A Cloud Based Service for Internet Users to Track Privacy-Related Data in Web

Nitin Singh Chauhan[1], Ashutosh Saxena[2], JVR Murthy[3]

[1]Infosys Ltd. Hyderabad, India
Email: nitin_chauhan01@infosys.com
[2]Department of Computer Science, CMR Technical Campus, Hyderabad, India
Email: saxenaaj@gmail.com
[3] Department of Computer Science, JNTU Kakinada, Kakinada, India
Email : mjonnalagedda@gmail.com

Abstract— *In the online world, data has become equivalent to currency of the real world. Search engines, e-commerce sites, online social networks, advertisers, fraudsters, spammers etc. are in thirst of data of users, more specifically Personally Identifiable Information (PII), which can be used for genuine business gains or malicious purposes. In most cases, sharing of user's data by a website to its partners is subjected to legal terms and conditions of the site. However, once data moves from a user's browser to the Internet, there is no mechanism to track the data. In this paper, we propose a system which aims to track and detect privacy violation on the web. System assists users of the web to maintain their own record of data they share with each website and also alert the user about the possible privacy violations using a client side plugin.*

Keywords—privacy, web, browser, data security.

I. INTRODUCTION

In today's digital era, online presence has become a commonplace. Almost all activities of the real world such as collaboration, shopping, discussions, banking etc., have moved online and each of them require personal information of end users. With the outburst of social apps, mobile apps and cloud based frameworks, assuring privacy on the modern web is a challenging task [1].

In the context of web, "right of the user to have control over the data collected by websites and reveal the identity based on his/her wish." commonly referred to as Web privacy. Sharing of user's data by a website to its partners is usually protected by legal terms and conditions of the site. Once data moves from a user's browser to the Internet, there are no technical mechanisms to track the data or detect possible privacy violation. To an extent, some applications contribute towards protection of privacy by preventing third party cookies from following users on the web [2] or by providing means to clean public databases via their APIs. However, these techniques do not assist in detecting how user's data has been leaked to the public or which site violated their privacy agreement.

Our invention aims to track and detect privacy violation on the web. An advisory system is developed which assists users of the web to maintain their own record of data they share with each website. The invention has two main components, one of which is a browser plugin called Privacy Tracker Service Plugin, while the other is a cloud based system called Privacy Information System, which has a Privacy Tracker Database and a Privacy Check Processor. During a browsing session, if the system encounters user's data which is not in its database, the privacy tracker service plugin alerts the user about the possible sites which have violated user's privacy and suggests suitable actions. Our system can be used on multiple form factors/devices.

II. RELATED WORK

Since online privacy failures can occur at several places right from visible IP address, unencrypted traffic, insecure applications, online social networks etc., there are technologies which attempt to protect privacy in each of these specific areas, which are different from our proposal. Tor browser bundle allows users to browse the web anonymously by encrypting network traffic and routing through complex network nodes [3].

Certain browser extensions which route information through proxy servers so that third party cookies (which track users) can be blocked. There are browser extensions designed to help users in understanding and taking control of the data they share on specific sites such as Facebook, Twitter, Gmail etc. Tools which help users in understanding who can see their profiles on social networks like Facebook and what data will be visible to the public [4]. Though not a privacy protection feature, web browsers store a history of sites visited by users, sometimes along with form data, and this may be used for manual inspection of visited sites.

Baviskar, et al [5] presented a secure way of interaction between browser objects through browser API'S and the JavaScript to protect user's information which is transferred over the network to the web server without the user's consent. Monjas, Miguel A., et al [6] presented a user-centric schema for self-managed privacy that enables users to use a dashboard to find out which user information a social network provider has shared with and to rule the way such a sharing procedure is done. However, this scheme is only limited to social network.

There are no systems or methods to track information submitted on webpages and check against it later to detect privacy violations. Existing technologies may have feature to store submitted pages but data field's storage cannot be selective or personalized. Present methods of page information storing even retains the submitted data. Storage of such information or sharing it to third party could lead to privacy violations. However, the proposed invention only stores data fields instead of actual data.

The browser's native history maintenance technique resembles the functionality of Privacy Tracker Database in our system to some extent. However, it is designed only to assist users in navigation and not as a privacy tracking/privacy violation detecting mechanism. Even otherwise, it has shortcomings such as: The "Clear History" option in browsers completely erases all browsing history of users. Changing/Reinstalling browsers will erase browsing history information. Since history data is stored locally in the machine, it is not available when users change their machines or it cannot be segregated when multiple users use the same machine.

Some applications contribute towards privacy protection by preventing third party cookies from following users on the web or by providing means to clean public databases via their APIs. Techniques such as self-destructing emails, anonymous browsing, data encryption, analysis of privileges in mobile devices etc. contribute towards privacy preservation. However, these techniques neither assist in detecting how user's data leaked to the public nor inform which site violated their privacy agreement.

III. PROPOSED SYSTEM

Proposed system aims to track and detect privacy violation on the web. The architecture of our system is explained in the figure-1. It consists of two main components, one of which is a browser plugin called Privacy Tracker Service Plugin, while the other is a cloud based system called Privacy Information System. The latter has two subsystems- Privacy Tracker Database and Privacy Information Processor.

Browser: A browser is a software application for retrieving, presenting and traversing information resources on the web. This application receives URL as input and access the information resource available on web. In context of our invention, end users access certain webpage through browser and submit their data or perform various activities, which could lead to generation of user related private data. This data is submitted to web sites of information collecting entity to meet users or business interest.

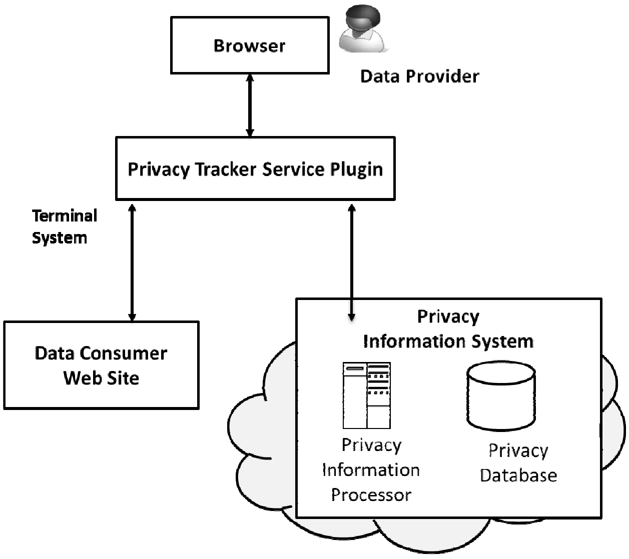

Browser
Data Provider
Privacy Tracker Service Plugin
Terminal System
Data Consumer Web Site
Privacy Information System
Privacy Information Processor
Privacy Database

Fig. 1: Overall architecture of Privacy Tracker Service

Privacy Tracker Service Plugin: Privacy Tracker Service Plugin is one of the components which can reside as browser component/plugin and gets activated when user opens any webpage to submit data to Data Consumer Website or perform some activity on the website. Privacy Tracker Service Plugin provides option to user to create his personalized privacy profile. User can define personal data fields which are sensitive and private. User can also define type of data and activity on website that should be logged in proposed system when users submit the data or perform activity. This customized information is captured by PTSP and stored in Privacy Tracker Database (PTD). Whenever user submits any form on website, PTSP retrieves the personalized profile and it identifies user defined personal data fields for which details are being submitted. PTSP stores this information along with website details, time stamp in PTD as privacy data history.

PTSP has another role when users access some website and identifies some sensitive personal data pertaining to him. User submits the identified data field to PTSP, which checks user's privacy profile history stored in PTD. It alerts for privacy violation if website reflecting the data value is never being provided with specific information. PTSP has following sub-modules.

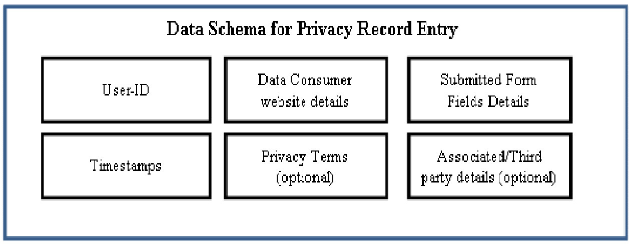
Web details Capturing Module: This module captures the information about data field and user activities from web page as per user's privacy profile.

Online Privacy Detection Module: This module enables user to verify any potential privacy violation, while accessing third party website. In case user related personal information is displayed on third party webpage, user can probe the privacy information system to check from users browsing history stored in Privacy Tracker Database. This module also raises alert after checking and display information about potential website that might have compromised user's privacy details. This module can be manual or automated based on implementation.

Connectivity Module: This module enables connectivity to various other components to send and retrieve the information.

Privacy Profile Management Module: This module enables user to create his/her personal profile/ preference/actions related to privacy. Users can add, modify, and delete the profile as per their requirement. Users can define data fields which are sensitive and contain private data.

Privacy Tracker Database (PTD): Privacy Tracker Database stores user privacy profile and privacy data history in is database. Data field schema includes User ID, data consumer website details, Submitted form field and Timestamps etc. Users have option to customize the information based on privacy requirements. Information related to privacy agreements/terms can also be stored in privacy tracker database if it is made available to user during data submission. Agreements may mention details of third party with whom data can be shared by data collecting web site. Database will have option to record those details as well.



Data Schema for Privacy Record Entry

User-ID	Data Consumer website details	Submitted Form Fields Details
Timestamps	Privacy Terms (optional)	Associated/Third party details (optional)

Fig. 2: Data Scheme for Privacy Record

Privacy Information Processor(PIP): Whenever user queries about some suspected data field, PIP retrieves the user profile and also checks websites where sensitive data was submitted. Based on search result, PIP provides result to user through PTSP.

Data Consumer Web Site (DCWeb): Data consumer website is owned by business/enterprise/ organization/individuals who provide the option for user to submit their details. These details are submitted as forms and used by enterprise to process this information for business or user interest.

IV. SYSTEM FLOW

The below steps explain how our system works and assists in tracking and detection of privacy violation:

Workflow-1: When a user accesses a website and submits personal information. Figure-3 describes the flow.

1. User installs PTSP which sits in the browser and routes all HTTP traffic through it.
2. When a data consumer website presents a form, user fills it with his details.
3. The PTSP keeps track of this data fields and saves it into PTD, which is a sub-system of our cloud based privacy information system.
4. The schema of our PTD is outlined in the Figure-2. Such a schema helps in aggregating and processing information relevant to user's session.

Workflow-2: When a user accesses any third party website to which he/she never submitted any personal information, but the site shows user's personal information. Figure-4 explains the workflow.

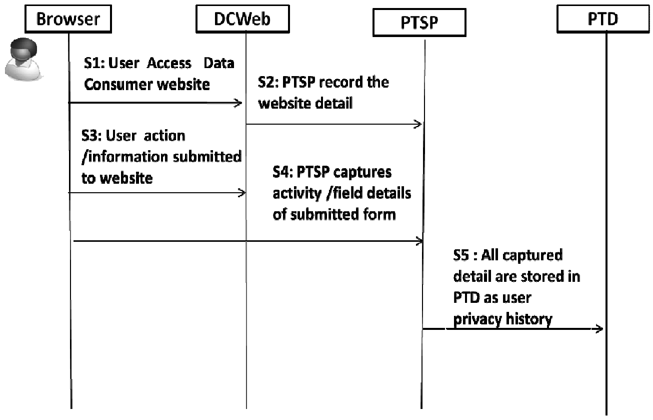

Browser
DCWeb
PTSP
PTD
S1: User Access Data Consumer website
S2: PTSP record the website detail
S3: User action /information submitted to website
S4: PTSP captures activity /field details of submitted form
S5 : All captured detail are stored in PTD as user privacy history

Fig. 3: Activity Sequence -When a user accesses a website and submits personal information

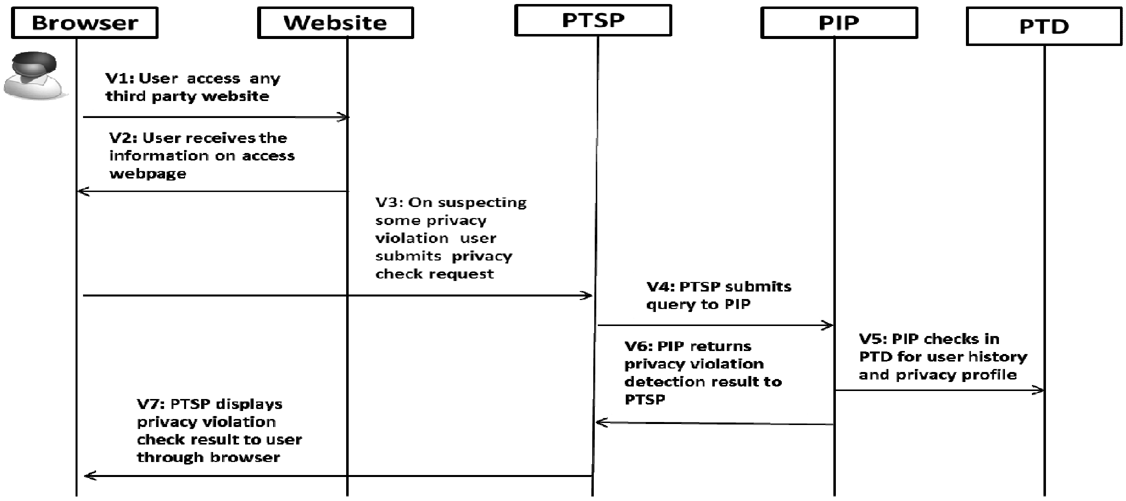

Browser
Website
PTSP
PIP
PTD
V1: User access any third party website
V2: User receives the information on access webpage
V3: On suspecting some privacy violation user submits privacy check request
V4: PTSP submits query to PIP
V5: PIP checks in PTD for user history and privacy profile
V6: PIP returns privacy violation detection result to PTSP
V7: PTSP displays privacy violation check result to user through browser

Fig. 4: Activity Sequence-When a user accesses any third party website and detects potential privacy violation

When the user browses another site (third part site) which happens to display the information entered by the user in one of his previous sessions, the Privacy Tracker Service Plugin sends the page to the privacy information processor. The processor analyzes the data by matching it with the records in its database and alerts the user which site in its database violated privacy of the user by leaking the information to a third party. In this way, our system assists users in tracking their information and thereby detecting privacy violation and exposure of their data on the web.

V. CONCLUSION

Privacy problem has escalated in new challenging environment of cloud and big data. Widespread use of social networking sites had increased the opportunity of privacy exposure. In the present era of web based services, users provide personal information to many websites. It's practically challenging to keep track of these sites and data fields submitted to them manually. Therefore, this system helps in creating privacy fingerprint for user by collecting details of web based activity where personal information is shared with third parties. There are possibilities that data collecting agencies/enterprise may share user data to third party for their business benefit, without taking users consensus. If user notice, such information is used by third party and represented on its website, user can identify data collector who might have involved in privacy violation.

Disclaimer: The views expressed here are the authors personal opinions; they might not represent the view of associated organizations. Authors acknowledge suggestions provided by Krishnachaityana T. during this work.

REFERENCES

[1] Mather, Tim, Subra Kumaraswamy, and Shahed Latif. Cloud security and privacy: an enterprise perspective on risks and compliance. " O'Reilly Media, Inc.", 2009.

[2] Soltani, Ashkan, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. "Flash Cookies and Privacy." In AAAI spring symposium: intelligent information privacy management, vol. 2010, pp. 158-163. 2010.

[3] https://www.torproject.org/projects/torbrowser.html.en , Last Accessed: 30 November 2016.

[4] Krishnamurthy, Balachander, and Craig E. Wills. "Characterizing privacy in online social networks." In Proceedings of the first workshop on Online social networks, pp. 37-42. ACM, 2008.

[5] Baviskar, Sanket, and P. Santhi Thilagam. "Protection of web user's privacy by securing browser from web privacy attacks." International Journal of Computer Technology and Applications 2.4 (2011): 1051-7.

[6] Monjas, Miguel A., et al. "Privacy Delegate: a browser-based tool for privacy self-management in social networks." Ericsson Position paper: W3C Workshop on identity in the browser. 2010.