PyP - Protect your Privacy

Swetha Tatavarthy, Devendra Lad, Indraneel Bende, Meenal Khandelwal
Maharashtra Institute of Technology
S.No.124, Paud Road, Kothrud
Pune 411 038
{t.swetha.mit, devendralad22, meenal.khandelwal17} @gmail.com, indraneelbende@yahoo.com

Kumar Vidhani, Akhil Dixit, Manish Shukla, Vijayanand Banahatti, Sachin Lodha
Tata Research Development & Design Centre,
Tata Consultancy Services, 54 B, Hadapsar Industrial Estate, Hadapsar, Pune 411 013
{kumar.vidhani, akhil.dixit, mani.shukla, vijayanand.banahatti, sachin.lodha} @tcs.com

ABSTRACT
Many of us share the sensitive information on various websites, due in part to the increased reliance on the internet for day-to-day tasks. This sensitive information might be misused or exploited. There is a need to enable internet users to specify and rank the sensitive information belonging to her. We propose Protect your Privacy (PyP), which collaborates with other internet users and generate alerts whenever user shares the sensitive information online.

1. INTRODUCTION
Many of us have accidentally revealed some sensitive information (For e.g. cell phone number etc.) on the social media such as Facebook, Twitter, LinkedIn, WhatsApp etc. Online Social Networks (OSNs) do not offer foolproof mechanism to protect the users’ sensitive information. This was particularly highlighted by the open source collation of e-government data and networks system where on entering the name of person and the area where she resided; critical information such as PAN card number, Voter ID could be fetched using public APIs [6]. Also OSNs routinely publish data of interest to third parties but in doing so they also often reveal relationships, such as a friendship or contractual association that a person with malicious intent can exploit. The user might not care about the posted information, such as a date of birth, but a person can still leverage that post to compromise user identities [2]. The information leakage is not limited to OSNs. According to an analysis (Fig. 1) conducted by the Wall Street Journal [4] on the personal information that popular websites share about users, it tested seventy one popular websites that request a login and found that more than a quarter of the time, the sites passed user’s personal information like real name, email address or other personal details to third-party companies.

![Figure 1. Websites sharing the personal information](image)

Currently, there are privacy add-ons such as Web of Trust (WoT) which works as warning mechanisms to users [3]. In WoT, the user is alerted about website reputations by displaying graphic icons next to search results while using Google, Yahoo!, Bing or any other search engine. It is basically built to protect the users against harmful websites only. In order to educate the user about her privacy behavior on the internet, to alert her about the risk associated with sharing specific information on websites and to eventually change the privacy habits of the user we have created a browser extension [5] that acts as privacy monitoring layer and provide a real-time warning to a general user or a user of a specific web page, about the sensitivity of the posted content. PyP unlike WoT, is concerned with users’ sensitive information and intends to warn the user from sharing the same. If a user chooses to post the content despite the warning, she would be permitted to do on a personal profile. We shall evaluate the impact of such privacy controls on the accidental information sharing.

2. SYSTEM DESCRIPTION

![Figure 2. System architecture of Protect your Privacy (PyP)](image)

The PyP system consists of five components: User Interface (UI), Event detector, Message processing unit, Risk calculator and Local storage.

UI: UI has screens for login and signup. User can submit the sensitive items and assign rank to them through UI screens. User can see her privacy history through charts.

Event detector: Event detector detects the DOM events such as KeyPress for a webpage and sends the data to message processing unit.

Message processing unit: Message processing unit fetches user’s sensitive data stored on local storage and processes the message.
**Risk calculator:** It runs on the server and calculates the risk score for each sensitive item.

**Local storage:** It stores the sensitive information of user locally along with risk scores calculated by risk score calculator.

Once the PyP is loaded in the browser, user can see the icon on top right corner. Clicking on that user can login (if she has already signed up) or create an account. While creating account she will be asked to fill in her sensitive information like Name, Email Id, Pan card no etc. (Fig. 3). PyP comes with eight predefined fields however the user can add more fields based on the requirement. The information entered by the user will be stored on chrome's local storage in an encrypted form. The encryption is achieved through Stanford Javascript Crypto Library [7]. The user is also required to rank all the sensitive fields mentioned on the signup page. The user can rank sensitive field from zero to the total number of fields minus one. The field with lower rank is considered more sensitive than with higher rank. If a user deems her Email Id as the most sensitive among all then she shall rank it as zero. Users can later edit the sensitive fields and their ranks. The sensitive fields and their corresponding ranks are passed on to the server for computing the risk score.

The PyP monitors user’s activity and every time she enters her sensitive data, she is given a warning along with the relevant risk score for that sensitive data. Furthermore, she can learn about her privacy activities over the internet for specified duration. There is a tabular and graphical representation of her activity along with the details of sensitive fields, number of times it has been entered, date and time and the URL of the website where the sensitive fields were entered.

![Figure 3. UI screen for entering sensitive fields' values](image)

**2.1 COMPUTATION OF RISK SCORE**

The risk score computation takes place on the server. The risk calculator maintains a matrix where each row represents a user and each column represents one sensitive field (Name, phone number etc.). This matrix stores the ranks (given by the user to that field) by dividing it with the total number of fields minus one. The formula for calculating the sensitivity of field \( j \) is [1]:

\[
\beta_j = \left( N - |R_i| \right) / N
\]

Here, \( N \) is total number of users.

\( |R_i| \) = Summation of all the cells of the column corresponding to sensitive field \( \beta_j \).

**Risk score** = \( \beta_j \times \text{website rank} \).

Website rank is computed using WoT APIs. The order of matrix will be of \( N \times M \) where \( 1 \leq i \leq N \) and \( 1 \leq j \leq M \) (M: total number of sensitive fields). The risk score is not specific to the rank provided by one user rather it is a generalization of all the ranks. This formula helps capture the perception of all the PyP users about particular sensitive field and thereby captures general user perception about that sensitive field. It also takes into account the authenticity of the website provided by WoT.

For example, as shown in Table, there are five PyP users whose ranking of their phone numbers is: 1, 2, 3, 0, 2. Let the total number of sensitive fields captured by PyP be five. The matrix will store the values 1/4, 2/4, 3/4, 0/4, 2/4 in “Phone no” column as shown below:

<table>
<thead>
<tr>
<th>User Id</th>
<th>Name</th>
<th>Phone no</th>
<th>DOB</th>
<th>Email Id</th>
<th>Pan card no</th>
</tr>
</thead>
<tbody>
<tr>
<td>User1</td>
<td>0.75</td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>User2</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>User3</td>
<td>0.5</td>
<td>0.75</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>User4</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.75</td>
<td>0.25</td>
</tr>
<tr>
<td>User5</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.25</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Here \( R_i = 2 \). Therefore the sensitivity according to the formula will be, \( \beta_j = 5-2/5=0.6 \). This will be multiplied with the website rank determined using WoT and will be sent from the server to the user. The user will be alerted when she enters her phone number on that particular website.

**3. CONCLUSION AND FUTURE SCOPE**

We have developed PyP browser extension which alerts users when they enter sensitive data on websites. Over time PyP also makes users conscious about their privacy habits. Currently PyP is chrome specific but can be implemented for other browsers too. The computation of risk score can also include other parameters such as the frequency with which the user shares her data, age group of user etc. New users could be given suggestions related to ranking for sensitive fields. Furthermore, there could be a provision to notify the user when another PyP user tries to reveal her information.

**4. REFERENCES**


