2010

Gaining system access through information obtained in Web 2.0 sites

Arsenio Guzmán

Gaining system access through information obtained in web 2.0 sites

By

Ing. Arsenio Guzmán

Thesis submitted in partial fulfillment of the requirements for the

degree of Master of Science in

Networking and Systems Administration

Rochester Institute of Technology

B. Thomas Golisano College
of
Computing and Information Sciences

[04/05/2010]

Rochester Institute of Technology

B. Thomas Golisano College
Of Computing and Information Sciences

Master of Science in Networking and Systems Administration

~ Project Approval Form ~
~

Student Name: Arsenio de Jesús Guzmán

Project Title: Gaining system's access through information obtained in Web 2.0 sites

Project Area(s): (circle one)
□ Networking □ Systems Administration
√ Security □ Other _____

~ MS Project Committee ~

Name Signature Date

Prof. Charles Border

_____
Chair

Prof. Bo Yuan

_____
Committee Member

Prof. Arlene Estevez

_____
Committee Member

Project Reproduction Permission Form

Rochester Institute of Technology

B. Thomas Golisano College
of
Computing and Information Sciences

Master of Science in
Networking and Systems Administration

Gaining system access through information obtained in web 2.0 sites

I, Arsenio Guzmán, hereby grant permission to the Wallace Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part. Any reproduction must not be for commercial use or profit.

Date: _____ Signature of Author: _____

Table of Contents

Abstract

Along with growth of the internet in recent years, a related phenomenon is also growing: the complaints by internet users that their accounts have been hacked by others. What are the reasons behind this? Are those hackers really that good that they can find out our passwords and almost magically gain access to our systems? This project focuses on the ways in which personal information is propagated by users on the Internet by web 2.0 sites like social networks as well as on the problems inherited by password reset methods, which are how we inadvertently provide others with easy access to our systems. Our findings clearly expose how the lack of interest by most users in interaction with web 2.0 sites, especially password handling, causes trouble. We conclude that gaining system access through accumulating personal information is entirely possible and does not require an expert. If we are aware and take security measures, however, we can make it very difficult for people with malicious intentions to gain access to our systems.

Introduction

Background

For many years security for system access has been a nightmare for every kind of application, especially for web applications that have a great number of users accessing their resources, thus are more exposed. One of the biggest cases to hit the news was in 2008, when Sarah Palin's email account in Yahoo (the Senator of Alaska) was hacked for revealing information. In my experience as web developer for Pontificia Universidad Católica Madre y Maestra (PUCMM), I cannot count how many times we have heard from users who come to us saying that they "lost" their password and they think that someone has stolen it. The situation seems to be worse for systems like ours and Yahoo, public email systems where tons of users create and access their accounts every day.

But why has system access security been compromised for so long? To answer this question, we have to look at both sides of the coin. First, we have to explore how applications are developed around system access and we also need to know how users behave when it comes to security. There are two other aspects of system access that are related to the two previous ideas: How strong are the users' passwords? And what are the password recovery mechanisms for these applications? All these exploratory queries can be summed up in one crucial question: How is the information handled that deals with system access security?

Information handling seems to be the key to the success or failure of system access security. Why? It is quite simple. Passwords are the information that users provide to check their identities in a particular application. The words that users choose for passwords range from simple to complex and, as the name "password" suggests, sometimes these words provide an easy "pass" access.

There are varied forms for recovering passwords, but the oldest and best known is where information about the users is checked to probe if the user is who he or she claims to be. What is the big deal about this handling of information? In fact, not much information today is secure and private. Thanks to the Internet, the contrary is true--in most cases information is wide open to the public. For example, in the case of Sarah Palin's hacked email account, what occurred was that a hacker found out that the password recovery mechanism was a secret question about her birth date.

How does a user's personal information become public in the Internet? Part of the responsibility can be attributed to the growing popularity of web 2.0 sites. In general, we can define web 2.0 sites as the most popular technologies in the web. They are based on a new way of interacting and collaborating between users. This new application brings a new experience to the end user, an experience based on client interface, thanks to technologies such as AJAX, which add to web interactions what was previously only possible for Desktop applications. (Pilgrim, 2008)

The web 2.0 applications are very attractive to both teachers and learners in modern educational applications because mechanisms can be established so that students can share their ideas with classmates and teachers, enabling and facilitating the active participation of each user. Many of the activities that are accomplished by web 2.0 applications and services are the publishing and storing of textual information by individuals (blogs) and collectively (wikis), as well as audio recordings (podcasts), video material (vidcasts), pictures, etc.

Today most educational institutions are implementing web 2.0 applications as part of the set of tools available to students for sharing information. Wikis, blogs, and social bookmarking are now commonly used in learning, but in many cases the information that is exposed on those applications, especially on social networks like Facebook, My Spaces, and others, reveals important user data.

Many efforts are internally addressed by educational institutions trying to protect the data of the students; these actions, in many cases, obey federal standards to establish security requirements. But when students and professors start using many of these web 2.0 sites, there is no longer any guarantee that important information will not be leaked that could be used for people whose purposes are malicious.

One of the biggest security problems is a direct consequence of this leak of information, which puts system access to these web 2.0 sites in jeopardy, along with the institutions' other systems like LMS and Email.

Research Focus

My project consisted of establishing a scenario where I sought out the information that users expose through those web 2.0 applications, identifying what the user tendency is with regard to publishing personal information and seeing if applications encourage users to provide information (that could be used for malicious purposes) in their profiles. I also analyzed how the system access security is handled in terms of password complexity and how secure the password recovery systems are for web applications. At the end of the study, with information recollected for the web 2.0 sites, I established the relationship between how information is handled and how system access is gained.

This project was based on students and teachers from various universities, but especially from the Pontificia Universidad Católica Madre y Maestra, and the web 2.0 sites were narrowed down to the most popular social networks in the web.

Research Aims and Individual Objectives

The aim of this project is certainly not to instruct anyone on how to gain access to others' accounts. My interests are to expose the problem caused when, via a series of simple steps; it is possible to break into the system access security of various web sites, and to provide a solution--guidelines for developers and users for avoiding these access problems.

The project's individual objectives are:

1. Identify the consequences of information leaks in web 2.0 sites
2. Analyze password security and recovery mechanism of web 2.0 sites.
3. Examine what information could be collected and what methods made possible a massive collection of information in web 2.0 sites.
4. Analyze how is possible to gain access to a web 2.0 site with the collected information.
5. Formulate the necessary recommendations for the secure handling of information and system access for web sites developers and users.

Literature Review

Password Handling and Recovery

Password Handling

Burnett establishes a series of recommendations for use of good passwords, exposing an important aspect for the selection of passwords, that they should not be a simple "pass" word. He recommends building a strong password for individual or system access based on the following rules: Complexity, Uniqueness and Secrecy. (Burnett, Perfect Passwords: Selection, Protection and Authentication, 2005)

In relation to the information that we normally use to choose our passwords, it is recommended to take into consideration the rules of complexity and uniqueness and apply the following recommendations for building passwords:

Complexity:

- Use a combination of at least 3 elements from characters, numbers, symbols, words and phrases.
- Use uppercase in a position different from the first letter.
- Use one or two numbers, not at the beginning or end of the password
- Use punctuation or other symbols in the password

Uniqueness

- Do not use common passwords.

- Do not use the same password more than once.
- Do not use personal information or information related to us.
- Use different patterns or sequences.
- Refresh our password from 3 to 6 months.

Uniqueness is a dilemma. After all, we need a password that is easy to remember, so we do not find it easy to follow those recommendations.

Password Recovery

Another undeniable reality is that users tend to forget or lose their passwords sooner or later, so most systems, especially web systems, have to deal with recovery mechanisms so that users can get access again. Password recovery is actually a password reset, letting users assign their own passwords through the system, thus avoiding the overhead of a more personalized method like temporary passwords via phone calls, faxes, or direct presence of the user. As Burnett said, companies with high security still demand a personal interaction for recovering passwords, but this is not the case for most web systems, especially web 2.0 sites. (Burnett, Hacking the Code : ASP. NET Web Application Security, 2004)

Password recovery is basically handled by following methods:

- Sending information via email.
- Using a secret question.
- Assigning a temporary password.
- Using a password hint.

The password recovery method of using a secret question is based on a challenge--the knowledge of the user about him or herself, in most cases asking for personal information that is expected to be in the private context of the user. Burnett exposes in the following list of facts why these secret questions can be unsecure:

- An attacker can often discover the information with casual research.
- The answer to the question is usually a fact that will never change.
- Users reuse the same secret questions and answers across multiple web sites.
- Someone close to the individual may know the answer to many of the questions.
- People rarely, if ever, change their secret questions.
- The answers are often case-insensitive and usually contain a limited character set.
- Some questions have a limited number of answers.
- With some questions, many people will have the same common answers.

Also, although many sites have a list of secret questions that could be considered secured for a user to select among, some sites offer the possibility of creating our own secret question, adding a potential insecurity for users who are not very aware of the repercussions of these kinds of selections.

Information Revealed in Web 2.0 Sites

Web 2.0 Sites

"Evolved from traditional web practices, and greatly influenced by the consumer-oriented economy, Web 2.0 is end-user focused architecture. It has been driven by the information and communication explosion that transcends geographical, networking, and cultural boundaries". (Noureddine & Damodaran, 2008)

This new architecture of interaction with the web transforms the users into active contributors, customizing media and technology, sharing content, and participating in the web design and programming. (Mircea Kristaly, 2008)

While some time ago, web applications were easily distinguishable from their desktop counterparts due to their design and point-click-wait interaction, today's web 2.0 applications are often recognizable as being web applications only at second glance. Due to techniques such as Ajax and Flash, responses from the user interface now behave similar to desktop applications (as long as a fast Internet connection is available). (Ullrich, 2008)

In conclusion, some of the most important concepts in web 2.0 are:

• Usability: the application must be easy (natural) to use;

• Standardization: compliance with current standards in web applications development (i.e. XHTML for source code, WAI for accessibility, etc.);

• Design: a web application must have a pleasant, yet practical, look-and-feel for its user-interface;

• Participation: the application allows the user to intervene at content level (i.e. edit, tagging, sharing, etc.);

• Convergence: the web application makes use of many different technologies, yet it must present to the user a unique, coherent interface. (Mircea Kristaly, 2008)

Now, even if we entirely take away the positive things that web 2.0 brings to the educational sector, hackers are still often interested in targeting these applications to break into the security of the companies at both the clients' computers and the servers' computers. The reason is that both sides of the Internet connection hold personal and business information that lures hackers. (Noureddine & Damodaran, 2008)

Methods of Collecting User Information

Web 2.0 services employ different measures for increasing user contributions and participation, for instance by building trust (e. g., offering users to leave with an export of their data), by explicit licenses (often open licenses such as Creative Commons), and, paradoxically, by making content accessible through RSS syndication and APIs (Application Program Interfaced). Behind the user provided data of web 2.0 lies the Semantic Web with its vision to make the data currently hidden in databases available for use by machines. (Ullrich, 2008)

For the following traditional web 2.0 sites, Facebook, Twitter, MySpace, MSN Live Spaces, and Hi5, we can confirm that an API exists to reiterate the user information that can be personally accessed. Most of these APIs are based in REST-Like interface, where methods are exposed through HTTP GET and Post calls over the internet to the server. In addition to the REST interface, the social network Hi5 offers a SOAP API, leaving the information exposed.

Each one of these sites exposes the information about their users and friends (or contacts), along with the methods used to query that information, in the following parts of its API:

Web 2.0 Sites	API methods for Data
Facebook	Data Retrieval Methods: GetComments, GetFriends, GetNotes, GetUserInfo.
Twitter	User Methods: userShow
MSN Live Spaces	**Live Services User Data APIs -> Windows Live Contacts API - Beta 1.0: ContactView**
MySpace	People Api: **GET v1 users userId details**
Hi5	People Schema

Security Issues in Web 2.0 Sites

As Noureddine shows, there are known security and reliability issues in web 2.0. We must take into consideration that, today, web browsers are capable of performing complex networking operations while rendering a page on the user's screen, but these capabilities also open up many doors and windows for hackers to enter and gain unauthorized access to user- and business-sensitive data. (Noureddine & Damodaran, 2008)

One of the problems of web 2.0 is the integration of the content generated by users into web sites. Anyone could upload content, including hackers, spammers, and people with malicious intent. Content with questionable quality is also wide spread. Noureddine exposes the primary causes of such unreliable content, which include the lack of filtering by web service administrators. Filtering content based on quality requires massive human resources; therefore it is often unjustified, economically.

Another cause of unreliable content is user tolerance. Users tolerate low quality content because information on the web is free, accessible, and disposable. (Noureddine & Damodaran, 2008)

Another fact that we have to keep in mind is that web 2.0 services reach for a wider range of clients than the PC browser. They allow access from and dissemination of data to devices such as mobile phones, PDAs, game consoles, etc. These devices are often less secure that the PC and easier for hackers to break into. (Ullrich, 2008)

It is suggested that developers who want to help prevent these security problem must analyze Usage Scenarios and apply Security Use Cases. Usage scenario analysis and modeling is critical in the web development area because they allow for the breakdown of security of the system as a whole. Developers must know their users, whether or not the application requires users to provide credentials. (Noureddine & Damodaran, 2008)

Research Methods

After establishing the project's first objective in the introduction and literature review chapters, my empirical research addresses and completes, in combination with the literature review, the second, third and fourth objectives of the research.

For the second objective, I investigated how users and web sites actually handle passwords and use their password recovering systems, contrasting these recommendations from authors in the Literature Review.

For the third objective, I presented what information users tend to publish in those webs 2.0 sites and what information web sites capture from users to format their profiles. This leads to the fourth objective, where the relationship between the current information that can easily be found and how that information could be used to break system access is presented.

Research Strategy

My research strategy was experimental research that exposed the problem of how information from web 2.0 sites could be used to gain system access, analyzing what the factors are that help to produce this problem and also presenting recommendations to try to stop or prevent users and websites from doing so. As mentioned earlier, I am a web

developer from PUCMM in the Dominican Republic, and, on a daily basis, I have to face and address this reality of account stealing and how to strengthen our recovery system.

Data Collection

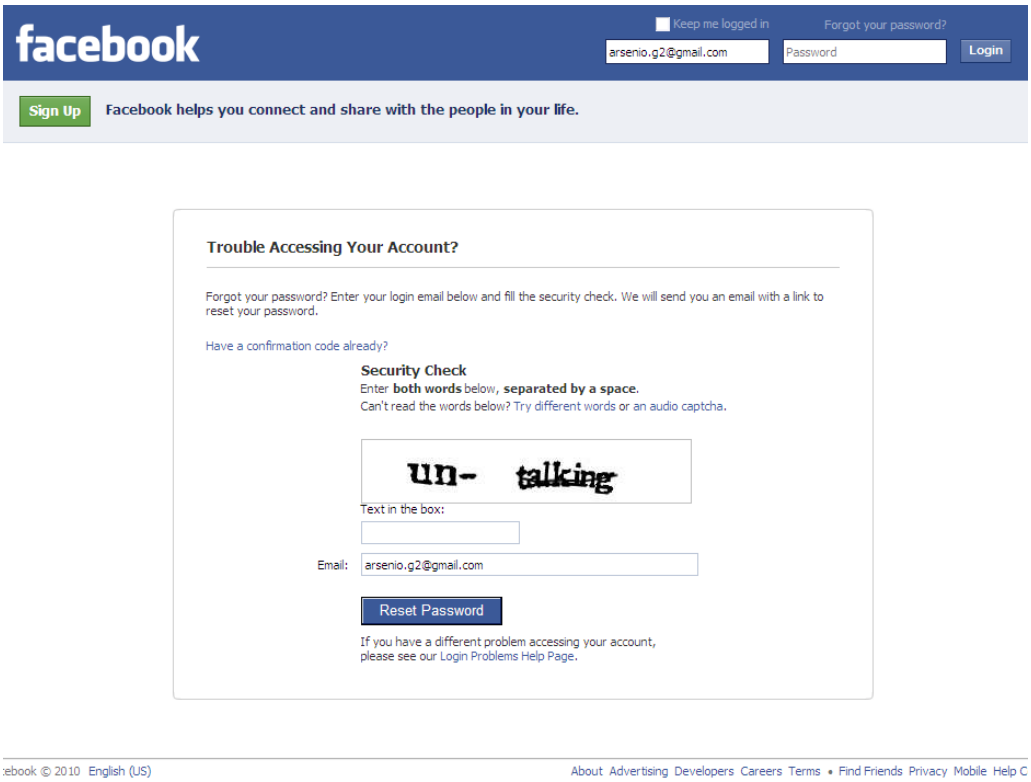
In order to collect the information to fulfill the project's second and third objectives, I conducted a manual inspection of the password security and password recovery processes, and the information published in the profiles of the users and friends or contacts of the users from each of the principal social networks in the web. I also conducted a survey to triangle the collected information about user practices and applications, regardless of the information collected and password handling. To fulfill the fourth objective, data obtained from the profiles was used as input in the investigation of how to gain access to the system.

Manual Inspection–First Part

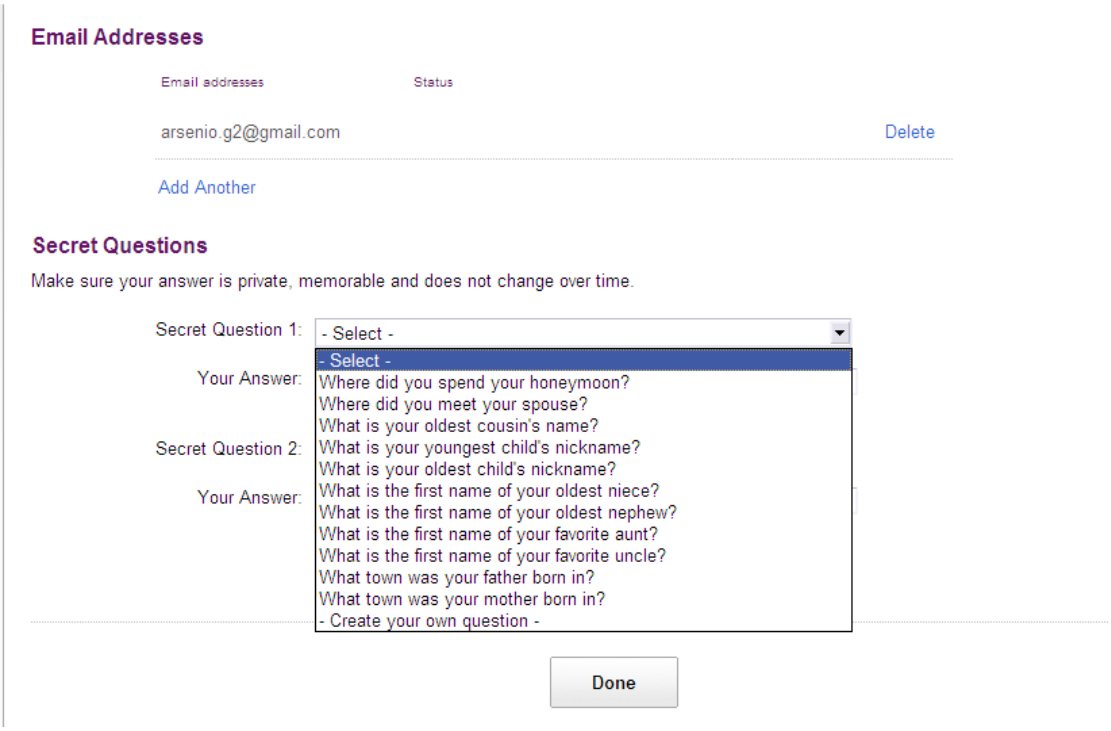
In the first part of the manual inspection, I created a new account and filled in the necessary data in the profile of each one of the principal social networks. For each site the following information was gathered (Appendix B):

- Data needed at registration.
- Password requirement.
- Preferred password recovery process.
- Protocol issues in authentication and password recovery.

Account Creation in Face Book



Password Recovery Method of Facebook



Password Recovery methods and Secret Questions for the Yahoo email account.

Manual Inspection – Second Part

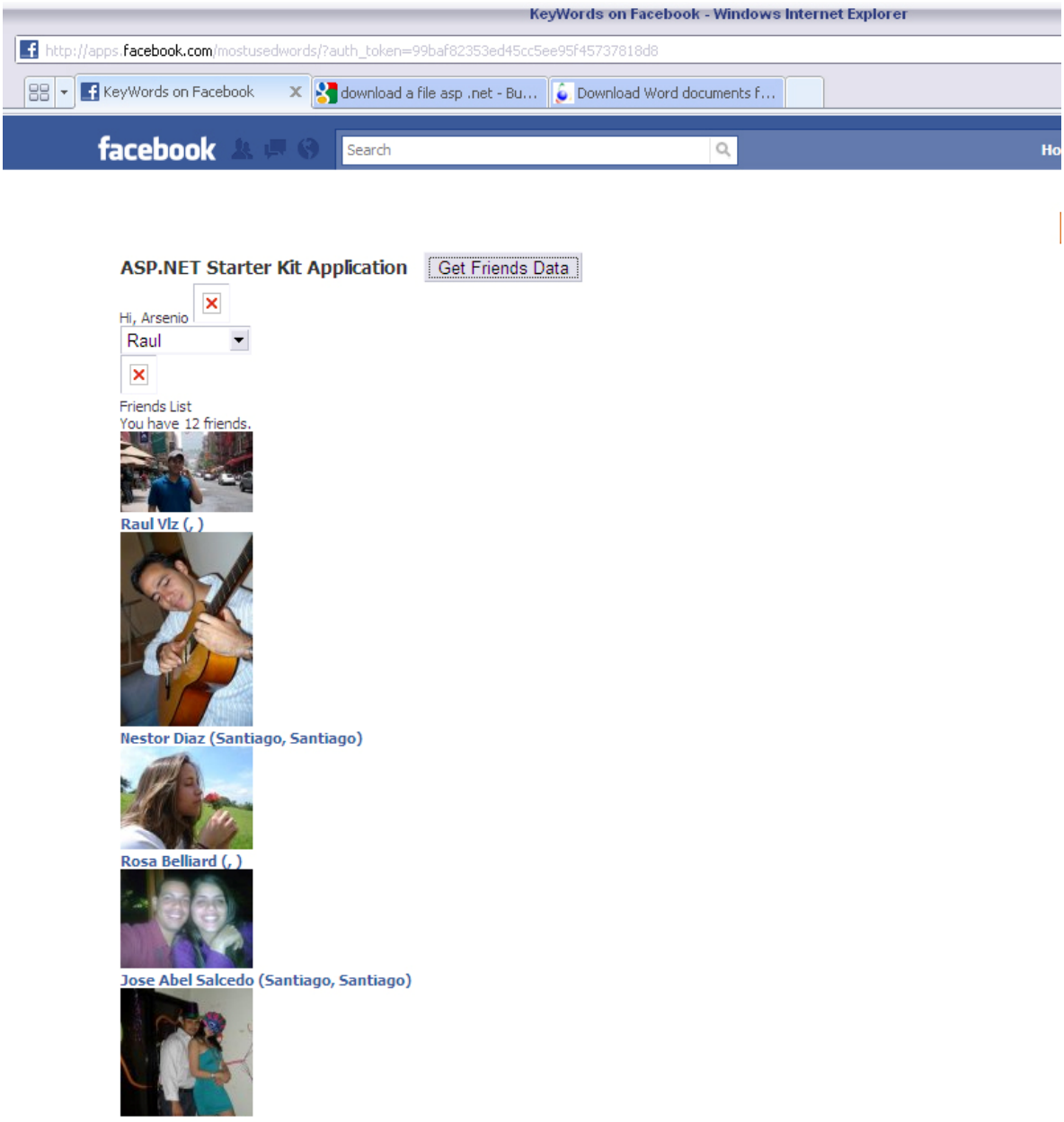
The second part of the manual inspection was accomplished by elaborating a web application to collect profile information for user through the API available from this network. Facebook was selected as the preferred network for the analysis, with its respective API.

The developed application was a Facebook web application. The API was an open abstraction of the Facebook API adapted to the .Net Framework obtained for the Codeplex project Facebook Developer Toolkit at the following address:

http://facebooktoolkit.codeplex.com. In order to get the Facebook application working, in our Facebook profile was active the developer option with the current information:

- **Application Name:** Keywords.
- **Canvas Page Address:** http://apps.facebook.com/mostusedwords
- **Canvas Callback Address:** http://localhost:4349/facebookapp1/

When everything was completed, I received an Application ID, Application Key, and Secret Code. The software used for development was Visual Studio 2008. Requirements from that application are .Net Framework 3.5 and an Application Key and Secret from the Facebook Application. The data collected from Facebook was structured as an XML file with user structure from the Facebook API (Appendix C).



Survey

To conduct the survey, I installed, configured and used, an open source web-based software called LimeSurvey (http://www.limesurvey.org) in our local environment that is capable of

providing us with the processed data to be exported as a CSV, XLS file for Excel or data files for SPSS.

This survey was conducted for all teachers and students of our campus, representing a potential universe of 5,000 people in a period of two (2) weeks. To determine sample size, we used a Sample Calculator for the website www.surveysystem.com with the following parameters:

Confidence level: 95%
Population: 5000
Confidence interval: we set this value bellow of 2 digits. (8.6)
Sample size: 128

Also I checked that the Confidence interval was set by a percentage of a 50% of the selection of an answer.

As previously mentioned, the survey covered the following points and objectives:

- The expertise of users in the internet
- Preferred web 2.0 social networks and configurations
- Information that users regularly expose
- If user has safety habits with password
- If users have forgotten and recovered a password

Investigation for Gaining System Access

This was also a two part process:

a) Data used and transformed to possible passwords for a specific user.
b) Data used to response the secret question and answer in cases where this was the selected method for password reset.

In order to transform the data, a command line application based on .Net Framework 2.0 in the development software Visual Studio 2008 was created. The application was invoked using:

C:\>getPassData inputfile.xml output.txt

The first parameter was the xml obtained from our previous Facebook application, and the second parameter the output file with all the password combinations in each line. The input file was processed with the next approach:

Information	**Data Type**
Birthdate	Number (ddmmyyyy)
Full Name	String (First Name, Last Name)
Phone	Number
Occupation	String
Academic Level	String
Music	String
Address	String and Number

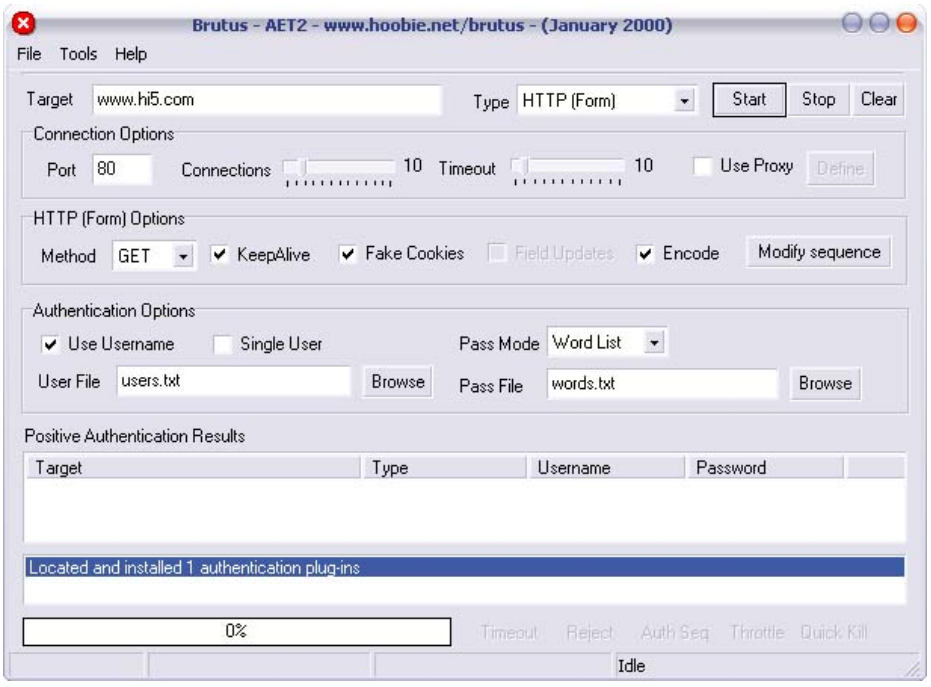
For password combinations, the next criteria were applied:

1. Information of type string where minimum length was 6 characters.
2. Information of type number where minimum length was 6 characters.
3. Combination of type string with number at the beginning and then with number at the end where the combination minimum length was 6 characters.

Examples:

Arsenio	03101982	arsenio1982	1982arsenio
Coldplay	coldplay34 (34 could be a number from the Address)		

The words dictionary could be the input for the password file of the application, like Brutus to perform a Brute Force Attack and confirm the password for a specific or range of users.



For procedure B, I proceeded to a manual process of password recovery.

Analysis of Information

To analyze the survey information, the data was output as SPSS files.

With the SSPS software ver. 16, the data was used in a series of crosstabs analyses with the options Chi Square and Regression. The result from these analyses was triangle with the other manual inspection in order to reach the conclusions.

Limitations

One of the limitations due to the manner in which the survey was conducted is that invitations were sent by email to the participants, even though at the moment the University is in a transition of the email system, so that users do not have 100% access to check information.

Another possible limitation (I did not actually do the procedure) could be that many web sites use active CAPTCHA methods to avoid brute force attacks after a defined number of authentication failures.

Observations and Findings

Manual Inspection

In Facebook, the following information could be obtained for the user and friends of the user through their API used in our Facebook web application: [About Me], [Activities], [Affiliations], [Birthdate], [Birthdate Date], [Books], [Current Location], [Education], [Interest], [Movies], [Music], [Political], [Quotes], [Relationship], [Religion], [Sex], [TV], [Work History]. (Appendix C)

Our Facebook application is in our local environment, and basically consists of a single button called **Get Friends Data,** that allows us to get the profile information of all friends or for a current selected friend.

In addition to this information for the current user, I could also get [Comments] from his or her posts. I also found that when friends have the profile Private, some of this information is not available to query.

It appears that only email providers today use the SQSA method for password recovery. In my inspection, I found that all the big mail providers like Google, Hotmail, and Yahoo used SQSA as a second alternative (Appendix B), but the first alternative, as in web 2.0 social networks, is an alternative address. In Yahoo‘s case, you have to provide the SQSA two times.

For Google, if you have an alternative address, you have to wait for 24 hours before you can try to recover your password through a secret question.

For Hotmail, in my case, something strange happened. My secret question was: **What is my favorite art?** But this secret question is no longer in the pool of questions available to select.

If we reveal the information that is exposed in the Facebook profile of our friends, the following Secret Questions seem to be insecure in cases when they are selected and match complete information.

Google:

- What was your first telephone number?
- Write your own question.

Hotmail:

- What is the birth date place of your mother?
- Who is your best friend from childhood?
- What is your favorite person from history?

Yahoo:

(Part 1)

- What is your youngest child's nickname?
- What is your oldest child's nickname?

- What town was your father born in?
- What town was your mother born in?

(Part 2)

- Who is your favorite author?
- What is your favorite book?
- What is the last name of your favorite musician?
- Who is your all-time favorite movie character?

Also, Google and Yahoo let you write your own secret question, which could insecure, particularly in the case of careless and inexpert users.

From among 15 people with Windows Live accounts (Hotmail), 12 with Google accounts, and 5 people with Yahoo accounts, I got the following results matching their selected secret questions with possible information collected from their profile:

	Total accounts	Possible Password Reset
Windows Live	15	3
Google	12	1
Yahoo	5	0

Among the evaluated sites, the only sites that do not use a certificate from a valid web authority (https) for authentication and password recovery are MySpace and Hi5.

Table of Security Parameters for account creation and access

	Facebook	Twitter	MySpace	Live Spaces	Hi5	Google	Yahoo
Minimum password length	6 - 50	6	6 - 50	6	6 - 20	8	6 - 32
Web Protocol for Authentication	https	https	http	https	http	https	https
Personal Information in Registration	Gender, Birthdate	none	Gender, Birthdate	Country, Birthdate, Gender	Birthdate	none	Country, Gender, Birthdate
Captcha Mechanism in Authentication	yes	no	no	no	no	yes	no
Number of permitted failure attempts	3	no limit	no limit	no limit	no limit	7	no limit
Password Recovery Mechanism	Alternative mail address	Alternative mail address	Alternative mail address	Alternative mail address, Secret Question	Alternative mail address	Alternative mail address, Secret Question	Alternative mail address, Secret Question
Captcha Mechanism in Password Recovery	yes	no	yes	yes	no		yes

Survey Findings

Social Networks and Information Revealed

The following results were obtained from the survey analysis in the SSPS software:

Most Used Social Networks

Number	Website	Rank
1	Facebook	104
2	Twitter	45
3	MySpaces	47
4	Windows Live Spaces	41
5	Hi5	77
6	Others	10

- Most users who considered themselves experts in the use of Internet left their profile Public in their social networks.
- The only information that users are afraid to provide is their address.
- Regardless, the network users are disposed to post similar information with small variance, like Twitter having the highest rate for comments and MySpace for music.
- Twitter has the highest value of people who enter daily or more and Hi5 the least.
- People with private profiles post, in most cases, more information, with the exception of their birth date, than could more found in public profiles.

Password Handling and Password Recovery

- Most users use the same password for their social networks, ranging from 50% for experts users to 83% for less experienced users.
- A third of the users has forgotten or lost their password.
- 22% percent of the users use words or numbers for passwords, 29 % use a combination of words and numbers, and 48% use passwords with random letters and numbers (principally expert users).
- A total of 17% of the users think that someone has hacked their accounts, wherein 33% are less experienced users and 14% expert users.
- A person who feels insecure in these social networks tends to post 10% less than people who feel secure.
- No matter the social networks to which a user belongs, at least 33% have their profile public.
- Among people who use the same password, 29% use words or numbers, 35% a combination of the previous two, and 36% a random combination of words and letters.
- Most people who have lost their password recovered it with the method of ”instructions sent by mail.”
- Among people who use the same password, 65% thought that theirs were hacked
- Among people who thought theirs were hacked, 20 % used words or numbers, 40% a combination, and 40% random letters.
- Among people who recovered their password with a Secret Question, only 33% were from MSN

Final Results and Conclusions

For the previous Observation and Finding chapters, I correlated the information with my conclusions from the Literature Review, and for each objective I presented the current results, also taking into consideration the perspective of users and the web site developers.

Consequences of leak of information in web 2.0 sites.

The biggest problem that was concluded in our project is that personal information posted in web 2.0 sites is in most cases one of two: plain password or part the password that users has for authentication or the answer to the Secret Question exposed for password recovery in email system.

Password Handling and Password Recovery

In the order of password handling it is concluded that user are in most cases careless in relation to password security. When the recommendation for building strong passwords is compared to how users actually build their passwords, most users do not follow the principles of complexity and uniqueness.

For complexity, the rule recommends using at least three elements for constructing a password, including words, numbers, symbols and others, but my finding reveal that 22 % of users use a simple word or number as password, and another 29% a combination of word

and numbers. It is also highly possible that the numbers, when used in combination with a word, are at the beginning or end of the word, which is also not recommended.

For uniqueness, the rule is not to use the same password for more that once site, but the survey reveals that 64% of the users use the same password over and over again. Among users who use the same password, the complexity is lower, ranging from 51% low complexity to 64% low complexity. Additionally, checking the complexity of passwords, it can be predicted that most of the passwords that are combinations of a single word and numbers are, indeed, personal information or information related to the user. As might be expected, the people who think their passwords were hacked are mostly those who always use the same password.

For password recovery or password reset, I have concluded that only most popular web 2.0 sites adopt the most serious methods in password security. I also could notice that most social networks and other systems, unlike email providers, have only one method for password reset, which is to send instructions to the user by email. A small percentage of social networks use the method of a Secret Question as an alternative, and all evaluated email providers use a second alternative for password reset. Secret Questions are used for password recovery in the case of those 33% of users who belong to Windows Live Spaces, which also has a Hotmail account.

Email providers use the method of Secret Question as an alternative since they cannot be certain that a user only has the email account provided by them.

Information Revealed in Web 2.0 sites

As these findings show, many users leave their profile public facilitating the dissemination of information. Users show information like their Birth Date in public profile, probably to disseminate this information widely. Users who have a private profile post 10% more information than those who have a public profile.

Also, any kind of personal information could be found in web 2.0 sites. Depending on the social networks where users subscribe, they tend to publish specific kinds of information. In MySpace, for example, users like to publish all of their favorite preferences, whereas Twitter users more frequently write comments.

Popular web 2.0 sites count with API that facilities access of information. As we exposed before there is a wide gamma of services and libraries specialized in access information from our profiles, including also pictures, comments and other activities.

Gaining System Access through the Information

For this object it is concluded that it is possible to gain system access using personal information for password guessing or brute force attack. Since 52% of users have weak passwords, a percentage that rises to 64% if you add in the people who consistently use the

same password, you can understand why, using a Brute Force attack scenario, as mentioned earlier, some users' passwords can be easily guessed.

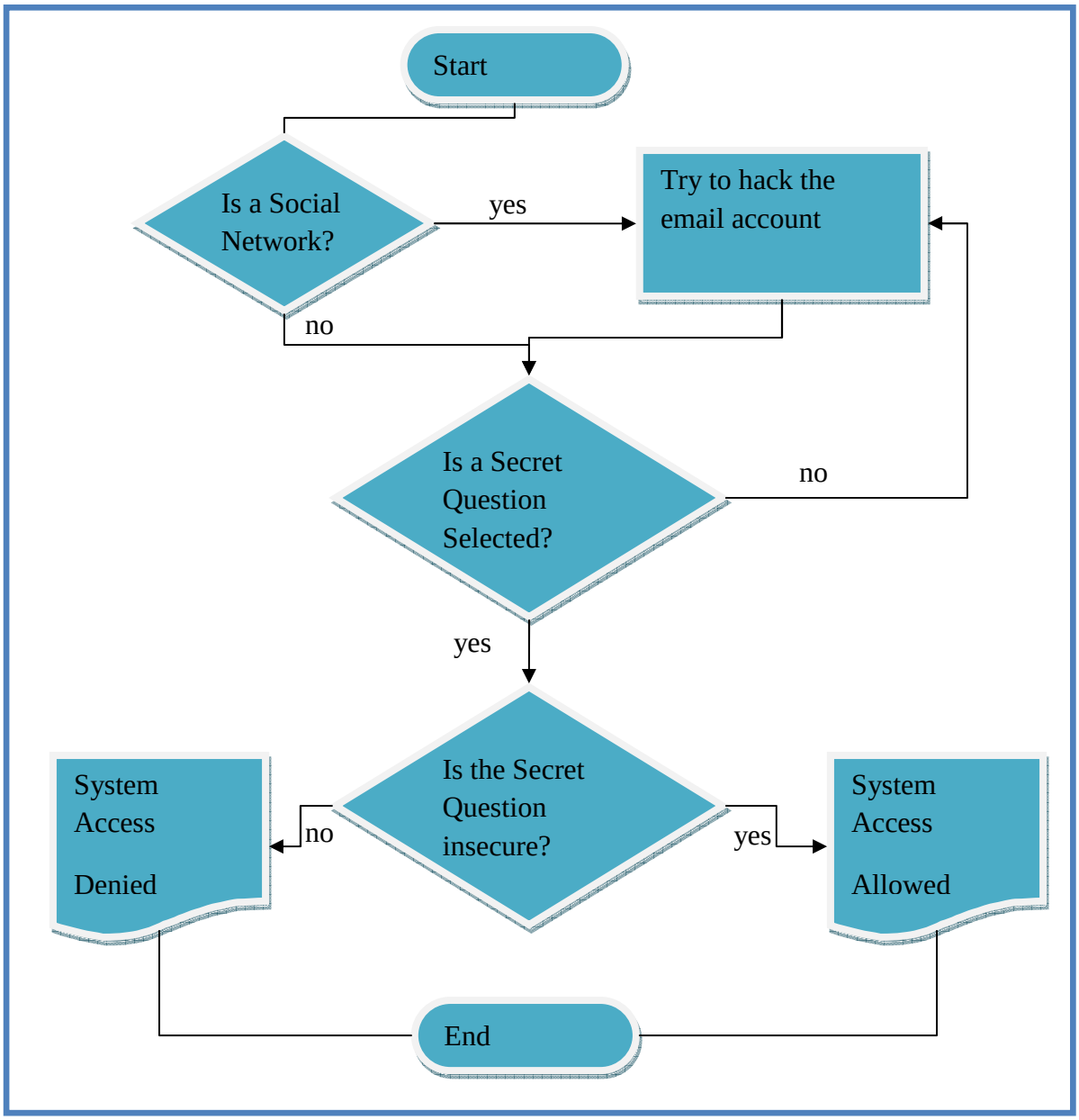
The previous statistic of users that use the same password for more than one system also leave to the following conclusion, if we guessed a password from one system, then we will have access to many other system, probably more in other system of the same category, specially web 2.0 sites where numbers also show that users are careless.

Also, it is possible to gain system access using personal information as answer in the password recovery system of Secret Question. Most of the social networks use instructions sent through an email account as the password reset method, so a hacker who wants to gain access to a social network account first needs to hack the user's email account. For gaining access to a user's email account, the hacker will most likely go though their password recovery system. The user probably selected the Secret Questions method of password recovery, and, since many users choose popular questions whose responses are frequently posted in a public version of the user's social network profile, the hacker can gain access to the email account.

Finally, it is concluded that gaining access to an email system would give access to other system based in that password recovery method. As I said before more web 2.0 sites and in general most actual system use instructions sent through an email account as the password

reset method, then all the system that are relying in a specific email account that has been hacked would be compromised by this relationship.

Gaining System Access to Password Reset Mechanism Diagram



Recommendations

For Users

In password handling:

- ✓ Follow the previously revealed rules for password complexity:
    a) Use a combination of a least 3 elements from characters, numbers, symbols, words and phrases.
    b) Use uppercase in a position different from the first letter.
    c) Use one or two numbers, but not at the beginning or end of the password.
    d) Use punctuation or other symbol in the password.
- ✓ Follow the principles of uniqueness for password construction:
    a) Do not use common password.
    b) Do not use the same password more than once.
    c) Do not use personal information or related information.
    d) Use different patterns or sequences.
    e) Refresh your password every 3 to 6 months.

Especially the rule that recommends the use of punctuation or symbols could be an excellent way to use something unique that is easy to remember; this is the technique of replacing some characters with symbols that look similar, for example, replacing the letter "a" with the symbol "@."

In password reset:

- ✓ Always try to select **send instructions to reset to an email account**, if available as a password reset method, instead of a Secret Question.

- ✓ When choosing a Secret Question for an email account, take into consideration the information that is already posted in your social network profiles and try to avoid questions that use this posted information as an answer.
- ✓ Try to frequently change your Secret Question, especially if there is any suspicion of a security break.

In Web 2.0 Sites:

- ✓ Set your profile as Private to restrict your personal information to your friends or contacts.
- ✓ Verify your friends and contacts before accepting them.

For Developers

In password handling:

- ✓ Present to users an indicator of password complexity when users are creating or changing their passwords.
- ✓ Avoid accepting weak passwords in your system.
- ✓ Present to users the recommendations for password complexity and uniqueness, especially the suggestion of changing passwords every 3 to 6 months.
- ✓ Activate a CAPTCHA mechanism in case of continuous failure of password authentication (3 or 5 times).

In password reset:

- ✓ Implement the method for password reset **send instructions to an alternative email account**; this will create an extra step for hackers because they will have to first try to gain access to the user's alternative email account.
- ✓ Remind users in the instructions sent for password reset to follow the previous recommendations for password handling.

Future Work

This project could be expanded to include the following research topics:

- Analyze if the collection of information in web 2.0 sites could be in contradiction to any law.
- Focus on other web 2.0 applications like Wiki and Blogs and their methods of user interaction (API).
- Investigate further the reasons why users are careless about web security.

Bibliography

Burnett, M. (2004). *Hacking the Code : ASP. NET Web Application Security.* Rockland, MA, USA: Syngress Publishing.

Burnett, M. (2005). *Perfect Passwords: Selection, Protection and Authentication.* Rockland, MA, USA: Syngress Publishing.

Facebook API. (2009, 12 01). Retrieved from Facebook Developer Wiki: http://wiki.developers.facebook.com/index.php/API

Hi5 API (beta). (2009, 12 01). Retrieved from Hi5 Developer Platform: http://api.hi5.com/

Live Services User Data APIs. (2009, 12 01). Retrieved from MSDN: http://msdn.microsoft.com/en-us/library/cc305075(v=MSDN.10).aspx

Mircea Kristaly, D. (2008). *Web 2.0 technologies in web application development.* Transilvania University of Brasov.

MySpace RESTful API. (2009, 12 01). Retrieved from MySpace Open Platform: Documentation WIKI: http://developerwiki.myspace.com/index.php?title=RESTful_API

Noureddine, A., & Damodaran, A. M. (2008). *Security in web 2.0 application development.* Microlead Business Solutions University of Houston.

Pilgrim, C. J. (2008). *Improving the Usability of Web 2.0 Applications.* Hawthorn, Australia: Swinburne University of Technology.

Scambray, J. (2006). *Hacking Exposed Web Applications (2nd Edition).* Emeryville, CA, USA: McGraw-Hill Osborne.

Twitter API Documentation. (2009, 12 01). Retrieved from Twitter API Wiki: http://apiwiki.twitter.com/Twitter-API-Documentation

Ullrich, C. (2008). *Why Web 2.0 is Good for Learning and for Research: Principles and Prototypes.* Shanghai Jiao Tong University.

Appendix A - Survey

Social Network Use Survey

There are 15 questions in this survey

General

1 Please select your main role at the University *

Please choose **only one** of the following:

- ○ Student
- ○ Teacher

2 What is your experience level in Internet use? (1 - A few, 5 - Expert) *

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

3 To which social networks do you belong? *

Please choose **all** that apply:

- ☐ Facebook
- ☐ MSN Live Spaces
- ☐ Twitter
- ☐ Hi5
- ☐ MySpaces
- ☐ Other:

4 What is the main status of your profile in those networks? *

Please choose **only one** of the following:

- ○ Public
- ○ Both
- ○ Private

5 What information do you have in your profile? *

Please choose **all** that apply:

- ☐ Birthdate
- ☐ Favorite Music
- ☐ Occupation
- ☐ Address
- ☐ Academic Level
- ☐ Favorite food

☐ Other:

6 How frequently do you access these networks? *

Please choose **only one** of the following:

○ Many times per day
○ Weekly
○ Daily
○ Many times per month
○ Many times per week

7 Do you have to provide some information about your profile at registration? *

Please choose **only one** of the following:

○ Yes
○ No

8 What are your regular activities in these networks? *

Please choose **all** that apply:

☐ Upload Pictures
☐ Play games
☐ Write comments
☐ Other:

Access to the Network

9 Have you used the same password for some of these networks? *

Please choose **only one** of the following:

○ Yes
○ No

10 Generally, your password is made up of? *

Please choose **only one** of the following:

○ Words easy to remember
○ Combination of 1 and 2
○ Numbers
○ Random combination of letters and numbers

11 Have you forgotten your password for any of these network? *

Please choose **only one** of the following:

○ Yes
○ No

12 If you forgot your password, which method do you use to recover it?

Only answer this question if the following conditions are met:
°Answer was 'Yes' at question '11 [B02]' (Have you forgetter your password of one of these network?)

Please choose **only one** of the following:

- ○ Send instructions by email
- ○ Answer a secret question
- ○ I got a tip for my password
- ○ I couldn't recover it
- ○ Other

13 How do you rate the process of recovering your password?

Only answer this question if the following conditions are met:
°Answer was 'Yes' at question '11 [B02]' (Have you forgetter your password of one of these network?)

Please choose **only one** of the following:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5

14 Do you think that someone has stolen your password? *

Please choose **only one** of the following:

- ○ Yes
- ○ No

15 How secure do you feel using these networks? *

Please choose **only one** of the following:

- ○ Very secure
- ○ Insecure
- ○ I don't care so much

Submit your survey.

Thank you for completing this survey.

Appendix B – Manual Inspection of Application Security

Facebook: In the case of Facebook, for a new account it is necessary to provide the gender and birth date. They state that the reason for this is to provide age-appropriate content. Password requirements are from 6 to 50 characters

Facebook offers a password recovery link. For password recovery, they establish a process that starts with your email address and CAPTCHA verification. An email with a verification code it is sent via email. They use a certificate (https) for authentication and password reset.

Twitter: For account creation, it is not necessary to provide any personal information. Password requirements are 6 characters. After the account is created, they asked for the topics in which you are interested. To complete the process, it is necessary to activate the new account by email.

Twitter offers a password and username "forgot" link. For password recovery they first ask for username or email address. After that, they send an email with a link to provide the new password. They use a certificate (https) for authentication and password reset.

MySpace: For account creation your username is your email address and you also have to provide your gender and birth date. Password has to be 6 to 50 characters

MySpace offers a password and username “forgot” link. For password recovery they first ask for username or email address. After that, they send an email with a link to provide the new password with a CAPTCHA validation. They use a certificate (https) for authentication and password reset

Live Spaces: See Hotmail (Windows Live ID)

Hi5: The username is your email address. In the registration process it is necessary to provide your birth day. Password requirements are from 6 to 20 spaces.

For password recovery they ask for your email to provide a password change. This network does not use https for authentication of either password reset.

Google: At registration you have to provide a Secret Question and an alternate address. The minimum length for account creation is 8 characters.

For password recovery they offer several options, like an alternate email address, response to a secret question, and establishing a direct communication with their support team. They use a certificate (https) for authentication and password reset

Secret Question:

What is your number for Frequent Flier?

What is your Library ID number?

What was your first telephone number? **

What was the name of your first teacher?

Write you own question **

Hotmail (MSN or Live address): For account registration is necessary to provide the following information: alternative email address, country, state, postal code, gender, and birth date. The password requires at least 6 characters and the answer to the secret question 5 characters.

For password recovery they first ask you for your Windows Live ID and a CAPTCHA. As options you can provide your information of residence and a secret question or have instructions sent to an alternative email. They use a certificate (https) for authentication and password reset.

Secret Question:

What is the birth date and place of your mother? **

Who was your best friend from childhood? **

What is the name of your first pet? **

Who is your favorite professor?

Who is your favorite person from history? **

What is your grandfather's occupation?

Yahoo: For account creation you have to provide the following information: Gender, birth date, country, an alternate email address, and two secret questions. Password requirements are from 6 to 32 characters.

For password recovery they ask you for your Yahoo ID and a CAPTCHA. They use a certificate (https) for authentication and password reset. They also offer the option to send an email with instructions to reset or a secret question that you can access via email:

Secret Questions:

(First Selection)

Where did you spend your honeymoon?

Where did you meet your spouse?

What is your oldest cousin's name?

What is your youngest child's nickname? **

What is your older child's nickname? **

What is the first name of your oldest niece?

What is the first name of your oldest nephew?

What is the first name of your favorite aunt?

What is the first name of your favorite uncle?

What town was your father born in? **

What town was your mother born in? **

Create your own question.

(Second Selection)

Who is your favorite author? **

What is the name of your best man at your wedding?

What is the name of your maid of honor at your wedding?

What is your favorite book? **

What is the last name of your favorite musician? **

Who is your all-time favorite movie character? **

What was the make of your first car?

What was your first pet's name?

What is the name of your favorite sports team?

What was your favorite food as a child?

What is the last name of your favorite teacher?

What is your main frequent flier number?

Create your own question.

Appendix C – User Structure from Social Network API

Facebook User Data through API

```
<?xml version="1.0" encoding="UTF-8"?>
<users_getInfo_response xmlns="http://api.facebook.com/1.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://api.facebook.com/1.0/
http://api.facebook.com/1.0/facebook.xsd" list="true">
    <user>
      <uid>8055</uid>
      <about_me>This field perpetuates the glorification of the ego.  Also, it has a character
limit.</about_me>
      <activities>Here: facebook, etc. There: Glee Club, a capella, teaching.</activities>
      <affiliations list="true">
        <affiliation>
          <nid>50453093</nid>
          <name>Facebook Developers</name>
          <type>work</type>
          <status/>
          <year/>
        </affiliation>
      </affiliations>
      <birthday>November 3</birthday>
      <books>The Brothers K, GEB, Ken Wilber, Zen and the Art, Fitzgerald, The Emporer's
New Mind, The Wonderful Story of Henry Sugar</books>
      <current_location>
        <city>Palo Alto</city>
        <state>California</state>
        <country>United States</country>
        <zip>94303</zip>
      </current_location>
      <education_history list="true">
        <education_info>
          <name>Harvard</name>
          <year>2003</year>
          <concentrations list="true">
            <concentration>Applied Mathematics</concentration>
            <concentration>Computer Science</concentration>
          </concentrations>
        </education_info>
      </education_history>
      <family list="true">
        <family_elt list="true">
          <family_elt_elt>mother</family_elt_elt>
          <family_elt_elt>1394244902</family_elt_elt>
```

```
      </family_elt>
      <family_elt list="true">
        <family_elt_elt>sister</family_elt_elt>
        <family_elt_elt>48703107</family_elt_elt>
      </family_elt>
      <family_elt list="true">
        <family_elt_elt>brother</family_elt_elt>
        <family_elt_elt>1078767258</family_elt_elt>
      </family_elt>
      <family_elt list="true">
        <family_elt_elt>brother</family_elt_elt>
        <family_elt_elt>John Doe</family_elt_elt>
        <family_elt_elt/>
      </family_elt>
    </family>
    <first_name>Dave</first_name>
    <hometown_location>
      <city>York</city>
      <state>Pennsylvania</state>
      <country>United States</country>
    </hometown_location>
    <hs_info>
      <hs1_name>Central York High School</hs1_name>
      <hs2_name/>
      <grad_year>1999</grad_year>
      <hs1_id>21846</hs1_id>
      <hs2_id>0</hs2_id>
    </hs_info>
    <is_app_user>1</is_app_user>
    <has_added_app>1</has_added_app>
    <interests>coffee, computers, the funny, architecture, code breaking,snowboarding,
philosophy, soccer, talking to strangers</interests>
    <last_name>Fetterman</last_name>
    <locale>en_US</locale>
    <meeting_for list="true">
      <seeking>Friendship</seeking>
    </meeting_for>
    <meeting_sex list="true">
      <sex>female</sex>
    </meeting_sex>
    <movies>Tommy Boy, Billy Madison, Fight Club, Dirty Work, Meet the Parents, My
Blue Heaven, Office Space </movies>
    <music>New Found Glory, Daft Punk, Weezer, The Crystal Method, Rage, the KLF,
Green Day, Live, Coldplay, Panic at the Disco, Family Force 5</music>
    <name>Dave Fetterman</name>
    <notes_count>0</notes_count>
    <pic>http://photos-055.facebook.com/ip007/profile3/1271/65/s8055_39735.jpg</pic>
```

```
      <pic_big>http://photos-
055.facebook.com/ip007/profile3/1271/65/n8055_39735.jpg</pic_big>
      <pic_small>http://photos-
055.facebook.com/ip007/profile3/1271/65/t8055_39735.jpg</pic_small>
      <pic_square>http://photos-
055.facebook.com/ip007/profile3/1271/65/q8055_39735.jpg</pic_square>
      <political>Moderate</political>
      <profile_update_time>1170414620</profile_update_time>
      <quotes/>
      <relationship_status>In a Relationship</relationship_status>
      <religion/>
      <sex>male</sex>
      <significant_other_id xsi:nil="true"/>
      <status>
        <message>Fast Company, November issue, page 84</message>
        <time>1193075616</time>
      </status>
      <timezone>-8</timezone>
      <tv>cf. Bob Trahan</tv>
      <wall_count>121</wall_count>
      <website>http://www.example.com</website>
      <work_history list="true">
        <work_info>
          <location>
            <city>Palo Alto</city>
            <state>CA</state>
            <country>United States</country>
          </location>
          <company_name>Facebook</company_name>
          <position>Software Engineer</position>
          <description>Tech Lead, Facebook Platform</description>
          <start_date>2006-01</start_date>
          <end_date/>
        </work_info>
      </work_history>
    </user>
 </users_getInfo_response>
```

Twitter User Data through API

```
<user>
<id>1401881</id>
<name>Doug Williams</name>
<screen_name>dougw</screen_name>
```

<location>San Francisco, CA</location>
<description>Twitter API Support. Internet, greed, users, dougw and opportunities are my passions.</description>
<profile_image_url>http://s3.amazonaws.com/twitter_production/profile_images/59648642/avatar_normal.png</profile_image_url>
<url>http://www.igudo.com</url>
<protected>false</protected>
<followers_count>1031</followers_count>
<profile_background_color>9ae4e8</profile_background_color>
<profile_text_color>000000</profile_text_color>
<profile_link_color>0000ff</profile_link_color>
<profile_sidebar_fill_color>e0ff92</profile_sidebar_fill_color>
<profile_sidebar_border_color>87bc44</profile_sidebar_border_color>
<friends_count>293</friends_count>
<created_at>Sun Mar 18 06:42:26 +0000 2007</created_at>
<favourites_count>0</favourites_count>
<utc_offset>-18000</utc_offset>
<time_zone>Eastern Time (US & Canada)</time_zone>
<profile_background_image_url>http://s3.amazonaws.com/twitter_production/profile_background_images/2752608/twitter_bg_grass.jpg</profile_background_image_url>
<profile_background_tile>false</profile_background_tile>
<statuses_count>3390</statuses_count>
<notifications>false</notifications>
<following>false</following>
<verified>true</verified>
<status>
<created_at>Tue Apr 07 22:52:51 +0000 2009</created_at>
<id>1472669360</id>
<text>At least I can get your humor through tweets. RT @abdur: I don't mean this in a bad way, but genetically speaking you're a cul-de-sac.</text>
<source>TweetDeck</source>
<truncated>false</truncated>
<in_reply_to_status_id></in_reply_to_status_id>
<in_reply_to_user_id></in_reply_to_user_id>
<favorited>false</favorited>
<in_reply_to_screen_name></in_reply_to_screen_name>
</status>
</user>

MySpace User Data

Marital status
Orientation
Hometown
Body Type

Education
Religion
Smoke
Drink
Ethnicity
Zodiac sign
Children
Income
Here for

Hi5 User Data

In the profile you can also set the following information:

[Basic] Access options available: Gender, Hometown, Looking to, Status, Religion,

Languages, About Me and Ethnicity.

[Contact] Access options available: Cell Phone, Country, City and Address.

[Interest] Not access options: Interest, Favorite Music, Favorite Movie, Favorite TV Shows,

Favorite Books, and Favorite Quote.

Windows Live ID Contact Data

About: Gender, Birthdate, and Occupation

Preferences: Movie, Music, Books

Social: Interest, Current Location, Status

Education: Degree, School Name

Work: Company, Job, Profession

Contact Information: Birthdate, Address, Phone, Cell Phone, Country

Appendix D – Survey Results

Numbers of records in this consult:	128
Total of records in this survey:	128
Percent of total:	100.00%

Summary of field for A01
Please selects your main role in University

Option	Count	Percent
Student (1)	114	89.06%
Teacher (2)	12	9.38%
Without answer	2	1.56%
No completed	0	0.00%

Summary of field for A02
What is you experience level in the Internet use? (1 - A few, 5 - Expert)

Option	Count	Percent
1 (1)	0	0.00%
2 (2)	3	2.34%
3 (3)	13	10.16%
4 (4)	58	45.31%
5 (5)	52	40.63%
Without answer	2	1.56%
No completed	0	0.00%

Summary of field for A03
To which Social Networks do you belong?

Option	Count	Percent
Facebook (1)	104	81.25%
Twitter (2)	45	35.16%
MySpaces (3)	47	36.72%
MSN Live Spaces (4)	41	32.03%
Hi5 (5)	77	60.16%
Other	10	7.81%

Summary of field for A04
What is the main status of your profile in those networks?

Option	Count	Percent
Public (1)	35	27.34%
Private (2)	62	48.44%
Both (3)	26	20.31%
Without answer	5	3.91%
No completed	0	0.00%

Summary of field for A05
What information do you have in your profile?

Option	Count	Percent
Birthdate (1)	106	82.81%
Occupation (2)	91	71.09%
Academic Level (3)	77	60.16%
Favorite Music (4)	60	46.88%
Address (5)	15	11.72%
Favorite food (6)	19	14.84%
Other	14	10.94%

Summary of field for A06
How frequently do you access to those networks?

Option	Count	Percent
Many times per day (1)	44	34.38%
Daily (2)	21	16.41%
Many times per week (3)	38	29.69%
Weekly (4)	1	0.78%
Many times per month (5)	19	14.84%
Without answer	5	3.91%
No completed	0	0.00%

Summary of field for A07
Do you have to provide some information of your profile at registration?

Option	Count	Percent
Yes (Y)	106	82.81%
No (N)	17	13.28%
Without answer	5	3.91%
No completed	0	0.00%

Summary of field for A08
What are your regular activities in these networks?

Option	Count	Percent
Upload Pictures (1)	72	56.25%
Write Comments (2)	91	71.09%
Play Games (3)	32	25.00%
Other	35	27.34%

Summary of field for B01
Have you used the same password for some of these networks?

Option	Count	Percent
Yes (Y)	77	60.16%
No (N)	43	33.59%
Without answer	8	6.25%
No completed	0	0.00%

Summary of field for B07
Generally, your password at made of?

Option	Count	Percent
Works easy to remember (1)	13	10.16%
Numbers of some information (2)	14	10.94%
Combination of 1 and 2 (3)	35	27.34%
Random combination of letters and numbers (4)	58	45.31%
Without answer	8	6.25%
No completed	0	0.00%

Summary of field for B02
Have you forgetter your password of one of these network?

Option	Count	Percent
Yes (Y)	44	34.38%
No (N)	76	59.38%
Without answer	8	6.25%
No completed	0	0.00%

Summary of field for B03
If you forgot your password, which method do you use to recover it?

Option	Count	Percent
Sent instructions by mail (1)	34	26.56%
Answer a Secret question (2)	9	7.03%
I got a tip for my password (3)	0	0.00%
I couldn't recover (4)	1	0.78%
Other	0	0.00%
Without answer	82	64.06%
No completed	0	0.00%

Summary of field for B05
How do you rate the process of recover your password? (1 Very hard - 5 Very easy)

Option	Count	Percent
1 (1)	3	2.34%
2 (2)	3	2.34%
3 (3)	2	1.56%
4 (4)	15	11.72%
5 (5)	20	15.63%
Without answer	83	64.84%
No completed	0	0.00%

Summary of field for B04
Do you think that someone has stolen your password?

Option	Count	Percent
Yes (Y)	20	15.63%
No (N)	100	78.13%

Without answer	8	6.25%
No completed	0	0.00%

Summary of field for B06		
How secure do you feel using these networks?		
Option	Count	Percent
Secure (1)	18	14.06%
Insecure (2)	27	21.09%
I don’t care so much (3)	75	58.59%
Without answer	8	6.25%
No completed	0	0.00%