COP: A Step Toward Children Online Privacy

Wei Xu, Sencun Zhu, and Heng Xu

Pennsylvania State University
{wxx104,szhu}@cse.psu.edu, hxu@ist.psu.edu

Abstract. We propose COP, a client-side system for protecting children's online privacy and empowering parental control over children's information disclosure with little manual effort. COP is compliant with the Children's Online Privacy Protection Act (COPPA) and it implements acquisition of parental consent before any private information submitted online by children, e.g., registration to a Web service. Instead of restricting access to certain Web services or blocking sensitive data from websites, COP employs perturbation techniques over personal data with the goal of concealing the sensitive information while providing certain usability of the data to the websites. We address several challenges in the implementation of COP, e.g., perturbation of different types of data, parsing user input and retaining transparency to children without obstructing their normal Web surfing activities. We apply COP in registrations to 23 most popular websites. The results indicate COP's effectiveness as a privacy protection tool. We also discuss some potential security attacks against COP's design and provide our countermeasures.

1 Introduction

1.1 Background

The Internet has evolved into a platform for communicating, exchanging information, carrying out commerce, streaming media and social networking among many uses. Children, being part of the Internet users, have been given unprecedented opportunities to communicate online with one another. Like every other Internet user, children's online safety has been put at risk. Exposing themselves to the virtual world has caused great concerns, especially considering the growing cases of children's online abuses [1], online predators [2], online children pornography [3] and other such matters. Moreover, many operators of websites are also interested in collecting children's personal information such as their names, ages, email addresses and phone numbers for commercial purposes [4, 5]. Release of such data jeopardizes children's privacy. According to a study [6] over U.S. census data and its follow-up work [7], the combination of gender, 5-digit ZIP code and full date of birth can uniquely or nearly uniquely identify 63% of the US population. Hence, it is not hard to imagine how much private information will be given out once it falls into the hands of malicious parties.

Compared to adults, children are more vulnerable to threats like re-identification because they are not mature enough to realize the harm of privacy divulgence. A study

shows that almost half of teens (47%) do not even worry about others using their personal information [8]. Besides, children are not sophisticated enough to protect themselves against such information leakage. For example, by exploiting the naivety of children, some websites lure children to online prizes in exchange for their personal information [9]. Without rational judgment of the websites, children intend to submit their private information to access certain Web services. Due to these inherent vulnerabilities, children's online privacy protection has become an imminent and challenging task.

To take a step toward children's online privacy, we focus on personal information gathering processes, a representative case of which is online registration. We will design client-side privacy protection mechanisms for registration processes because it is not realistic to assume that website operators will not violate children's privacy in any way. We cannot hope for full parental supervision whenever their children are online since it places an extra burden on the parents who most likely do not have such time. This calls for an automated technical solution to facilitate parental control without bothering them or with little effort.

Our design of COP is compliant with Children's Online Privacy Protection Act (COPPA) [10], which governs the online collecting of personal information from children under the age of 13 and further distribution of such information in the United States. COPPA states that any website directed to children under 13 must post a link to their privacy policy at any place where it collects personal information from children. COPPA also requires the website to obtain verifiable parental consent for the collecting action and any further use and disclosure of children's personal information. The definition of children's personal information in COPPA includes individually identifiable information such as full name, physical address, email address (or other online contact information), telephone number, age, gender, social security number as well as other auxiliary information such as hobbies, preference and information collected through cookies.

1.2 Related Work

Most of the current technical solutions intend to protect users' online privacy in general. We will first introduce several such solutions and then focus on techniques for protecting children's privacy.

Cookies, a unique identifier that can be used for retrieving records from the databases, authenticating users and tracking users' activities, were seen as a major threat to users' online privacy [11]. COPPA recognizes cookies as privacy-invasive and disallows operators from collecting cookies that can be linked to a child. As a countermeasure, most Web browsers adopt cookie control to give users the option to disallow cookies from a website. These cookie blocking features are effective but they only address a very small portion of COPPA's requirements because these solutions can not prevent websites from explicitly collecting personal information from children under the age of 13.

Anonymizer [12] is a solution that protects user's privacy by providing a way for anonymous Web surfing. It redirects all Web traffic through intermediary proxy servers to hide the user's IP address. The Anonymizer serves as a good privacy solution to fight against phishing and pharming attacks. However, it is not sufficient for protecting children's online privacy because it cannot prevent websites from collecting personal

information during online registration process. In addition, anonymous browsing may encourage children to access objectionable materials once they are aware that they are not being identified as children.

Another popular approach to privacy assurance is through self-regulatory efforts which involve the setting of standards either by the website itself or an industry group and the voluntary adherence to the set standards or policies [13]. Under a self-regulatory approach to regulating children's online privacy, groups like TRUSTe [14] have been active as the third party entities policing children's privacy and promoting trustworthiness to websites through seals of approval. By becoming a member of these private watchdog groups, a website is permitted to post the seal of approval. These seal programs provide a means to guarantee that members abide by a set of clearly identified self-regulatory standards [15]. However, research has shown that, most privacy policies posted online are written in jargon and ambiguous language and thus readability is low [16–18]. For those parents who are not technically inclined or are unaware of COPPA, they usually fail to make informed decisions for their children's information disclosure. In addition, it has been found that few users recognize privacy seals [17]. Thus, we conclude that the self-regulatory approach to children's privacy through privacy policies or privacy seals cannot be adopted as a stand-alone solution but as an additional protection layer complimentary to technical enforcement of COPPA.

A few tools were dedicated to protecting children's online privacy. Parental Online Consent for Kids Electronic Transactions (POCKET) [19] is one of such tools designed to give parents control over children's personal information disclosure. POCKET requires both Web clients and merchant websites to maintain a privacy preference file (PPF) stating their privacy policies. POCKET enables a trusted third party (TTP) server to perform mutual authentication between clients and merchant websites. Children's privacy is preserved by comparing the PPF from a merchant website with a user's own PPF and disclosing only the mutual parts. Setup of a PPF on a client side reflects the parental consent on children's personal information disclosure. First problem with POCKET is its avoidance of HTML forms, which generates inconsistency with users' normal activities. For example, users normally submit their registration information by filling out HTML forms, but when POCKET is enabled, users do not have any control over what information will be submitted on a site by site basis. Another concern is that merchant websites may not always follow their privacy policies and it should be the users' responsibility to protect their own privacy. Moreover, TTP might be the single point of failure although it only works in the system registration phase.

Several other software packages have also been proposed to empower parental control over children's online behavior. One of them is Windows Vista's parental control [20], which is designed to help parents to manage what their children can do on computers. Another is Privo [21], which will suspend children's online registration and ask for parents' opinions if the websites require privacy information. Other tools such as icouldbe [22], Net Nanny [23] and Parental Control Bar [24] are also developed to protect children's online safety by filtering Web contents and blocking functions. The problem with these tools is that they usually filter out outbound user inputs or change them to asterisks to prevent children from divulging any privacy. This restraint on information disclosure during registration may hinder children from gaining access to nor-

mal services. Maintaining the balance between protecting children and retaining their accesses to appropriate Web contents has become the concern of industry practitioners and government agencies [25]. It is also one of COP's design goals.

1.3 Contributions

The main contributions of this paper include:

- We present COP, a light-weight client side solution to protect children's online privacy.
- We show that COP fulfills COPPA's requirements by implementing verifiable parental consent before collecting of children's personal information.
- We demonstrate that COP achieves the balance between children's privacy protection and usability of collected data set by leveraging concepts from privacy preserving data mining (PPDM).
- We evaluate the effectiveness of COP in preserving children's private information during online registration.

The remainder of the paper is organized as follows. Section 2 describes COP's design and implementation. Section 3 evaluates the effectiveness of COP serving as a children's privacy protection solution. Section 4 lists potential attacks from malicious websites and our countermeasures. Section 5 discusses the limitation of COP as well as our future work on COP. Section 6 concludes.

2 Design and Implementation of COP

COP is designed to prevent private information leakage in online registrations and is implemented as a client-side Web browser extension. We rule out the usage of trusted third party in COP to avert single point failure and to give parents maximal control over COP.

To begin with, we define two types of users in the design of COP, namely, parent user and child user. These two users are given different access rights to COP and involved in different phases of COP's operation. In the installation phase, only parent users who have the administrator privilege should install COP in their computers and set up their parent passwords to protect COP from being modified or disabled. After installation is done, COP works as a browser extension and keeps itself transparent to child users. The only information child users can see is the warnings prompted by COP prohibiting registration to certain websites. Since only parent users have access to COP's settings after they verify themselves as parents to COP by inputting their passwords, it is the parents' responsibility to keep the password from being disclosed to others such as their children.

Figure 1 illustrates the system level design of COP. Once COP has been installed, configured and activated properly by a parent user, it starts to monitor the outgoing traffic. When children intend to sign up in a website, the browser sends an HTTP request to the Web server which responds with a Web page containing a registration form. Children will fill out the form with their information such as age, firstname, lastname

and submit the form to the website (Here we assume that children always provide true information and we argue that if they provide false information either intentionally or by mistake, their privacy will not be jeopardized since no real personal information is disclosed). This submission will be intercepted by COP as shown in Figure 1. Then COP will adopt data perturbation on the registration information sent to this website. COP will also notify parents either immediately or at a pre-selected time depending on the preference of parents. Upon receiving the registration information, the website, which is assumed to conform to COPPA, would know the client is under 13, and it should send back its privacy policy. This policy will be intercepted by COP and showed to parents later. Once the registration is done, the name of the website as well as the perturbed information submitted to the website will be logged by COP.

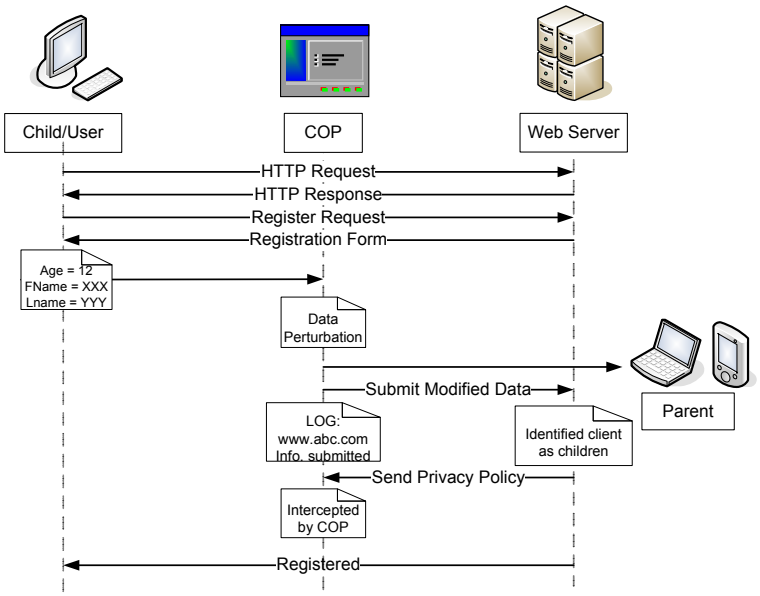

Fig. 1. System design overview

2.1 Privacy Preference

In this section we discuss the implementation of privacy preference in COP. Instead of applying simple rules such as allowing or prohibiting personal data being collected by a website, COP establishes one privacy preference entry for each website. Each preference entry indicates what categories of information can be collected and what cannot be collected by that website. These categories of personal information are defined by COPPA, e.g., 'Name', 'Age', 'Date of Birth', 'Gender', 'Phone Number', 'Address'. A preference entry is a reflection of parental consent on collecting of their children's

information by the website in the entry. Some entries are pre-defined; others are automatically generated by COP when that website is visited for the first time. For example, when a user connects to "www.example.com", COP first searches the preference entry list for an entry of www.example.com. If a match is found, COP will apply that entry to the personal information required by www.example.com in registration. If no match is found, a pre-defined default privacy preference will be used for www.example.com. Fully customizable feature of the preference list provides parents with a fine-grained control over disclosure of their children's personal information.

2.2 Privacy Preserving Data Perturbation

One feature distinguishing COP from other similar schemes is that COP treats personal information as a set of privacy metrics instead of one single piece of information. Since COP is a client-side solution, personal information is stored and processed in a distributed fashion. Without the knowledge of the population of a data set, existing privacy measurements such as k-anonymity can not be applied here. To embody the privacy protection from adopting COP in this work as well as to build a foundation for quantitatively analyzing privacy (discussed as future work in Section 5), we leverage the concept of privacy preserving data perturbation to generate data that appear to be genuine instead of random results and substitute these data for user inputs which contain protected personal information.

One advantage of data perturbation over blocking user inputs is the avoidance of failure in registration. Web servers always perform extensive user inputs validation check by either server-side examination or inline script functions. For example, when a registration form expects users to input an email address, it normally will not accept strings like "John" or "test.com". When a credit card number is required, a VISA card indicated by user must at least have a starting digit of "4" and a 16 digits length in order to pass the validation check. In some cases, only a valid credit card number can survive the examination (e.g., Luhn algorithm), not even one false digit is allowed and these validations will keep bothering users until a genuine data is input. To this end, data perturbation provides a solution to pass these checks without providing personal information.

Another reason to use data perturbation is to retain the statistical properties of submitted data when individually collected data records are put together as a data set by the Web server. As long as COPPA is not violated, we should allow websites to analyze collected data for their own purposes and consider users' privacy preserved at the same time.

In practice, challenges arise from introducing data perturbation into COP. First we need to process many different types of data, such as numerical, string and enumeration. Clearly, there is no single algorithm suitable for perturbing all these kinds of data. Second, as mentioned before, there exist constraints on the perturbed data for passing validation checks to be accepted by Web servers. Third, conflicts might appear between perturbed data, for example, the first three digits of a phone number might give differ-

ent geographical information than a ZIP code [1]. In cases where such conflicts impede user registrations, maintaining consistency between various perturbed data is a necessity. Last but not least, some parts of personal information such as email address cannot be automatically generated because a fake email account will not support further communication between a user and a website.

To address these challenges, we borrow some approaches from PPDM, namely additive perturbation [26], multiplicative perturbation [27, 28] and probability distortion [29]. In additive perturbation, noise is added to data in order to mask the attribute values of records. The noise added is sufficiently large so that the individual record values cannot be recovered from the perturbed data. Note that although there are known drawbacks of additive perturbation such as additive noise may be easily filtered out through correlation of the data points within a large data set, it will not be an issue for data processed in COP because a website only has one data entry from each child. There are two basic approaches to perform multiplicative perturbation. We only consider the first one. This method is based on generating random numbers that have a truncated Gaussian distribution with mean equal to one and a small variance. It multiplies each element of the original data by this noise. Unlike the previous two approaches, probability distortion perturbs the value of each data element (point distortion) and replaces it with another sample from the same (estimated) distribution. The merit of this approach is the difficulty to compromise perturbed data using repeated queries.

Table 1 shows four types of data that might be requested from a child during registration. For each data type, its related data items, potential perturbation methods, range or formats and special notes are listed. For non-format numerical type of data such as age, there are two perturbation options. The first one applies probability distortion, and the retention of existing distribution of ages is achieved by following that distribution when generating perturbed age data. The other option exploits the idea of multiplicative perturbation to conceal individual user's age in a normal distribution. By default, COP adopts the second perturbation approach for the reason that no such age distribution is known worth following. For formatted numerical type of data, we first confirm the potential information each format gives away, and then determine the extent to which COP's protection will cover. Take the ZIP code and phone number as examples, the first three digits of these two numbers indicate users' geographical information and COP's policy allow the disclosure of these information for the balance between user anonymity and data usability. Perturbation policy of email address differs from others for the consideration of possible usage of email account to retrieve password in future. Unlike other data, validity of perturbed email addresses can not be assured by COP. For enumeration type of data, it can be processed as numerical data with a certain range.

Another issue in data perturbation is the validation of perturbed data. For example, when a ZIP code is perturbed from "02108" to "02107", the perturbed value "02107" is not a validate ZIP code thus it can not pass the validation check. To avert this issue, COP prepare a list of all valid ZIP code, and the perturbation process can choose from the list according to the rules in Table 1.

[1] This discrepancy might also happens within true data because some people use phone number from other regions.

Table 1. Data Perturbation Approaches for different Data Types

Data Type	Data Item	Possible Perturbation Methods	Format	Notes
Numerical Value (non-format)	Age	1: Follow certain predefined distribution; 2:ϵ, normal distribution with $\mu = 1$ and $\sigma = 0.5$. $\alpha = (Age \times \epsilon) mod 13$, if$\geq 6$, $R(Age) = \alpha$, else $R(Age) = \alpha + 6$	$6 \sim 12$	1: Assume we know the distribution of ages from 6 to 12; 2:Multiplicative perturbation
Numerical Value	Phone Number	Reserve area code generate other 7 digits	123-XXX-XXXX	Certain Rules
	SSN Number	Randomly Perturb	XXX-XX-XXXX	Certain Rules
	Date of Birth	1: Year must be in accordance with age; 2:Month and day follow age's perturbation	XX-XX-($>$1996)	
	ZIP Code	Reserve the first 3 digits, perturb the last two digits	021XX	Consist with Address
	Credit Card	Follow the CCN rules, randomly perturb or use predefined dataset	16 or 15 digits	-
String	Name	Choose from certain data set like cartoon names	Mickey Mouse	-
	Username	Do not perturb unless real name used	-	-
	Address	Keep state name change door number street name and city name	1234, test street, fake city, PA	City name can be preserved if defined by parents
	E-mail	Change to parents' email address	-	-
Enumeration	Gender	change with probability a%	-	-

2.3 Parsing User Input

COP is designed to minimize its interference with users' normal Web activities. To this end, parsing user input in COP needs to distinguish online registration Web pages with other Web pages in the first place. Generally, a registration Web page always contains a form and a submit button, and the button is associated with an event handler (e.g., OnClick) to send out the form. Although many online shopping Web pages may also have the similar structures, we assume that a child user under the age of 13 is unlikely to shop online. Thus COP can discriminate registration Web pages from others by these characteristics.

If the current Web page is considered as a non-registration page, COP follows each user input and compares the content of the input with pre-stored personal information. Upon a match is encountered, COP will retrieve id/name attributes of the input field and the tag before the field. If this retrieved information indicates that the input field asks for personal information. COP will treat this input field as a potential leakage of privacy and perturb the user input. Otherwise, COP will just leave the input field unchanged.

In the case of registration Web pages, COP not only follows and compares each user input with pre-stored personal information, but also considers the data type of the input content. If the input is a string and matches one of the personal information stored as a string such as name, address and email, the input is considered as personal information and is perturbed. If the input is a numerical value with format and matches one the following personal information: ZIP, Credit Number, Phone Number, SSN and Date of Birth, it is also considered as personal information and is perturbed. However, if the input is a numerical value with no format and matches the pre-stored age information, for example, user inputs "12" and her age is also 12. In this case, COP will resort to tag and id/name to recognize the meaning of this input. If the tag and id/name indicate that this input field does ask for user's age, then COP will perturb this input to protect privacy. Otherwise, COP will not change this input because many elements in an HTML file are treated as input fields with a small numerical content. A simple example is a dropdown-list. The selection on the list will be stored as a numerical value representing the list index. If COP perturbs each input field that has the same numerical value as the user's age, a high false positive will be introduced and the user's normal Web activities will be interrupted.

However, there are cases where the input needs to be perturbed even the content of the input does not match any personal information. Considering a child lying about his age to access restricted content, the tag or id/name of the input field will indicate this field asks for user's age, but the content does not match stored age information. In this case, COP will perturb the input to a number less than 13 to prevent the child from lying.

2.4 Transparency to Children

Since COP is dedicated to protecting children under the age of 13, one of COP's design objectives is keeping their child-like innocence. We notice that children might confuse perturbation with deceiving because they are not mature enough to understand the

privacy protection purpose of perturbation. To prevent this, COP keeps the data perturbation operations unobserved to child users. After parsing a user's input, COP generates the perturbed data and marks the input field without changing its content immediately. Only when the submission action is triggered, COP will change all the marked contents with perturbed ones and submit the form. Obviously, reentry would be a problem for users if their usernames have been changed by COP without notifying them. Since username is not treated as personal information by COPPA, under normal circumstances, COP will not interfere with username. However, there are cases where children use their real names as username for login, the real names would be randomly perturbed by COP and this would cause failure in logging into users' accounts. COP addresses this problem by implementing a logging facility. When children visit a website that has been recorded in the log, COP first looks up the related perturbation data for this website. Once found, COP will use these data when required instead of creating new perturbed data. This means if a username has been perturbed during registration, COP will provide it when users try to log in. With the assistance of logging function, COP can preserve transparency to child users.

2.5 Verifiable Parental Consent

COPPA requires that "prior to collection, use, and/or disclosure of personal information about a child, an operator must obtain from a parent of the child verifiable parental consent..." [10]. Obtaining such content every time when information is required from a child may get cumbersome both for the child and the parents. Therefore, we need a way of delegating this responsibility to the tool for approving website policies. Note that since COP will be enforcing the privacy preferences set by the parents, irrespective of the website's policy, COP is always ready to approve the website's policy and to release information that has been pre-approved by parents. The only technical issue is to provide parental consent to the Web server in an automated way, which is not yet implemented in the current version of COP. Nevertheless, in reality, automated approval is not always necessary. Since websites with age censorship tend to acquire this consent directly from parents (see Section 3.1). Common practice of websites is sending email to parents and letting them click a link to give their approval.

2.6 Browser Extension Implementation

We implement COP as a Firefox extension (Firefox version 2.0.0.20). As an add-on to Firefox's functionality, COP is integrated into the browser's main frame once installed. As discussed at the beginning of Section 2, only parent users with administrator privilege should install COP. The installation involves three important steps, which are setting password, specifying privacy preference and filling in children's information. Password is used to identify parent users before they are allowed to modify COP's configuration.

Figure 2 shows the panel where parents can specify privacy preferences in COP. Parents can add any websites they know of into the list and set up the appropriate policy for each of them. Specifically, parents can indicate which categories of their children's information should be withheld from the website by selecting the check-boxes below

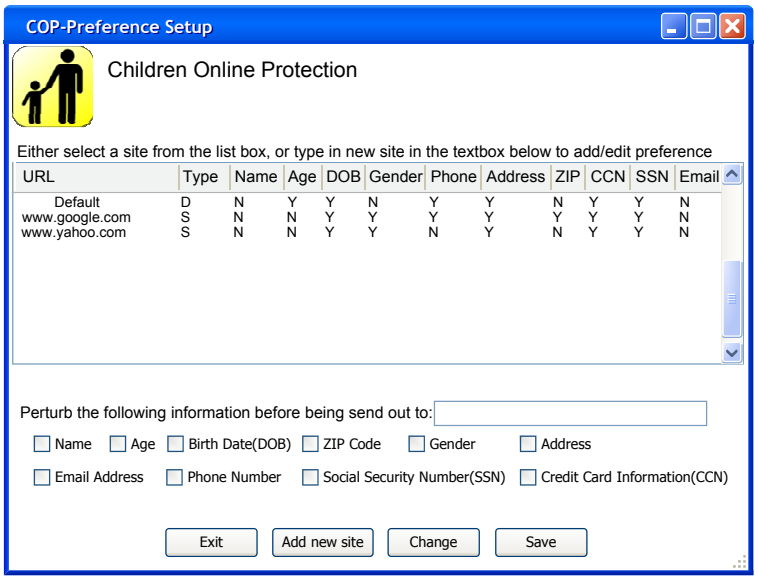

Fig. 2. Installation phase: specifying privacy preference

the list. A checked one means COP will perturb this information before sending it out, which is marked as 'Y' in list (otherwise 'N'). For unknown websites, COP will apply the default policy.

Figure 3 illustrates the user-interface for parents to add their children's information. This information is used as the reference for parsing user input as discussed in Section 2.3. Parents need to fill out the form in this panel for each of their children. If they have more than one child, they need to use "Add Child" button to generate a new form for another child. All the information gathered in this step is considered to be private and COP will prevent children from releasing any of this information to websites, depending on the privacy preferences.

After installation, COP will be activated with an indication appearing in Firefox's status bar. Figure 4 shows the menu list of COP. "Activate/ Deactivate" option is password protected. This ensures COP can only be disabled by the parents but not the children. "Parent Identification" option is used by parents to change the identification information entered during the installation phase. "Preferences Setup" option enables parents to change preferences as showed in Figure 2. "Log" option gives parents access to logged activities, which include the websites visited by children and the perturbed information submitted to those websites.

3 Evaluation

In this section, COP is evaluated for its effectiveness of protecting children's online privacy.

Fig. 3. Installation phase: filling in children's information

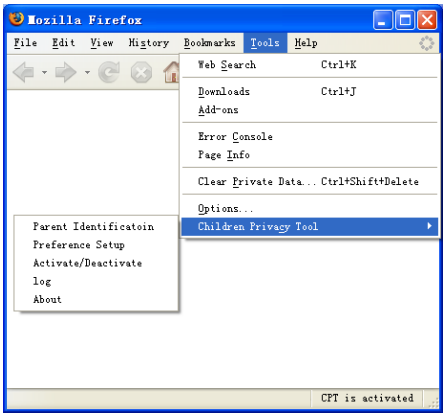

Fig. 4. COP works as a Firefox extension

We test COP with registrations on 23 websites, as listed in table 2. Twelve of these websites are selected from the top site list for kids and teens (suggested by Alexa [30]), e.g., Skyrock, GameSpot, Hyves, Nick. Other sites are the representatives of most popular Web services, e.g., Yahoo, Google, MySpace and Facebook.

We examine the results from the following aspects. First, we look into the amount of personal information required for registration. Numbers in the second column of Table 2 indicate which pieces of information are required by a website among a total of 11 categories. The average number of categories of information in this column is 4.48. The most wanted information is email, (82.6% of visited websites ask for email) followed by full name (65.2%), date of birth (60.8%) and gender (56.5%). Considering the threat of cross identification studied in [7, 6], the required registration information from 5 websites is enough for a malicious party to identify a person, which highlights the importance of privacy protection scheme for online users especially children.

Second, we evaluate the effectiveness of COP protecting personal information. By using default privacy preference setting, all the 11 categories of personal information are protected by COP. Data perturbation shields real information from disclosure. However, in some rare cases such as KidsCom [31], where users provide personal information by clicking pictures figuring preselected answers, COP could not intercept this information. We consider this case as an example of covert information access and will discuss the details in Section 4.1.

We take the registration process on www.yahoo.com as an example to demonstrate how COP works. After COP has been installed in Firefox, "COP is activated" shows up in the status bar. When the browser opens the registration page of Yahoo, personal information such as first name, last name, gender, birthday, country, ZIP, alternative email are required in a form. After we finish filling out all the fields in the registration form and click "Create My Account", COP intercepts this action, performs data perturbation based on the local privacy preference setting, places all the perturbed data into the corresponding fields and sends out the form. Since Yahoo has age censorship,

Table 2. Popular Websites Visted in COP's Evaluation

Website	Information Required for Registration [a]	Age Verification
www.yahoo.com	2,4,5,6,9	Yes
www.live.com	2,3,4,5,6,9	Yes
www.google.com	4,5	No
www.kidscom.com	4,5,6,10	No
www.gzkidzone.com	3,4,5,10,11	No
dashboard.aim.com	1,2,3,4,5,6	No
www.facebook.com	1,2,3,4,5	Yes
www.livejournal.com	1,3	Yes
www.youtube.com	1,2,3,6,9	Yes
www.myspace.com	1,2,3,4,5	Yes
www.blogger.com	3	No
www.hi5.com	1,3,4,5	Yes
www.wordpress.com	3,4,5	No
www.skyrock.com	1,2,3,4,5,6	Yes
www.gamespot.com	1,2,3,4,5,6,7,8,9	Yes
www.hyves.nl	1,2,3,4,5,6,9	No
www.gamefaqs.com	1,2,3,4,5,6,7,8,9	Yes
www.neopets.com	1,2,4,6,7,9	No
www.nick.com	1,2,3	No
www.everythinggirl.com	3	No
www.stardoll.com	1,2,3	No
www.lego.com	1,3	No
www.timeanddate.com	3,4,5	No

[a] 1:birth date; 2:gender; 3:email; 4:first name; 5:last name 6:country; 7:state; 8:address; 9:ZIP; 10: age; 11:parent email

upon receiving this registration request from a user under 18, Yahoo requires an adult's Yahoo ID to proceed. We provide it to finish the registration. All the information Yahoo gets from this registration is a subset of what we, as parent users, allow to release. The most detailed information in this case is an area code in ZIP. This very limited private information disclosure can effectively protect our identities online. From the perspective of child users, this registration process is no different than the one without COP, only their privacy is preserved.

During our test, we notice that age censorship is quite common among current websites. However, these censorships can be bypassed by children lying about their ages to be over 18. Since in configuration COP already has the real age of the child, it can change the filled age to a number less than 13 to make sure those censorships will be invoked.

4 Attacks from Malicious Websites

Some websites may not favor COP for its data perturbation design. If COP is widely deployed, those "untrustworthy" websites might try to exploit any design vulnerabilities in COP. In this section, we discuss some of the attacks that malicious websites would take against COP and our countermeasures.

4.1 Covert Information Access

Although COP spends a lot of efforts on preserving the usability of perturbed data for collectors, it is not surprising to see that some websites which are aware of the usage of COP would try to bypass COP. For example, in the KidsCom case, the website acquires personal information from children by letting them choose from a group of preselected pictures. In this way, no recognizable user input happens, and COP will not be able to intercept any private information. Solution to this problem requires image recognition techniques, which is beyond the scope of the current design of COP. On the other hand, we believe that this information collection fashion is not a general practice among websites, because it is inconvenient for both the users and the websites.

4.2 Embedded Code

One potential attack comes from special embedded JavaScript code. In normal cases, a button control is embedded in a form; once clicked, the Web browser composes an http request by putting together all user input data and sending it in the request. In this case, COP is able to intercept this standard request. However, with JavaScript, a webpage may link a button to an embedded JavaScript code, for example, <input key="send" onclick="SendData()" value="sendsecretly" type="button">, where SendData() is an inline JavaScript function. This function can easily read all the data the user has provided so far, encode it in a specified secret way understandable to the Web server alone, and finally send the data to the server. Thus, the browser (and COP) will not see the original data fields as the values look random. This in turn could make perturbation of the data an impossible task. To address this attack, we consider automatically adding a hook

JavaScript method before the JavaScript data submission method. For example, we may implement COP inside the Web browser layout engine (e.g., Gecko for Firefox). When it detects the above HTML source code, it can modify the code to <input... onclick= "ICheckFirstHook(); SendData()"> instead. Here ICheckFirstHook() is a JavaScript function, added by COP to the HTML source to check or perturb the user input data before it is passed to the original JavaScript function defined by the Web server.

5 Discussion

COP's mission is to reduce divulgence of children's personal information to websites. Despite COP's effective protection, it is not realistic to solely rely on COP, a scheme focus on registration process, to eliminate the possibility of any invasion of privacy. A simple example would be a child posting his or her name, hobbies and maybe photos on an online blog after registration. In another case, a child may release sensitive information such as addresses to others during online chatting. These activities obviously violate the privacy preserving requirements, but their diversity makes it challenging for a single scheme to prevent all of them. Moreover, children might install other browsers to bypass COP or they might be able to disable or uninstall the Firefox extension. To prevent COP from being circumvented, a comprehensive implementation such as a proxy is envisioned. The proxy will intercept all Web traffic. It is installed by system administrator and can only be uninstalled or disabled by users with the same privilege as the administrator. As long as children are not given such privilege, the proxy scheme can not be circumvented.

COP suggests default privacy preference avoiding exposure of "ZIP", "Gender" and "Date of Birth" information to the same website to defend against cross-identification attacks. However, other forms of privacy intrusion might happen due to availability of other auxiliary information. Solutions to this intrusion will include understanding of the mechanisms of such attacks and a more conservative default privacy preference that prevents disclosure of almost all the personal information.

One of our future works, as discussed in Section 2.2, is to propose an approach for quantitatively analyzing data privacy under the distributed model, in which sensitive data are collected from individuals where privacy control (E.g., COP) are placed. The challenge in this problem is the definition and measurement of privacy from an individual's perspective without the knowledge of data possessed by others. The approach adopted in this work, which considers each piece of personal information as a separate privacy metric might suggest a possible direction for solving this problem. In our future effort, we also plan to conduct a comprehensive survey with a sufficient large sample size. The sample should consist of parents as well as their children to deliver more representative feedbacks on the effectiveness of COP.

6 Conclusion

COP offers a novel solution to address the ever-growing concern on children's online privacy divulgence. By concentrating on online registration process, the most common

way of user private information collection, COP preserves children's privacy from releasing to websites while trying to preserving their normal activities. COP can also fulfill the requirements of COPPA and maximize parents control over their children's online activities. Moreover, parents' real-time involvement is little if any except the easy configuration in the installation phase. COP manages to achieve these goals by implementing functions like privacy preserving data perturbation, user input parsing and transparency to children. We demonstrate COP's effectiveness in serving its privacy protection purpose by evaluating its performance in popular websites. We also believe COP is one of the directions for protecting children from online threats. With further improvements, COP can help parents to protect their children in a more effective way.

Acknowledgement: We thank the reviewers for their valuable comments and suggestions. This work was partially supported by NSF CAREER 0643906.

References

1. Katz, L.: When 'digital bullying' goes too far. http://news.cnet.com/when-digital-bullying-goes-too-far/2100-1025_3-5756297.html. (June 2005. Retrived Jan 2010)
2. : Online predators: Help minimize the risk. (Jan 2007. Retrived Jan 2010)
3. BBC: Net blamed for rise in child porn. http://news.bbc.co.uk/1/hi/technology/3387377.stm (2004. Retrived Jan 2010)
4. FTC: Xanga.com to pay 1 million for violating children's online privacy protection rule. http://www.ftc.gov/opa/2006/09/xanga.shtml. (2006. Retrived Jan 2010)
5. FTC: mbee.com settles ftc charges social networking site for kids violated the children's online privacy protection act; settlement includes 130,000 civil penalty. http://www.ftc.gov/opa/2008/01/imbee.shtml. (2008. Retrived Jan 2010)
6. Sweeney, L.: Uniqueness of simple demographics in the u.s. population. Technical report, LIDAPWP4, Carnegie Mellon University, Laboratory for International Data Privacy,, Pittsburgh (2000)
7. Golle, P.: Revisiting the uniqueness of simple demographics in the us population. In: 2006 Workshop on Privacy in the Electronic Society, ACM Press (2006) 77–80
8. : Cox communications teen internet safety survey wave ii. Technical report, Teen Research Unlimited (March 2007)
9. Youn, S.: Teenagers' perceptions of online privacy and coping behaviors: A risk-benefit appraisal approach. In: Journal of Broadcasting & Electronic Media. Volume 49., Routledge (March 2005) 86–110
10. : Children's online privacy protection act (1998. Retrived Jan 2010)
11. Center, E.P.I.: Pretty poor privacy: An assessment of p3p and internet privacy. http://epic.org/reports/prettypoorprivacy.html. (June 2000. Retrived Jan 2010)
12. : How anonymizers work. http://www.livinginternet.com/i/is_anon_work.htm (2007. Retrived Jan 2010)
13. Zwick, D., Dholakia, N.: Models of privacy in the digital age: Implications for marketing and e-commerce. Technical report, Research Institute for Telecommunications and Information Marketing (RITIM), University of Rhode Island (1999)
14. : Truste. http://www.truste.org/ (Retrived Jan 2010)
15. Culnan, M.J., Bies, R.J.: Consumer privacy: Balancing economic and justice considerations. Journal of Social Issues **59**(2) (2003) 104–115

16. Milne, G.R., Culnan, M.J.: Strategies for reducing online privacy risks: Why consumers read(or don't read) online privacy notices. Journal of Interactive Marketing **18**(3) (2004) 15–29
17. Hsiao, M., Belanger, F., Hiller, J., Aggarwal, P., Channakeshava, K., Bian, K., Park, J.M.: Parents and the internet: Privacy awareness, practices and control. In: Proceedings of Americas' Conference on Information Systems. (2007)
18. Culnan, M.J., Milne, G.R.: The culnan-milne survey on consumers & online privacy notices: Summary of responses. In: Proceedings of Get Noticed: Effective Financial Privacy Notices, Washington, DC, A Federal Trade Commission Workshop (2001)
19. Crossler, B., Belanger, F., Hiller, J., Aggarwal, P., Channakeshava, K., Bian, K., Park, J.M., Hsiao, M.: The development of a tool to protect children's privacy online. In: Annual Workshop on Information Security and Assurance, Montral, Canada (2007)
20. : Parental controls in windows vista. http://www.microsoft.com/windows/windows-vista/features/parental-controls.aspx (Retrived Jan 2010)
21. : Privo. http://www.privo.com/ (Retrived Jan 2010)
22. : icouldbe. http://www.icouldbe.org/ (Retrived Jan 2010)
23. : netnanny. http://www.netnanny.com/alt_rotate (Retrived Jan 2010)
24. : Parental control bar. http://www.parentalcontrolbar.org/ (Retrived Jan 2010)
25. Thierer, A.: Social networking and age verification: Many hard questions; no easy solutions. Progress & Freedom Foundation Progress on Point Paper No. 14.5 **14**(5) (2007)
26. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: On the privacy preserving properties of random data perturbation techniques. In: ICDM '03: Proceedings of the Third IEEE International Conference on Data Mining, Washington, DC, IEEE Computer Society (November 2003) 99–106
27. Kim, J.J., Kim, J.J., Winkler, W.E., Winkler, W.E.: Multiplicative noise for masking continuous data. Technical report, Statistical Research Division, US Bureau of the Census, Washington D.C (2003)
28. Krishnamurty Muralidhar, D.B., Kirs, P.J.: Accessibility, security and accuracy in statistical database: The case for the multiplicative fixed data perturbation approach. JSTOR-Management Science **41**(9) (1995) 1549–1564
29. Liew, C.K., Choi, U.J., Liew, C.J.: A data distortion by probability distribution. ACM Trans. Database Syst. **10**(3) (1985) 395–411
30. : Alexa-top sites by category. http://www.alexa.com/topsites/category/top/kids_and_teens (Retrived Jan 2010)
31. : Safe kids chat rooms. http://www.my.kidscom.com/ (Retrived Jan 2010)