**Title:**

**Privacy on the Web: Facts, Challenges, and Solutions**

**What is the research problem?**

There has been exponential growth on number of people using Internet and the Web services available today. Although there has been regulatory and technical efforts put to safeguard the privacy of the end user but still every now and then incidents are privacy theft are reported. The papers highlights the important issues and proposed solution, focusing on the generic user’s web privacy.

**Defining Privacy:**

* Definition of privacy varies according to context, culture and environment.
* As early as 1890 privacy we defined as “the right to be left alone”.
* In 1967 it was defined as, the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude and their behavior to others.
* With the recent trends now it’s defined as “right to determine what (personal) information is communicated to others” or “the control an individual has over information about himself or herself”.

**Overview/main points of the proposed approach/architecture**

* Internet has made information ubiquity a reality.
* However this ease of information availability coupled with the ready availability of personal data, also made it easier and more tempting for people with wrong intent
* Researchers have proposed a range of techniques to preserve Web user’s privacy. However, despite considerable attention, Web privacy continues to pose significant challenges due to Differences and incompatibilities in privacy regulations and standards
* This paper emphasizes on resolve the above said difference in standards and highlight the threats on user privacy and focus on its prevention.
* As per the paper the two major factors contributing to the privacy issues on the web are:
  + Inherently open, nondeterministic nature of the Web: unlike enterprise network which has only known user with pre-defined set of privileges, Web is an open environment in which there are numerous unknown users accessing information
  + Complex, leakage-prone information flow of many Web-based transactions that involve the transfer of sensitive, personal information: personal information that a Web user submits to a given party might, as a result of the application’s intrinsic workflow, be disclosed to one or more other parties.
* In many case customer submit their personal information to companies which under certain circumstance might have to share it thus exposing the customer.
* A study suggests that the loss due to privacy concerns had reached up to 18 billion in 2002 itself.
* Even the government agencies collecting user information can be compromised at times.

**Understanding Web Privacy**

* A Web transaction is any process that induces a transfer of information among two or more Web hosts
* Web user’s information are classified into following three categories
  + Personal data: name, marital status, mailing and email addresses, phone numbers, financial information, and health information
  + Digital behavior: refers to Web users’ activities while using the Web, including the sites they visit, frequency and duration of these visits
  + Communication: includes Web users’ electronic messages, postings to electronic boards, and votes submitted to online polls and surveys
* The paper specifies different dimensions of Web Privacy such as Information Collection, Information Usage, Information Storage, Information dis-closure, Information Security, Access Control, Monitoring and Policy changes.

**Sources of Privacy Violation**

* Unauthorized Information transfer: Personal information is increasingly viewed as an important financial asset. For benefit often, information is transferred without an individual’s explicit consent.
* Weak Security: Web’s inherently open nature has led to situations in which individuals and organizations exploit the vulnerability of Web-based services and applications to access classified or private information.
* Data Magnets: are techniques and tools that any party can use to collect personal data such as
  + Information gathering using online registration
  + Identifying user IP address based on user’s requests
  + Software Download: Asking users to provide information for unique identification
  + Cookies
  + Trojan horse
  + Web beacons: to monitor user activities
  + Screen Scraping
  + Federated Identity
  + Indirectly collecting information

**Privacy Preserving Solutions**

* Based on the Source of Violation the Solution are categorized into
* Technology enabled Solution:
  + Client based Solutions: include protecting personal data stored on a personal computer, protecting email addresses, deleting any trace of Web access, and hiding Web surfers’ real identities. A firewall or hardware to private network should work. Proxy based anonymizer, etc.
  + Server Based Solution: Virtual private networks (VPNs) and firewalls are two mechanisms that have been particularly effective in protecting security and privacy at an enterprise scale.
  + Client-server based Solution: Negotiation and encryption based solutions.
* Regulation-enabled solution:
  + Self-Regulation: refers to the information keepers’ ability to voluntarily guarantee data privacy.
  + Mandatory regulation: refers to legislation aimed at protecting citizens’ privacy while they transact on the Web.

**Conclusion:**

This paper intents to highlight the need of more effective techniques necessary to protect high quality Web information from illegitimate access and use. The objective is to empower

User’s to have better control over the access and the use of their data.