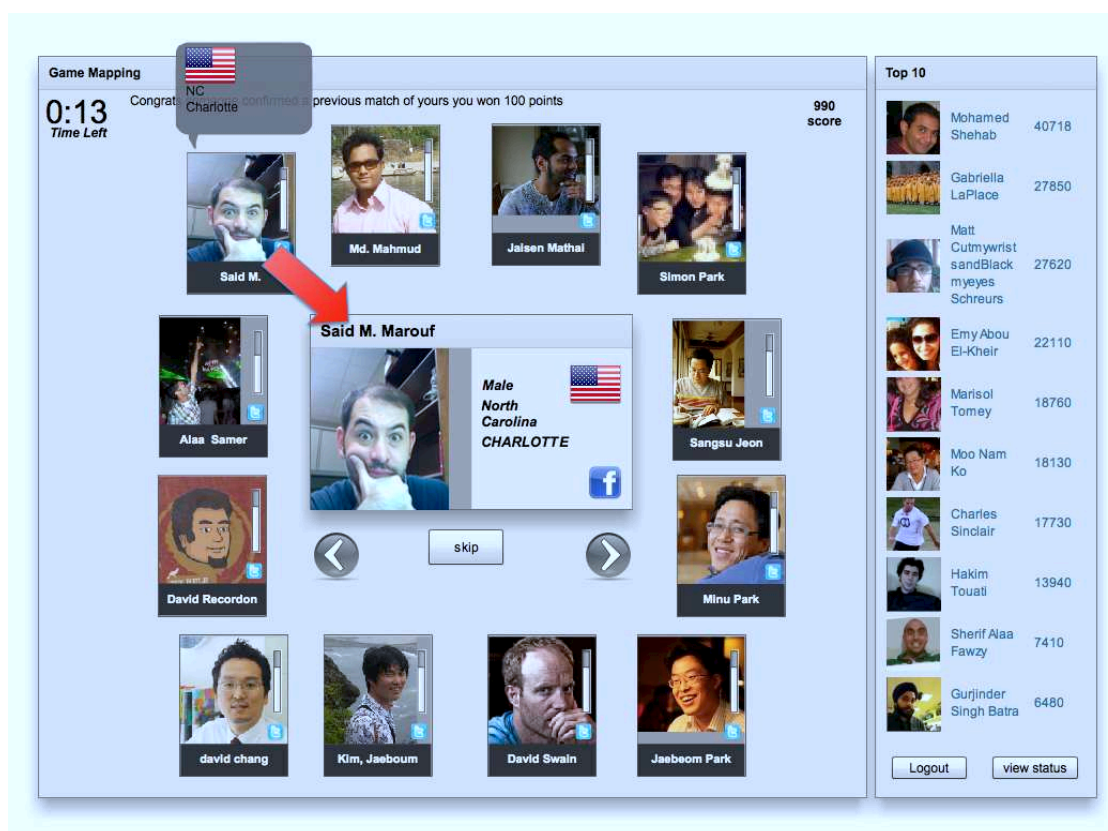


Figure 6.1: The GameMapping Screen Shot

motivate players into making correct decisions of either mapping or skipping, the game awards the player 10 points for any provided map, 100 bonus points if the provided map is confirmed by another player, and 30 bonus points if a skip is confirmed by another player.

In order to maximize the points (reward), a player should focus on providing the mappings that will most probably be confirmed by other players. When a player starts the game, the player first plays the game with the player own network dataset. In other words, the player maps friend's profiles. After the player is done mapping his local network, the player plays the game with a game dataset that is randomly selected. It ensures that players provide mappings towards multiple local profile mappings and at the same time ensure the game continuity.

6.4 Recommendation Generation

Given a player u who owns profiles u_i and u_j , and the neighborhoods \mathcal{N}_u^A and \mathcal{N}_u^B the focus user u_f is selected randomly from the neighborhood that has the smaller number of nodes, which we refer to as the focus network. This design choice was made as the maximum number of possible mappings is equal to $\min(|V_u^A|, |V_u^B|)$. Figure 6.2 shows both neighborhoods and the focus user u_f . Lets assume the focus user u_f is selected from \mathcal{N}_u^A .

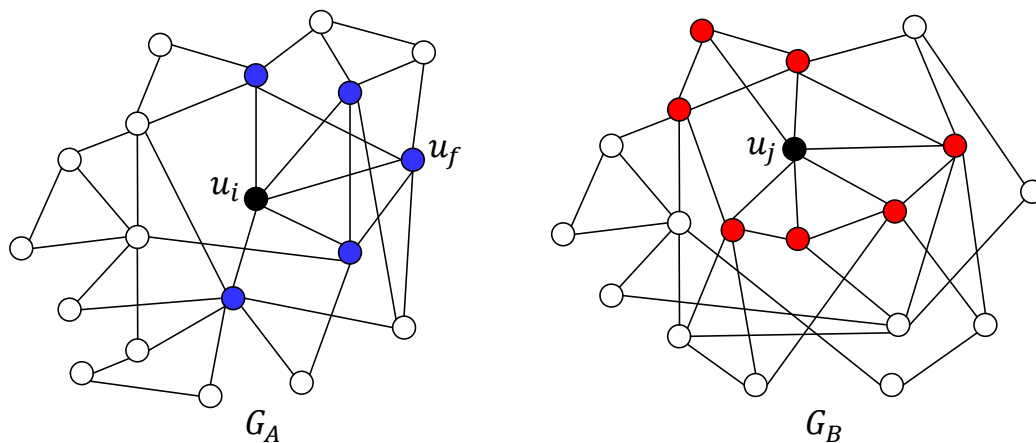


Figure 6.2: Neighborhood and Focus User Recommendations.

Given the focus user the mapping recommendation is generated by ranking the user profiles in \mathcal{N}_u^B based on their similarity to the focus user. The similarity between two profiles is

computed as a weighted sum of distances between the different user profile and network attributes. The profile attributes include first name, last name, gender, age and address. The network attributes include the centrality, betweenness, hit rate, degree and eigen values [12, 51]. We investigated several vector distances which include the Chebychev and Minkowski distance for numerical attributes, Cosine and Levenshtein distance for nominal attributes, and the Euclidian distance for the numerical attributes (i.e. age) and the Levenshtein distance for nominal attributes (i.e. gender, name) [42]. The weight of each attribute was computed based on a linear regression classifier trained using the knowledge collected from our initial experiments [77]. The recommendation set R is the sorted list of proposed user profiles based on their computed similarities with the focus user. As indicated in Figure 6.1, the game presents the user with the top 12 recommended mappings select from the recommendation set R following the Top-k Fagin’s algorithm [25]. The selected recommendations are shuffled randomly then displayed in a clock-wise fashion around the focus user. This randomization is required to ensure that players put some effort in finding the possible profile mapping among the displayed 12 recommendations. Moreover, by randomizing the recommendation set R this would avoid possible collusion between different players as each player is presented with the same 12 recommendations but not in the same location on the screen.

6.5 Game Theoretic Analysis

In this game, the players do not communicate and each player does not know the action taken by the other players. The game can be modeled as a two player extensive game with incomplete information. In this game the players are provided with a focus user u_f and a

set of recommended mappings $R = \{u_1, \dots, u_n, \phi\}$. Each player has a set of $n + 1$ actions of the form $a_k = \mathbf{map}(u_f, u_k)$ where $u_k \in R$. Note, the action $a_{n+1} = \mathbf{map}(u_f, \phi)$, which is equivalent to the **skip**(u_f). The set of actions $A_1 = A_2 = A$, and the utility (δ_i) of player i is selected to satisfy the following conditions:

- $\delta_1 = \delta_2 = \delta$,
- $\delta(a_i, a_j) = \delta(a_j, a_i)$,
- $\delta(a_i, a_i) > \delta(a_i, a_j)$ for all $i \neq j$,
- $\delta(a_i, a_i) > \delta(a_{n+1}, a_{n+1})$ for all $1 \leq i \leq n$

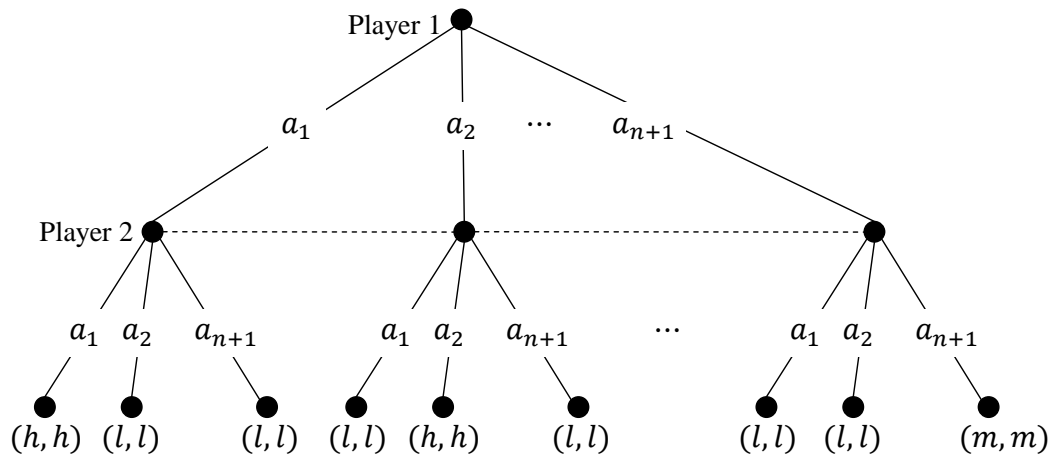


Figure 6.3: Game Tree with Imperfect Information.

Figure 6.3 shows the extensive game tree, where nodes represent players and edges represent player actions. The payoffs for players 1 and 2 are shown at the terminal nodes. The values of h and l are chosen such that $h > l$, this ensures that $u(a_i, a_i) > u(a_i, a_j)$ for all $i \neq j$. This game is a coordination game in which the each player is trying to make the same choice as the other players to maximize their utility.

Rational players intend to maximize their expected game payoff. Note that the payoff from agreeing on a map is higher than the payoff from agreeing on a skip ($h > l$), this motivates rational players to try to find possible maps between the focus user and one of the recommendations and to skip if they can not find a suitable map. The Nash equilibrium is a commonly used equilibrium notion that provides an equilibria such that no player can profitably deviate from and enhance their payoff with the belief that other players will not deviate [54]. Referring to the game representation in table form in Figure 6.4, The game has $n + 1 = |A|$ pure Nash equilibria represented by the set S where $S = \{(a_i, a_i) : a_i \in A\}$, that is the strategy that would result in maximizing the user payoff is when both users make the same action.

		Player 1				
		a_1	a_2	\dots	a_n	a_{n+1}
Player 2	a_1	(h, h)	(l, l)	\dots	(l, l)	(l, l)
	a_2	(l, l)	(h, h)	\dots	(l, l)	(l, l)
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	a_n	(l, l)	(l, l)	\dots	(h, h)	(l, l)
	a_{n+1}	(l, l)	(l, l)	\dots	(l, l)	(m, m)

Figure 6.4: Game Nash Equilibria Indicated in Grey

Since the game has multiple equilibria, it is still not clear what action strategy with a rational player act upon. Given that each player does not know the action taken by the other player, the question that each player asks themselves is that given $\{u_f, R\}$ “what would other players do if they are presented with the same $\{u_f, R\}$?” and by the theory of focal points [48] players will usually coordinate at points that in some sense stick out from

the others (focal points). A player game strategy can be described based on the probability of selecting an action a_i from the action set A given the focus user and recommendation set $\{u_f, R\}$. The probability $p(a_i|\{u_f, R\})$ represents the probability of choosing an action a_i conditioned on the game parameters $\{u_f, R\}$, which can be represented as $p(a_i|\{u_f, R\}) = p(a_i) \times r(a_i, \{u_f, R\})$. Where $r(a_i, \{u_f, R\}) = \frac{p(a_i, \{u_f, R\})}{p(a_i) \times p(\{u_f, R\})}$ is the relevance of action a_i to the set $\{u_f, R\}$. According to the focal point analysis, a rational player would choose the action that maximizes the $p(a_i|\{u_f, R\})$ which is the action that is most relevant to the current $\{u_f, R\}$ set, which is described as follows:

$$a^* = \arg \max_{a_i \in A} p(a_i) \times r(a_i, \{u_f, R\})$$

By choosing action a^* players maximize their chance of being matched by other players in the system and ultimately gaining the payoff $\delta(a^*, a^*)$.

Assuming players are rational, and they will choose the action that is most relevant for the given focus user and recommendation set, a dominant strategy that ensure that players coordinate and maximize their expected utility is attained when players follow the same actions selection probability $p(a_i|\{u_f, R\})$ [70]. This implies that players will be motivated to provide a map when they recognize a map and will prefer to choose skip if a map does not exist.

6.6 Implementation Details

The game is implemented as an online game ¹. The game server is responsible for retrieving user profiles from social networking sites, generating focus user and recommendation

¹Visit at <http://liispapps.uncc.edu/gamemapping>

datasets, and storing all the mapping information. To support these features, we implemented social web application tools and APIs in the game server.

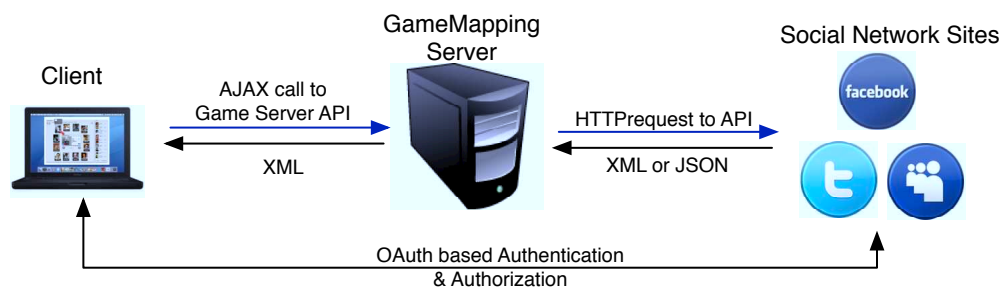


Figure 6.5: The Architecture of GameMapping

Figure 6.5 depicts the architecture of our system. The game server connects to the each social networking site using social web application tools such as Facebook Connect, MySpaceID, and TwitterID. These tools allow our game server to interact with the APIs of each social networking site on behalf of game players. Facebook Connect is based on OAuth 2.0 specification while MySpaceID and TwitterID are based on OAuth 1.0a specification. We also implemented social plugins such as Like Button and Invitation to enhance the popularity and adoption of our game through the friend of friend invitations and word of mouth. We implemented a polling mechanism to enable the retrieval of user's profile information that is based on both server and client technologies (Ajax).

6.6.1 Collusion and Irrational Behavior

It is possible that some players map different profiles intentionally. Based on the game theoretical discussion in Section 6.5, rational users are able to maximize their payoff by selecting the correct actions (map or skip). Irrational players are players who attempt to play the game and provide inaccurate mappings in the hope of gaining high points or simply

affecting our mapping accuracy. Although our game system does not provide a chatting feature, players might collude using another communication channel such as AIM or MSN chat, in order to provide some inaccurate mappings to the game. To prevent collusion among players, our game displays randomly selected datasets to different players, who are allowed to play each game dataset only once. Another irrational behavior is a player providing inaccurate mappings continuously by guessing, and getting l points for each provided map or skip. The game scoring mechanism ensures that rational players converge to a high score faster than guessing players.

In addition, we insert detection datasets into the normal game datasets to detect the irrational players. The detection game datasets are normal dataset that do not contain any correct mapping. If a player provides many mappings for the detection game dataset, there is a high probability the player is an irrational player. We also recorded the amount of time taken by players in making each mapping to detect the irrational players and robots. If a player is an irrational player or a robot, the player might spend less time in each single mapping than rational players since the irrational players might provide mappings without comparing profiles. The game provides a CAPTCHA if the response rate is above the normal rate to prevent robots from playing the game. Finally, we applied mapping confirmation strategy. If an irrational player provides inaccurate mappings, there is a low chance the inaccurate mapping gets a confirming map from other rational players.

6.7 Experiments



To evaluate our approach, we recruited participants who have accounts in multiple social networks by inviting users from MySpace, Twitter, and Facebook. As an incentive to play

the game, we held a two week game competition to encourage people to participate in our research and distributed 10 iTunes gift cards to the top 10 players and an iPod Nano to the top player. One hundred and twenty-four players agreed to play the game, of which 80 were male, 32 female and 12 did not indicate their gender. There were two kinds of game the Facebook-MySpace (FB-MS) game for mapping user profiles between Facebook and MySpace and the Facebook-Twitter (FB-TW) game to map Facebook to Twitter. The FB-MS game was played by 30 players, and 94 players registered and played the FB-TW game. Perhaps users favored playing the FB-TW game due to the increasing popularity of both Facebook and Twitter. During the two weeks game competition, we collected 38,532 Facebook profiles, 8,452 MySpace profiles, 11,775 Twitter profiles and 7,411 profile mappings between user profiles. The collected profiles were used to generate the game datasets which were presented to the players to provide mappings between profiles in different networks.

We manually verified all the profile mappings results. We designed a simple web tool that generates a comparison result of two mapped profiles. The tool compares the last name, first name, age, and gender automatically and requests the inspectors to input a comparison result for profile pictures and countries as Figure 6.6. For each profile mapping, we compared the profile pictures and categorized them into one of 5 types which include, Same, Similar, Different, Picture present only in one site, and None (picture is not present). In case of address and location information, geocoding distances were used to compare both profiles. If the profile information was not enough to make a decision, the inspectors visited profile page in each social networking site to compare both profiles.

Profile Verification

Current status : 1/2055

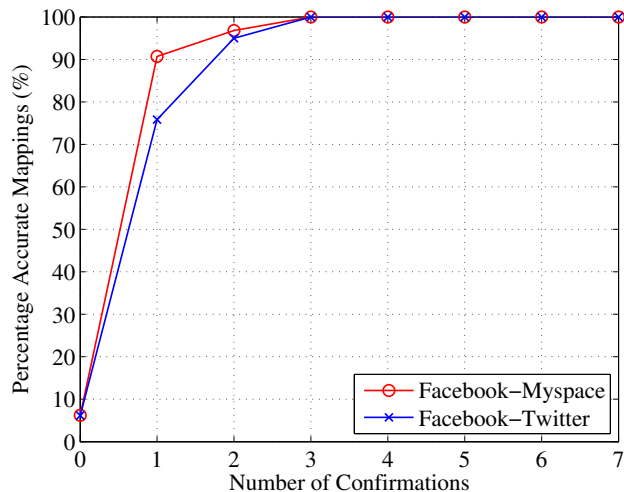
Attribute	Facebook	MySpace	Result
Profile Picture	current profile pic:  profile	 profile	Select radio button: <input type="radio"/> Same <input checked="" type="radio"/> Similar <input type="radio"/> Different <input type="radio"/> Picture is only one site <input type="radio"/> NONE
Last Name	Ko	Ko	Levenshitein Value: 0
First Name	Moo Nam	Moo Nam	Levenshitein Value: 0
Age	39	39	Age match: match
Gender	male	male	Gender match: match
Country	United States		Select radio button: <input type="radio"/> Match <input type="radio"/> Not match <input checked="" type="radio"/> NONE or One site

Match count: 20
Your final decision is :
 Match or Not match

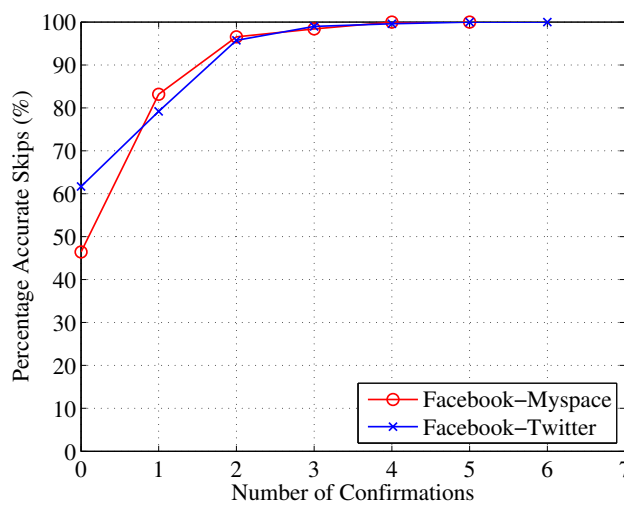
Figure 6.6: The Verification Tool for Profile Mappings

6.7.1 Evaluation of Mapping Results

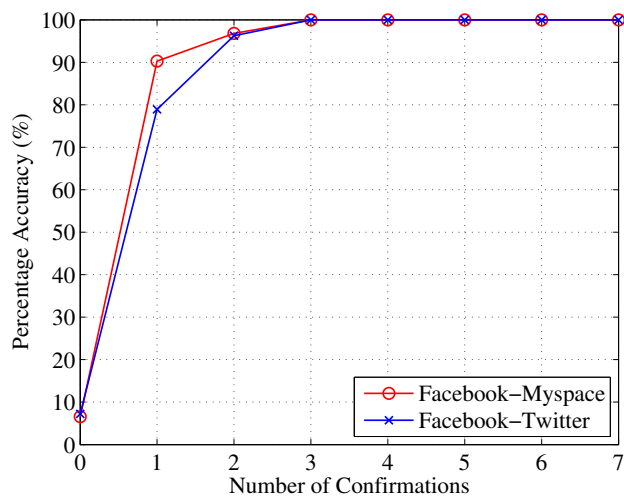
We analyzed the number of player confirmations required for accurate profile mappings and skipings by comparing the mappings provided by the players with the mappings verified manually. Figure 6.7(a) presents the mapping accuracy for different number of confirmations for both kinds of games (FB-MS and FB-TW), as shown the mapping accuracy increases as the number of confirmations increase. Note that, the mapping confirmation plateau's at 100 % after 3 confirmations, which indicates that we need at least 3 confirmations to support 100 % accuracy and 2 confirmations for 95 % mapping accuracy. Figure 6.7(b) presents the skipping accuracy, which follows a similar pattern as the mapping accuracy as it also plateau's at 100 % accuracy after 3 player confirmations for both FB-MS and FB-TW games. The FB-MS mapping and skipping results show a higher accuracy



(a) Mapping Accuracy



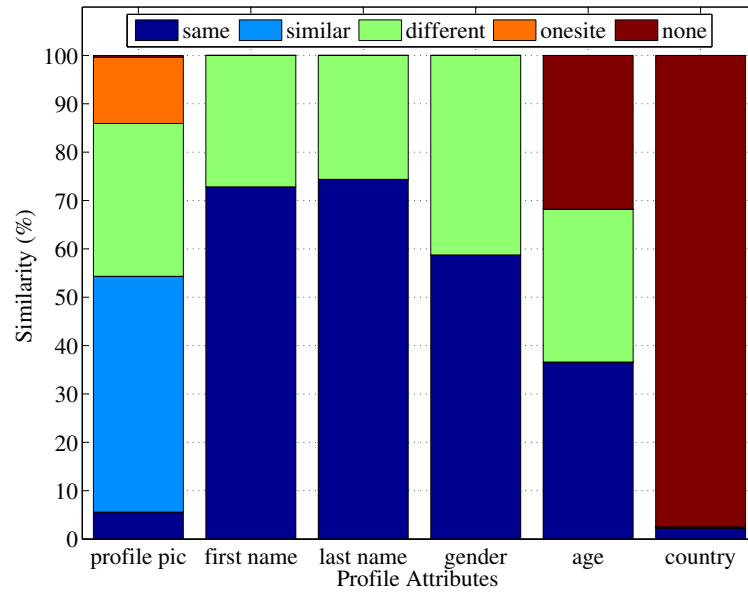
(b) Skip Accuracy



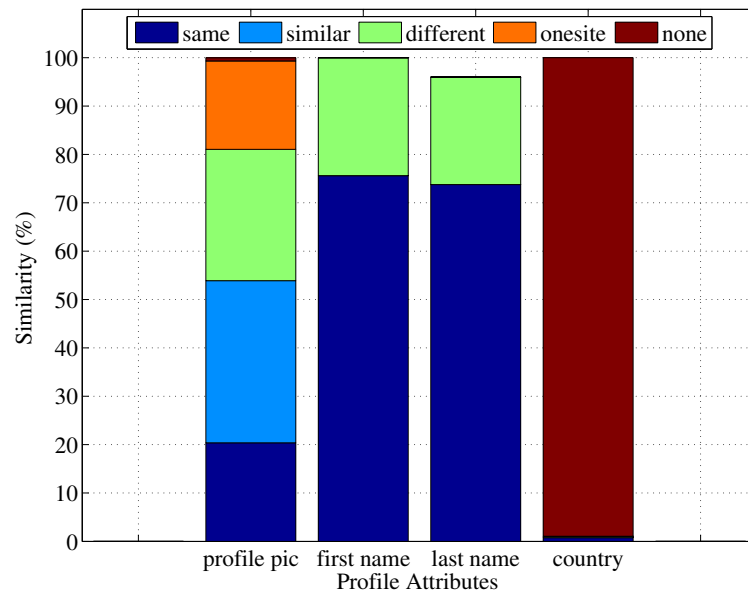
(c) Overall Accuracy

Figure 6.7: Accuracy of Mapping Results.

when compared to the FB-TW case. We believe the reason is the FB-MS dataset provides more user profile information to the player such as gender, age, address and other attributes. It may help players in easily locating similar profiles accurately. Figure 6.7(c) shows the over all confirmation accuracy for both the map and skip cases, which also plateau's at 3 confirmations. Figure 6.8(a) depicts the contribution of each profile attribute in verified FB-MS mapping results. Six attributes such as profile picture, first name, last name, gender, age, and country were used in comparing the profiles in the game. Note that, only 5.6 % of users post the same profile picture and 96.4 % of users do not use a same profile picture (48.7 % use similar pictures, 31.6 % use different pictures, 13.7 % of users have a profile picture in only one site, and 0.4 % of the users do not have profile pictures). This shows that players mapped the same profiles based on other knowledge such as friendship information even if the two profiles did not use the same profile pictures. Last name and first name are important attributes in attribute based mapping. Our results show that 74.4 % of the users have the same last name, and 72.8 % users have the same first name. Which indicates that if the profile mapping is performed by comparing the name attributes, we expect about 73 % matching accuracy. In other words, our game based mapping approach with confirmation is able to detect profile mappings for none matching profile names and provide a 27 % improvement over name based mapping. If gender and age are considered in attribute based mapping, the mapping result is not expected to increase as this usually missing or is low quality. Figure 6.8(b) depicts the contribution of each attribute in the verified FB-TW profile mapping results. In Twitter, only four attributes are used to compare the profiles in the game which include, profile picture, first name, last name, and country. The game datasets are generated from the player's network, Friend of Friend (FOF) network,



(a) Attribute Similarity in FB-MS



(b) Attribute Similarity in FB-TW

Figure 6.8: Attribute Similarity of Mapping Results

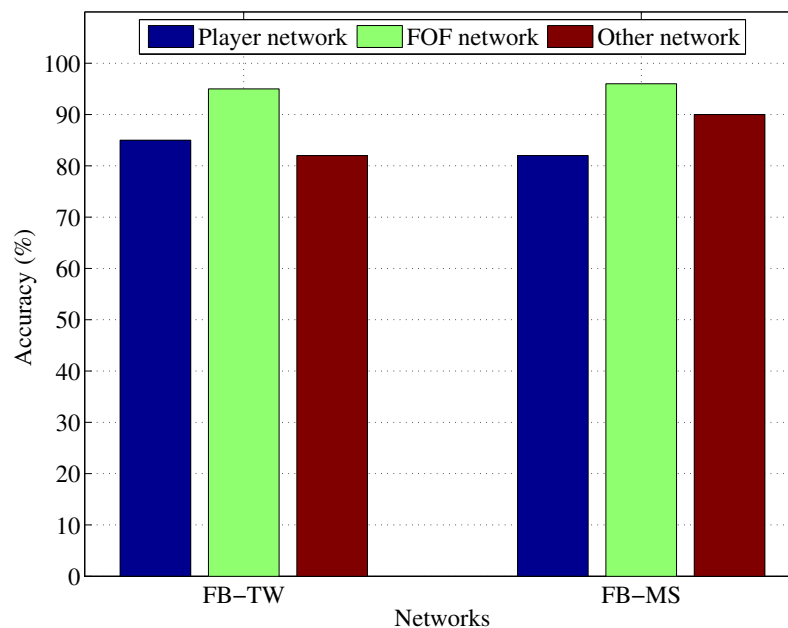


Figure 6.9: Accuracy for Different Networks

and other user's network data. Figure 6.9 depicts the average accuracy of mapping results for different network types. For both FB-MS and FB-TW games, the results show that the accuracy of player network is lower than the accuracy of FOF network. The results did not meet our expectation that the accuracy of player network is higher than the accuracy of FOF network, which would be in turn higher than the accuracy of other network, since the players have more knowledge about their friends. We investigated the whole process of the game to answer the question why the accuracy of player network is lower than the accuracy of FOF network. First, we found that most players did not watch the video tutorial that is on the game homepage before they started the game. It made the players start the game without the knowledge about the game. Second, the players first played the game for their network dataset. Therefore, the players learned how to play the game while they were making incorrect or correct mappings on their network dataset. Then, they were able to play better when they played on the FOF network or other user's network game datasets.

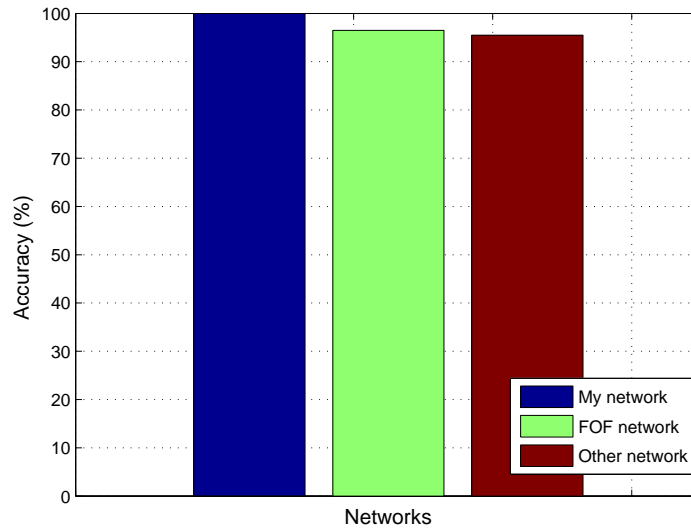


Figure 6.10: Accuracy of Knowledgeable Players

To confirm our discovered cause, we also investigated the mapping data. Figure 6.10 depicts the accuracy of knowledgeable players who knew how to play the game before starting the game. The knowledgeable players provided 100% accuracy on their network, 96.5% accuracy on FOF network, and 95.5% accuracy on other networks. It shows the players' friend relation influence on the accuracy of mapping results. The players provided higher accuracy on their friend profile mappings than unknown people's profile mappings.

To understand how other network based approaches perform in matching the collected profile data. We used the similarity flooding graph matching approach [49], which matches profiles based on both profile attributes and network neighborhood similarity. The algorithm takes two labeled graphs (game datasets) as input and produces as output a mapping between matching profiles. We applied the collected game datasets to the similarity flooding algorithm and the generated an average matching accuracy of 47 %. This result is far less than our proposed game mapping approach. The low accuracy generated by the similar-

ity flooding approach could be attributed to the nature of our dataset. As indicated in Figure 6.8(a) and 6.8(b) profile attributes used in different social networks have a low degree of similarity, users do not always provide correct data or data is missing, attribute similarity is important in similarity flooding as it is used in initialization and flooding phases of the similarity flooding algorithm. In addition, the neighborhood graph information for users in different social networks do not have considerable similarity in friendship connections and neighborhoods which tends to reduce the effectiveness of the flooding based similarity. On the other hand, our proposed approach provides higher accuracy due to the fact that player's map profiles not only based on the profile attributes but also based on the player's implicit knowledge about the profiles.

In the presented experimental results, we show that the game based profile mapping approach is able to generate over 25 % improved profile mapping results when compared to attribute based profile mapping approaches. Moreover, we show that with 3 or more mapping confirmations we are able to generate 100 % accurate profile mappings. Friend relation knowledge influences on the accuracy of mappings for different network types. Our approach shows that human computation and wisdom of crowds can generate accurate user profile mappings across social networking sites.

6.7.2 Evaluation of Irrational Player Detection

In the initial stage of game design, we considered the irrational players and designed prevention and detection strategies as described in Section 6.6.1. To identify the irrational players, we calculated the mapping accuracy distribution of players as presented in Figure 6.11.

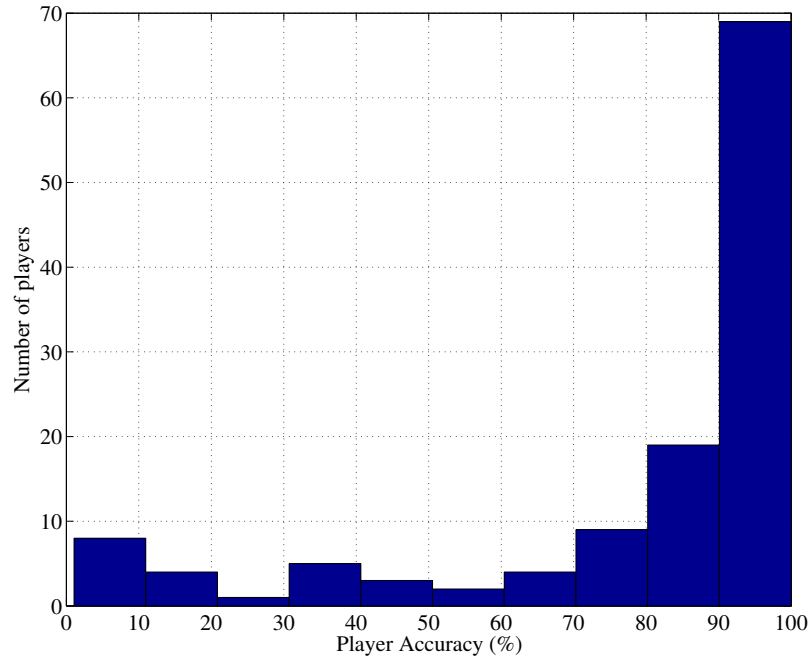


Figure 6.11: Accuracy Distribution of Players.

In our game period, 69 players provide over 90 % mapping accuracy (18 players provided 100 % mapping accuracy), and 8 players provided less than 10 % mapping accuracy. We classify irrational players as either passive or active irrational players. A passive irrational player is a player that provides a small number of mapping, which is lower than the average mapping of all the game players (105 mappings), and has an accuracy of 20 % or less. On the other hand, an irrational player is considered active if he provides more than the average number of mappings and has 20 % accuracy or less. Based on this classification, we discovered 12 irrational players, with 9 passive and 3 active irrational players. The passive irrational players provided 14 mappings on average, which implies that most passive irrational players did not spend much time in playing the game and left it shortly after their registration stage. There might be several reasons behind the reason for their low accuracy. One possible reason is that they did not understand the game and decided to test

it out by providing random mappings. Table 6.1 show a summary of the results extracted from the 3 active irrational players.

Table 6.1: Active Attackers

Active irrational players	Mapping	Accuracy	Average Time	Detection game sets
Player 1	130	6.15 %	7 sec.	played
Player 2	551	3.62 %	0.55 sec.	played
Player 3	2643	1.05 %	1.65 sec.	played

The player 1 spent on average 7 seconds to map each profile and provided 130 mappings with 6.15 % accuracy. The player 2 spent 0.55 seconds to map each profile and provided 551 mappings with 3.62 % accuracy. Both players have low accuracy, and it is evident that player 2 did not review the focus user data or the recommend user profiles instead he preferred to randomly map or skip the presented user. All the three players played the detection game. They provided mappings randomly for the detection game. Therefore, all the above 3 players were detected by the detection game strategy. Another detection strategy was based on comparing the average mapping time, where the average mapping time of the players who have accuracy above 90 % was 6.7 seconds. On the other hand, the average mapping time for the irrational players was 3 seconds. This implies that rational players spend more time to map profiles when compared to irrational players. Moreover, most mapping results from the irrational players did get a few confirmations, and they were

not in the top 10 players.

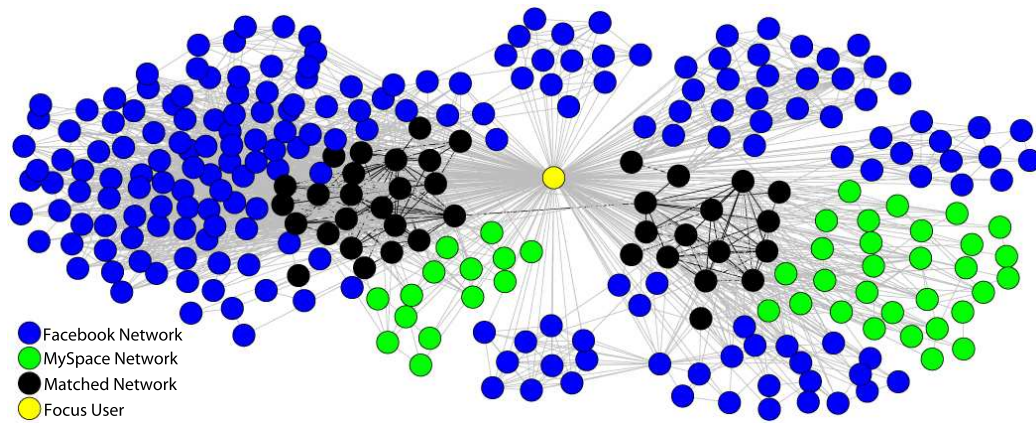
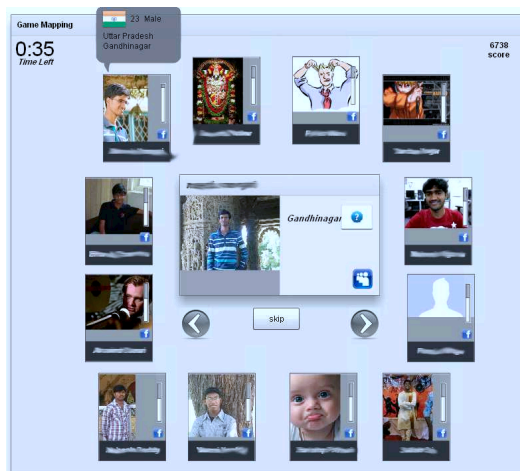
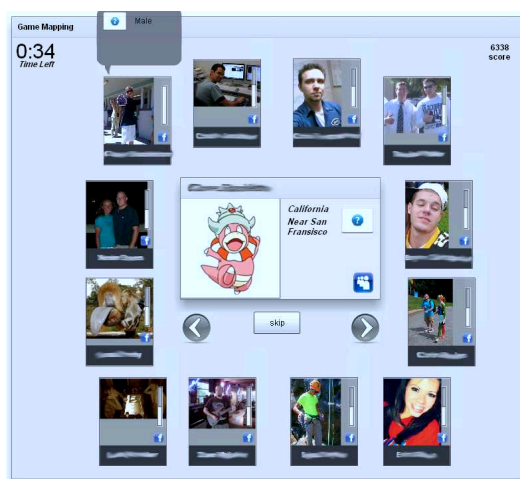


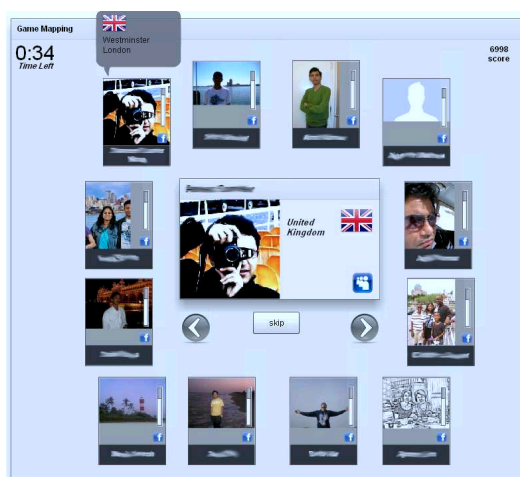
Figure 6.12: Example Matched Network



(a) An Example of Similar Profile Pictures



(b) An Example of Different Profile Pictures



(c) An Example of Same Profile Pictures

Figure 6.13: GameMapping Experimental Results.

CHAPTER 7: CONCLUDING REMARKS

7.1 Summary

In this research, we proposed a new cross-site interaction framework that manages resource sharing and access control across social networking sites. We provided a cross-site access control policy, which enables users to specify policies that allow/deny access to their shared contents across social networking sites. We also proposed the policy levels to provide more flexible choice of cross-site policy enforcement to the content owner. Moreover, we demonstrated the feasibility of the *x-mngr* framework by implementing a photo sharing application *MyCrossAlbum* between Facebook and MySpace. The user study results show that the participants had a positive attitude for the *x-mngr* framework, specified the cross-site policy easily, and understood the sharing status well using the *MyCrossAlbum* interfaces.

We also propose identity-mapping approaches that map users identities across social networking sites. The partial mapping approach based on a supervised learning mechanism provides users identity mapping refer to a small subset of the profile mappings. We provide mechanisms to enable users to fuse identity-mapping decisions that are provided by their friends or others on the social network. The experimental results show that the proposed partial mapping approach provides both high accuracy and precision in performing profile and exception list matching. Furthermore, we propose a Game With A Purpose (GWAP)

approach that provides identity-mappings using a social network game. We provide two types of games: Facebook-MySpace (FB-MS) game and Facebook-Twitter (FB-TW) game. To detect irrational player who provide incorrect mapping intentionally, we also designed and applied an irrational player detection strategies to our game system. In our experiments, the proposed detection strategies detected irrational players effectively. It discovers the active irrational player spent 50 % less time than rational players for mapping, and their most mapping results did not get the agreement from other players. The evaluation of mapping results shows our proposed mapping approach generate higher mapping accuracy (FB-MS: 27 % improvement, FB-TW: 25 % improvement) than the name based mapping results. We also observed that users are able to map their friends, friend of friend, and other network profiles accurately. Finally, we showed that accurate mappings could be concluded if 3 or more rational players agree on it.

7.2 Future Work

This section outlines possible future research directions based on this dissertation.

7.2.1 Assertion Based Cross-Site Interaction Framework

The current proposed framework allows users to share their content with friends on other social networks via *x-mngr*. To implement this framework, it needs a trusted third party that operates the *x-mngr* between different social networking sites. Another possible cross-site interaction framework model is an assertion based cross-site interaction model. Different social networking sites directly interact with each other without a trust third party that operates the *x-mngr*. A social networking site issues an assertion about their user's relationship and other social networking sites make an access control decision based on the issued asser-

tion. For instance, a user in Facebook want to interact with her friends that have accounts in MySpace can use the fact that they are friends on Facebook to access their resources on MySpace.

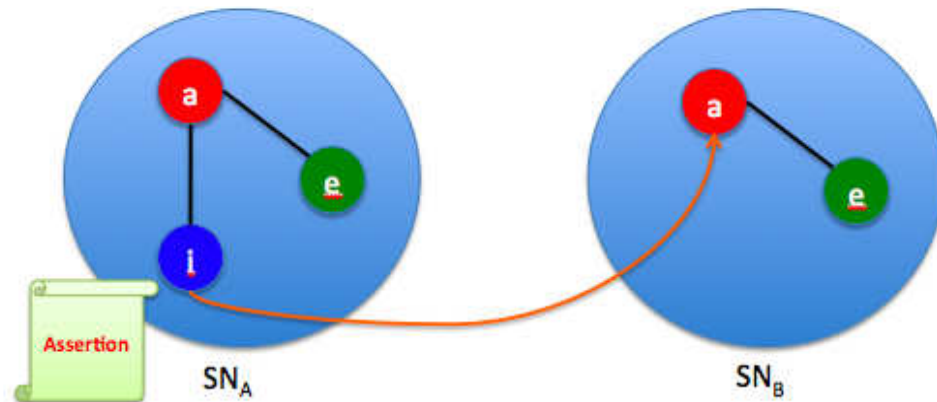


Figure 7.1: Assertion based Cross-Site Interaction

Figure 7.1 shows the social relationships established between users in SN_A and SN_B respectively. Although user i does not have an account in SN_B , the assertion based cross-site interaction model will allow user i in SN_A to access user a 's profile in SN_B using an assertion issued from SN_A . The challenging task of this model is to design the SAML assertion or equivalent functions on the REST services that are used by most social networking sites. Generally, the SAML assertion standard is designed to operate on the SOAP web service. There is no current specification that describes how to add SAML to REST web services. According to [17], theoretically sending SAML assertion on the REST services is possible but parsing or validating the SAML response is not guaranteed since each vendor adopts custom limits on URL length and this will result in truncating long SAML responses. Therefore, designing an assertion based framework on the REST web services is challenge.

7.2.2 Portable Social Graph with Policy

In Web 2.0, various identity management architectures are coexistent, and many websites allow users to select their preferring identity providers in login and registration process. Based on user's preference, each user can select their identity providers such as Facebook, Twitter, Yahoo, Google and MySpace. Although each identity provider uses different identity management technologies such as OpenID and OAuth, they put the user in the middle of the transaction, and allow the user to control their identities. In such competitive environment, several social networking sites have become major identity providers. For instance, the number of new users who select Facebook and Twitter as their identity providers in registration of TypePed had risen rapidly from June 2009 to September 2009, refer to Figure 7.2. Several social networking sites are becoming major identity providers enabling

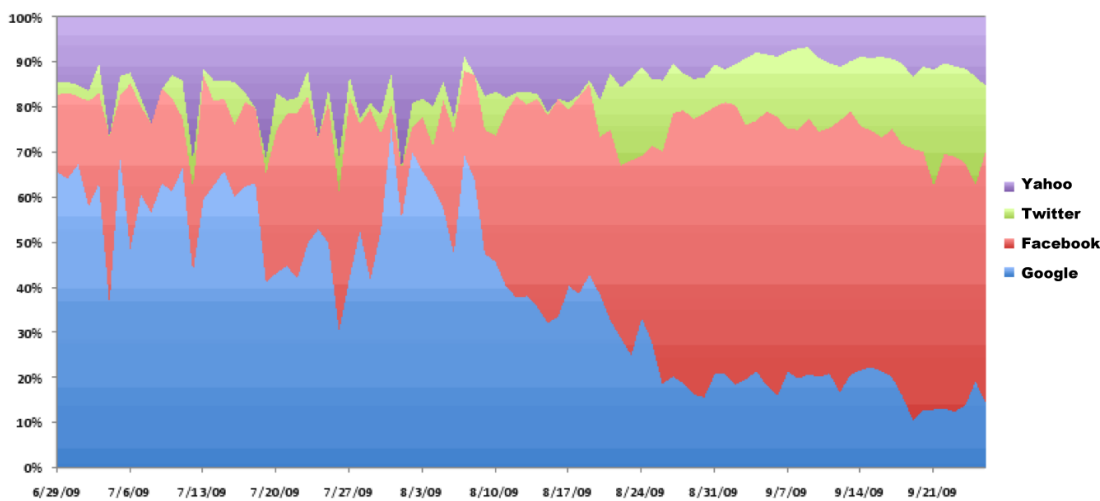


Figure 7.2: Trend of Identity Provider Selection(image source: [38])

users to manage their profile, friends, contents and privacy setting in one place. Moreover, users are able to reuse these social features in other sites via Social Network Connect

Service. However, there are no appropriate privacy protect mechanisms between social networking sites and other sites. For example, users can export their friends from a social networking site to other sites and re-connect friends but can not export the privacy setting with their friends. Exporting a well-managed social graph with its privacy settings will provide better privacy protect to users across the sites because a consistent privacy setting for the same friends is applied across the sites. Therefore, we will investigate a group based access control model that allows the user to craft different privacy settings for different groups of friends and export group of friends with privacy setting together across sites.

BIBLIOGRAPHY

- [1] A. M. Abdulkader. Parallel algorithms for labelled graph matching. *PhD thesis, Colorado School of Mines.*, 1998.
- [2] Gail-Joon Ahn and Moonam Ko. User-centric privacy management for federated identity management. In *CollaborateCom*, pages 187–195, 2007.
- [3] Gail-Joon Ahn, Moonam Ko, and Mohamed Shehab. Privacy-enhanced user-centric identity management. In *Communications, 2009. ICC 09. IEEE International Conference on*, June 2009.
- [4] Gail-Joon Ahn and John Lam. Managing privacy preferences for federated identity management. In *Digital Identity Management*, pages 28–36, 2005.
- [5] Alan Finder. For some, online persona undermines a Résumé. <http://www.nytimes.com/2006/06/11/us/11recruit.html>, June 2006.
- [6] Mansour Alsaleh and Carlisle Adams. Enhancing consumer privacy in the liberty alliance identity federation and web services frameworks. In *Privacy Enhancing Technologies*, pages 59–77, 2006.
- [7] Surapong Auwatanamongkol. Inexact graph matching using a genetic algorithm for image recognition. *Pattern Recognition Letters*, 28(12):1428 – 1437, 2007.
- [8] Luciano Barbosa and Juliana Freire. Combining classifiers to identify online databases. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 431–440, New York, NY, USA, 2007. ACM.
- [9] David A. Basin. A term equality problem equivalent to graph isomorphism. *Inf. Process. Lett.*, 51(2), 1994.
- [10] Vittorio Bertocci, Garrett Serack, and Caleb Baker. *Understanding windows cardspace: an introduction to the concepts and challenges of digital identities*. Addison-Wesley Professional, 2007.
- [11] Vincent D. Blondel, Anah Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM*, 2004.
- [12] Stephen P. Borgatti and Martin G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, October 2006.
- [13] Danah Boyd. *Taken Out of Context: American Teen Sociality in Networked Publics*. Phd dissertation, University of California-Berkeley, School of Information, 2008.
- [14] Ulrik Brandes and Thomas Erlebach. *Network Analysis: Methodological Foundations*. Springer, first edition, 2005.

- [15] Barbara Carminati, Elena Ferrari, and Andrea Perego. Rule-based access control for social networks. In *OTM Workshops (2)*, pages 1734–1744, 2006.
- [16] CBC News. 4 charged after school protest over Facebook suspensions. <http://www.cbc.ca/technology/story/2007/03/23/protest-birchmount.html>, March 2007.
- [17] CBC News. How to use SAML with REST Web Services. <http://tarlogonjava.blogspot.com/2008/05/how-to-use-saml-with-rest-web-services.html>, May 2008.
- [18] Roberto Cesar, Endika Bengoetxea, and Isabelle Bloch. Inexact graph matching using stochastic optimization techniques for facial feature recognition. *Pattern Recognition, International Conference on*, 2:20465, 2002.
- [19] Roberto M. Cesar, Jr., Endika Bengoetxea, Isabelle Bloch, and Pedro Larra naga. Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recogn.*, 38(11):2099–2113, 2005.
- [20] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. The platform for privacy preferences 1.0 (p3p1.0) specification. Technical report, Available at <http://www.w3.org/TR/P3P/>, April 2002.
- [21] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. Use of a p3p user agent by early adopters. In *WPES*, pages 1–10, 2002.
- [22] Declan McCullagh. Privacy groups assail Facebook changes. http://news.cnet.com/8301-13578_3-20006220-38.html, May 2010.
- [23] Anhai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. Ontology matching: A machine learning approach. In *Handbook on Ontologies in Information Systems*. Springer-Verlag, 2004.
- [24] Xin Dong, Alon Halevy, Jayant Madhavan, Ema Nemes, and Jun Zhang. Similarity search for web services. *VLDB*, 2004.
- [25] Ronald Fagin, Amnon Lotem, and Moni Naor. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences* 66, 2002.
- [26] Kuo-Chin Fan, Jeng-Ming Lu, and Gwo-Dong Chen. A feature point clustering approach to the recognition of form documents. *Pattern Recognition*, 31(9), 1998.
- [27] FarmVille Game. Zynga game network inc., <http://www.facebook.com/FarmVille>, 2010.
- [28] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, January 1979.
- [29] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1990.

- [30] Kiran K. Gollu, Stefan Saroiu, and Alec Wolman. A social networking-based access control scheme for personal content. *Proc. 21st ACM Symposium on Operating Systems Principles (SOSP '07)*. Work in progress, 2007.
- [31] L. Gong and X. Qian. The Complexity and Composability of Secure Interoperation. In *SP '94: Proceedings of IEEE Symposium on Security and Privacy*, pages 190–200. IEEE Computer Society, 1994.
- [32] L. Gong and X. Qian. Computational Issues in Secure Interoperation. *IEEE Transaction on Software and Engineering.*, 22(1), Jan 1996.
- [33] Google Inc. Google Maps API Services. <http://code.google.com/apis/maps/>, 2009.
- [34] H. Jenkins and D. Boyd. Discussion: MySpace and Deleting Online Predators Act (DOPA). Available at <http://www.danah.org/papers/MySpaceDOPA.html>, 2006.
- [35] Josef Kittler, Mohamad Hatef, Robert P.W. Duin, and Jiri Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
- [36] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.
- [37] Moonam Ko, Gorrell P. Cheek, Mohamed Shehab, and Ravi S. Sandhu. Social-networks connect services. *IEEE Computer*, 43(8):37–43, 2010.
- [38] Leah Culver. Log in or sign up with OpenID. <http://blog.leahculver.com/2009/11/log-in-or-sign-up-with-openid.html>, November 2009.
- [39] A. Lenhart and M. Madden. Teens, Privacy & Online Social Networks. *Pew Internet & American Life Project*, April 2007.
- [40] Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. Technical Report 8, Soviet Physics Doklady, 1966.
- [41] Liberty Alliance Project. <http://www.projectliberty.org/>.
- [42] Bing Liu. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer, first edition, 2007.
- [43] Marc Langheinrich Lorrie Cranor and Massimo Marchiori. A p3p preference exchange language 1.0 (appel1.0). Technical report, Available at <http://www.w3.org/TR/P3P-preferences>, April 2002.
- [44] Jayant Madhavan, Philip Bernstein, Kuang Chen, Alon Halevy, and Pradeep Shenoy. Corpus-based schema matching. In *In ICDE*, pages 57–68, 2003.
- [45] Eve Maler. Controlling Data Usage with User-Managed Access (UMA), <http://www.w3.org/2010/policy-ws/papers/18-Maler-Paypal.pdf>, 2010.

- [46] Maps of World. Social Networking Websites Popularity Map, . <http://www.mapsofworld.com/world-top-ten/social-networking-websites-popularity-map.html>, 2010.
- [47] Matt McKeon. The Evolution of Privacy on Facebook. <http://mattmckeon.com/facebook-privacy/>, 2010.
- [48] Richard H. McAdams. A focal point theory of expressive law. *Virginia Law Review*, 86(8):pp. 1649–1729, 2000.
- [49] Sergey Melnik, Hector Garcia-molina, and Erhard Rahm. Similarity flooding: A versatile graph matching algorithm. In *Ontology handbook*, pages 117–128, 2002.
- [50] M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132+, June 2001.
- [51] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [52] Nick O'Neill. 10 New Privacy Settings Every Facebook User Should Know. <http://www.allfacebook.com/facebook-privacy-new-2009-12>, December 2009.
- [53] OpenID. <http://openid.net/>.
- [54] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, July 1994.
- [55] P. Boutin. Investors hot on social shopping: Firstgroupon, now livingsocial, <http://venturebeat.com/2010/04/29/livingsocial-funding>, 2007.
- [56] Alex Patriquin. Connecting the Social Graph: Member Overlap at OpenSocial and Facebook, . <http://blog.compete.com/2007/11/12/connecting-the-social-graph-member-overlap-at-opensocial-and-facebook/>, November 2007.
- [57] Johan Perols, Kaushal Chari, and Manish Agrawal. Information market-based decision fusion. *Manage. Sci.*, 55(5):827–842, 2009.
- [58] Birgit Pfitzmann. Privacy in enterprise identity federation - policies for liberty single signon. In *Privacy Enhancing Technologies*, pages 189–204, 2003.
- [59] Pingdom.com. Study: Ages of social network users, . <http://royal.pingdom.com/2010/02/16/study-ages-of-social-network-users/>, February 2010.
- [60] Liberty Alliance Project. Liberty architecture framework for supporting privacy preference expression language (ppels). White paper, Available at <http://www.projectliberty.org>, November 2003.
- [61] E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB*, 2001.

- [62] Randall Stross. How to lose your job on your own time. <http://www.nytimes.com/2007/12/30/business/30digi.html>, December 2007.
- [63] David Recordon and Drummond Reed. Openid 2.0: a platform for user-centric identity management. In *DIM '06: Proceedings of the second ACM workshop on Digital identity management*, pages 11–16, New York, NY, USA, 2006. ACM.
- [64] David S. Rosenblum. What anyone can know: The privacy risks of social networking sites. *IEEE Security & Privacy*, 5(3):40–49, 2007.
- [65] SAML. Security Assertions Markup Language (SAML). <http://xml.coverpages.org/saml.html>, August 2004.
- [66] M. Shehab, E. Bertino, and A. Ghafoor. Secure Collaboration in Mediator-Free Environments. In *CCS '05: Proceedings of the 12th ACM conference on Computer and Communications Security*. ACM Press, Nov 2005.
- [67] Shibboleth. <http://shibboleth.internet2.edu/>.
- [68] James F. Short and Lorine A. Hughes. *Studying youth gangs* / edited by james f. short, jr. and lorine a. hughes, 2006.
- [69] Anna C. Squicciarini, Ayca Azgin Hintoglu, Elisa Bertino, and Yucel Saygin. A privacy preserving assertion based policy language for federation systems. In *SACMAT '07: Proceedings of the 12th ACM symposium on Access control models and technologies*, pages 51–60, New York, NY, USA, 2007. ACM Press.
- [70] Robert Sugden. A theory of focal points. *Economic Journal*, 105(430):533–50, May 1995.
- [71] San-Tsai Sun, Kirstie Hawkey, and Konstantin Beznosov. Secure web 2.0 content sharing beyond walled gardens. In *Proceedings of the 2009 Annual Computer Security Applications Conference, ACSAC '09*, pages 409–418, Washington, DC, USA, 2009. IEEE Computer Society.
- [72] Amin Tootoonchian, Kiran Kumar Gollu, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. Lockr: social access control for web 2.0. In *Proceedings of the first workshop on Online social networks, WOSP '08*, pages 43–48, New York, NY, USA, 2008. ACM.
- [73] Zeynep Tufekci. Can you see me now? audience and disclosure regulation in online social network sites, 2008.
- [74] Stephen Volda, W. Keith Edwards, Mark W. Newman, Rebecca E. Grinter, and Nicolas Ducheneaut. Share and share alike: exploring the user interface affordances of file sharing. In *CHI*, pages 221–230, 2006.
- [75] Luis von Ahn and Laura Dabbish. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326, NY, USA, 2004.

- [76] Luis von Ahn, Ruoran Liu, and Manuel Blum. Peekaboom: a game for locating objects in images. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, NY, USA, 2006.
- [77] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, second edition, 2005.
- [78] Haiyuan Wu, Qian Chen, and Masahiko Yachida. Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(6):557–563, 1999.
- [79] Laura Zager. *Graph Similarity and Matching*. Master dissertation, MIT, 2005.

APPENDIX A: SURVEY RESPONSE

The results consist of 306 responses provided between June 23, 2010 and August 12, 2010.

Part 1. Tell us about your self

Are you male or female?

Female (64)	20.9 %
-------------	--------

Male (242)	79.1 %
------------	--------

(N=306, one response allowed)

What is your age range?

under 18 year old (14)	4.5 %
------------------------	-------

20-29 year old (108)	35.3 %
----------------------	--------

30-39 year old (90)	29.4 %
---------------------	--------

40-49 year old (57)	18.6 %
---------------------	--------

50-59 year old (28)	9.2 %
---------------------	-------

over 60 year old (9)	2.9 %
----------------------	-------

(N=306, one response allowed)

What is the highest level of education you have completed?

Less than High School (8)	2.6 %
High School (42)	13.7 %
2 year College (44)	14.4 %
4 year College (99)	32.4 %
Master's Degree (76)	24.8 %
Doctoral Degree (26)	8.5 %
Other (11)	3.6 %

(N=306, one response allowed)

Part 2. Social networking site experience

Do you have accounts in multiple social networking sites?

Yes- I have accounts in several social networking sites such as Facebook, MySpace and so on (268)	87.6 %
No- I have an account in only one social networking site (36)	11.8 %
No- I don't use social networking sites (2)	0.7 %
(N=306, one response allowed)	

For what reasons are you using multiple social networking sites?

To get different services (205) (Facebook: fun, LinkedIn: professional)	76.5 %
Friends are scattered (137) (College friends: Facebook, Indian friends: Orkut)	51.1 %
To meet others who have similar hobbies (65)	24.3 %
For curiosity (71)	26.5 %
Other (16)	6.0 %
(N=268, multiple responses and manual input allowed)	

Please select two(2) social networking sites that you are mainly using.

Facebook (220)	82.1 %
Myspace (14)	5.2 %
Orkut (11)	4.1 %
LinkedIN (41)	15.3 %
Twitter (218)	81.3 %
Youtube (25)	9.3 %
Flickr (12)	4.5 %
Other (8)	3.0 %

(N=268, multiple responses and manual input allowed)

How often do you access social networking sites?

Constantly (131)	42.8 %
A few times a day (125)	40.8 %
One a day (23)	7.5 %
Once or twice a week (21)	6.9 %
Once a month or less (3)	1.0 %
No answer (3)	1.0 %

(N=306, one response allowed)

What percentage of your friends is duplicated between two social networking sites that you selected in the prior question?

0 % duplicated friends (24)	9.0 %
10 % duplicated friends (107)	39.9 %
20 % duplicated friends (40)	14.9 %
30 % duplicated friends (40)	14.9 %
40 % duplicated friends (10)	3.7 %
50 % duplicated friends (17)	6.3 %
60 % duplicated friends (5)	1.9 %
70 % duplicated friends (12)	4.5 %
80 % duplicated friends (5)	1.9 %
90 % duplicated friends (3)	1.1 %
100 % duplicated friends (5)	1.9 %
(N=268, one response allowed)	

Have you ever used any social applications in social networking sites?

Yes (205)	67.0 %
No (101)	33.0 %
(N=306, one response allowed)	

Have you ever used Facebook Connect, MySpaceID,
or OpenID to register or login to other 3rd party sites?

Yes (199)	65.0 %
No (107)	35.0 %

(N=306, one response allowed)

When you register for a new site, do you like to use
social connect services (Facebook Connect, MySpaceID,
or OpenID) or fill up a registration form and create
an account?

I'd like to use the connect service (121)	60.8 %
I'd like to fill up the form and create an account (78)	39.2 %

(N=199, only answer this question if answered "Yes"
to the previous question)

Part 3. Privacy preference

What kinds of relationships are between you and
your friends on social networking sites?

Family (266)	86.9 %
School friends (260)	84.9 %
Co-workers (239)	78.1 %
Acquaintance (208)	68.0 %
Neighbor (73)	23.9 %
Other (34)	11.1 %

(N=306, multiple responses and manual input allowed)

How do you set up your privacy settings on
the social networking site that you mostly use.

MyStatus and profile picture

Everyone (126)	41.2%
Friends of friends (38)	12.4 %
Friends only (138)	45.4 %
Myself (4)	1.3 %

Bio

Everyone (84)	27.5 %
Friends of friends (46)	15.0 %
Friends only (165)	53.9 %
Myself (11)	3.6 %

Photo album and video

Everyone (40)	13.1 %
Friends of friends (44)	14.4 %
Friends only (200)	65.4 %
Myself (22)	7.2 %

Birthday

Everyone (50)	16.3 %
Friends of friends (39)	12.7 %
Friends only (174)	56.9 %
Myself (43)	14.1 %

Family and relationship

Everyone (40)	13.1 %
Friends of friends (32)	10.5 %
Friends only (191)	62.4 %
Myself (43)	14.1 %

Email address or IM

Everyone (31)	10.1 %
Friends of friends (26)	8.5 %
Friends only (189)	61.8 %
Myself (60)	19.6 %

Phone number and address

Everyone (14)	4.6 %
Friends of friends (19)	6.2 %
Friends only (144)	47.1 %
Myself (129)	42.2 %

(N=306, one response allowed)

Do you use similar privacy settings for
other social networking sites?

Yes (247)	92.2 %
No (21)	7.8 %

(N=268, one response allowed)

Do you organize your friends into customized
groups (Friend Lists) to apply different privacy settings?

Yes (108)	35.3 %
No (198)	64.7 %

(N=306, one response allowed)

Why don't you organize your friends into groups and
apply different privacy policy?

I don't know this group function (61)	30.8 %
I want to use it but I'm lazy (45)	22.7 %
My friends are not many so I don't need it (61)	30.8 %
Other (31)	15.7 %

(N=198, only answer this question if answered "No"

the previous question)

How many customized groups do you have for
managing your friends?

0 group (3)	2.8 %
1 - 3 groups (52)	48.1 %
4 - 6 groups (34)	31.5 %
7 - 10 groups (18)	16.7 %
11 - 15 groups (1)	0.9 %

(N=108, only answer this question if grouped friends)

How do you categorize your friends into
groups (Friend List)?

Based on friends' affiliation (65) (same school or same company)	60.2 %
Based on friendship (68) (best friends or just friends (acquaintance))	63.0 %
Based on location or nationality (7)	6.5 %
Based on common interest (36)	33.3 %
Based on common features (12) (gender, religious, or relationship status)	11.1 %

(N=108, multiple responses and manual input allowed)

Do you think customized groups help your
privacy protection on social networking sites?

Yes (98)	90.7 %
No (10)	9.3 %

(N=108, only answer this question if grouped friends)

Do you block any people on social networking sites?

Yes (206)	67.3 %
No (100)	32.7 %

(N=306, one response allowed)

Have you ever excepted some friends when you
share a content on social networking site?

Yes, I have excepted some friends (121)	39.5 %
No, I have not excepted some friends (126)	41.2 %
I know it but I have not used it (59)	19.3 %

(N=306, one response allowed)

Part 4. Content sharing experience

Have you ever uploaded a same content to
multiple social networking sites to share it with
scattered friends?

Yes (136)	44.4 %
No (170)	55.6 %

(N=306, one response allowed)

I think a content sharing service between
social networking sites is necessary

Strongly Agree (56)	18.3 %
Agree (118)	38.6 %
Neither Agree nor Disagree (91)	29.7 %
Disagree (28)	9.1 %
Strongly Disagree (13)	4.2 %

(N=306, one response allowed)

I want to share content with a specific
group of friends on other social networking sites

Strongly Agree (50)	16.3 %
Agree (118)	38.6 %
Neither Agree nor Disagree (97)	31.7 %
Disagree (33)	10.8 %
Strongly Disagree (8)	2.6 %
(N=306, one response allowed)	

Have you ever used any sharing services to share your
favorite content such as videos, articles, and photos
with your friends?

Yes (232)	75.8 %
No (74)	24.2 %
(N=306, one response allowed)	

When you share a content such as photo, video, and others
with friends, what service do you prefer to use?

Social network services such as Facebook and MySpace (141)	46.1 %
Content sharing services such as Flickr and Youtube (50)	16.3 %
Email Services such as Hotmail and Gmail (28)	9.1 %
Micro blog service such Twitter (82)	26.8 %
Personal blog services such as LiveJournal and Blogger (5)	1.6 %

(N=306, one response allowed)