Profiling High-School Students with Facebook: How Online Privacy Laws Can Actually Increase Minors' Risk

Ratan Dey
Polytechnic Institute of New York University
Brooklyn, New York
ratan@cis.poly.edu

Yuan Ding
Polytechnic Institute of New York University
Brooklyn, New York
dingyuan1987@gmail.com

Keith W. Ross
Polytechnic Institute of New York University
Brooklyn, New York
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 displaying only minimal information in registered minors' public profiles, not listing minors when searching for users by high school or city, and banning young children from joining altogether.

In this paper we show how an attacker can circumvent these precautions. We develop efficient crawling and data mining methodologies to discover and profile most of the high school students in a targeted high school. In particular, using Facebook and for a given target high school, the methodology 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. Such profiles can be used 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 approach. 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. Our analysis strongly suggests there would be significantly less privacy leakage if Facebook did not have age restrictions.

Categories and Subject Descriptors

E.m [**Data**]: Miscellaneous

Keywords

Facebook; COPPA; Privacy; Minor; Policy; High School

IMC'13, October 23–25, 2013, Barcelona, Spain.

ACM 978-1-4503-1953-9/13/10 ...$15.00.
http://dx.doi.org/10.1145/2504730.2504733.

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) [6] is designed to protect the privacy of children under 13 years of age. It details what a website operator must include in a privacy policy, how to seek verifiable consent from a parent, and what responsibilities an online service provider has to protect children's privacy and safety, including restrictions on the marketing to children under 13. Many consumer, privacy and child advocacy groups continue to actively lobby governments to provide better privacy protection for minors [7]. The US Congress is currently considering new bills to strengthen online safeguards for children and teens [10, 12].

Online Social Networks (OSNs) additionally take measures to protect the privacy of minors. Facebook, for example, treats minors and adults with distinctly different policies [5]. 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 an attacker (third party) can circumvent these precautions to discover and profile most of the high-school students in a targeted geographical area. In particular, using Facebook and for a given target high school, we construct an efficient methodology 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. For each discovered student, the additional information minimally includes 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 further 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. By profiling all the high schools in a city, a third-party can discover and develop profiles for most of the minors, ages 14-17, in that city. As discussed in Section 2, the third-party 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 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 to obtain complete lists of students attending a given target school directly from OSNs. As of April 2013, and documented in this paper, Facebook takes explicit actions 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 a third party to *directly* use Facebook's search services (including Facebook's Graph Search) to collect the names of the students at any target high school and attempt to profile them. Google+ also takes similar measures to protect the privacy of minors, as described in the Appendix A of this paper.

Ironically, the third-part privacy leakages described in this paper are indirectly exacerbated by part of the COPPA law, which was designed to protect minors' privacy. As part of the COPPA law, children under thirteen are required to obtain verifiable parental consent before joining an online service. Given economic costs, social concerns, and technical issues, most online services — including Facebook and Google+ — choose to avoid this COPPA requirement 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 methodology 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 [19]. 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. A smaller fraction of these "lying" minors will indicate in their public profiles that they currently attend the target high school.

4. By identifying the minors returned by the search results, crawling their friend lists 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, a component of the COPPA law and the fact that OSNs do not verify the age of its users have together inadvertently set the stage for widespread discovery and inference of minors' private information.

To demonstrate the feasibility of high-school profiling, we applied the methodology to one small private high school and two relatively large public high schools, located in different geographical regions in the USA. 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 smaller 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. For the larger high schools, we do not have complete ground-truth information to evaluate the approach. Instead, we obtain limited ground-truth information for a small set of students by scraping Facebook, and develop an evaluation methodology based on the limited ground-truth information.

We also develop a methodology to estimate how much privacy leakage would occur in a world where the OSN lets all users join, independent of age, and compare the estimate to the extent of leakage in our current world with bans on children under 13. Our results suggest that a third-party not only can discover more minors, but can also build more extensive profiles than what would be the case in a world without age restrictions. Thus, in terms of third-party privacy intrusion, this component of COPPA actually puts minors at greater risk than they would be if the law had never been enacted. We emphasize, however, that COPPA has many components, and does help to protect the privacy intrusion from first party services (such as OSNs). 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.

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 methodology described here.

This paper is organized as follows. In Section 2 we briefly outline the consequential threats resulting from the high-school profiling. In Section 3 we provide definitions of terms used throughout the paper, discuss Facebook's policy for minors, and discuss ethical issues associated with the collection and analysis of our data. We present the details of the high-school profiling in Section 4. We evaluate the success of the attack for three high schools in Section 5. In Section 6 we investigate to what degree minors can be profiled using the attack. In Section 7 we estimate the extent of leakage that would occur in a world without COPPA. In Section 8 we consider one promising countermeasure, namely, disabling reverse lookup. In Section 9 we discuss relevant prior work, and in Section 10 we summarize our conclusions.

2. CONSEQUENTIAL THREATS

Suppose a third party, 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 collection of these pro-

[1]A public Facebook profile for a minor at most contains name, profile picture and gender. Thus the constructed profiles addition-

files the *high-school profiles*. Moreover, suppose the third party 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 [13, 11]. 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.

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" [9]). A pedophile use the approach himself, using the acquired profiles to prospect for victims.

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.

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" [2], 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. Facebook recently introduced "Graph Search" which provides a natural way to search for content, people, pages, and so on. In particular, for a given target high school HS1, a third party can now search for users who "study at HS1 in/after/before 2013" or for "current students at HS1 who live in city1" and many other combinations. As with the Find Friends Portal, using the ground-truth data, our experiments found that *Facebook does not return registered minors when a stranger searches with Graph Search*.

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. Google+ takes similar measures, as described in the Appendix A.

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 the methodology in

ally contain current city, high school name, graduation year, high-school friends, inferred birth year, and for many students much more information.

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		✓		✓

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 three high schools, the crawling was not particularly aggressive and did not 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 methodology 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 schools involved. The data is encrypted, password protected and lies behind a firewall. In the future, we will be destroying and whitewashing all the collected data.

4. HIGH SCHOOL PROFILING

We now describe our basic version of the high-school profiling methodology. The third party 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 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 third party to be an OSN friend, or a friend-of-a-friend, of any of the students in M, that is, the third party may be a stranger to all the students in the high school. With sufficient computational resources, the methodology could be applied to hundreds or even thousands of high schools.

4.1 The Basic Methodology: 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 methodology in its most basic form is as follows.

1. The third party 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 third party 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 third party 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 third party 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 third party 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:

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

Clearly each $G_i(u) \subseteq F(u)$. Note that to obtain the $G_i(u)$'s, the third party *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 third party 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 third party calculates

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

6. The third party rank orders the users in K according to their $x(u)$ values, from highest to lowest. The third party 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 third party 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 third party 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 third party has a set of OSN users H believed to be students at the target high school. The third party has also classified all the students in H by graduation class year. For each student, by knowing the high school, the third party knows the current city; by knowing both the student's last name and current city, the third party can often determine the student's home address from voter registration records. The third party can also estimate birth year from the graduation year. As described in the Section 6, the third party can further determine each student's high-school Facebook friends (even though this is also not available in the public profiles), and can obtain significantly more information for about half of the high-school students.

Note that the methodology relies on the ability to obtain a small set of core users, that is, finding a set of users for whom the third party 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 methodology is passive, that is, without attempting to establish friend links with any of the students.

4.2 Performance of Methodology

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 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 we 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 each of the three test high schools.

4.3 Enhanced methodology

We now describe an important enhancement, which requires a relatively small amount of additional crawling. In the *enhanced methodology*, 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, use interaction graphs [26], or consider the evolution of the activity between users [25] to optimize the results. As the purpose of our research is to demonstrate the feasibility of the methodology 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 methodologies, we also examine filtering out some of the candidate users. This filtering variation, as with the enhanced methodology, requires the third party to download the public profiles of the first $(1+\epsilon)t$ users in the candidate set. After downloading these profiles, the third party 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 Measurement Effort

Most OSNs employ anti-crawling techniques to protect the data of their members and maintain the performance of their sites. Typically, if a member behaves suspiciously (for example, if the member tries to access many user profiles in a short time), the member's account will be temporarily, or permanently, disabled. Therefore the measurement effort is an important consideration.

For the Basic Methodology, the measurement 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 Methodology, 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 methodologies.

5. RESULTS FOR THREE HIGH SCHOOLS

5.1 Data Sets

In order to estimate the success of the methodologies, we applied them to three US high schools, which we refer to as HS1, HS2, and

HS3. We collected the data for HS1, HS2, and HS3 in March 2012, June 2012, and June 2012, respectively. 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 methodologies. 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 methodology 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 methodology 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.

HS2 is a public suburban high school on the East Coast with a much larger student body of approximately 1,500 students. The school has diverse economic and racial demographics, with about 15% of the students being African-American, 10% Asian, and 10% Hispanic. HS3 is a public high school in a small city in the Midwest, also with approximately 1,500 students. Although neither for HS2 nor HS3 were we able to obtain complete ground-truth information, we develop a methodology to evaluate the approach based on partial ground-truth information mined from Facebook.

5.2 Initial Seed Set

We obtained initial seed sets from Facebook's Find Friend portal, using two accounts for the smaller HS1 and four accounts for each of the larger high schools HS2 and HS3. Table 2 provides a summary of the data collected for the three schools. As shown in Table 2, for HS1, HS2, and HS3, we found 18, 70, and 46 core users (with friend lists) and 6,282, 14,317, and 11,736 candidates, respectively. For the enhanced methodology, we obtained 22, 152, and 178 (extended) core users for each of three high schools. For each high school, the number of core users is roughly 5% of the number of students in the school.

5.3 Measurement Effort

Table 3 summarizes the approximate measurement effort required to collect the data sets for the three high schools in Table 3. 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 methodology, and about five times the number of students in the target high school for the enhanced methodology.

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 methodologies, 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 methodology). 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 methodology with filtering gives the best results; for the top 500 case, the enhanced methodology without filtering gives somewhat better results than the enhanced methodology with filtering.

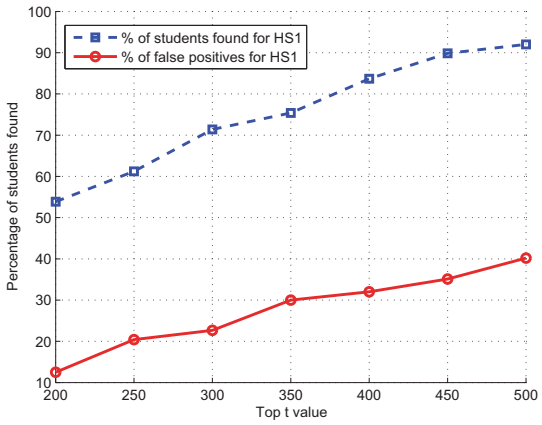

Figure 1: Overall performance of enhanced methodology for HS1

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 third party decides to use the enhanced methodology with filtering, and considers the top 400 users as students in HS1. Examining the column for 400 students in Table 4, we see that with this choice of threshold, 272 (84%) of the 325 students are included in the set. So with this threshold, the third party 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 third party wants to reduce the false positives, the third party 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 third party 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 methodology 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

Table 2: Seeds, core users, and candidates for the three high schools

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
HS2	1,500 (approx)	N/A	1,559	70	14,317	152
HS3	1,500 (approx)	N/A	1,532	46	11,736	178

Table 3: Measurement effort

	Facebook accounts used	HTTP requests for seeds	Profile pages	Requests for friend lists	Total requests for basic methodology	Total requests for enhanced methodology
HS1	2	34	352	360	746	1,576
HS2	4	101	1,559	1,400	3,060	7,700
HS3	4	90	1,532	920	2,542	8,182

if they are former students or not (although most likely are since they have a large number of friends in HS1).

5.5 Results for HS2 and HS3

For each of the two large public high schools, in order to evaluate the performance of the basic and enhanced methodologies, we collected a first set of seeds with four Facebook accounts and a second set of seeds with an additional four accounts. We use the first set of seeds in the methodologies; we use the second set for evaluation. Specifically, for HS2, from the second set of seeds we obtained 43 users who specify they are currently at HS2 and are not included in the first set of seeds. To evaluate the methodologies, we check to see which of these 43 test users are in our inferred set, and which of those are classified in the correct graduation year. For HS3, we obtained 47 such test users.

We now describe our methodology for evaluating performance with limited ground-truth information. Let x_t be the number of test users found in the top t. For the basic methodology, the set of actual high-school students discovered for a threshold t has two disjoint groups: (i) the core users; and (ii) the non-core high-school students who are discovered. The fraction of non-core high-school students who are discovered is given by $p =$ non-core HS students discovered/non-core HS students. This fraction can be estimated by $(x_t/\#$ test users$)$. Thus, an estimate of the number of students in the high school found with a threshold t is:

$$\#\text{ of core users} + \frac{x_t}{\#\text{ test users}} \times (\text{HS size - }\#\text{ of core users})$$

(For the enhanced methodology we replace the number of core users with the number of extended core users.) To estimate the percentage of high-school students found for threshold t, we divide the above by the high-school size. To estimate the number of false positives for a threshold t, we use

$$t - \frac{x_t}{\#\text{ test users}} \times (\text{HS size - }\#\text{ of core users}),$$

since the false positives are those users among the top-t minus the expected number of students found in the the top-t (excluding the core users). To estimate the percentage of false positives for a threshold t, we divide the above by $\#$ of core users $+ t$. We show these estimates for the enhanced methodology with filtering in Figure 2. For example, for HS2, the top 1,652 users ($t = 1500$ plus the extended core users), the third party can obtain 85% of all the HS2 students with 22% false positives in the set of 1,652 users. On the whole, the results in Figure 2 for HS2 and HS3 are similar to those in Figure 1 for HS1.

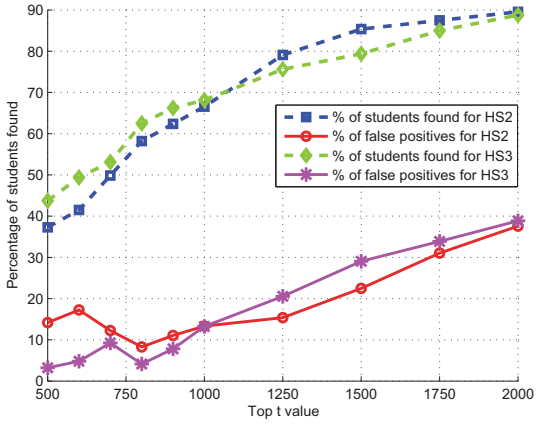

Figure 2: Overall performance of enhanced methodology for HS2 and HS3

5.6 Summary of Results

As discussed in Section 2, 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 a third party, with relatively little crawling effort, can discover the majority of the students at the target high school. For example, we obtained 83%, 85% and 79% of all the students in HS1, HS2, and HS3, respectively, with false-positive rates of 32%, 22% and 29%. Moreover, for each high school student in the list, the third party can determine the student's graduation year with a high-level of accuracy. A third party 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, as described in the Appendix, our preliminary analysis indicates that the attack applies to Google+ as well.

6. EXTENDING THE PROFILES

Recall that when a stranger visits the Facebook page of a minor, in the philosophy of Facebook's current privacy policy, the stranger should see minimal information, which at most includes the minor's full name, profile photo, gender, and networks. However, due to statistical inference and many minors registering as adults, a third party can leverage OSNs to significantly extend the profiles. We now quantify the amount of additional profile information is

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

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

Table 5: Extending the profile for minors registered as adults

	HS1	HS2	HS3
# minors registered as adults	112	700	795
entire friend list public	73%	77%	87%
avg # of friends for users who make friend list public	405	960	908
public search enabled	71%	80%	86%
Message link	89%	86%	91%
relationship info	15%	26%	34%
interested in	13%	20%	33%
birthday	9%	4%	6%
average # of photos shared	19	51	57

readily available to the third party. We do this separately for two classes of minors: those who are registered minors, and those who are registered adults. Again, we do this for the users classified in the first three years of high school (since some of the fourth year students are adults).

6.1 Extending Profiles of Registered Minors

Suppose the attacker applies the methodologies in this paper to all the high schools in a city. Then for most of the high-school students in that city, in addition to the minimal information (full name, profile photo, gender, and occasionally networks), we have shown that a third party can infer current high school and graduation year. From the inferred high school, the third party can also infer hometown and current city, and from graduation year the third party can further estimate the minor's birth year.

Moreover, the third party can use "reverse lookup" to obtain partial friend lists of all the discovered high-school students. Specifically, after obtaining a set H of (likely) current students at the target high school, the third party downloads the friend lists for all users in H whose friend lists are publicly available. A student in H, say Alice, without a public friend list, will typically be in the friend lists of other students in H who make their friend lists public. With this information, the third party can determine at least a portion of Alice's friends. We applied reverse lookup to the registered minors in HS1 and the inferred registered minors in HS2 and HS3. For each of these minors we were able to create friend lists. In particular, we found on average 38, 141, 129 friends per registered minor in HS1, HS2, and HS3. (On average HS1 students have fewer high-school friends since it is a much smaller high school.)

In summary, for each registered minor discovered through high-school profiling, the third party can create a profile with full name, profile photo, gender, a large subset of friends, high school, graduation year, hometown, and an estimate of birth year. This is substantially more information than what Facebook makes publicly available, no matter how the registered minor configures her privacy settings. As discussed in Section 2, this information can serve as a base for creating more comprehensive profiles, by matching the information with public records and other online sources.

We briefly remark that reverse lookup will only obtain a subset of the registered minor's high school friends. In particular, it will not identify friendship relations between two registered minors, since the friend lists of both minors is unavailable to a stranger. Although not explored in this paper, it is possible to infer a friendship link between two registered minors by examining the number of common friends they have (with the common friends determined by reverse lookup). For example, for two registered minors Alice and Bob, let F_A and F_B, respectively, be their friends obtained by reverse lookup. We can then compute the Jacquard index between Alice and Bob:

$$J(A,B) = \frac{F_A \cap F_B}{F_A \cup F_B}$$

If $J(A,B)$ is high, Alice and Bob are very likely to be friends. In this way, we can also find the hidden friendship links among registered minors, and more generally between any two users for which their friend lists are not publicly available.

6.2 Extending Profiles of Minors Registered as Adults

We now examine here the additional information that a third party can obtain for a minor who is registered as an adult. In this case, significantly more information is often directly available, depending on how the user has configured her privacy settings. Possible additional information that can be collected by a stranger — beyond the information a third party can obtain and infer for a registered minor — includes shared photos, photo tags, full friend list, relationship info, interested in, wall postings, likes, favorites, political views, religious views, videos, links, website, birthday and contact info (such as personal email address, IM screen name, address, phone number). For HS1, HS2, and HS3 we determined how much additional information is available for some of these attributes. The results are shown in Table 5.

A stranger can obtain a significant amount of information about a registered minor, as we saw in the previous subsection. But the stranger can often obtain even more information for a minor registered as an adult. For example, as shown in Table 5, a third party can obtain on average over 50 shared photos in the two large high schools, and has access to the "Message" link for more than 86% of the minors registered as adults in all three high schools.

7. HOW ONLINE PRIVACY LAWS CAN ACTUALLY INCREASE MINORS' RISK

We now estimate how much privacy leakage there would be for minors in a world without the COPPA law. Without COPPA, there would be no age restrictions so that children under 13 could create accounts without having to lie about their ages. Although some children might still lie as a joke, the frequency of such occurrences would likely be much smaller. In this without-COPPA analysis, we will assume all users register with their actual birth dates. We also assume the OSN maintains the same privacy policy for minors as it does today — in particular, (i) when searching for users who attended a target high school or live in a target city, the OSN does not return registered minors; (ii) the OSN displays only minimal information on a minor's public profile page.

We now address two questions in the context of a world without COPPA. First, for a given target high school, can a third party still find a set of OSN users such that (a) most of the students in the set attend the target high school (low false positive rate), and (b) the list contains most of students attending the high school (high coverage)? Second, for the students in the set, can a third party create profiles that go significantly beyond the minimal public profiles? As we have seen, in a world with the COPPA law, the answer to both of these questions is yes. But to what extent is it also true in a world without COPPA?

In high-school profiling (with COPPA), a key component is discovering a set of "core" users who currently attend the target high school and have friend lists. Finding such a core set is facilitated by the fact that some minors are registered adults since they lied about their ages when creating accounts (when they were under 13). Because Facebook treats these minors as adults, not only are they easier to find, but their friend lists are often public, thereby also making their classmates easier to find. *However, in our modified world without COPPA, such core users become more difficult to locate, since no minors would appear in search results.*

7.1 A Natural Approach in a COPPA-less World

Nevertheless, even in a world where everyone registers with their actual birth date, it would still be possible to locate some candidate minors. Here we discuss one natural mechanism to do this. Because many young adults (18-20 years old) will likely have friends who are a few years younger than them, if the third party can find the young adults who recently graduated from the target high school and collect their friends, the third party could create a list that contains minors in the target high school. Specifically, we suppose the third party takes the following natural approach:

1. Obtain a set of users who are adults and have recently graduated from the target high school (or are adults in the last year of high school). Call the subset of these users who make their friend lists publicly available the core users.
2. Obtain the public profiles of all the friends of all the core users. Call the union of all these friends the candidate set. Most likely, the candidate set would contain many minors in the target high school.
3. All the minors (and some adults) will have minimal public profiles. To narrow down the candidate set, filter out all users who do not have minimal public profiles.
4. To further narrow down the candidate set, additionally filter out all users who have fewer than n friends in the core set. The third party then considers this filtered set H as the minors in the target high school.

7.2 Apples-to-Apples Comparison

We now evaluate how successful this approach would be for a world without COPPA, and compare the success rates with those obtained in Section 5 for the world with COPPA. One of the challenges in this evaluation is that we are not able to collect data for the world without COPPA. If we apply the above heuristic to our existing data, then the above heuristic will not find any of the minors registered as adults; however, if we were applying the heuristic to actual without-COPPA data, more minors would be found since more minors would have minimal public profiles. In order to make a mostly apples-to-apples comparison, we therefore compare the number of minimal profile students obtained by the above heuristic for the without-COPPA case with the number of minimal profile students obtained in Section 5 for the with-COPPA case. We make this comparison for HS1, for which we have ground truth information. Recall HS1 has 325 students with Facebook accounts. Of these 325 students, 148 have minimal public profiles (22, 47, 45, and 34 students in 2012, 2013, 2014, and 2015, respectively). We now investigate how many of these 148 students are discovered in the two cases.

First consider the without-COPPA case. We will use students who have graduated in 2010 and 2011 to discover students graduating in 2012-2015, using the same data we used in Section 5, which was collected in March 2012. Specifically, using the HS1 data obtained from the high-school search, we find 52 users who indicate their graduation year as either 2010 or 2011 for HS1 and publicly make available their friend lists. These 52 users become our "core set". We then apply the heuristic above. For $n = 1$, we have $|H| = 4,572$ filtered candidates and among these candidates, 92 of them are in the ground truth data set. Thus, $n = 1$ gives 62% of the minimal-profile ground-truth students with $4,480$ false positives. The results for $n = 1, 2, 3$ are shown in Figure 3.

Now consider the with-COPPA case. For any t value, we need to determine the number of minimal profile students found and the number of false positives. Let M_t be the set of top-t users from Section 5 who have minimal profiles. The number of minimal profile students, z_t, is the number of students from M_t who are in the ground-truth set. The number of false positives is given by $|M_t| - z_t$. For $t = 300$, 165 users have minimal profiles of which 95 are in the ground truth data set. Thus, $t = 300$ gives 64% of the minimal-profile ground-truth students with 70 false positives. The results for $t = 300, 400, 500$ are shown in Figure 3.

As shown in Figure 3, we see that without-COPPA, for obtaining the same number of minors as with-COPPA, the third party has *many more false positives*. For example, with-COPPA gives 64% of the HS1 students with 70 false positives; without-COPPA gives 62% of the HS1 students with $4,480$ false positives. Similarly, for the same number of false-positives, without-COPPA *finds significantly fewer students*. For example, with-COPPA gives 91% of the HS1 students with 170 false positives; without-COPPA only gives 29% of the HS1 students with 186 false positives.

7.3 Profile Creation in a COPPA-less World

Having shown that, without COPPA, the third party finds significantly fewer students for the same number of false positives, we now address the second question, namely, for the students in the guess set H, can the third party create profiles that go significantly beyond the minimal public profiles that Facebook displays for registered minors? Recall that the minimal profile at most contains name, profile picture and gender. In the without-COPPA case, using the heuristic above, the third party would be able to augment this minimal profile with high school using the target high school, although his level of confidence would be significantly less

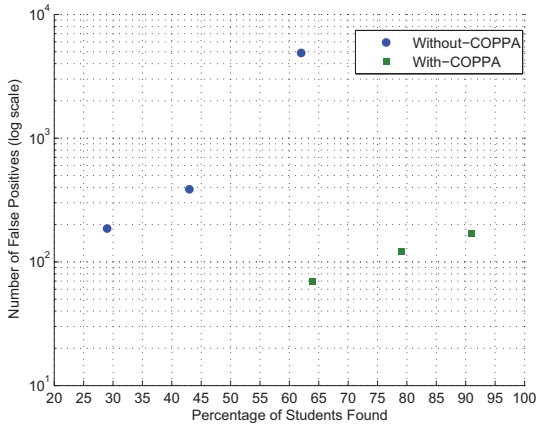

Figure 3: For HS1, False Positives Found With-COPPA vs. Without-COPPA

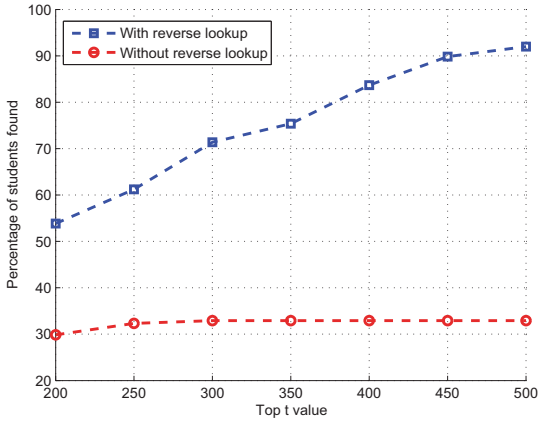

Figure 4: Percentage of HS1 students found with and without reverse lookup

because of the large number of false positives. Moreover,the third party would not be able to easily determine the student's graduation year, and the third party would not be able to create a friend list that includes students in the same year, as is the case for with-COPPA. Thus, in the without-COPPA case, using the techniques in this paper, the third party would not be able to construct a profile beyond the minimal profile plus (a low-confidence guess of) the high school. In the with-COPPA case, a third party can obtain additional profile information for all minors: specifically, graduation year and high-school friend lists for registered minors; graduation year, high-school friend lists and often much more information (e.g., complete friend lists and shared photos) for minors registered as adults. Furthermore, for the without-COPPA case, a third party would *not* be able to send Facebook messages to any of the minors (unlike the with-COPPA case).

In summary, we can conclude that an indirect result of COPPA is that a third party not only can discover more minors, but can also build much more extensive profiles than what would be the case in a world without COPPA.

8. COUNTERMEASURES

In an ideal world, policymakers would enact laws and OSNs would take measures so that (i) it would be difficult for third parties to discover minors in targeted geographical regions and construct detailed profiles of those minors; while at the same time, (ii) providing a highly usable service for minors and adults alike. Designing and evaluating all combinations of possible laws and measures is a major research problem on its own. In this paper we examine just one promising countermeasure — namely, disabling reverse lookup — and quantify the reduction in privacy leakage.

With reverse lookup disabled, if a user's friend list is hidden from strangers (either because the user has configured his friend list as such or because the user is a registered minor), then that user would not be visible to strangers in any other user's friend list. If the OSN takes this countermeasure, users with hidden friendlist will not be found using reverse lookup, thereby reducing privacy leakages. To evaluate the effectiveness of this measure, we estimate its impact on HS1, for which we have ground-truth information. Out of 325 ground-truth users of HS1, 112 of them make their friend list publicly available. For the remaining users, the friend list is not publicly available to strangers, and these users, by assumption, cannot be found via reverse lookup. So in evaluating this countermeasure, we remove these users from the candidate set and then proceed with the methodology steps.

Figure 4 shows the results of the enhanced methodology with filtering for HS1 with and without reverse lookup. From the results, it is clear that performance of high-school profiling methodology decreases significantly. For example, by removing reverse lookup, the percentage of students found in the top-500 decreases from the 92% to 33%.

9. RELATED WORK

There are several earlier studies on the usage of OSNs by minors, both for minors (under 18) and underage users (under 13). In 2010, Pew Research released a report stating that 73% of online American teens ages 12 to 17 used an OSN website [20]. In 2011, Pew released another report, with collaborators at Cable in the Classroom and the Family Online Safety Institute, where it was found that 44% of online teens admit to having lied about their age so they could access a Web site or sign up for an online account [19]. Similar results have been reported for European teens [21]. Boyd et al. [15] provided survey data showing that many parents know their underage children are on Facebook in violation of the site's restrictions, and that they are often complicit in helping their children join the site. More recently, Pew Research found that teens and adults have no significant variations for their privacy settings [22]. These reports collectively provide great insight into behavioral characteristics of minors and their parents. None of these reports, however, address automated discovery and profiling of minors. Our measurement work shows that because minors often lie to circumvent the age restriction, they put themselves and their non-lying high-school friends at risk for a variety of potential online and in-person abuses.

There is substantial previous work on using statistical inference to infer private information about OSN users. Zheleva and Getoor [28] 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 [17] demonstrated a method for accurately predicting the sexual orientation of Facebook users by analyzing friendship associations. Other papers have also examined inferring private information from social networks. Thomas et al. examined scenarios where conflicting privacy settings between friends will

reveal information that at least one user intending to remain private [24]. Becker and Chen [14] 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. [16] examined a large dataset and develop a methodology to estimate ages of Facebook users. Mislove et al. [23] 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. Recently Kosinski et al. [18] showed how Facebook likes can be used to predict a range of highly sensitive personal attributes automatically.

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 the third party. This is also the first paper to measure the additional third-party privacy disclosure risk for minors due to the enactment of the COPPA law.

10. CONCLUSION

In this measurement study we have shown how a privacy law for protecting children's privacy (along with the fact that OSNs do not verify users' ages) 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 a methodology to profile the current students in the high school. The methodology 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. As described in Section 6, the profiles of about half students (those who have lied about their ages) also include a varying amount of additional information, including shared photos and wall postings.

We estimated how much privacy leakage would occur in a world without an age restriction and compared the estimate to our measured results for current world with the age restriction. Our results show that a third party not only can discover more minors, but can also build more extensive profiles than what would be the case in a world without an age restriction. Thus, in terms of third-party privacy intrusion, a component of the COPPA law actually puts minors at greater risk than they would be if the law had never been enacted.

Although the age restriction in 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.

11. 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.

12. REFERENCES

[1] Age requirements on Google Accounts. http://support.google.com/accounts/bin/answer.py?hl=en&answer=1350409, accessed May 7, 2013.

[2] Find Friends Portal. https://www.facebook.com/find-friends/browser/, accessed May 7, 2013.

[3] Google+ Search by School Portal. https://plus.google.com/circles/school, accessed May 7, 2013.

[4] Google+ Teen Safety Guide - Features for teens. http://support.google.com/plus/bin/answer.py?hl=en&answer=2409072, accessed May 7, 2013.

[5] How Does Privacy Work for Minors? http://www.facebook.com/help/?page=214189648617074, accessed May 7, 2013.

[6] Children's Online Privacy Protection Act, 1998. http://www.ftc.gov/ogc/coppa1.htm, accessed May 7, 2013.

[7] Groups Make Recommendations for Kids' Facebook, Adweek, June 18, 2012. http://www.adweek.com/news/technology/groups-make-recommendations-kids-facebook-141195.

[8] Creating a Google Plus Account Now Requires You to Enter Your Birthday, August 27, 2011. http://techie-buzz.com/social-networking/google-age-restrictions.html.

[9] 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.

[10] Do Not Track Kids Act of 2011, May 13, 2011. http://www.gpo.gov/fdsys/pkg/BILLS-112hr1895ih/pdf/BILLS-112hr1895ih.pdf.

[11] 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.

[12] Update Urged on Children's 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.

[13] On the Web, Children Face Intensive Tracking, The Wall Street Journal, September 17, 2010.

[14] J. Becker and H. Chen. Measuring Privacy Risk in Online Social Networks. In *Proceedings of W2SP 2009: Web 2.0 Security and Privacy*, 2009.

[15] D. Boyd, E. Hargittai, J. Schultz, and J. Palfrey. Why Parents Help Their Children Lie to Facebook: Unintended Consequences of the 'Children's Online Privacy Protection Act'. *First Monday*, 16(11), November 7, 2011.

[16] R. Dey, C. Tang, K. W. Ross, and N. Saxena. Estimating age privacy leakage in online social networks. In *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA*, pages 2836–2840, 2012.

[17] C. Jernigan and B. F. T. Mistree. Gaydar: Facebook Friendships Expose Sexual Orientation. *First Monday*, 14(10), 2009.

[18] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, Apr. 2013.

Table 6: Google+: 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, Profile Picture	✓	✓	✓	✓
Gender, Employment, HS, Hometown, Current City		✓	✓	✓
Home and Work Phone			✓	✓
Relationship, Looking			✓	✓
Birthday			✓	✓
Photos			✓	✓
Public Search		✓	✓	✓
In Your Circles		✓	✓	✓
Have You in Circles		✓	✓	✓

[19] A. Lenhart, M. Madden, A. Smith, K. Purcell, K. Zickuhr, and L. Rainie. 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.

[20] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr. Social media and mobile Internet use among teens and young adults, Pew Internet, February 3, 2010. http://www.pewinternet.org/~/media//Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf.

[21] S. Livingstone, K. Olafsson, and E. Staksrud. "Social Networking, Age and Privacy", EU Kids Online, 2010. http://www2.lse.ac.uk/media@lse/research/EUKidsOnline/ShortSNS.pdf.

[22] M. Madden. Privacy management on social media sites, Pew Internet, February 24, 2012. http://www.pewinternet.org/~/media/Files/Reports/2012/PIP_Privacy_management_on_social_media_sites_022412.pdf.

[23] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. 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*, pages 251–260, 2010.

[24] K. Thomas, C. Grier, and D. M. Nicol. Unfriendly: multi-party privacy risks in social networks. In *Proceedings of the 10th International Conference on Privacy enhancing technologies*, pages 236–252, 2010.

[25] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online Social Networks*, pages 37–42, New York, NY, USA, 2009. ACM.

[26] C. Wilson, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. Beyond social graphs: User interactions in online social networks and their implications. *ACM Trans. Web*, 6(4):17:1–17:31, Nov. 2012.

[27] X. Zhao, A. Sala, C. Wilson, X. Wang, S. Gaito, H. Zheng, and B. Y. Zhao. Multi-scale dynamics in a massive online social network. In *Proceedings of the 2012 ACM conference on Internet Measurement Conference*, pages 171–184, 2012.

[28] E. Zheleva and L. Getoor. 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*, pages 531–540, 2009.

APPENDIX

A. GOOGLE+ AND REGISTERED MINORS

Although the focus of this paper is on Facebook, we briefly mention here that Google+ is also susceptible to high-school student profiling. Like Facebook, to create a Google+ account, the user must register as 13 years or older [1, 8]. Google+ also provides a mechanism for searching for users associated with a high school [3].

Unlike Facebook, Google+, which uses circles, has asymmetric friendship links. For example, for Alice there is one set of users in her circle; and there is a second set of users who include her in their circles. Google+ also provides various safety guidelines for teens [4]. Google+ registered minors also have different default privacy settings than do registered adults, as shown in Table 6. It would also be of interest to examine the privacy settings of minors in RenRen, a large Chinese social network [27].