A Privacy Awareness System for Facebook Users

Charles Hélou, Ala Eddine Gandouz, Esma Aïmeur
Université de Montréal. Montreal, Canada
{helou, gandouza, aimeur}@iro.umontreal.ca

ABSTRACT: *Social networking sites have experienced a steady and dramatic increase in the number of users over the past several years. Thousands of user accounts, each including a significant amount of private data, are created daily. As such, an almost countless amount of sensitive and private information is read and shared across the various accounts. This jeopardizes the privacy and safety of many social network users and mandates the need to increase the users' awareness about the potential hazards they are exposed to on these sites.*

We introduce Protect_ U, a privacy protection system for Facebook users. Protect_ U analyzes the content of user profiles and ranks them according to four risk levels: Low Risk, Medium Risk, Risky and Critical. The system then suggests personalized recommendations designed to allow users to increase the safety of their accounts. In order to achieve this, Protect_ U draws upon both the local and community-based protection models. The first model uses a Face book user's personal data in order to suggest recommendations, and the second seeks out the user's most trustworthy friends to encourage them to help improve the safety of his/her account.

Keywords: Facebook, Online privacy, User-profile, Social network sites, Recommender system, Classification

Received: 19 October 2011, Revised 7 January 2012, Accepted 21 January 2012

1. Introduction

Social networks enable users to keep in touch, discuss and share information with friends. They additionally provide users with interaction and communication tools to enable them to expand their social circles and interact via third-party applications [14].

The main purpose for using social networks is to interact and meet new friends. However, some Internet users, owing to their carelessness and lack of knowledge when it comes to publishing personal information on their accounts, might be inadvertently jeopardizing their privacy. Quite often, they are not aware that their profiles are readily available to basically every other user. They periodically post details about their private lives, mistakenly believing that this information will remain well-protected within the virtual environment. This phenomenon is, in fact, encouraged by social networking sites, which offer users extremely detailed fill-in fields querying a great deal of their personal information. In some cases, users are forced to enter extremely confidential information (Facebook requires new users to enter their *date of birth* and *gender*, and anyone wishing to create applications on its platform must enter a valid phone number or a credit card number[1]). Therefore, it is extremely important to make Internet users fully aware of the hazards related to social networks.

In this paper, we present a *community-based* privacy protection system for Facebook users called *Protect_ U*. Section 2 features

[1] https://www.facebook.com/press/info.php?statistics

the main hazards currently faced by Internet users. Section 3 reviews related research that has dealt with this very issue. Sections 4 and 5 describe the methodology we apply and the features of our implementation, as well as the results of our validation experiments. Section 6 presents a brief conclusion, discussion and proposed directions of future work.

2. Risk with social networks

The most widely known risks nowadays can be summarized as follows [2, 3, 25, 26]:

• A misleading concept of community: many social network suppliers claim to have transposed the "*real*" world's communication structures into "*cyber space*". For instance, they claim that publishing personal data on their platforms is risk-free, since it is much like information sharing among friends, as if people were face-to-face. If users do not know how their profile data is shared or how to protect it, this notion of a virtual "*community*" may lead them to share personal information they would have otherwise kept private.

• Difficult to close a user account: once published, data can remain stored for a long time, even after the user deletes them from the original site. Copies may exist on other third-party sites, or even on other users' computers. In addition, some service suppliers actually ignore any data or profile deletion requests submitted by their users.

• Secondary data collection by service suppliers: social network suppliers can gather a great deal of secondary information about their users, such as their geographic location, habits, tastes or preferences, in order to personalize their services. Such data can be used to target, discriminate and transfer information to third parties for resale purposes.

• Facial recognition: the pictures published by users on social networking sites can be used to seed automatic biometric identifiers. Facial recognition software applications have experienced an impressive rate of improvement over the past few years, partially due to large readily available online dataset. Once a face is tagged in a picture, it becomes relatively straight forward to retrieve the name from other pictures.

• Identity theft: the increased availability of personal data in user profiles and the potential theft of such data by unauthorized third parties generate a greater risk of identity theft. Such data might be used to create counterfeit profiles that are likely to harm the reputation of the identity theft victims.

• Phishing: hackers can urge users to download files or visit infected sites, posing as their friends on the social network. These users are particularly vulnerable to script attacks, which automate the injection of virus-containing links and automatically collect the published personal data they have access to.

Thus, we assert that a social network users' behaviour often inadvertently jeopardizes their own privacy and that of their family and friends. This motivates the need for a secure tool to protect the confidentiality of a user's personal information on a social network.

3. Related work and comparison

The large expansion of social networks and the exchange of vast amounts of personal and private information have been facilitated to an extraordinary degree due to almost ubiquitous Internet access. Previous work has shown that social networks can pose a significant threat to users' privacy [4, 6, and 11]. The increasing popularity of social networking sites and ever increasing user subscription rate has prompted many studies related to the security and privacy protection aspects of these networks [1,2, 27]. Several works aim to understand the significance of the risks that are involved by focusing on examining identity leakage due to individual attacks or unwanted data mining [5, 15].

In order to protect social network users, many studies, such as *Protect_ U*, applied the *Privacy Feedback and Awareness* (*PFA*) [16] approach. This approach aims to enhance users' understanding of the privacy implications of their system use. Moreover, it assists users in specifying, comprehending and maintaining a high level of privacy, for example with the *Privacy Wizard* by Fang and Le Fevre [10]. The goal of this tool is to automatically configure a user's privacy settings with minimal effort and interaction from the user. Similarly, Mazzia and Adar [20] introduce *PViz*, a system that allows users to understand the visibility of their profiles according to natural sub-groupings of friends, and at different levels of granularity. *PViz* relies on a graphical output model that illustrates the user's social network. Patil and Kobsa [22] propose the *PRISM (PRIvacy-Sensitive Messaging)* system, providing Internet Messaging users with various informative visualizations as well as mechanisms for presenting

oneself differently to various groups of contacts. Lipford et al. [17] introduce and evaluate the *Audience View* interface which allows a user to view his profile as it appears to an individual friend or to a manually-specified sub-group of friends. A variation of this interface is currently applied by Facebook. The uses of expandable grid interfaces were suggested by Reedar et al. [23] and they allow a system administrator to visualize and modify access control settings. In a similar vein, some studies [18, 19] propose a methodology for quantifying the risk posed by a user's privacy settings. A risk score reveals to the user how far his/her privacy settings are from those of other users. It provides feedback regarding the state of his/her existing settings. However, it does not help the user refine his/her settings in order to achieve a more acceptable configuration. On the other hand, Carmagnola et al. [9] recently presented *So NARS*, an algorithm for social recommender systems. *So NARS* targets users as members of social networks, suggesting items that reflect the trend of the network itself, based on its structure and on the influence relationships between users'.

While much work has focused on tools for understanding and adjusting existing privacy settings, *Protect_ U* uses machine learning techniques to recommend privacy settings based on a user's personal data and trustworthy friends. *Protect_ U* analyzes user profile contents and ranks them according to four risk levels: Low Risk, Medium Risk, Risky and Critical. The system then suggests personalized recommendations to allow users to make their accounts safer. In order to achieve this, it draws upon two protection models: local and community-based. The first model uses the Facebook user's personal data in order to suggest recommendations. It is up to the user to accept or reject these recommendations. The second model seeks the user's trustworthy friends to encourage them to help improve the safety of their counterpart'saccount.Table1 highlights the main characteristics of the aforementioned systems compared to our proposed *Protect_ U* solution.

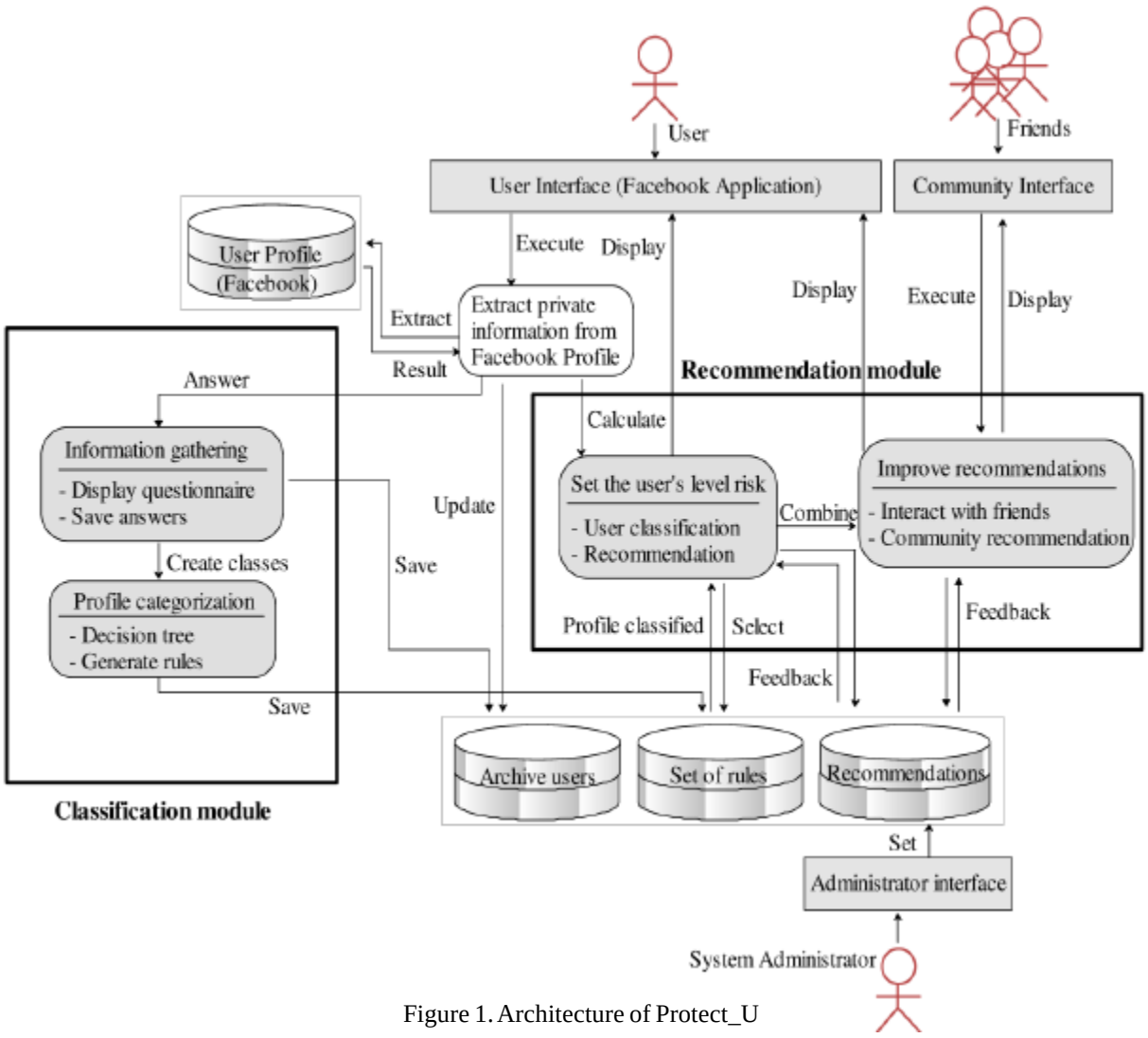


Figure 1. Architecture of Protect_U

Name	Techniques applied	Visibility[2]	Transparency[3]	Access control[4]	Recommendations[5]	Friends implication[6]	Average Accuracy
Protect_U	- Questionnaire - Machine learning model - Recommendation model - Community-based filter	Yes	Yes	Yes	Yes	Yes	- Recommendations: 70.18% - Trusted friends recognition: 93.75% for critical profiles and 79.49% for risky profiles
Privacy Wizard	- Machine learning model - Questionnaire	Yes	Yes	Yes	No	No	- User's settings configuration: 90% (if a user labels 25 friends of over 200)
PViz tool	- Modularity optimization - Fruchterman-Reingold layout algorithm - Questionnaire	Yes	Yes	Yes	No	No	-Helpfulness: 100% -Enjoyment: 80% -Likely to personally use:100%
PRISM	- Questionnaire - Instant Messaging (IM) plugin	No	Yes	No	No	No	- Users satisfaction: 81.82%
Audience View	- New interface for managing privacy settings in Facebook based on an audience point of view - Questionnaire	Yes	Yes	Yes	No	No	- Increase in accuracy comparing to Facebook privacy settings interface: 27.7%
SoNARS	- Social Networks-based Algorithm	No	No	No	Yes	Yes	- Recommendations: 80% - Precision: 67%

Table 1. Main awareness-characteristics of systems in Section 3

Age	Gender	Number of friends	Number of publications per week	Number of groups	Number of pictures	Percentage of private pictures	Sensitive data
23	Male	112	13.46	268	43	0	Yes
15	Female	115	19.44	80	194	0	Yes
22	Female	122	12.5	31	60	0	No
28	Female	124	21.88	165	383	98.96	yes

Table 2. Overview of some data collected during the classification model

4. Methodology

The overall system architecture of *Protect_ U* is illustrated in Figure 1.

In order to protect users, *Protect_ U* collects the following personal data: *age*, *gender*, *number of friends* (added in his/her list of friends*)*, *number of postings per week* (posted on his wall per week), *number of groups(*discussion groups and fan pages to which the user is subscribed*)*, *number of pictures* (public and private pictures uploaded by the user), *percentage of private pictures in the user's account(*compared to the number of pictures in the user account) and *sensitive data*(religion and political

[1]Visibility: the system gives feedback on the visibilities of one's historical data or of one's historical/current settings which may have implications for future data.

[2]Transparency: the system provides the user with up-to-date information about his privacy online and about the information he is sharing.

[3]Access control: the system supports the user in making informed decisions and helps him managing data disclosures in different social contexts.

[4]Recommendations: the system applies a recommendation model to make users more aware about the danger of exposing sensitive information on their accounts.

[5]Friends implication: the system encourages friends to help users in order to make their accounts more secure.

orientation). Such parameters are important because, when put together, they can reveal important trends which may lead to a potential overexposure to the privacy hazards on the social network [21].Table 2 shows an example setting of these parameters for four users.

The values of these parameters are stored in the *Archive user's* database. The set of data gathered at this level will serve as an input for *Protect_ U*'s two modules: the *classification module* and *the recommendation module.*

4.1 Classification module

The classification module involves two steps. The first step is Information gathering, and consists of submitting a questionnaire of 18 questions to participants based on various Facebook privacy invasion tactics. To create this questionnaire, we relied on a study conducted by Statistics Canada in 2009 (General Survey - Social Networks) [28]. This survey is based on 210 questions related to social networks and privacy.

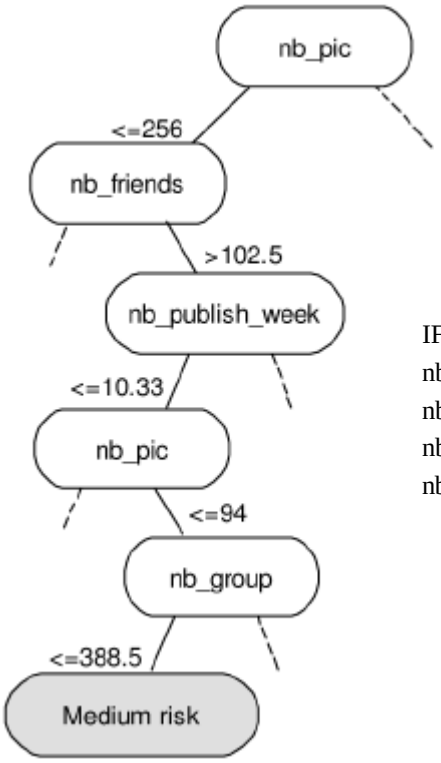
According to the study conducted by Aïmeur et al. [2], privacy invasions have been distributed among four risk levels*: low risk, medium risk, risky* and *critical*. The 18 questions are divided into three groups representing the three **highest** risk levels listed above. The questions in *Group 1* are flagged *medium risk*. These questions will help determine whether the user's account contains sensitive data that may not be harmful for the time being but that might, once accumulated together, jeopardize the user's privacy. An example is:

How is your date of birth posted?

- In full
- Day and month only
- Hidden

The questions in *Group 2* are flagged *risky*. These questions help determine whether the user has experienced problems because of the information published on Facebook. An example is:

Did you ever receive indecent comments on your wall or in one of your postings?



IF nb_pic <=256 AND
nb_friends >102.5 AND
nb_publish_week <=10.33 AND
nb_pic <=94 AND
nb_group <=388.5 THEN risk = Medium risk

Figure 2. Example of rules

- Yes
- No

The questions in *Group 3* are flagged *critical*. These questions help determine whether the user has ever been physically or morally harassed owing to the publication of disturbing content An example is:

Have you ever been exposed to harassment by another member?

- Yes
- No

We rank every participant in the class matching the highest risk level identified according to their answers to the survey. Participant profiles were then categorized according to the following four classes:

1. *Low risk* profile if a participant answered every question negatively.
2. *Medium risk* profile if a participant answered at least one of the questions in *Group 1* affirmatively.
3. *Risky* profile if a participant answered at least one of the questions in *Group 2* affirmatively.
4. *Critical* profile if a participant answered at least one of the questions in *Group 3* affirmatively.

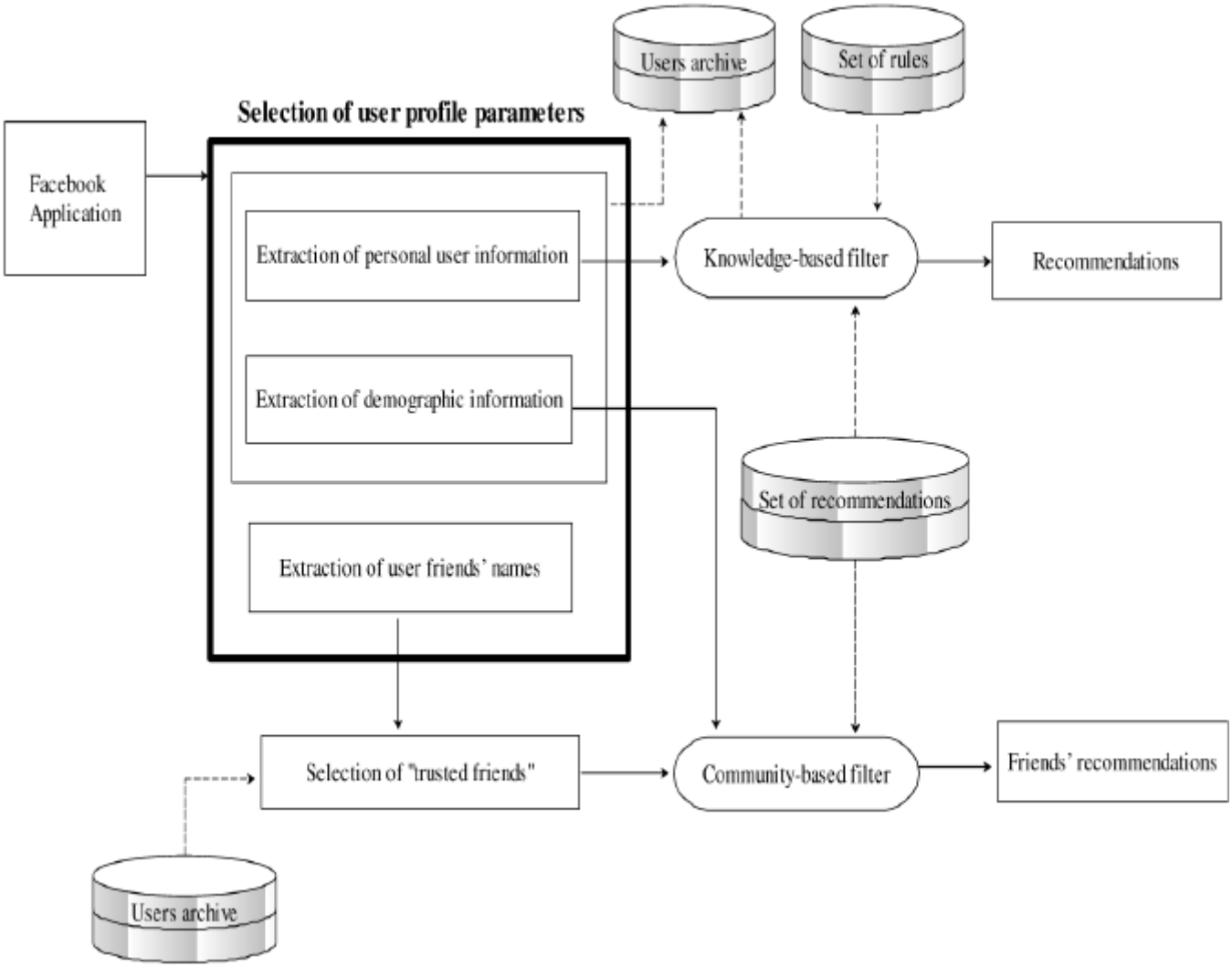


Figure 3. An overview of the recommendation module's structure

Age	Gender	Number of friends	Number of publications per week	Number of groups	Number of pict.	Percentage of private pictures	Sensitive data	Level of risk
23	Male	112	13.46	268	43	0	Yes	Risky
15	Female	115	19.44	80	194	0	Yes	Critical
22	Female	122	12.5	31	60	0	No	Critical
28	Female	124	21.88	165	383	98.96	yes	Risky

Table 3. Categorization of users' risk level

Table 3 presents an example of participant classifications for the same four participants in Table 2. The first user has been placed in the Risky level because he answered "*yes*" to the question, "*Do you post your phone number or address in your Facebook profile?*" which belongs to the set of Risky questions.

The second step (*Profile categorization*) involves applying the *decision tree* [13] and the *C4.5* algorithms [24] to extract *classification rules*. Each path on the decision tree is a rule associated with a class. These rules are saved in the *Set of rules* database. Figure 2 shows an example of such rules.

To validate our algorithm's performance, the *leave-one-out cross-validation* method was applied. This method is the most suitable when the amount of training data is small and is a specific case of the *K- fold cross-validation* method, which randomly distributes the initial training data sample into *k* subsets. The leave-one-out method applies the same procedure, except that *k* is set as the total number of participants in the initial sample set [7].

Once the *classification rules* are specified, a *set of recommendations* are created and entered by an administrator, while taking into consideration the conditions and thresholds generated by the decision tree application. This set of *recommendations* is stored in the *Recommendations* database.

4.2 Recommendation module

The architecture of the recommendation module is illustrated in Figure 3.

When a user executes our application, *Protect_ U* gathers their *personal data*, their *demographic data* (age, nationality and family situation), as well as the list of their *friends' names*. The recommendation module uses this collected data to suggest recommendations adapted to the users' profiles and encourage them to protect themselves better. The recommendation module applies two filters: the *Knowledge-based filter* and the *Community-based filter*.

4.2.1 Knowledge-based filter

The Knowledge-based filter uses users' *personal data* and the *rules* stored in the *set of rules* database to determine the users' profile class and an appropriate recommendation. It uses the rule that most closely matches the profile by applying the *vector similarity function*. For profile A and rule B, the *similarity function* is expressed as:

$$Similarity(A,B) = \sum_{i=1}^{N} \frac{v_{A,i}}{\sqrt{\sum_{i=1}^{N} v^2_{A,i}}} * \frac{v_{B,i}}{\sqrt{\sum_{i=1}^{N} v^2_{B,i}}} \quad (1)$$

Where N being the maximum number of parameters considered, $v_{A,i}$, being the value of parameter i pertaining to user A, and $v_{B,i}$ being the value of parameter I pertaining to rule B. The rule that yields the smallest value of *Similarity(A, B)* will be assigned to profile A.

Example 1: Suppose that we are attempting to find the best matching rule for the profile of user A featuring the following parameters:

$v_{A,1}$ = Age = 15, $v_{A,2}$ = Number of pictures published = 300, $v_{A,2}$ = Number of friends = 1200 and suppose also that we have the following three rules:

Rule B1: if $v_{B1,1} < 18$, $v_{B1,2} > 100$ and $v_{B1,3} > 50$ then *Risky* profile,

Rule B2: if $v_{B2,1} > 40$, $v_{B2,2} < 20$ and $v_{B2,3} < 10$ then *Low risk* profile, and

Rule B3: if $v_{B3,1} < 25$, $v_{B3,2} > 30$ and $v_{B3,3} > 100$ then *Critical* profile.

The user profile matches rules B1 and B3. The highest value determines which one of these two rules is closest to the profile. Since *Similarity(A, B1) = 0.644* and *Similarity(A, B3) = 0.97*. One can deduce that the profile of user A is closer to rule B3 i.e., *Critical*.

After determining the rule that is closest to the user profile, the Knowledge-based filter searches through the set of recommendations database to find those recommendations that match the conditions imposed by this rule. For example, the rule represented in Figure 2 associated with a Risky profile yields the following recommendation:

"***Reduce the number of publications posted on your wall.*** A large amount of information published on your wall increases the risk of personal information being gathered from your profile. Be mindful of the content that you share. Always remember that some friends might be affected or disturbed by the videos or images that you publish. This might affect your relationship with them."

4.2.2 Community-based filter

The Community-based filter seeks to improve the quality of recommendations by making use of the user's friend network. Even if Facebook gives its users the possibility to create a list of close friends, most of them will not use it or may not necessarily add a trusted friend to it. For this purpose, *Protect_ U* automatically selects *trusted friends* who are suspected to have a personal knowledge of the user. A *trusted friend* is a friend belonging to one of the following groups: *family members, close friends, people with whom the user communicates most frequently* and *individuals who are tagged in the pictures posted on the users' profile*. Obviously, trusted friends could exist across different geographical locations.

By answering a small questionnaire, these friends help refine the information already gathered regarding the user and help target the weaknesses on his/her profile. The questionnaire essentially seeks to determine the risk level, all while taking into consideration the following important privacy aspects: the *contents of posted pictures*, the *contents of the comments* and *publications* (video, external links, etc.). The questions we ask are chosen based on the information gathered through *demographic selection* (*age, nationality* and *family situation*). Thus, the questions asked to an adult will be different from those asked to a minor.

On the other hand, the system checks whether said friends have ever executed *Protect_ U* by assessing whether the *Users archive* already contains the risk level related to their profiles. If so, their risk levels will be taken into consideration when determining whether they should be considered as trusted friends.

To select trusted friends, each rating is attributed a *Weight* (see Equation 2), for each one of the user's friends. This weight is based on the following parameters: the friend's *knowledge relationship* with the user (C_1), the *number of messages exchanged* with the user (C_2), the *number of times that the friend is tagged in the user's account* (C_3), and the *risk level of the friend's profile* (C_4).

If the friend's *risk level* is known, C_4 will be worth 2 for a *low risk* profile, 1.5 for a *medium risk* profile, 0.5 for a *risky* profile, or 0 for a *critical* profile. A value of zero means that if a friend's profile is found to be critical, the friend will be removed from the trusted friends list. If *Protect_ U* is not able to determine a friend's risk level, the system will automatically assign a value of 1 to C_4, which means that this parameter will no longer be taken into consideration. Thus C_4 can take on the following discrete settings:{0, 0.5, 1, 1.5, and 2}.

To give parameters more weight according to their order of importance, we have assigned a value of 3 to the *knowledge relationship*, a value of 2 to *the number of messages exchanged* and a value of 1 *to the number of times the friend has been tagged*. With such values, priority is given to users who are *family members* or *close friends* because we consider that, on average, they are likely to have a closer (and thus "*better*") relationship to the user. On the other hand, a friend who exchanges many messages with the user is supposed to know the user without necessarily being part of his "*knowledge relationship list*".

Thirdly, a person tagged in one of the user's pictures may have a close relationship with the user, hence the function:

$$P(C_i) = \begin{cases} 3 & if\ i = 1 \\ 2 & if\ i = 2 \\ 1 & if\ i = 3 \end{cases}$$

Afterwards, a weighting coefficient α_1 is assigned to each parameter C_i (with $1 \leq i \leq 3$) in order to refine the general weight, where α_1 has a value of 2 if the friend is a family member, 1 for a close friend and zero otherwise, α_2 is the integer component of $\frac{3 * C_3}{TAG}$, where *MAX* is the maximum number of messages exchanged by all friends, and α_3 is the integer component of $\frac{3 * C_2}{MAX}$, where *TAG* is the maximum number of pictures where the user is tagged. Thus, $\alpha_1 \in \{0, 1, 2\}$, $\alpha_2 \in \{0, 1, 2, 3\}$ and $\alpha_3 \in \{0, 1, 2\}$ and the weights are calculated according to:

$$Weight = C_4 \sum_{i=1}^{3} \alpha_i P(C_i)$$

Example 2: Consider a friend who is a family member of a user in need of protection. Suppose that they have exchanged 20 messages and that the total number of messages exchanged between the user and his friends is 300, and that this friend's profile has been tagged 7 times by the user (out of a total amount of 10 tagged pictures). Moreover, assume *Protect_ U* has found that this friend's account features a medium risk profile. The weight assigned to this friend via Equation 2 is calculated as: 1.5 (2 * 3 + 0 * 2 + 2 * 1) = 12.

Thus, *Protect_ U* assigns a weight to each friend using Equation 2 and only the top 10 friends (according to their computed weight) will be identified as *trusted friends*. We note that *Protect_ U* also makes it possible for the user to manually edit the list of trusted friends. This is an important feature since a user may sometimes prefers to add friends who are not on the list generated automatically by *Protect_ U*. The last step consists in sending the user his profile risk level and the related recommendations. The following section presents the details surrounding the implementation of *Protect_ U* as well as the results of our validation experiments.

5. Implementation and validation

Protect_ U is a Facebook Application (executed directly on the Facebook website).It was developed with the Facebook platform API which allows users to interact with the system while giving access to networking functions, shared information and user friends lists. We will present the *predictions* obtained at the classification module level, as well as the results yielded by the validation of the recommendation module.

5.1 Classification module

165 users were involved in the classification module rule creation stage. They were contacted on an individual basis via Facebook and data collection lasted 2 months: March and April 2011. This data purposefully includes a sampling of users from various age groups and geographic locations. Any given user could only participate once and all participants who failed to fully answer the questionnaire were disregarded. At the end, only the 131 participants who answered on all the questions were considered.The entire set of rules obtained via the application of algorithm C4.5 are shown in Figure 4.

A *confusion matrix* was calculated to determine prediction results yielded by the decision tree, as well as an *accuracy evaluation* to assess the result's accuracy. The values of the *confusion matrix* are shown in Table 4. The line related to the *critical* level reveals the following predictions: 79.3% of critical profiles will be detected correctly,13.8% will be considered to be risky,6.9% will be considered to be medium risk, and 0% will be considered low risk. Risky profiles will be accurately detected 61.7% of the time, medium risk profiles will be properly identified in 58.5% of the cases, and low risk profiles will be identified 85.7% of the time.

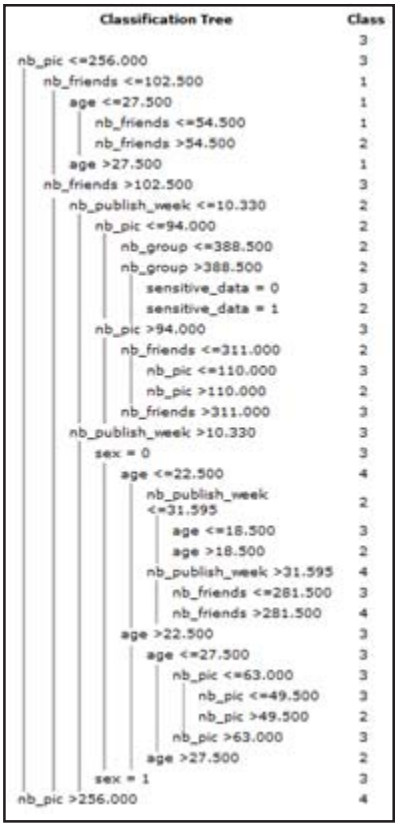


Figure 4. Set of rules extracted from the decision tree

		Predicted classes			
		Low	Medium	Risky	Critical
Actual classes	Low	85.7%	14.3%	0.0%	0.0%
	Medium	7.3%	58.5%	31.7%	2.4%
	Risky	0.0%	27.7%	61.7%	10.6%
	Critical	0.0%	6.9%	13.8%	79.3%

Table 4. The entries of the confusion matrix

	Classification accuracy	Sensitivity	Specificity	Precision
C4.5	0.6718	0.8571	0.9744	0.8000

Table 5. Accuracy evaluation

The *accuracy evaluation* table (see Table 5) allows us to determine that the *average estimation* of the *classification*, when applied to a new sample, was 67.18%. It would be interesting to investigate the variation of this percentage when considering a larger sampling of users.

Sensitivity and *Specificity* reveal that the classification module is able to identify positive results 85.71% of the time and negative results 97.44% of the time.

The *Precision* indicates that the probability of obtaining the same results under unchanged conditions is 80%, which is a promising result.

The following section presents the results of the classification module, knowledge-based filter and community-based filter validation process.

5.2 Recommendation module

163 (new) Facebook users agreed to take part in the *Protect_ U* validation process. None of these participants had been approached during the classification process. By means of a small questionnaire integrated in the application(consisting of five questions), participants were asked to submit their comments respecting the relevance of the recommendations provided by our system, as well as the accuracy of our automatic trusted friends identification results. The questionnaire also checks whether they are willing to change their behavior on Facebook beyond the magnitude proposed in the recommendations they received through our system. An overview of this questionnaire follows below:

1- How many people whose pictures are shown above are among your close friends?

- ☐ No picture is shown
- ☐ None
- ☐ Some
- ☐ Most
- ☐ All

2- Do you think that the risk level shown applies to your profile?

- ☐ No
- ☐ Yes
- ☐ I don't know

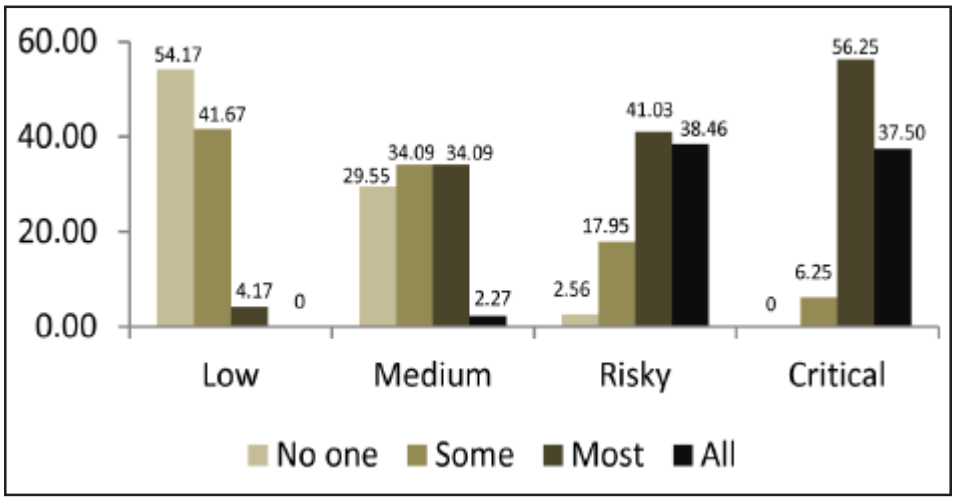


Figure 5. Precision of selected close friends per risk level

3- Do you feel that these recommendations apply to you and might increase the safety of your profile?

- ☐ No recommendation is shown
- ☐ No
- ☐ Yes
- ☐ I don't know

4- Are you willing to apply such recommendations?

- ☐ No recommendation is shown
- ☐ No
- ☐ Yes
- ☐ I don't know

5- How many people whose pictures are shown above are among your close friends?

- ☐ No
- ☐ Yes
- ☐ I don't know

The distribution of these participants' risk level shows that 15.30% have a low risk profile, 40% have a medium risk profile, 25.88% have a risky profile and 18.82% have a critical profile. Figure 5 shows that *Protect_ U* has accurately recognized 79.49% [41.03% + 38.46%] of the risky profiles' trusted friends and 93.75% [56.25% + 37.50%] of critical profiles' trusted friends.

However, for medium risk users, *Protect_ U* determined trusted friends in only 36.36% [34.09%+2.27%] of the cases. This percentage decreases to 4.17% for low risk profiles. This last result is not surprising considering that we assign the greatest weight to parameters such as the knowledge relationships (family and close friends) and messages exchanged (Equation 2). Generally, the values of these two parameters are not high in the case of medium risk or low risk profiles. Given the weak level of detection of trusted friends in these two cases, *Protect_ U* chooses the first ten friends having obtained the greatest weight by applying Equation 2, even though there may be individuals among these who are not considered to be trusted friends.

The answer to the question – *Do you think that the risk level shown applies to your profile?* – has allowed us to draw the results appearing in the *Found* values of Figure 6. These clearly show that the results obtained for the low risk profile considerably exceeds our predictions. On the other hand, the values for the critical profile do not reach their predicted value. This result can

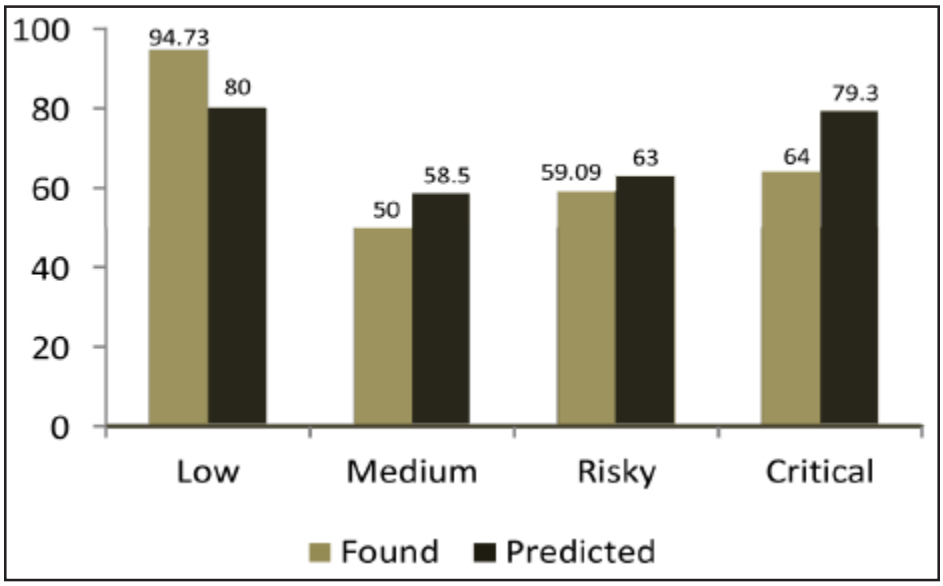


Figure 6. Found profiles versus predicted profiles

be interpreted as follows: since users were pleased to see a low risk profile diagnosis they easily confirm it without hesitation. In the case of critical profiles however, not all participants appreciated or accepted the fact that their profile was considered to be *critical*. However, *medium risk* and *risky* profile results remain relatively similar to their predictions.

With this in mind, over 70.18% of respondents found, *on average*, that the suggested recommendations were useful. However, only 58.38% of these respondents were willing, *on average*, to take these recommendations into consideration. In addition, those who are most reluctant to apply our recommendations are users with a critical profile. This may happen because the number of recommendations to apply was high and may have discouraged many of them from adopting them.

Another group of reluctant users are the users with low risk profile assessments. This reluctance may be due to the fact that they received few recommendations from *Protect_ U*, or that they did not feel the need to act on the recommendation, given their low risk rating.

Our questionnaire also enabled us to observe that 71.93% of participants, on average, were willing to change their behavioral habits on Facebook after seeing the recommendations. This result hints towards the fact that social network users are not always aware of the hazards they face when publishing sensitive information. However, some need to be directly called upon to react. This is a role that a sound privacy protection system can and must play in the future.

6. Conclusion

We have presented *Protect_ U*, a system to improve the security of private information on Facebook. The system is based on two modules: a classification module and a recommendation module. For the first module, we surveyed 131 Facebook users, while gathering key parameters from their accounts, in order to create a database that enabled us to extract a set of rules to classify the users in either a low risk, medium risk, risky, or critical security classification. Based on these rules and classes, we then matched one or several recommendations to the users, in order to help them increase their account's security. For the recommendation module, the classification rules were applied to 163 new participants and recommendations were sent to them. On average, 70.18% of participants found our recommendations to be relevant and 71.93% were willing to change their behavior. In order to better personalize our recommendations, this second module leverages an approach called community protection that encourages a user's trusted friends to participate and observe the content of the user's account in order to report any suspicious data exchange. For this purpose, *Protect_ U* automatically constructs a list of these trusted friends, based on their knowledge relationship with the user, the number of messages exchanged, the number of tags on the friend, the risk level of the friend's profile, and whether the user has executed Protect_ U before.

Our automatic process recognizes trusted friends in most cases. For critical profiles it yielded a 93.75% accuracy and, for risky profiles, 79.49%.The results we presented reveal, in some respects, the extent to which some Facebook users tend to protect their privacy. For large-scale experimentation purposes, it would be possible to adjust the *Protect_ U* parameters so that the system might be executed on the Open Social development platform used by social networks like *Google+*, *MySpace*, and *Friendster*.

It would also be interesting to expand the functionalities of *Protect_ U* to enable the system to protect users from suspicious friends by analyzing, among other metrics, the contents of published pictures and offending texts. We are currently investigating these and other additional modules that might enable the system to reach this objective in the near future. Finally, we certify to respecting a strict ethical and legal obligation throughout our entire research process: all information gathered throughout our studies has remained anonymous.

7. Acknowledgments

This research has been supported in part by funding from Canada's Natural Sciences and Engineering Research Council (NSERC).

References

[1] Aïmeur, E., Brassard, G., Fernandez, JM., Onana, FSM., Rakowski, Z. (2008). Experimental Demonstration of a Hybrid Privacy-Preserving Recommender System. *In:* International Conference on Availability, Reliability and Security (ARES).Barcelona , Spain. p. 161-170.

[2] Aïmeur, E., Gambs, S., Ai, H. (2010). Towards a Privacy-Enhanced Social Networking Site. *In: International Conference on Availability, Reliability and Security (ARES).* Krakow, Poland. p. 172-179.

[3] Aïmeur, E., Schönfeld, D. (2011) The ultimate invasion of privacy: Identity theft. *In:* Privacy, Security and Trust(PST). Montréal, Canada. p. 24-31.

[4] Bilge, L., Strufe, T., Balzarotti, D., Kirda, E. (2009). All your contacts are belong to us: automated identity theft attacks on social networks. *In:* The international conference on World wide web (WWW). Madrid, Spain. p. 551-560.

[5] Bin, Z., Jian, P.(2008). Preserving Privacy in Social Networks Against Neighborhood Attacks. *In:* International Conference on Data Engineering (ICDE).Cancún, México. p. 7-12.

[6] Bonneau, J., Bonneau, J., Preibusch, S. (2009). The Privacy Jungle: On the Market for Privacy in Social Networks.*In: The Workshop on the Economics of Information Security (WEIS).*London, UK. p. 1-45.

[7] Bramer, M. (2007). Principles of Data Mining. London, UK. p. 82-84.

[8] Statistic Canada (2009). General Social Survey - Social Networks (GSS). p. 4-143.

[9] Carmagnola, F., Vernero, F., Grillo, P. (2009). SoNARS: A Social Networks-Based Algorithm for Social Recommender Systems.*In:* International Conference on User Modeling, Adaptation, and Personalization (UMAP). Trento, Italy. p. 223-234.

[10] Fang, L., LeFevre, K.. (2010). Privacy wizards for social networking sites, *In:* The international conference on World wide web (WWW).North Carolina, USA. p. 351-360.

[11] Fogel, J., Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior.* 25 (1) 153-160.

[12] Hussey, J. (2011). Twitter in higher education from application to alumni relations. Higher education administration with social media: including application in student affairs, enrollment management, alumni relations, and career centers. Bingley, UK. p. 249-272.

[13] Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods and Algorithms. NY, USA.p. 69-198.

[14] Kelsey, T. (2010). Social Networking Spaces: From Facebook to Twitter and Everything In Between. p. 5-10.

[15] Korolova, A., Motwani, R., Nabar, S., Xu, Y. (2008). Link privacy in social networks. *In:* The Conference on Information and Knowledge Management (CIKM). Napa Valley, California, USA. p. 289-298.

[16] Lederer, S., Hong, I., Dey, K., Landay, A. (2004). Personal privacy through understanding and action: five pitfalls for designers. *Personal Ubiquitous Comput*.8 (6) 440-454.

[17] Lipford, HR., Besmer, A., Watson, J. (2008). Understanding privacy settings in facebook with an audience view. *In:* Usability, Psychology, and Security (UPS). San Francisco, California. p.1-8.

[18] Liu, K., Terzi, E. (2010). A Framework for Computing the Privacy Scores of Users in Online Social Networks. *Transactions on Knowledge Discovery from Data.* 5(1) 1-30.

[19] Maximilien, ME., Grandison, T., Sun, T., Richardson, D., Guo, S., Liu, K. (2009). Privacy-as-a-Service: Models, Algorithms, and Results on the Facebook Platform. *In:* Web 2.0 Security and Privacy (W2SP). Oakland, CA, USA. p. 351-360.

[20] Mazzia, ALK., Adar, E. (2011). The PViz comprehension tool for social network privacy settings. University of Michigan Technical Report (UMTech Report). Michigan, USA. p. 1-8.

[21] Ninggal, M., Izuan, H., Abawajy, J. (2011). Privacy Threat Analysis of Social Network Data. *In:* International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP). Melbourne, Australia. p. 165-174.

[22] Patil, S., Kobsa, A. (2010). Enhancing privacy management support in instant messaging. *Interacting with Computers*. 22 (3) 206-217.

[23] Reeder, RW., Bauer, L., Cranor, LF., Reiter, MK., Bacon, K., How, K. (2008). Expandable grids for visualizing and authoring computer security policies. *In:* Conference on Human Factors in Computing Systems (CHI). Florence, Italy. p. 1473-1482.

[24] Soman, KP., Diwakar, S., Ajay, V. (2006). Insight into Data Mining: Theory and Practice.p. 67-68.

[25] Timm, C., Perez, R. (2010). Seven Deadliest Social Network Attacks. p. 14-127.

[26] Ryan, N., Lavoie, P-E., Dupont, B., Fortin, F. (2011). Note de recherche no. 13. La fraude via les médias sociaux. p. 7-13

[27] Yan, J., Ahmad, ASE. (2008). A low-cost attack on a Microsoft captcha. *In:* Computer and communications security (CCS). Alexandria, Virginia, USA. p. 543-554.