# The High-School Profiling Attack: How Online Privacy Laws Can Actually Increase Minors' Risk

Ratan Dey, Yuan Ding, Keith W Ross

Polytechnic Institute of New York University, Brooklyn, New York
ratan@cis.poly.edu,dingyuan1987@gmail.com,ross@poly.edu

**Abstract.** Lawmakers, children's advocacy groups and modern society at large recognize the importance of protecting the Internet privacy of minors (under 18 years of age). Online Social Networks, in particular, take precautions to prevent third parties from using their services to discover and profile minors. These precautions include banning young children from joining, not listing minors when searching for users by high school or city, and displaying only minimal information in registered minors' public profiles, no matter how they configure their privacy settings.

In this paper we show how an attacker, with modest crawling and computational resources, and employing data mining heuristics, can circumvent these precautions and create extensive profiles of tens of thousands of minors in a targeted geographical area. In particular, using Facebook and for a given target high school, we construct an attack that finds most of the students in the school, and for each discovered student infers a profile that includes significantly more information than is available in a registered minor's public profile. An attacker could use such profiles for many nefarious purposes, including selling the profiles to data brokers, large-scale automated spear-phishing attacks on minors, as well as physical safety attacks such as stalking, kidnapping and arranging meetings for sexual abuse.

Ironically, the Children's Online Privacy Protection Act (COPPA), a law designed to protect the privacy of children, indirectly facilitates the attack. In order to bypass restrictions put in place due to the COPPA law, some children lie about their ages when registering, which not only increases the exposure for themselves but also for their non-lying friends.

## 1 Introduction

It is generally recognized that protecting the Internet privacy of minors (under 18 years of age in the US) is important, with modern society manifesting this concern in many ways. The US government, through the Children's Online Privacy Protection Act (COPPA) [3], requires commercial Web sites to obtain affirmative consent from parents before children under 13 can create an account. Many consumer, privacy and child advocacy groups continue to actively lobby governments to provide better privacy protection for minors [4]. The US Congress is currently considering new bills to strengthen online safeguards for children and teens [6], [8].

Online Social Networks (OSNs) additionally take measures to protect the privacy of minors. Facebook, for example, treats minors and adults with distinctly different

policies related to their public profiles, how members can find each other, and how they can contact each other [2]. Facebook currently bans young children (under 13) from joining, does not list minors when searching for users by high school or city, and displays only minimal information in registered minors' public profiles, no matter how they configure their privacy settings.

In this paper we show how a third party, with modest crawling and computational resources, and employing data mining heuristics, can circumvent these precautions and create extensive profiles of tens of thousands of minors in a targeted geographical area. In particular, using Facebook and for a given target high school, we construct an attack which finds most of the students in the school, and for each discovered student infers a profile which includes significantly more information than is available in a registered minor's public profile. The additional information minimally includes, for each discovered student, the student's current city, current high-school, graduation year, inferred birth year, and list of school friends. The generated profiles of about half of the identified minors also include varying amounts of additional information, including shared photos and wall postings. The information is collected *passively*, that is, without attempting to establish friend links with any of the students. As discussed in Section 2, an attacker could use such profiles for many nefarious purposes, including selling the profiles to data brokers, large-scale automated spear-phishing attacks on minors, as well as physical safety attacks such as prospecting candidate children for stalking, kidnapping and arranging meetings for sexual abuse.

Using off-line channels, it is difficult for an attacker to obtain complete lists of students attending a given target school. For example, in the course of the research for this paper, while seeking ground-truth data, we contacted administrators of four high schools and asked them to provide us with a list of names of all students currently attending their schools, with assurances of keeping the lists entirely confidential as well as not mentioning the names of the schools in this study. But the administrations of these high schools would not provide the lists, even with such assurances, fearing potential lawsuits from parents or other legal actions. High-school websites today also do not publicly provide lists of current students.

It is also difficult for an attacker to obtain complete lists of students attending a given target school directly from OSNs. As of May 2013, and documented in this paper, Facebook takes explicit measures to prevent people from obtaining school lists directly from its site. Although Facebook allows its members to search for other members who are associated with any given high school or city, *the search results returned by the service do not include registered minors*; for a high school search, they only include members who are registered as currently being 18 years or older, with the vast majority of the results being alumni of the high school. Because of this measure, it is not possible for an attacker to *directly* use Facebook's search service to collect the names of the students at any target high school and attempt to profile them.

Ironically, the privacy leakages described in this paper are indirectly caused by the COPPA law, which was designed to protect minors' privacy. Given economic costs, social concerns, and technical issues, most online services — including Facebook and Google+ — choose to avoid the COPPA obligations by banning users younger than 13. Upon creating an account, these sites ask users for their birth date to determine if they

are 13 or older. If the user indicates being under 13 years of age, the site prevents the user from creating an account. The key observations and ingredients behind our attack are the following:

1. In order to circumvent the age restriction (due to COPPA), many under-13 users lie about their age to gain access to online social networks when creating their accounts [15]. For example, in order to gain access to Facebook, an 11-year-old boy may say he is 13 years old or may even say he is over 18 when registering.
2. Several years later, when the lying minor enters high school, his registered age very possibly will be 18 or older. The OSN will therefore consider him an adult although he is actually a minor.
3. When searching for users by high school, Facebook (and Google+) only returns members who are registered adults. But a small fraction of these registered adults will in truth be minors.
4. By identifying the minors returned by the search results, and performing statistical processing on their friend lists, we show it is possible to discover most of the students in the target high school and, for each discovered student, create a profile that contains significantly more information than should be available in a minor's public profile.

Thus, the COPPA law has inadvertently set the stage for widespread discovery and inference of minors' private information.

To demonstrate the feasibility of the high-school profiling attack, we carried it out on one high school. Our institution provided us with an IRB to perform the research under the condition that we keep private all collected and inferred information about individuals and only release aggregated results. For the target high school, we were able to obtain, through a confidential off-line channel, ground-truth information including the names of all the students in the high school and their graduating classes.

To our knowledge, this is the first paper that $(i)$ identifies the third-party privacy leakage problem in OSNs for minors, $(ii)$ quantifies the extent of the privacy leakage, and $(iii)$ using measurement and analysis, investigates and quantifies the impact of a privacy law on privacy leakage. As part of responsible disclosure, we informed both Facebook and Google about the attack. This paper is a short version of a full paper, which additionally analyzes three high schools, studies privacy leakage when there is no age restriction, studies one defense, and characterizes the information available in profiles [11].

## 2 Consequential Threats

Suppose an attacker, with modest crawling and computational resources, for a given target high school, is able to determine all the students in the school and profile them, with the profiles containing a varying amount of information, but minimally including full name, profile picture, gender, current city, high school name, graduation year (i.e., grade), high-school friends and inferred birth year[1]. For a given high-school, we call the

[1] A public Facebook profile for a minor at most contains name, profile picture and gender. Thus the attacker's constructed profiles additionally contain current city, high school name, graduation year, high-school friends, inferred birth year, and for many students much more information.

collection of these profiles the *high-school profiles*. Moreover, suppose the attacker has a means to send messages directly to many of the students, and can send friend requests to all of the students. We now describe some of the consequential threats.

The first major threat is that of data brokers collecting high-school profiles and selling them to others, such as advertisers, college recruiters, and employment agencies. Because the teen market surpasses US$200 billion in the US alone, it is not surprising that data brokers are already seeking to compile dossiers on children [9] [7]. By leveraging the information in the high-school profiles, data brokers can also enhance the profiles by linking them with other personal data available online and from public records. For example, by obtaining voter registration records (which most states make available for a small fee), the data broker can use the last name and city in the high-school profiles to link the students to parents in the voter registration records, *thereby determining the street address of many of the students*. For those students with friend lists in the high-school profile, if a parent appears in the friend list, then the street-address association can be done with greater certainty. As another example, for many students, the first name, last name and city in the high-school profiles can be linked with Skype profile information in the Skype directory, thereby augmenting the profile with a means of calling and videoconferencing with the teenager.

The second major threat is that of a pedophile, who seeks to use the Internet to arrange sexual encounters with children. For example, recently a man allegedly used Facebook to arrange meetings and have indecent contact with seven different girls, ranging in age from 13 to 15. The district attorney for the case stressed the importance of minors "not sharing personal information online, like full names, ages, addresses, phone numbers and school information" [5]). A pedophile could launch the high-school profiling attack himself, using the acquired profiles to prospect for victims. As a first step, the attacker could use the profiles to narrow down the candidates in the target community. The attacker could then leverage the profile information to perform social engineering attacks and establish online contact with the candidates.

Finally, the profiles could also be used to fuel a large-scale and highly personalized spear-phishing attacks against minors. Messages could automatically be generated which mention the target students' high schools, graduation years, and friends, tricking the targets into installing malware on the family computer, for example.

## 3 Preliminaries

Throughout this paper we define a *minor* to be any person who is currently under 18 years old. Anyone 18 years or older is said to be an *adult*. Note that most students currently attending a high school are minors. (A fraction of the final-year students may be adults, with the fraction increasing each month in the school year.) OSNs typically require users to specify their birth date (day, month, and year) when they register. As discussed in the Introduction, some users may lie about their birth dates when creating accounts in order to circumvent the minimum age requirement. A user is said to be a *registered minor* if the OSN believes the user is currently a minor based on the registered birth date. We define a *registered adult* in a similar manner. In the context of Facebook, we say a user (say, Alice) is a *stranger* to another user (say, Bob) if all the following conditions are satisfied: $(i)$ Alice is not a friend of Bob; (ii) Alice is not a friend of

friend of Bob (that is, Alice and Bob have no mutual friends); and $(iii)$ Alice does not belong to any of Bob's school or work networks.

### 3.1 Facebook and Registered Minors

In Facebook, registered minors have a different experience with privacy than do registered adults. We now highlight the differences that are relevant to the current study. Table 1 shows the information about a user available to a stranger for when the user keeps the default settings and for when the user configures the setting for maximum sharing (worst case). A check in the box means the information is available to the stranger for the specific scenario. As shown in Table 1, when a stranger visits a registered minor's profile page, only a limited amount of information is available to the stranger: at most the user's name, profile photo, networks joined, and gender are available. (Typically less, depending on how the user configured her privacy settings. For example, typically less than 10% of registered minors specify network.) Further, the "Message" button will never be visible to a stranger. We say that *only minimal information is available about a user* (registered minor or adult) if a stranger, when visiting the user's public profile, sees at most name, profile photo, networks joined, and gender, and the "Message" button is not available. It follows that if a stranger visits a user's public profile and more than the minimal information is available, then the user must be a registered adult.

**Table 1.** Facebook: Default and worst-case information available to strangers

| | **Default for Reg. minors** | **Default for Reg. Adults** | **Worst-case for Reg. Minors** | **Worst-case for Reg. Adults** |
|---|---|---|---|---|
| Name, Gender, Networks, Profile Photo | ✓ | ✓ | ✓ | ✓ |
| HS, Relationship, Interested In | | ✓ | | ✓ |
| Birthday | | | | ✓ |
| Hometown, Current City, Friendlist | | ✓ | | ✓ |
| Photos | | ✓ | | ✓ |
| Contact Information | | | | ✓ |
| Public Search | | ✓ | | ✓ |

OSNs typically provide a friend-search feature, allowing its users to find new friends from different parts of their past and current lives, including friends from previous high schools. Facebook provides this feature in its "Find Friends Portal" [1], where a user can search for potential friends by inputting either hometown, current city, high school, mutual friend, college or university, employer, or graduate school. When a stranger does a high school search by the high school name, Facebook returns a few hundred users who are associated with the target high school. The stranger can also attempt to obtain additional users by creating additional fake accounts. We wrote a script that collects users in this manner. The script takes as input the target high school's Facebook ID, a username and password for a fake account, and outputs several hundred unique Facebook user IDs. We observed in the course of experiments that *Facebook does not*

*return any registered minors when a stranger searches with the Find Friends Portal*. We verified this claim by carrying out an experiment with a high school for which we have the complete list of current students at the high school, as well as the complete list of recent alumni.

*In summary, in an attempt to act responsibly towards minors, Facebook takes some precautions to protect minors' privacy. We observed and verified that Facebook does not return registered minors when a stranger searches by high school. Also, when a stranger visits a registered minor's public profile page, only limited information is made available, no matter how the minor configures the privacy settings. In particular, a minor's high school, graduation year, and friend list are never directly available to a stranger.*

### 3.2 Legal and Ethical Considerations

To perform the research described in this paper, we implemented customized crawlers that visit public Web pages in Facebook and download the HTML source code of each Web page. Our parser then extracted relevant data from the HTML source code and stored the data in an SQL database.

Crawling data in OSNs is an ethically sensitive issue. One question that arises is if it is ethically acceptable and justifiable to conduct crawling experiments in social networks? We believe that the only way to reliably estimate success rates of attacks in the real-world is to use realistic experiments. We nevertheless took several precautions while crawling. First, we only accessed user information that was publicly available. Second, by implementing sleeping functions and limiting our study to one high school, the crawling was not particularly aggressive and didn't perturb the performance of Facebook.

We also obtained IRB approval for this work from our university. As part of responsible disclosure, we informed both Facebook and Google about the attack in October 2012. Because of the sensitive nature of the information we gathered and inferred, we will not be making our data sets public and we will not explicitly identify the high school involved.

## 4 The High School Profiling Attack

We now describe our basic version of the high-school profiling attack. The attacker begins by selecting a target high school. Let $M$ be the set of all the students currently attending the target high school with active accounts in the OSN. The goal of the attack is to find most of the students in $M$ and obtain (or infer) as much profile information as possible about each of those students. We do not require the attacker to be an OSN friend, or a friend-of-a-friend, of any of the students in $M$, that is, the attacker may be a stranger to all the students in the high school *throughout the duration of the attack.* With sufficient computational resources, the attack could therefore be launched against hundreds or even thousands of high schools.

### 4.1 The Basic Attack: Exploiting Lying Minors

For any user $u$ in the OSN, let $F(u)$ be the user's current set of friends. For some users, $F(u)$ will be visible on the user's public profile; for other users $F(u)$ will not be publicly available. The attack in its most basic form operates as follows.

1. The attacker inputs the name of the target high school into the OSN's high-school search function. The search function returns a list of members who are associated with the target high school. The attacker may use a script to automatically scroll down the page (thereby sending additional HTTP requests with AJAX) in order to get a longer list of members. The attacker may also use multiple accounts when searching. We refer to the set of all the members found in this manner as the *seeds* and denote the set by $S$.
2. The attacker uses a crawler to download the public profile pages for each of the seeds, parses the pages, and determines the users who indicate they currently attend the target high school (by listing their high school as the target high school and providing a graduation year that is the current year or a future year). Let $C'$ be the subset of seeds who explicitly indicate (in their public profiles) that they are currently students in the target high school. (Most of the users in $C'$ will be minors who, several years earlier when under 13, lied about their age during registration.) Let $C$ be the subset of users in $C'$ who make their friend lists public. We refer to $C$ as the *core set*. As we will see, the number of core users is typically fairly small, on the order of 5% of the number of students in the high school. For each user in set $C$, we know the user's graduation class year. Assuming that the high school is a four-year school, denote $C_1$, $C_2$, $C_3$, and $C_4$, for students in the first, second, third, and fourth school years in the core set $C$.
3. For each student $u \in C$, the attacker downloads the friend list, $F(u)$, from the OSN. Let $K$ be the set of all friends obtained from the core users, that is,

   $$K = \cup_{u \in C} F(u).$$

   We refer to $K$ as the *candidate set*. Our experiments show that the number of candidates will approximately be one order of magnitude greater than the target high school size.
4. We expect some of the users in $K$ to be current students in the target high school. We now try to determine which ones. For each candidate $u \in K$, we use *reverse lookup* to determine its friends in the core. Specifically, for each $u \in K$, we determine the set of friends in the core set for each of the four graduation years:

   $$G_i(u) = \{v \in C_i \ : \ u \in F(v)\}, \qquad i = 1, 2, 3, 4. \tag{1}$$

   Clearly each $G_i(u) \subseteq F(u)$. Note that to obtain the $G_i(u)$'s, the attacker *does not* have to obtain the profile pages or friend lists of any of the users in the large candidate set $K$. In fact, user $u$'s friend list may not even be directly available to strangers.
5. For each candidate $u \in K$, the attacker calculates the fraction of users in each of the core class sets with whom the candidate is friends, and then calculates the

maximum of these four fractions. Specifically, the attacker calculates

$$x(u) = \max_{1 \le i \le 4} \frac{|G_i(u)|}{|C_i|} \tag{2}$$

6. The attacker rank orders the users in $K$ according to their $x(u)$ values, from highest to lowest. The attacker chooses a threshold $t$ in the vicinity of the total number of students attending the high school (which can typically be found from Wikipedia or some other source). The attacker then considers the first $t$ students as current students in the target high school (as well as the students in the set $C'$). Let $T$ denote the set of $t$ students and $H = T \cup C'$. The attacker also classifies each such student $u \in T$ into a graduating year according to the highest $|G_i(u)|/|C_i|$ value, $i = 1, 2, 3, 4$.

At the end of these steps, the attacker has a set of OSN users $H$ believed to be students at the target high school. The attacker has also classified all the students in $H$ by graduation class year. For each student, by knowing the high school, the attacker knows the current city; by knowing both the student's last name and current city, the attacker can often determine the student's home address from voter registration records. The attacker can also estimate birth year from the graduation year.

Note that the attack relies on the attacker's ability to obtain a small set of core users, that is, finding a set of users for whom the attacker knows with certainty that the users are in the high school and knows their graduation year. Because the search function only returns registered adults who make their high school public, *a priori* the core set will have no students in the first three years of high school and few in the last year. However, because a significant high-school students lied about their birth dates when creating accounts when they were under 13 (in order to circumvent the age restriction due to the COPPA law), it is indeed possible to obtain a core set from the search function including students distributed across the four years. Also note that the attack is passive, that is, without attempting to establish friend links with any of the students.

### 4.2 Attack Performance

The set $H$, and the classification of its members by graduation year, is obtained by statistical inference and therefore may contain errors. For example, some of the users in $H$ may be false positives, that is, they are not current students at the target high school. Furthermore, $H$ may not contain all of the students in $M$. Two important measures for the performance of the attack are the *fraction of students from $M$ found*, given by $|H \cap M|/|M|$, and the *number of false positives*, given by $|H - M|$. Note that by varying the value of the threshold $t$ the attacker can trade off these two performance measures: increasing $t$ should increase the fraction of students found but should also increase the number of false positives. In this paper we estimate these measures for the test high school.

### 4.3 Enhanced attack

We now describe an important enhancement of the attack, which requires a relatively small amount of additional crawling. In the *enhanced attack*, after rank ordering the

$x(u)$'s and selecting a threshold $t$, we download the public profile pages of the first $t(1+\epsilon)$ users. (In this paper, we use $\epsilon = 1$ throughout.) Denote this set of users by $T+$. For each user $u$ in $T+$, we then check the user's profile to see if he indicates he is currently a student in the target high school. If so, we move $u$ from $T+$ to $C$, thereby increasing the size of the core set. After doing this for all $u \in T+$, we recalculate $G_i(u)$ for each $u \in T+$ and $i = 1, 2, 3, 4$, and proceed from Step 5 in the Basic Approach.

In addition to these approaches, there are many possible heuristics one may construe based on the $G_i(u)$ data. It is also possible to explore traditional machine learning approaches. As the purpose of our research is to demonstrate the feasibility of the attack rather than fully optimize it, we do not pursue these optimizations here.

### 4.4 Filtering

In order to possibly improve the performance of the basic and enhanced attacks, we also examine filtering out some of the candidate users. This filtering variation, as with the enhanced attack, requires that the attacker download the public profiles of the first $(1+\epsilon)t$ users in the candidate set. After downloading these profiles, the attacker applies filtering rules to eliminate candidates who are likely former students at the target high school (and have transferred out or have already graduated). We used the following filter rules:

- *Graduate School:* The candidate specifies a graduate school in the public profile page.
- *Different High School:* The candidate provides *one* high school and that high school is different from the target high school.
- *High school graduation year:* The candidate provides a high-school graduation year that is not in the current year or in the subsequent three years.
- *Current city:* The candidate provides a current city other than the city in which the high school resides.

### 4.5 Estimating the Crawling Effort

Most OSNs employ anti-crawling techniques to protect the data of their membersand the performance of their sites. Typically, if a member behaves suspiciously (for example, if he tries to access an overly large amount of user profiles in a short amount of time), the member's account will be temporarily, or permanently, disabled. Therefore another important measure is the crawling effort required to perform the attack.

For the Basic Attack, the crawling effort has three components: $(i)$ the number of HTTP GETs sent to obtain the IDs of the seed users $S$ (Note that with AJAX, multiple HTTP GETs may need to be sent to get the entire page.); $(ii)$ the number of HTTP GETs sent to obtain the public profile pages of the seed users in $S$; $(iii)$ the number of HTTP GETs sent to obtain the friend lists of each of the core users (again sending multiple GETs via AJAX). The approximate number of HTTP GETs sent is therefore given by $A \cdot R + |S| + |C| \cdot f/p$, where $A$ is the number of accounts used, $R$ is the number of HTTP GETs sent per account when gathering the seed list, $f$ is the average number of friends a student has, and $p$ is the number of friends gathered with a single HTTP request. (Currently, Facebook uses $p = 20$).

For the enhanced attack, we additionally $(i)$ download the profile pages of an additional $(1+\epsilon)t$ users, where $t$ is roughly the number of students in the target school, and $(ii)$ download the friend lists for the augmented core set. In Section 5 we will show that the total number of requests for a typical school is small for both the basic and enhanced attacks.

# 5 Results for High School

## 5.1 Data Sets

In order to estimate the success of the attack, we applied it to one US high school, which we refer to as HS1. We collected the data for HS1 in March 2012. HS1 is a small private urban high school with about 360 students. For this high school, we were able to obtain, through a confidential channel outside of Facebook, the complete student lists (segmented by graduation year) for the high school, and also complete alumni lists for recent graduation years. These lists enable us to evaluate the success of the attack. HS1 has a relatively high churn rate, with 10-20% of the students transferring in and out of the high school every year. Because of the high churn rate, it is a challenging problem to determine an accurate estimate of the current snapshot of the student body. However, we will see that even with this high churn rate, the basic attack provides good results.

For the HS1 students in the 2012, 2013, 2014, and 2015 graduating classes, we were able to find the Facebook IDs and public profile pages for $|M| = 325$ students. We did this essentially by running the basic attack on HS1, finding the users who were ranked the highest, and checking for their names in the ground truth list. We were not able to find the Facebook IDs for about 10% of the student body at HS1. Most of these remaining students most likely do not have Facebook accounts. A small number of them may have accounts with alias names that we could not match to the ground-truth list. The 325 students are roughly evenly distributed over the four years; for 112 students (34%) their friend lists are publicly available.

## 5.2 Initial Seed Set

We obtained initial seed sets from Facebook's Find Friend portal, using two accounts for HS1. Table 2 provides a summary of the data collected for this high school. As shown in Table 2, for HS1, we found 18 core users (with friend lists) and 6,282 candidates. For the enhanced attack, we obtained 22 (extended) core users for this high school. The number of core users is roughly 5% of the number of students in the school.

**Table 2.** Seeds, core users, and candidates for the target high school

| High school | # of students | # of students on Facebook | # of seeds | # of core users | # of candidates | # of extended core users |
|---|---|---|---|---|---|---|
| HS1 | 362 | 325 | 352 | 18 | 6,282 | 22 |

### 5.3 Crawling Effort

Table 3 summarizes the approximate crawling efforts required to collect the data sets for HS1. Note that the effort is quite small, with the number of HTTP requests sent being about twice the number of students in the target high school for the basic attack, and about five times the number of students in the target high school for the enhanced attack.

**Table 3.** Crawling effort

| | **Facebook accounts used** | **HTTP requests for seeds** | **Profile pages** | **Requests for friend lists** | **Total requests for basic attack** | **Total requests for enhanced attack** |
|---|---|---|---|---|---|---|
| HS1 | 2 | 34 | 352 | 360 | 746 | 1,576 |

### 5.4 Results for HS1

Recall that for HS1 there are 325 students having Facebook accounts. Also recall that we have the complete ground-truth information for HS1 (i.e., the Facebook IDs and graduation years for all of the 325 students). The results for both the basic and enhanced attacks, with and without filtering, are shown in Table 4 for thresholds $t$ ranging from 200 to 500. The set of users in each column includes the core users (or extended core users for the enhanced attack). In the notation $x/y$, $x$ is the number of users from the set of 325 students that are found; and $y$ is the number of users, from the set of $x$ users, that are classified in the correct classification year. We see for the top 200, 300, and 400 cases, the enhanced attack with filtering gives the best results; for the top 500 case, the enhanced attack without filtering gives somewhat better results than the enhanced attack with filtering.

**Table 4.** Results for HS1 (which has 325 Facebook users)

| | **Top 200** | **Top 300** | **Top 400** | **Top 500** |
|---|---|---|---|---|
| Basic attack without filtering | 140/112 | 206/162 | 271/224 | 301/254 |
| Basic attack with filtering | 148/122 | 196/165 | 259/227 | 299/264 |
| Enhanced attack without filtering | 169/155 | 231/211 | 261/239 | 304/281 |
| Enhanced attack with filtering | 175/158 | 232/211 | 272/250 | 299/276 |

We see that the filtering indeed reduced the number of false positives for the threshold of top 200, top 300, and top 400 users. But for the larger threshold, the filtering actually increased the number of false positives. This can be explained as follows. On one hand, when we increase the threshold beyond 400, we add mostly false positives, since there are not many true positives remaining. On the other hand, the filtering also accidentally filters out some of the true positives, giving an overall decrease in performance.

As an example, let us suppose that the attacker decides to use the enhanced attack with filtering, and considers the top 400 users as students in HS1. Examining the column

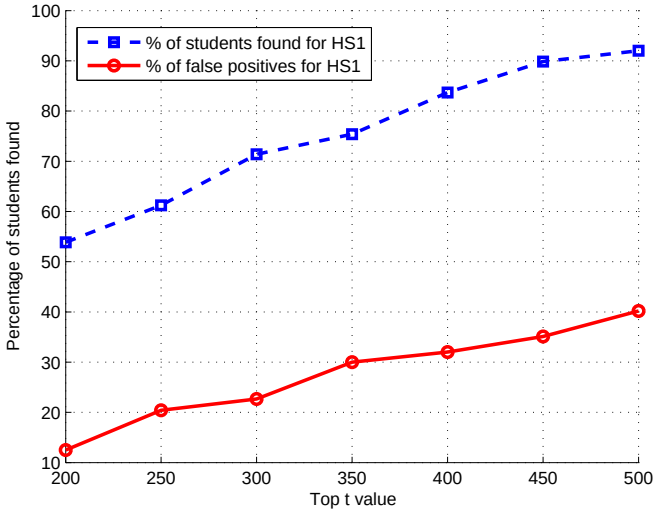


**Fig. 1.** Overall performance of enhanced attack for HS1

for 400 students in Table 4, we see that with this choice of threshold, 272 (84%) of the 325 students are included in the attacker's set. So with this threshold, the attacker finds 84% of the high school student body (having Facebook accounts) with 128 false positives (32%). Moreover, of these 272 students, 250 (92%) have been classified in the correct graduation year. If the attacker wants to reduce the false positives, the attacker can declare only the top 200 users as students, in which case there are only 25 (13%) false positives, with 54% of the students found, of which 90% are classified in the correct graduation year. If the attacker can accept a larger number of false positives, he may instead choose the top 500 students, which would include 92% of the high school student body having Facebook accounts. We show these estimates for different choices of threshold $t$ for the enhanced attack with filtering in Figure 1.

The results of obtaining 84% of the students in the high school, of which 92% are classified in the correct year, with 32% false positives are remarkable, particularly when considering the 10-15% annual churn rate at the high school. Many students attend HS1 for a short period of time. They make friends with the other students during their period of study, then their families move to another city. We manually inspected the 128 false positives (from the set of top 400 users) and found that about half of them were former students at HS1. For the other half of the false positives, they make very little public information available, so it is difficult to determine if they are former students or not (although most likely are since they have a large number of friends in HS1).

### 5.5 Summary of Results

As discussed in Section 3, when using Facebook's Find Friends Portal to search for users in a target high school, Facebook takes precautions to protect minors by not returning any registered minors. We have shown that an attacker, with relatively little

crawling effort, can discover the majority of the students at the target high school. For example, we obtained 83% of all the students in HS1, with false-positive rates of 32%. Moreover, for each high school student in the list, the attacker can determine the student's graduation year with a high-level of accuracy. An attacker can then create profiles with varying degrees of information for the high-school students, as well as a variety of means to contact the students.

## 6 Related work

There is substantial previous work on using statistical inference to infer private information about OSN users. Zheleva and Getoor [20] proposed techniques to predict gender and political views of users in four real-world datasets (including Facebook) using general relational classification and group-based classification. Jernigan and Mistree [14] demonstrated a method for accurately predicting the sexual orientation of Facebook users by analyzing friendship associations. Other papers [19,13,17] have also examined inferring private information from social networks. Thomas et al. examine scenarios where conflicting privacy settings between friends will reveal information that at least one user intending to remain private [18]. Becker and Chen [10] inferred many different attributes of Facebook users, including affiliation, age, country, degree of education, employer, high school name and grad year, political view, relationship status, university and zip code using the most popular attribute values of the user's friends. Dey et al. [12] examine a large dataset and develop a methodology to estimate ages of Facebook users. Mislove et al. [16] proposed a method of inferring user attributes by detecting communities in social networks, based on the observation that users with common attributes form dense communities.

All of the above studies focus on inferring information about adults. To our knowledge, this is the first paper that identifies the privacy problem in OSNs for minors, and also the first paper to quantify the extent of the privacy leakage. The problem is challenging since, for registered minors, little information, including friend lists, is available to an attacker. The attack makes use of two key properties in modern OSNs: $(i)$ many minors lie about their age and are therefore considered adults by the OSN; and $(ii)$ using reverse lookup, an attacker can construct a user's friend list even if the user hides her friend list to everyone.

## 7 Conclusion

In this paper we have shown how a privacy law for protecting children's privacy can inadvertently increase minor's exposure to third parties. Facebook and other Online Social Networks (OSNs) take precautions to prevent strangers from using their services to extensively profile minors. But because a significant fraction of minors lie about their ages, we show how many of the precautions can be circumvented, putting both lying and truthful minors at risk. For a given target high school, we described an attack of using an OSN to profile the current students in the high school. The attack finds the majority of the students in the school, and for each student builds a profile that includes information that is not normally available to strangers, including current city, current school, graduation year, high-school friends, and estimated birth year.

Although the COPPA law indirectly exacerbates the third party privacy problem for minors, we are certainly not arguing that governments should abandon enacting laws to protect the online privacy of children. We believe, however, that the laws must be carefully designed and consider leakages to third-parties as well as to first-parties.

## 8 Acknowledgment

This work was supported in part by the NSF (under grant 0966187). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of any of the sponsors.

## References

1. Find Friends Portal, https://www.facebook.com/find-friends/browser/, accessed May 7, 2013
2. How does privacy work for minors?, http://www.facebook.com/help/?page=214189648617074, accessed May 7, 2013
3. Children's Online Privacy Protection Act (1998), http://www.ftc.gov/ogc/coppa1.htm
4. Groups Make Recommendations for Kids' Facebook (Adweek, June 18, 2012), http://www.adweek.com/news/technology/groups-make-recommendations-kids-facebook-141195
5. Attorney General Kelly announces criminal charges in elaborate "Facebook" false identity scam targeting young girls for sex (February 10, 2012), http://www.attorneygeneral.gov/press.aspx?id=6431
6. Do Not Track Kids Act of 2011 (May 13, 2011), http://www.gpo.gov/fdsys/pkg/BILLS-112hr1895ih/pdf/BILLS-112hr1895ih.pdf
7. Senator Opens Investigation of Data Brokers (The New York Times, October 10, 2012), http://www.nytimes.com/2012/10/11/technology/senator-opens-investigation-of-data-brokers.html
8. Update Urged on Childrens Online Privacy (The New York Times, September 15, 2011), http://www.nytimes.com/2011/09/16/technology/ftc-proposes-updates-to-law-on-childrens-online-privacy.html
9. On the Web, Children Face Intensive Tracking (The Wall Street Journal, September 17, 2010)
10. Becker, J., Chen, H.: Measuring Privacy Risk in Online Social Networks. In: Proceedings of W2SP 2009: Web 2.0 Security and Privacy (2009)
11. Dey, R., Ding, Y., Ross, K.W.: The High-School Profiling Attack: How Online Privacy Laws Can Actually Increase Minors Risk. Tech. rep. (2012), http://cis.poly.edu/~ross/papers/HighSchool.pdf
12. Dey, R., Tang, C., Ross, K.W., Saxena, N.: Estimating age privacy leakage in online social networks. In: Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA. pp. 2836–2840 (2012)
13. He, J., Chu, W.W., Liu, Z.V.: Inferring privacy information from social networks. In: Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics. pp. 154–165 (2006)
14. Jernigan, C., Mistree, B.F.T.: Gaydar: Facebook Friendships Expose Sexual Orientation. First Monday 14(10) (2009)

15. Lenhart, A., Madden, M., Smith, A., Purcell, K., Zickuhr, K., Rainie, L.: Teens, kindness and cruelty on social network sites (Pew Internet, November 9, 2011), http://pewinternet.org/~/media/Files/Reports/2011/PIP_Teens_Kindness_Cruelty_SNS_Report_Nov_2011_FINAL_110711.pdf
16. Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: inferring user profiles in online social networks. In: Proceedings of the third ACM International Conference on Web Search and Data Mining. pp. 251–260 (2010)
17. Tang, C., Ross, K., Saxena, N., Chen, R.: What's in a name: a study of names, gender inference, and gender behavior in Facebook. In: Proceedings of the 16th International Conference on Database Systems for Advanced Applications. pp. 344–356 (2011)
18. Thomas, K., Grier, C., Nicol, D.M.: Unfriendly: multi-party privacy risks in social networks. In: Proceedings of the 10th International Conference on Privacy enhancing technologies. pp. 236–252 (2010)
19. Xu, W., Zhou, X., Li, L.: Inferring Privacy Information via Social Relations. In: 24th International Conference on Data Engineering Workshop. pp. 154–165 (2008)
20. Zheleva, E., Getoor, L.: To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th International Conference on World Wide Web. pp. 531–540 (2009)