Hyves for criminals - A Case Study

Showing the privacy risks of social networks

J. Hofste
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, Enschede, The Netherlands
j.hofste@student.utwente.nl

ABSTRACT
Social networks are very popular today and the number of members is increasing. People post a lot of information on these networks and it’s relatively easy to get the information out of it. It is only possible if a profile is public or you have a direct or indirect friendship with the person. The networks contain a lot of privacy-sensitive information about the members. They post for example messages on their profiles about what they did in the past, or they plan to do in the future.

A risk is that people can misunderstand the visibility of their profile and of course the information on the profile. If people make their profile public it’s visible for everyone in the world. The second risk is that some members like to have a lot of friends and they simply accept every invitation from others. We try to expose that these risks are serious.

The contribution of this paper is to unveil the risks of social networks and demonstrate how these networks can be abused. We lay out in a case study how a specific social network, Hyves, can be used to make it interesting for criminals.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - Information Filtering; H.3.4 [Information Storage and Retrieval]: Systems and Software - Information networks; K.4.2 [Computers and Society]: Public Policy Issues - Privacy

General Terms
Experimentation

Keywords
Social Networks, Privacy, Telephone book, Information filtering

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to post on servers or to redistribute to lists requires prior specific permission and/or a fee.

12th Twente Student Conference on IT January 22, 2010, Enschede, The Netherlands.
Copyright 2010, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

1. INTRODUCTION
Social networks are increasingly popular and attract more and more users [1]. Some examples of social networks are Hyves, Facebook, LinkedIn and MySpace. In every network you can choose the visibility of your profile. It seems that not everybody knows the risks and the future effects of a public profile [16]. For example, your future employer can view your profile. They can read about your interests, see your photo’s and so on. So he/she can get an idea of who you are and what you’re doing in your life. But not only your future employer. If you send a work related message to one of your friends and the information is negative about your current employer or other colleagues, then they can also read the message with serious consequences. It’s obvious that social networks can have an influence on your daily life and work.

In the social network Hyves people post so called "Who What Where” messages, in Dutch "Wie Wat Waar” messages (WWW). In these messages people talk about what they did in the past, or they plan to do in the future. It looks like Twitter-messages [30]. This information seems very useless, but if this information is filtered for the word “holiday” for instance, people are perhaps in fact on holiday at that moment; especially if they post a message like “Tomorrow on holiday till the 20 of August”. It is possible that the person is not at home from tomorrow till the 20 of August. It only depends whether they go on holiday with everyone in the house, or just with acquaintances or friends who live somewhere else.

To acquire the location of people, we search the home addresses of people in the telephone book. We also show how easy it is to get the exact or unique address of someone, otherwise the people cannot be located.

So if the information will be combined with the addresses of people, a potential list of houses or addresses where potentially nobody is home can be made. Social networks contain a lot of (privacy-sensitive) information and this case study will show which information can be combined to give us valuable information. That means that we can generate a list with potential empty houses.

In the literature you can read [7] [2] about social networks and their privacy consequences. There are many applications being made for these networks, for both Facebook [4] and Hyves [12]. But there is no application which couples multiple systems with each other, to get some privacy-sensitive information and show the risks of this networks. In this paper we will focus on the social network Hyves, because this social network is the most popular in The
In the Netherlands [25].

In the next paragraph we enumerate the research questions.

**Research Questions.** The main research question we examine in this paper is:

- Is it possible to get a list with potentially empty houses if we combine Hyves with other systems?

To answer the main research question we propose the following sub-questions:

- Which information can be extracted from Hyves and which information can be combined with other applications?
- What is the ratio of unique hits in the telephone book, searching on the last name?
- How can the WWW-messages be analyzed?
- Is it possible to get more information if we try to make a friendship with the top 10 popular people in a city and show that popular people simply accept every invitation?
- How can the results (potential addresses) be filtered, so that the list is more accurate?

**Approach.** To answer the research questions we will propose the following method. We will try to link the social network Hyves to the Dutch telephone book. The information of people can be extracted from Hyves, then for each person we search the address in the Dutch telephone book. After that the WWW-messages will be obtained from Hyves for each person. Finally the WWW-messages will be analyzed. That is briefly explained our approach. If we follow this procedure, we can determine the statistics and the further results. The organization of this paper will be described below.

**Organization.** Section 2 starts with a short overview of the current social networks. In section 3, the State of Art, the current research in social networks will be described, focussing on the network Hyves. Further, in section 4, the method how we try to couple these systems to get this information will be explained. In section 5, the results and statistics of the research will be shown. After that, in section 6, the sub-questions above will be validated. In section 7, the conclusions will be drawn.

## 2. CURRENT SOCIAL NETWORKS

In The Netherlands there are a lot of social networks used. We discuss some of them below; namely Hyves, Facebook, LinkedIn, MySpace and Twitter.

**Hyves.** Hyves is a very large social network in The Netherlands. It is the most visited social network, according to the STIR [25]. Nowadays it has more than 9 million members. Every month there are 8 million unique visits on Hyves from The Netherlands [8]. That means that 48.5% of the residents in The Netherlands (total 16.56 million) [31] visit Hyves every month. Furthermore, they serve 6.5 billion page views every month[11]. In addition, the name of this network is the most searched word on Google.nl [10].

**Facebook.** Facebook is an application that is used all over the world and is stationed near San Francisco in the United States. It has 350 million members world wide. The statistics of this network are interesting, the average user spends more than 55 minutes per day on Facebook. The site has 70 translations and has 350.000 applications [3].

**LinkedIn.** LinkedIn is a more business oriented social network. It has 53 million members in over 200 countries and territories [18]. It’s primarily intended to summarize your professional expertise and accomplishments. LinkedIn has 385 employees and it is located in California.

**MySpace.** MySpace is also stationed in California and has 250 employees[20]. This network has declined in popularity over the last years.

**Twitter.** Twitter is a real-time short messaging service and is located near San Francisco. It has 104 employees and millions of people use it every day [29].

There are five popular social networks listed above. The only one which is totally different from the others is Twitter. In Twitter you can only post messages with a maximum length of 140 characters [29]. In the other networks you can make a full profile, with photos, larger messages and specify more information about yourself.

## 3. STATE OF ART

Today there are a lot of papers about social networks [2] [7] [21]. Research has been conducted about the different social groups in social networks, but also around the different ages [21]. The most social networks have a privacy policy. It contains a description of how to protect your privacy [14] [5]. The papers give attention to children who are younger than 18. They may use social networks, but only if they have consent of their parents or guardian. Research has also been done concerning privacy issues in social networks. In the paper of Gross and Acquisti [7], 4.000 users of Facebook have been analyzed. They conclude that only a small number of members change the default privacy preferences in their profile. That means that the majority has an open profile, so it’s easy for others to make a digital dossier of someone.

An other study about privacy in two social networks has been done by Dwyer, Hiltz and Passerini [2]. They’ve made a comparison between the networks MySpace and Facebook. The conclusion in their paper is that people trust Facebook more then MySpace. That concludes that people have placed more identifying information about theirselves on Facebook than on MySpace.

In the privacy policies of the social network sites [14] [5] [17] [19] it is also explained what they use your data for. For example, in the policy of Hyves is written that they use your data to adapt Hyves to your wishes and needs, which can be considered a rather vague description. You don’t know what you’re data will be used for. Furthermore, it is explicitly said that they protect your information by encrypting your password [14]. But if the databases will be compromised by perhaps an DB administrator, personal
The choice for the social network is Hyves. Facebook shows in its privacy policy that they encrypt the data with an SSL technology, which seems save, but that’s only the connection and not the storage [5]. Such large networks would give more attention to these risks, and implement something like Type Based Proxy Re-Encryption [26]. With such crypto techniques it’s not so easy to read the information if the databases will be compromised. These kind of techniques are based on public and private keys and a semi-trusted proxy (e.g. the database server) who encrypt the data so that only people with the right keys can decrypt the data. The proxy itself can’t read the information, because it doesn’t has the keys. This technique is maybe an exaggeration for this purpose, but if in the privacy policy only the encryption of passwords is guaranteed and no further protectionist measures, it seems easy to compromise the database and extract information.

Applications. For Hyves some applications have been developed, which also use the API [12]. That are applications which calculate the average age of your socialnetwork-friends, or calculate the ratio male/ female of your friends. There’s also a GoogleMaps application available which displays your friends on a map, based on their location. One more interesting implementation is the “Twwwitter” application [22]. This application makes it possible to put your Twitter messages automatically on your hyvespage, in the “Wie Wat Waar” section. Thus, there can be more information in the “Wie Wat Waar” section, if users use this application. If there’s more information in this section then it’s possible that there is a larger number of results. Some applications have been developed for social networks, but there’s no application which shows the privacy risks in this networks.

4. METHOD

In this section we explain our method which allows us to answer the research questions. The main idea is to couple two systems, Hyves [13] and De Telefoongids [27]. De Telefoongids is a Dutch application for searching telephone numbers from people. It’s the Dutch publisher of the telephone book. There are also other telephone book applications, like ZoekEnBel.nl [32] and Gebeld.nl [6]. From some samples it seems that they use very different data sources, because the number of results are different. We chose for De Telefoongids because that is the official publisher of the telephone book and they publish every year an updated version of the telephone book. Therefore the data has a maximum age of 1 year.

The choice for the social network is Hyves because it is the most popular social network in The Netherlands [28]. In addition, if we want to check our results in practice it is easier to do so in the neighborhood than in foreign countries.

For this case study we focus on a medium-sized city in The Netherlands, Enschede. The city Enschede has just 156,000 residents [23] and is located in the east of The Netherlands.

Now the chosen method will be described in several steps:

Step 1. First of all, we subtract all people from Enschede out of the social network Hyves [13]. For this subtraction we used the Hyves API [15]. We saved everything we can get from the profiles: user ID, nickname, first name, last name, gender, birthday and friends count. This can be used if the system will be extended or coupled to other systems.

Step 2. For all the profiles we pick the last name and search this in the telephone book. On the website of the telephone book [27] you enter a last name and a city and subsequently the possible addresses will be shown. All the possible addresses will be saved for each profile. If there is just one result, we assume that it’s the correct address for the profile. If there are multiple addresses found for a last name, another search can be done according to the initials and filter a possible address, but that is more complicated and due to time constraints not possible in this case study.

Step 3. After searching in the telephone book, we get the WWW-messages for each profile. We make a web-service call to the Hyves API with the specific user ID. So if we are allowed to view the messages we save them in the profile row in the database.

Step 4. After this step the WWW-messages can be analyzed. When the messages are filtered on the Dutch words for holiday and October, we may find profiles of people who are possibly on holiday during the month October (Dutch autumn holiday). After the filtering there’s a list with people who are possibly on holiday. If the list is filtered for people whose address is known, we have a list with profiles and addresses where maybe nobody is at home.

After step 4 we repeat the steps 1 till 4 again but with some adjustments. The messages we filter are only mes-
sages which are publicly available or are available because the profiles are connected (friend relationship or friend of friend relationship) with the specific Hyves account. The next steps, which all have a prefix (1.), are listed below.

**Step 1.1.** To receive more messages than we have received up to now, we try to establish a friendship with the 10 most popular females in the city, based on the friend count. The ten most popular women we try to get connected to are between 15 and 20 years old. This category is chosen, because our own profile says "male" and "21 years old" and the females with that age like to have as many friends as possible, especially men. Hopefully they accept our invitation. If they ask why we try to get a connection, we say we’ve met each other in a pub.

We selected these people by this SQL query, out of the table with all Hyves-users from Enschede:

```
SELECT *
FROM 'hyvers'
WHERE 'gender' = 'female'
AND 'birthday_year' <= 1994
AND 'birthday_year' >= 1989
ORDER BY friendscount DESC
LIMIT 10
```

We wait a few days and then we repeat step 1. We subtract all the people from Enschede out of Hyves.

**Step 1.2.** For all profiles we acquire in step 1.1, the possible addresses will looked up in the telephone book, same as step 2.

**Step 1.3.** For each profile, a webservice call will be made to get the WWW-messages. These will be saved in the profile row.

**Step 1.4.** Now the messages we received after step 1.3 will be checked for indicator words. At the same time the analyzed messages will be checked for the accuracy of the results. Otherwise they will be filtered according to the date the messages were created.

Next, the results with and without the ten extra friends will be compared. Because there are also people who have set the visibility of their profile to "Vrienden van vrienden". That means that friends of friends can view your profile.

We use this method, because now we can answer the questions consecutively. After step 1, we can answer the first research question. After step 2 the second question about the ratio, searching on the last name in the telephone book, can be answered. The third research question can be answered after step 4. The answer of research questions 4 and 5 can be given after step 1.4. So the method is derived from the research questions.

5. RESULTS

In this section the results will be shown. This case study is focussed on the profiles in the city of Enschede. This gives us 51.693 results, that means that 33% of all people in Enschede have a Hyves profile, assuming that everybody has one profile. The profiles we inspect are not only the public profiles, but also the private profiles, which we are allowed to see with our own profile. From the total number of profiles there are 4.117 profiles without a last name. Thus we can use just 47.576 profiles. For each profile we get the last name and the city (always the same in this case study), and search in the telephone book. We found 5.187 profiles with a unique address, which means finding only one result in the telephone book search matching the user/owner of the profile. Thus 10% from the total number of profiles have a unique address.

After analyzing the WWW-messages, we arrive at a total of 3.741 profiles where these messages are available. That’s 7.2% of the total profiles, where these messages are available. The total number of WWW-messages is 5.349, that’s an average of 1.4 message per profile. To the other profiles we didn’t get access to view the messages (40% of the total profiles). The rest of the profiles didn’t have any WWW-messages.

In total there are 374 people whose addresses are known and who have WWW-messages. That’s a ratio of 0.7% of the total users count 51.693. The 374 people have a total number of 554 messages which can be analyzed.

The query described in the section ‘Method’ by step 5, gives us more then 3.000 results (query without the last row LIMIT 10). We try to establish a social connection with these people, to add them as friend on Hyves. In the third column you see the people who have accepted our invitation (Y) or not yet (-).

<table>
<thead>
<tr>
<th>id</th>
<th>friend count</th>
<th>accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1494</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>1046</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>1031</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>1000</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>997</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>979</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>959</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>929</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>913</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>886</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 1: Table with top 10 friend counts of women born between 1989 and 1994

The sum of the friend counts who have accepted our invitation is 6.365. Six of the 10 people we’ve invited accepted our invitation. After accepting the invitations the WWW-messages will be gathered again. The result is 395 people compared to the previous result of 374. The number of people who have WWW-messages and where the address is known has increased by 5.6%. And our friend count has risen from 204 to 2010, a percentage change of 2.9%.

The total number of messages has a lower increase to 561 compared to the old value of 554. The rise of the number of messages is just 1.2%.
6. VALIDATION

In the chapters 4 and 5 the used methods and the results have been shown. In this chapter we validate the listed sub-questions.

One of the sub research questions is: Which information can be extracted from Hyves and which information can be combined with other applications?

Now we have seen that it is quite easy to get information from Hyves [15]. The API allows us only to get a maximum of 100 results by each webservice call. If we want to get all the information of the 51,693 profiles, we have to make 517 webservice calls. But they can be generated automatically, therefore this is not a problem. In principle, the information which is available on the frontend for example blogs, pictures and profile information can be gathered by the API. So, as is shown, it is possible to combine last names with a telephone book application to produce an identity profile. If other applications are also linked to this system, perhaps Google, then it’s possible to make better identity profiles.

The second question is: What is the ratio of unique hits in the telephone book, searching on the last name?

As shown in the Results section, of the 47,576 profiles with last names, there are 5,187 profiles with a unique hit in the telephone book. That means that the ratio of searching a last name in the telephone book is: 5,187 / 47,576 * 100 = 10.9%. So approximately 1 of the 9 search queries has a unique result. Enschede is a medium-sized city and this percentage gives an estimate.

The third question is: How can the WWW-messages be analyzed?

All the WWW-messages will be matched with a regular expression. We match 2 fields of each "Wie Wat Waar" record.

Array

```php
$array = array(
    'wwid' => '042d9c626df338f2',
    'emotion' => 'Morgen op vakantie',
    'where' => 'Dubai',
    'url' => 'http://***.hyves.nl/wiewatwaar/***/',
    'userid' => 'c25e00618a3baed7',
    'visibility' => 'superpublic',
    'created' => 1246649982,
)
```

Figure 2: Example of an "Wie Wat Waar" record

These are the emotion and the where fields, see Figure 2. The other fields are not interesting for word filtering, because those fields can’t be filled in by the user himself. The "Wie Wat Waar" records are filtered on the words below. These are all words in Dutch, English and German for amusement parks, holiday, weekend break, hotel, countries and capitals of European countries.

So the total number of messages which contain one or more matching words from the array in figure 3 is 40 on the emotion field and 15 on the where field. So a total count of 55 messages. That is a percentage of 9.9% from the 554 WWW-messages. So now we have seen that 1 out of 10 messages can be interesting, because they contain one or more matching words from the array above.

The fourth question is: How can the results (potential addresses) be filtered, so that the list is more secure?

The list can be filtered on the field created, that is the unix timestamp when the WWW-message is created. Thus if we filter the most recent items, our list will be more accurate. We filter the results on the last month, there are 5 messages left.

The last five messages are manually analyzed. Two of them were posted in the last week and we’ve called them. On of them did not answer, the other person we’ve called answered his phone. We assume that the person who didn’t answer the calls was not home at that moment. For a graphic presentation, see Figure 4.

Figure 3: Words, which the messages will be filtered for.

Figure 4: Filtering the WWW-msgs

The last subquestion is: Is it possible to get more information if we try to make a friendship with the top 10 popular people in a city and show that popular people simply accept every invitation?

We’ve tried to invite the ten most popular women with an age between 15-20 years who have the most friends. Six of them accepted our invitation and through this the count
of the people who have WWW-messages and whose addresses are known has increased by 5.6%. This increase is admirable because or friend count has just grown by 2.9%. But if we compare this with the total number of messages it’s different, that number is only increased with 1.2%. This can be explained by the point in time when the first sample was taken, just before the Autumn Holidays in The Netherlands. The second sample, with the extra friends, is taken during a normal week in November. Maybe people post more messages if they have nice things in perspective.

7. NEW DEVELOPMENTS

At the end of November 2009 Hyves got a new design for their homepage [9]. It’s now possible to label your WWW-messages with a location. Now you can see a GoogleMap with posted WWW-messages from your friends. If people label their messages with a location and if they have a public profile as well or are directly or indirectly connected to you, you can also make a list when filtering these messages. Then it’s not necessary to search the telephone book, because the users identify their current “Wie Wat Waar” location by themselves.

If this new development becomes more popular and this method will be extended to also use that information, the number of results we get will be higher and maybe more accurate.

8. CONCLUSIONS

Social networks are very popular and people share more and more information about themselves on these platforms. In this case study we’ve shown how two systems can be combined to get valuable information. From a medium-sized city, Enschede, we get a list with 55 addresses of potentially empty houses. These are profiles with WWW-messages which content match one of the $swords$ to filter and where the address is also known. If the list will be filtered on the last month we get just five results. After that we’ve filtered on the last week and we get two results. Then we called the two people and one answered the call and one not. So we may conclude that this research has demonstrated that it’s a risk to be careless with your information, see also Figure 4.

This research was only based on one city, but if we extend these to multiple cities or to the whole country, the number of results in the list will be much higher. Perhaps if we did the research during another time, we could get more results. During the summer months much more people go on holiday. Last year there was a recession[24] and people had less money to go on holiday, but when the economy is restored more people might go on vacation and the result would be more numerous.

So we can conclude that people don’t know which risks there are if they place information about themselves on the internet. Especially how to protect their information. Finally we can say that it’s possible to create a list with potentially empty houses.

Future work. As future work the prototype can be extended. It’s possible to create a realtime system which saves all the information about the profiles on Hyves. For each profile the addresses will be searched in the telephone book. All the profiles which have a unique address can be filtered. For each profile the WWW-messages will be received, every day or every hour/every minute. The system is then more interesting. If people post “work tonight” or “tonight birthday of...” then you instantly know if people are at home during the next few hours.

Furthermore, the system can also be extended to categorize the people. Maybe people 30 years and older have more valuable household and personal belongings than a student of 20 years. Then it’s more lucrative to go to the house where 30 year old men or women live.

Besides, this system can also be used for other social networks like Facebook. The difference is that there are no WWW-messages but there the equivalent is called: status updates. Another phone book would have to be chosen as well.

Acknowledgments. Many thanks to all the people of the Bachelor Referat track: ”Broadband for All”, for their critical reviews and comments.

9. REFERENCES
