A Study on Privacy enhancing Technologies for the Internet

Kriti Arora

1

Abstract – **It is rightly said that we are living in the Information age. Internet has become an integral part of our lives nowadays. This pervasive use of Internet in every walk of life has put a big question mark on Individual privacy today as we share more and more personal information online. Nothing is hidden in today's fast age of connectivity and Internet as all information posted online is available to so many prying eyes and companies. When your every move or mouse click is being tracked or monitored you are constantly under the scanner. In such an environment it becomes more and more difficult for individuals to say what they feel, read what they want, express their views openly and do online activities without fear of being attacked by hackers or getting their personal information violated. Some powerful technologies are available today that can help protect our online privacy. This paper discusses some currently available privacy enhancing technologies and tools; and analyses their effectiveness in ensuring online privacy. For each tool it discusses what technology they use, what situations they are helpful in and what threats they can handle, their effectiveness, limitations and drawbacks.**

Index Terms **– Anonymizers, Digital Currency, Privacy preserving Data mining, Remailers, VPN**

I. INTRODUCTION

Internet is all about sharing information, ideas and thoughts these days. Sharing personal information on the Internet - be it through social networking sites or through internet applications and forms is a favorite pastime as well as a necessity these days. The growing world of internet and its widespread use in all areas such as social networking, banking and financial applications, e-commerce, professional world, education, healthcare domain and government applications has put a big question mark on Individual privacy today. In such a well connected information age, protecting sensitive and critical information of an individual or organization from unauthorized access and malicious attackers becomes very important. Privacy is the ability or need of an individual, group or organization to withhold themselves, or information about themselves from public purview thereby expressing or revealing information about themselves selectively. In this paper we focus on Internet privacy solutions - which deal with controlling what information about a user is collected and disclosed over the internet, managing access to that information and regulating the purpose for which that information may or may not be used.

Many tools and technologies have been developed and are used today to protect individual privacy of Internet users. This paper is a study on the various techniques and technologies available today for protecting online data privacy. We first discuss some potential threats to Individual privacy through Internet. The next section analyses in detail some currently available tools and technologies that can be used to protect personal privacy on the internet and help the users access the Internet and do online transactions securely and anonymously without compromising their individual privacy. These technologies and techniques are generally known as Privacy Enhancing Technologies or PET. For each technique and technology the paper discusses what technology they use, what situations they are helpful in and what threats they can handle, their effectiveness, limitations and drawbacks.

II. PRIVACY THREAT MODELS

Through the use of various online applications like Social networking sites, online forms and applications, online surveys and e-commerce sites, we directly upload Personal Identifying Information or PII on the Internet or transmit personal data through Internet. Whereas several other type of Information like IP addresses, non personally identifiable profiling, what internet sites people visit, their interests, likes and dislikes can be gathered easily from their browsing history. Some common internet practices or methods through which individual privacy is threatened and compromised are mentioned briefly below –

1. HTTP cookies used by almost all websites these days can be easily used to track individuals and prepare detailed individual profiles by placing a personally identifiable tag together with their personal data like user id, IP address, browsing patterns and computer information.
2. Device fingerprinting helps maintain records of users by tracking their IP addresses, email addresses, location and user accounts. Every move one makes on the Internet can be watched and tracked by marketers or attackers using this method.
3. The increasing use of smart phones and tablets with their smart location tracking features help track your location around the world in a minute. When using a smart phone,

[1] ***Kriti Arora,*** *Department of Computer Science, Shyam Lal College, University of Delhi, Delhi, India.*

threats include geo location, meaning that one's phone can detect exact location where they are currently present and post it online for all to see.

4. Search engines with their ability to track the various searches made by users and link it to the user's account, IP address or computer thereby revealing personal information pose a major threat to individual privacy.

5. Dedicated organizations using Data mining and Big analytics techniques can build huge dossiers of user data collected from various sources, compile and analyze this data and prepare profiles of internet users. These profiles or detailed user data is sold or shared with other companies for market analysis and exploring new customer markets.

6. The widespread sharing of personal information and photographs on social networking sites causes privacy threats such as hacked accounts and profiles, data misuse, leaked photographs, fraudsters posing as genuine friends trying to access personal data besides other vulnerability threats.

7. Data stored on Clouds faces various security and privacy risks like Insider attacks on clouds, online hacking attacks and network security threats.

8. Embedded software or links to malicious websites can cause malicious programs like spyware, malware, spam mails and viruses to attack your computer and data stored in it during web browsing.

III. PRIVACY ENHANCING TOOLS AND TECHNOLOGIES

A. Anonymizers - freenets, mixnets, TOR

Technology – Anonymizers or proxies are the most commonly used tool for protecting online privacy and anonymity. It essentially involves the use of proxies or computer servers that acts as an intermediary between a client computer and the rest of the internet. It accesses the internet on the user's behalf, protecting the personal information of the user by hiding the user's identifying information. Using only one proxy is never sufficient to protect the identity and privacy of the user since it could be easily hacked, infiltrated or eavesdropped. To counter this problem, usually a chain or network of proxy servers often called mixnets or freenets are used. To provide further data and privacy protection, they use cryptographic techniques for encrypting the messages. These proxy servers form a relay network of proxies that pass the message form one proxy server to the next often encrypting the message in the process to further provide protection. A mixnet is a collection of nodes, usually individual computers on a network. All nodes have a public-key/secret-key pair whose public key is easily accessible and known to everyone. A node in the mixnet listens for encrypted messages; once it receives one, it decrypts the message using its secret key. The decrypted message reveals a chunk of ciphertext and instructions regarding to whom the node should forward the ciphertext. This process is followed at every node until the message is finally forwarded on to its final destination. Zero Knowledge Systems was the first anonymizer system based on network chain of proxy servers. Many anonymizers like ZKS also provide a Cookie manager that allows you to keep track of your cookies and separate them into separate folders. The cookie manager prevents unnecessary tracking that does nothing to enhance to browsing the experience. For example an advertisement by an ad server like DoubleClick.net, can easile be dispensed with, so it blocks all HTTP requests to ad servers so that they neither receive requests nor cookie information. This prevents web-based tracking and improves download time [1].

TOR or The Onion Router is the best and most widely used example of Anonymiser network. TOR routes or provides a unique path to transmitted data packets through a free, worldwide, volunteer network consisting of thousands of relays or proxy servers so that the sender and receiver's exact location and activities are concealed from anyone conducting network surveillance or traffic analysis. The term "onion routing" refers to application layers of encryption, nested like the layers of an onion, used to anonymize communication. Tor encrypts the data including the source and destination address multiple times and passes the message from one relay node to another through a random virtual circuit path. Each relay node decrypts a layer of encryption to reveal the address of next relay in the circuit to which the data needs to be passed next. Thus the various layers of encryption are decrypted one by one by the various relay nodes. When it reaches the final relay, it decrypts the innermost layer of encryption to retrieve original data which it sends to the destination. Each relay node in the path only knows the next node or next hop in the route and does not have complete route details. This eliminates the chances of tracking or tapping communication through network surveillance or other traffic analysis or tracking methods at any single node. TOR conceals the user's identifying information like IP address, location and browsing patterns from any attacker trying to do network surveillance or traffic analysis and other malicious users trying to break into the user's message or networks to access their personal information. Thus it makes it difficult to track the Internet activities and transactions to a particular user including visits to websites, online posts, instant messages, online transactions and other online activities. The main purpose of TOR is to protect the personal privacy of users by hiding their identifying information during Internet browsing thus providing them freedom and ability to conduct confidential online transactions and communication [2].

Effectiveness and Benefits – Proxy based anonymizers like TOR are very effective in protecting the user's online privacy and to stop any privacy violations or online activity tracking through network surveillance or traffic analysis. They also provide strong and effective protection against network hackers, ISPs system admins, website operators and malicious attackers trying to eavesdrop user's messages and communication. An eavesdropper trying to break into a proxy chain network will have to employ equipment and programs to watch all the proxy servers in the anonymizer's Internet network, which may be distributed at various locations around the world. Thus only a very strong adversary comprised of large organizations having huge dedicated resources can do the Internet network traffic interception and analysis required for this sort of eavesdropping. For all other casual attackers and small network hackers, anonymizers like TOR provide strong and effective protection for anonymous web browsing.

Limitations and Drawbacks – In a multi node proxy network, since the communication passes through a chain of proxies or computers, there is some degree of risk at each node for compromise of confidentiality, and this risk is directly proportional to the number of nodes. Proxy chain anonymizers have the same problem -- at each computer in the anonymizer chain there is a risk that it has already been compromised by the owner or an intruder and the communications can be tapped. Also anonymizers like TOR cannot and do not protect against monitoring of traffic at the boundaries of the TOR network. A small overhead experienced with the use of anonymizers is that relaying nodes can sometimes add a delay of few seconds if they are busy with too much traffic. To handle this issue some anonymizers keep a local cache of commonly accessed sites. Also anonymizers cannot protect the user from someone who has access to their computers at home or office. Though the TOR network was officially declared to be compromised in July 2014 by some attackers running a number of relays in the Tor network and the National Security Agency of the US government also seemed to be trying to get across TOR user's identities since quite some time, still TOR remains a strong favorite for the casual online users trying to escape from public tracking and scanning.

B. *Identity management systems –*

Technology – Identity management is a broad area that deals with identifying individuals in a system and controlling their access to resources within that system by associating user rights and restrictions with the established identity. It involves management of individual identities, their authentication, authorization, defining roles and privileges within or across the system. Identity management can involve three basic functions [3] –

a. Pure Identity function – this involves the creation, management and deletion of online identities of individuals. It is called pure identity in the sense that this model is not tied to any specific application context. A digital online identity may be created and provided to the user which may be a user-id or

b. User access (log-on) function – User access provides a specific digital identity to enable access to uses across specific or multiple applications. The use of a single identity for a given user across multiple systems eases tasks for administrators and users.

c. Service function – This involves providing stand alone, personalized, on-demand identity management services to users within an organization for internal as well as external communications.

Identity management systems include products following tools and technologies - directory services like Active Directories and Web Services, Service Providers, identity managers like Identity Providers, Access control and Digital Identities, Password Managers, authentication mechanisms like Single Sign-on, Security Tokens and OpenID, WS-Security, WS-Trust and Identity Protector tools. Active directory service is a directory service provided by Microsoft for controlling access, authentication and authorization for all computers and users in a Windows domain network. It provides a means for assigning and enforcing security policies for all computers and users in a network. Many identity management systems including Active Directory and Single sign-on employ LDAP or Lightweight Directory Access Protocol for user authentication. LDAP is an open, vendor-neutral standard application protocol used to provide a single sign-on where one common login and password for a user is shared by many services to provide him requisite access and authorities. Identity providers, Access control and Digital Identity systems provide Identity services to users on a network that include creation, modification and deletion of user identity data for all users. ID management systems provide administrators with the tools and technologies to change a user's role, to track user activities and to enforce compliance with corporate policies and government regulations on an ongoing basis. Regulating user access can involve a number of authentication methods for verifying the identity of a user, including passwords, digital certificates, tokens and smart cards. Password managers are applications that help users store and organize passwords by storing them in an encrypted format. Hardware tokens and credit-card-sized smart cards have traditionally served as one component in the two-factor authentication scheme, which combines something you know (your password) with something you have (the token or the card) to verify a user's identity. A security token is generally a physical device that may store cryptographic keys such as a digital signature or a biometric and that generates some security codes that can be used for user authentication. A smart card carries an embedded integrated circuit chip that can be either a secure microcontroller or equivalent intelligence with internal memory or a memory chip alone [4].

Latest Identity management tolls and systems can also provide online identity protection while using Cloud services and Mobile devices. Similar to Identity management tools are Identity protector tools that control the disclosure of an individual's true identity to various processes and applications within the system and Internet. It hides off personal identity details of the user from the Internet and other systems that do not require this information. It converts the user's real identity (like name, address, SSN, license number, account number, etc) into a pseudo-identity or an alternate identity that the user may use and adopt when using the system, Internet and applications. The Identity protector systems can take the form of a smart card or a smart token controlled by the user which could generate pseudo-identities as required, convert these pseudo identities to real or actual identity of the user at the application end thus helping to combat fraud and misuse of personal information.

An Identity protector system creates two domains – an identity domain in which the user's true identity is known and accessible and a pseudo-domain in which the true identity is hidden and not known to others. The Identity protector system keeps these two domains separate while deciding which user or system lies in which domain. The identity domain is kept as minimal and small as possible which means the true identity of user is revealed to a very few number of people or systems. The service provider and all other routine applications will record the user's privileges and activities in the pseudo domain under the pseudo identity. It helps to prevent frauds, identity thefts and misuse of personal information. [5]

Effectiveness and Benefits – Identity management systems help provide secure and controlled access to users within an organization at the same time protecting the users and systems against unauthorized access, frauds and cases of identity thefts. They also help to provide regulated access to users outside the organization without compromising the security of sensitive and confidential data of the organization. Thus it helps organization open doors to customers, partners, suppliers and contractors providing them access to their internal systems. This leads to increase in efficiency and lowering of operational costs as a common applications and database can now provide for both internal and external users simultaneously without compromising data security. Through their monitoring, auditing and reporting mechanisms ID systems helps to prevent cases of stolen identities, passwords, unauthorized accesses and hacked systems thus preserving data privacy and confidentiality.

Limitations and Drawbacks – Identity management systems are designed to authenticate and authorize authentic users. They are not designed for hacking and code breaking challenges. They are very effective against casual hackers and attackers trying to break into the company's private data but cannot protect against targeted attackers empowered with advanced computing systems and hacking techniques and technologies. Also they provide protection only for data stored within the company's data bases and applications, they cannot provide protection to data in transit that is when the data is being communicated or sent out of the company. Additional protection systems need to be employed for securing data networks and data in transit.

C. Privacy preserving data mining –

Technology - Data mining is an analytic process designed to explore and analyse large amounts of data in search of consistent patterns or systematic relationships between variables. The ultimate aim of data mining is prediction. By applying the findings from analysed data to new data, data mining models attempt to predict trends, outcomes and user preferences. Information gathering on the Web is pervasive mostly because usage tracking and data-mining technology are deeply integrated into most Web based software and application systems and websites these days. Through search engines, social networking sites and online applications, loads of customer data is collected and stored mainly with the aim of understanding customer needs and improving products and services. This all pervading trend of online data collection, storage and analysis becomes a privacy concern for many. Privacy preserving data mining lets business derive the understanding and trend information they need without invading the personal data boundaries. The basic approach followed in Privacy preserving Data mining is to let users provide a modified value for a sensitive data item or randomize sensitive customer data, so that it prevents the disclosure of any individual data but still can provide data for algorithms to analyze aggregate information, build mining models, and deliver actionable insights to businesses.

The proposed privacy preserving data mining techniques can be divided into two broad categories – Query Restriction and Data Pertubation or Randomization [6]. Query Restriction techniques include restricting the size of query result, controlling the overlap among successive queries, keeping an audit trail of all queries posted and answered by each user and constantly checking for any privacy compromises. Data Perturbation techniques modify or tweak the data so that it no longer remains real individual data. It includes methods like swapping values of attributes between records, replacing the original database by a sample from the original distribution, adding noise or randomizing factor to the values in the database, adding noise or randomize the results of a query and sampling the results of a query. Two randomization methods used for modifying and hiding sensitive data values are Value Class Membership and Value Distortion. Value class Membership is a kind of discretization or disassociation method in which the values for an attribute are partitioned into a set of mutually disjoint partition classes. Instead of storing the true attribute value, it stores the interval in which the value of the attribute lies thereby hiding the individual data value. The Value distortion method adds a randomization factor to hide or distort the true value of an attribute before storing it in the database. Another perturbation technique is data swapping i.e. exchanging data values between recods in ways that preserve certain statistics but destroy the real individual data.

For example consider a scenario in which an online merchant asks a Web site visitor for specific personal information such as age. The data entered by the customer is scrambled or randomized by a client-side software before sending it to the merchant for storage and analysis. The scrambling involves taking the entered number and adding or subtracting a random value. The software performs this randomization step independently for every user who enters this data. So, an entry of 30 might become 42, while an entry of 34 might become 28. The data base and data mining software used by the online merchants cannot determine the true age value of customers. It has access only to the randomized values and the randomization parameters. On the basis of this information only, the Privacy preserving data mining software can build a close approximation of the true distribution. This reconstruction of actual distribution will be accurate and sufficient fr data mining observations over thousands of people but it does not reflect correct data for single users thereby preserving individual privacy. [7]

Effectiveness and Benefits – Privacy preserving data mining techniques help us to employ data mining and make predictions and inferences on the basis of collected data without disclosing the private individual data. Thus, it provides a means whereby private data can be safely collected online through web applications and used for drawing conclusions and inferences without the risk of privacy violations. The key for privacy preservation is that you don't know which records are correct; you simply have to assume that the data doesn't contain real values. It disassociates the individual user from his original data or makes the data look like generic data. Thus it promises to provide a safe and secure environment to users where they can share their personal data without the risk and fear of any privacy violations. A study found that those organizations that posted privacy policies or notices including privacy preserving data mining on their web sites were able to put consumers more at ease when shopping online.

Limitations and Drawbacks – The downside of Privacy preserving data mining is that most of the work in this direction has been theoretical and it has not been implemented practically on a larger scale by most organizations and companies. Privacy preserving data mining has not generally been adopted by industry. The primary reason for its limited use is that any organization wishing to implement these methods must do a lot of ground work before implementing them. They need to setup a well defined and clearly stated privacy policy and privacy goals to exactly identify what data needs to be protected and till what extent it needs to be protected. Secondly data mining resources and processes must be defined clearly. Then PPDM algorithms need to be incorporated into existing data handling systems. Also the adoption and implementation of PPDM technologies may be quite expensive. Another main point to consider here is that PPDM technologies are devised to protect privacy while doing data mining, they are not data privacy methods, their primary focus is still on data mining and the best security and privacy they can offer may still be unacceptable to some customers and users. Finally we need to keep in mind that PPDM is not designed for thwarting hackers or preventing any security attacks on data.

D. Hippocratic databases –

Technology – Proposed by a team of researchers at IBM, Hippocratic databases ensure the privacy of data they manage by providing statistical high level information without disclosing sensitive personal information. Inspired by the privacy clause given in the Hippocratic oath of physicians, Hippocratic databases include the responsibility of data they manage by incorporating 10 fundamental privacy principles for protecting private data. These 10 principles are given below – [8]

1. Purpose Specification - For any personal information stored in the database, the purpose for which that information has been collected should be stored and associated with that information.
2. Consent - The donor or owner of the personal information should agree or give his consent to the purposes given in above point for storing personal information.
3. Limited Collection – As far as possible minimum and only necessary personal information should be collected for accomplishing the specified purposes.
4. Limited Use - The database should allow to execution of only those queries that are consistent with the purposes for which the information has been collected.
5. Limited Disclosure - The personal information stored in the database should not be communicated outside the database for any purpose other than those approved by the donor or owner of the information.
6. Limited Retention - Personal information should be retained only for as much time as it is necessary for achieving the purposes for which it has been collected and should be deleted soon thereafter.
7. Accuracy - Personal information stored in the database should be accurate and up-to-date.
8. Safety - Personal information should be protected and safeguarded against theft and other security attacks.
9. Openness - A donor or owner of personal information should be able to access all information about him stored in the database.
10. Compliance - A donor should be able to verify that the above mentioned principles are being adhered to. Also, the database should be able to address a challenge concerning compliance of above rules.

The Strawman architecture of Hippocratic databases implements the above principles by using **purpose** as the central concept to build privacy protection. It defines privacy metadata tables for each purpose and each piece of information. These tables store the following three pieces of information for each purpose and stored data item –

- external-recipients: the set of people with whom this information be shared with
- retention-period: how long the information is stored
- authorized-users: the set of users or applications who are allowed to access this information.

Thus, the privacy metadata tables define and describe who are the authorized users of data, who can access data and how long it will be stored on the database. These three pieces of information are stored further in two tables. The external-recipients and retention attributes for each data item are stored in the privacy-policies table, while the authorized-users attribute is stored and maintained in the privacy-authorizations table. The privacy-policies table captures the privacy policy, while the privacy-authorizations table captures the access controls that support the privacy policy. Thus all the privacy preferences of users related to their data are stored in these privacy tables. These privacy preferences are verified and authenticated whenever the related information is accesses or used.

Before the user stores any information in the database, the Privacy constraint Validator checks whether business privacy policy is acceptable to the user. The Validator takes the user's privacy preferences as input to verify its implementation correctly with the business. When user data is stored in the tables, a special attribute for each record called purpose is stored in the privacy tables defining the privacy preferences specified by the user for that data. When any user sends request to access the private data, the application will provide authentication details of the user and the associated purpose attached together with queries. Based on the purpose and credentials of the user, the system checks corresponding metadata stored in the database and ascertains whether the user issuing the query is a legitimate recipient of the requested private data for that purpose. Next, the Attribute Access Control of the database analyzes the query to find out if it is accessing any fields that are not explicitly listed for the query's purpose in the privacy policy. Finally, the Record Access Control of the database ensures that only records having a purpose attribute matching the query's purpose will be visible to the query. Before the query results are displayed to the user, the Query Intrusion Detector is run on the query results to find out any out of order queries or queries whose access pattern is different from the usual access pattern of queries received from that user. The detector builds a Query Intrusion Model for each authorized user by analyzing past queries and the purpose for each query and uses this model. Further the Data Retention Manager deletes data items whose purpose has been completed and they are superfluous and no longer required. If a certain data item was collected for multiple purposes, it is stored and maintained for the retention period of the purpose having the maximum retention time. An audit trail of all queries posted by each user is maintained to provide data for external privacy audits, as well as to address any challenges regarding compliance. [8]

Effectiveness and Benefits – Hippocratic databases can help organizations effectively manage the privacy and security of the data they collect and store. It is highly effective for protecting data privacy in medical, banking and financial applications, where privacy of personal data is very important. Hippocratic databases provide strong protection against accidental or intentional leakage of sensitive personal information from the data maintaining authorities. It is meant to protect all threats of data leakage at the source level i.e. through the repository of data or the database itself. It can also provide effective protection against hackers and attackers trying to infiltrate or break into the database using unauthorized means.

Limitations and Drawbacks – Hippocratic databases are aimed to provide data privacy protection at the database level where data is actually kept or stored. But it cannot protect against hackers or attackers that are trying to infiltrate or attack personal data while it is being transmitted and communicated through the networks. Also as all authentication and authorization provided by Hippocratic databases is tied to user ids, if the user id or password is hacked by an attacker or hacker then unauthorized data access cannot be prevented. Thus special care for keeping user ids and passwords is of utmost importance.

E. Privacy policy encoding –

Technology – P3P or Platform for Privacy Preferences is one of the most well known and widely used example of Privacy policy encoding. P3P is a standard or protocol developed by the World Wide Web Consortium or W3C to allow websites to declare their intended purpose and use of information they collect about web browser users and utilize it suitably while users browse the internet. Any website collecting user data and wishing to use P3P setup their privacy policy in a standard machine readable XML format. Users also specify their privacy goals or privacy policy stating what information they wish to be seen by the websites they visit in XML format. The privacy policy can be retrieved as XML files or can be included in the HTTP header of the webpage. When a user visits a site, the browser storing P3P policy will compare what personal information the user is willing to disclose, and what information the server wants to fetch – if the two do not match, it will inform the user by showing a message asking them if they are willing to proceed to the site to release more information or should restrict the website from taking further action. As an example, a user can store in the browser preferences that tracking information about which webpages he visited should not be collected. If the policy of a Website states that a cookie is used for tracking this information, the browser will reject the tracking cookie automatically. The

Privacy policy contains details of what type of user data is collected by the website (IP address, email address, name, etc.), for what purposes that information will be used (for regular navigation, telemarketing, personalization, tracking, etc.), who all can access this information (only the current company, third party, etc) and how long the data will be stored by the servers[9].

P3P provides users with a clear understanding of how their personal information will be used by a particular website. It provides an assurance to the users that their information will not be collected and misused without their approval and knowledge in any way. It provides an easy and convenient way for online users to decide whether they want to disclose information or not, what information they want to disclose and on which websites [5]. Browsers can this way develop and follow a predictable behavior when blocking content like cookies thus giving a real incentive to eCommerce sites to behave in a privacy friendly way. This avoids the current scattering of cookie-blocking behaviors based on individual heuristics imagined by the implementer of the blocking tool. Such properly documented and followed privacy policies can help in increasing user's confidence in online transactions as they are presented with more meaningful information and an easy method to control their online privacy and visibility.

Effectiveness and Benefits – P3P allows users and website owners to understand, state and implement their privacy policies in a simple, easy and organized manner rather than searching through the entire website. P3P promotes transparency and increases accountability on the part of website owners. By providing a system of notice and messages, P3P ensures that users know what information is collected about them and how it will be used beforehand. Providing the power to control your online privacy, P3P boosts the user's confidence in website security and provides them a hassle free browsing experience. The systematic procedure provided by P3P for stating privacy policy helps website owners uncover gaps in their existing privacy policies and become fully aware of its nuances and practical implications.

Limitations and Drawbacks – P3P is only a guideline and a standard, it is not an enforcing tool and its effectiveness in privacy protection depends hugely on its implementation. Many people consider it good only in theory but not practically effective since it does not set minimum standards for privacy, nor can it monitor whether websites follow their stated privacy procedures thoroughly. It just provides a language for stating the privacy policy and a guideline for enhancing privacy but it cannot enforce privacy. It cannot ensure that companies or organizations will actually follow the privacy policies stated on their websites. Addressing all of the complied, fundamental issues surrounding privacy on the Web require appropriate combination of technology, a legal framework and self-regulatory practices to be applied all together [5]. There has been criticism on P3P on account of being too complex and difficult for average user to follow. Another concern is that neither websites nor users are obligated to use P3P, hence the privacy offered by P3P depends on the proper implementation and understanding by users rather than being a guaranteed solution in itself.

F. Virtual private networks

Technology - VPNs or Virtual Private Networks provide a safe and secure way to communicate through a public network like the Internet. VPN allows private network administrators to utilize the public Internet services to provide the functionality and security of a private WAN connection for private communication at a lower cost. Most companies maintain VPNs so that employees can access files, applications, printers, and other resources on the private office network without compromising security through the Internet. Individuals can use VPNs to get access to network resources when they're not physically on the same LAN (local area network), or as a method for securing and encrypting their communications when they're using an untrusted public network. A VPN creates an encrypted private tunnel across a public network or Internet by establishing a virtual point to point connection through the use of dedicated connections. Users on remote computers wishing to connect to the company's private network have VPN client software installed on their machines. Any VPN client wishing to connect with the VPN network first initiates a connection to a VPN server using a Connection Manager by sending their authentication details. The VPN server then verifies the authentication details of the client through the Internet Authentication Service (IAS) server and authorizes a user session and maintains the connection until the communication is not complete. All local services available in the private network like file and print sharing, access to online applications and databases, Web server access, and messaging are made available to the remote client connected through the VPN.

Two main techniques used by VPNs to provide secure path through public networks are (1) virtual tunneling protocols like Internet Protocol Security (IPsec), Point to Point Tunneling protocol (PPTP) and Layer 2 Tunneling protocol (L2TP) and (2) encryption mechanisms like Transport Layer Security protocols (SSL/TLS). Tunneling provides a virtual point-to-point link between VPN clients and the private network by encapsulating or wrapping the data packets within a header. The header contains routing information that determines the path to be followed by the data packets through the public network to reach the destination. The encapsulated packets with the added headers are routed through the specified path to reach other end of tunnel over the network. The logical path through which the encapsulated packets travel through the network is called a tunnel. When these data packets reach their destination on the network, they are de-encapsulated by removing the header information and the remaining payload is sent to the final destination. Tunneling includes this entire process of encapsulation, transmission, and de-encapsulation of packets. Many protocols like Internet Protocol Security (IPsec) at the Network layer, Point to Point Tunneling protocol (PPTP) and Layer 2 Tunneling protocol (L2TP) at the data link layer can be employed to provide these tunneling and encapsulation services in a VPN. Further, in order to secure the data being communicated from being intercepted by unauthorized users, it is encrypted using cryptography. Transport layer protocols like SSL/TLS based on cryptographic techniques are used to encrypt data packets over the VPN tunnel. Encryption makes the data packets indecipherable without the encryption keys on the shared or public network so that even if someone taps the channel he cannot understand the actual data contents. The tunneled link in which the private data is encapsulated and encrypted is known as a VPN connection [10].

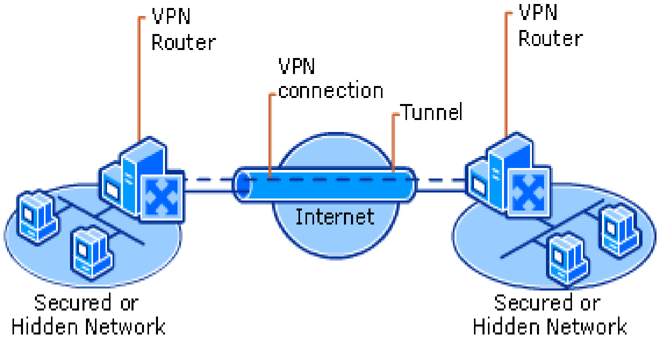

Fig. 1 – Tunneling in VPN [10]

A complete VPN solution frequently includes many advanced security technologies such as data encryption, encapsulation or tunneling, authentication, authorization and Network Access Quarantine Control. VPN clients typically need to logon to the VPN server by providing their authentication details (Login id and password) before making a VPN connection. The VPN server authenticates the VPN client and ensures that the VPN client has the requisite permissions before providing access to private network. Further L2TP or IPsec protocols can also perform computer level authentication to verify that the remote access client computer is trusted or not. To ensure data authentication and integrity VPN protocols can also include a cryptographic checksum based on an encryption key. Network Access Quarantine Control is used to delay remote access to a private network until the authentication of the remote access computer has been verified and approved. Only after the user's identity is verified, the VPN connection is established and the user is allowed to communicate with the private network. VPN security model ensures Authenticated access, Data Privacy, Confidentiality and Integrity of private data.

Effectiveness and Benefits – VPN provides a very reliable, secure and cost effective method for individual clients to connect to private networks through the Internet.
VPN can protect data against most of the network attacks like snooping, sniffing, spying and network hacking. Tunneling techniques helps to hide the IP address of VPN clients thus providing a means for anonymous, private and secure web browsing for individual users. It can help protect privacy of your sensitive data like bank account details, credit card details and passwords from snoopers while doing online transactions. VPN can also help to bypass filters and firewalls setup by network administrators to access all websites. It also provides protection against malwares and malicious websites by detecting and blocking malicious, phishing and spam sites from infecting the client computers. Companies frequently setup VPN as a means to provide remote access to employees to access their company's machine or intranet from home or while travelling outside the office securely. VPNs are also useful in connecting multiple private networks securely together. Most businesses big and small rely on a VPN to share servers and other networked resources among multiple offices or stores across the globe. Individual users can also use VPN to protect their online privacy while connecting to Internet through an untrusted network. VPN can provide cost savings for an organization by eliminating the need for expensive long distance leased lines.

Limitations and Drawbacks – Though VPN seems to be a panacea of all network attacks but it has certain limitations and some points need to be kept in mind while using VPNs. Proper and effective setup of VPN require detailed understanding of network security issues and careful configuration to ensure sufficient protection through a public network. Another point to note is that reliability and performance of Internet based VPN relies on the service provider and not directly under the control of an organization. The incompatibility of VPN solutions from different vendors may prove to be a limitation in connecting various private networks. Another important limitation of VPNs is that they are useful for point-to-point connections and do not support broadcast or multicast networks. The speed of VPN connection is determined by the slowest internet connection between the two end points. Thus unpredictable internet performance can create delays, disruption and degraded service. Also VPNs do not offer complete guarantee for Quality of Service and packet loss is variable and can be very high occasionally.

G. Anonymous Remailers –

Technology – Remailers are proxying tools used for sending emails anonymously. An anonymous remailer is a server that receives messages with embedded instructions on where to send them next and that forwards them without revealing where they originally came from. Different type of remailers like Proxy remailers, Cypherpunk anonymous remailers, Mixmaster anonymous remailers, and nym servers use different strategies for anonymously sending emails. Proxy remailers or Type 0 remailers, are simple proxies that accept mail from senders then strip off the mail's headers and forward the message on to the intendend recipient. The remailer stored a fixed mapping from pseudonyms (such as "an024601@anon.penet.fi") to real addresses (such as "alice@mit.edu"); if a message was sent to the anonymizer with an envelope address like "an024601@anon.penet.fi", the anonymizer would translate the address using its secret mapping, and then forward the message on to the real address. Type 0 remailers, which use only one proxy, suffers from a single point of failure: users must trust that the remailers won't be compromised, and won't divulge any information from its secret table of pseudonyms.

To address this issue, Cyperhunk remailers or Type1 remailers were designed that stripped the sender address from the mails and use a standard mixnet having remailer chains that prevent the single point of failure. Type 1 remailers use cryptography to send and receive encrypted mails that prevents eavesdroppers from linking incoming and outgoing messages. Type 1 remailers suffer from weaknesses that could allow an eavesdropper to link incoming and outgoing messages. One cannot reply to a message sent via a Cyperhunk remailer. Type 2 remailers or "Mixmaster" remailers addresses the weakness of Type1 remailers and counters traffic analysis attempts by adding a random delay in between receiving and forwarding messages and forwarding messages of a fixed size only. Type 2 remailers collect batches of messages, randomly reorder them, and then forward them to prevent eavesdroppers from linking messages simply by the order in which they arrive and leave, Type 2 remailers can also detect and remove duplicate messages, thus providing maximum anonymity to users. These remailers also allow replying to anonymous mails by including the reply address within the body of the email [1].

Type 3 or Mixminion Remailers further enhance the anonymity provided by type 2 remailers by addressing the challenges of replies, forward anonymity, replay prevention and key rotation, exit policies, integrated directory servers and dummy traffic. Mixminion uses a mix network architecture to provide strong anonymity, and prevent eavesdroppers and other attackers from linking senders and recipients. To send an anonymous message, mixminion breaks it into uniform-sized chunks (also called "packets"), pads the packets to a uniform size, and chooses a path through the mix network for each packet. At each mix server the encrypted message is decrypted, re-ordered and re-transmitted to the next relay in the path. The software encrypts every packet with the public keys for each server in its path, one by one. Every e-mail passes through several mixes so that no single mix can link message senders with recipients. Type 3 remailers provide almost complete anonymous solution to sending mails securely and anonymously [11].

Effectiveness and Benefits – The need and applicability of Remailers is huge since email is such a widely used and trusted means of communication these days. Type 0 or proxy remailers provide weak security but are quite easy to use and hence may be good for novices. Type 1 and 2 remailers are somewhat more difficult to use as they require knowledge of current remailer network as well as PGP procedures. Type 1 and 2 remailers are very effective against most of the attacks but still remain subject to several potential attacks. Type 3 remailers are more complex and need expert knowledge of relay networks and cryptography procedures. Type 3 remailers provide very effective security against all known attacks.

Limitations and Drawbacks – Use of Remailers is pretty much limited to sending anonymous mails only. Type 0 Remailers are quite useless due to weak security and are not used. Type 1 and 2 Remailers provide strong security and are used against usual adversaries. Type 1 Remailers are susceptible to attacks by network providers, or to anyone who can compel, bribe or convince these operators to divulge user's identity. Type 2 Remailers mitigate this risk but still are vulnerable to active attacks by dedicated adversaries. Type 3 Remailers provide the best of breed security features but need advanced knowledge and software for implementation. Finally all proxy chains and remailing networks are vulnerable to social attacks. If all the forwarding nodes or servers can be traced in the chain, the users can be traced easily. The security of users is as good as the integrity and security of each relay node. For maximum anonymity, users should use long mixnet chains with nodes located in different countries and run by reliable operators.

H. Digital Cash

Technology – Digital cash or Digital currency serve as secure and safe medium of transferring money while providing anonymity to payors at the same time. Digicash, InternetCash.com, Bitcoin, etc are examples of different types of Digital currency available today. Digicash was the first popular Digital currency. It allowed digital payments issued by a supporting bank using coins as paying medium. In Digicash the payor would choose few large random serial numbers and combine these random numbers with the denomination of the digital coin to be paid. These random numbers were then multiplied by a blinding factor to create few blinded number which were then sent to the supporting bank. The bank would choose one blinded random number and sign it with the Bank's secret key and return to payor. The payor is asked to surrender the serial numbers for other randon mumbers so as to authenticate his denomination. By using a special operation, the payor removes the blinding factor of the signed random number keeping the Bank's signature on the digital coin. This digital coin he then hands to the payee for making the anonymous payment. The Payee takes the coin to the Bank where t is authenticated and if it has not been used earlier, the requisite amount is paid to the Payee. If the coin has been redeemed or used earlier, the bank rejects the coin. This process is based on cryptographic techniques of private key encryption. [1]

Another type of Digital currency is the Crypto currency such as Bitcoin that uses cryptography for chaining together digital signatures of token transfers across many peer to peer networks. Bitcoin is a payment system based on open source software of a public ledger run on independent nodes in peer to peer networks. The public ledger also called the block chain is a software that records transactions in bitcoins and is maintained by a network of communicating nodes running the bitcoin software. Transactions of the form - Payor X sends Y bitcoins to Payee Z are sent and stored in this network using available software applications. Network nodes can validate these transactions, add them to their copy of the ledger and then broadcast these transactions to other nodes. Bitcoins are created as a reward for payment processing work, in which users offer their computing power to verify and record payments into the public ledger. This activity is called mining. Mining is a record-keeping service of the Bitcoins. Through this mining activity, the block chain is kept consistent, complete, and unalterable by various miners repeatedly verifying and collecting newly broadcast transactions into a new group of transactions called a block. The new block of transactions contains information that chains or links it to the previous block and gives the block chain its name. Each block in the block chain contains a cryptographic hash usually SHA-256 of the previous block. The owner of Bitcoins can spend Bitcoins associated with a specific bitcoin address using a private key. To spend Bitcoins, the payor must digitally sign the transaction using the corresponding private key. Only a user with knowledge of the private key can sign the transaction and spend bitcoins associated with it. The network verifies the signature using the public key. Crypto currency like Bitcoin offers fast and quick payments that can be done quite anonymously by paying a very small fee. The complete control and hiding of personal information in Bitcoin transactions protect users from any frauds, countefeits and thefts. [12]

There are other companies like InternetCash.com that sell anonymous debit cards. These cards sold anonymously at retailers offer another means of anonymous payments. These cards act like prepaid phone cards and are quite simple to use and create. Each card represents a balance and every time the card is used to purchase something, the account is debited. These cards are quite useful for anonymous buyers who don't want to disclose their identity while doing the transactions.[1]

Effectiveness and Benefits – Digital currency systems provide safe, secure and anonymous way of making payments and doing monetary transactions. Not only do they provide anonymity to the payer, but also makes it easier and faster for anyone to exchange currency by paying little. Both Digicash and Bitcoin helps customers to make anonymous payments that cannot be tracked by the bank, third parties, ISPs, website operators, law enforcement officials and malicious attackers. Bitcoin transactions are faster and cheaper than credit cards as they do not go through bank approvals, foreign exchange controls or other forms of human intervention. Transactions are processed by the bitcoin network immediately. Digital currency like Digicash and Bitcoins can't be faked or counterfeited and minimize the chances of identity theft. The identity of the payer as well as the receiver remain hidden and the transactions are completed automatically and instantaneously. It can be used for making very small monetary payments also. Such a technology may prove very useful in the future if pay-per-view websites or small licensing fees become popular. Digital cash would enable users to pay for such things without exposing their identity. If digital cash is built and used for the same transactions credit cards are used for, the technology should scale very well. Prepaid cards sold by InternetCash.com are the easiest to use and provides great benefit for people who do not have resource to set up financial accounts. Prepaid cards offer complete anonymity to users. Unless the consumer could be tied to some kind of pattern linking him to the cards, he is well protected.

Limitations and Drawbacks – The biggest drawback of Digital currency is that it can be used easily to do illegal transactions of large amounts. As these methods hide the identity of the payer they can prove a boon for illegal trading and selling. Crimes such as money laundering and tax evasion could become much easier with the widespread use of blinded coins. With respect to anonymity, Digicash provides good privacy against most of casual attackers but cannot protect against very strong opponents or local threats. The hacker or computational attacker may be able to identify the spender by somehow uncovering the blinding factor. They could accomplish this by infiltrating the spender's computer system or by some kind of brute force key attack. Bitcoin provides strong anonymity protection to most users, but a very strong adversary can break through the cryptographic security provided by bitcoin networks. There have been many cases of Bitcoin thefts where attackers have infiltrated the private keys of a victim's bitcoin address. Similarly, though prepaid cards provide sufficient anonymity, prepaid card transactions can be linked to a given card, and can also be linked to where that card was purchased revealing details of the payer. Though most of the digital currency systems provide strong and adequate privacy and anonymity to payers but still they are not very popular and widely used. Firstly, since they can pose quite a challenge to law enforcing authorities, so governments and authorities do not have encouraging policies for digital currencies. Secondly, as they can be used easily for illegal activities and evading laws, banks and financial institutions are a bit wary and too cautious in accepting Digital currency as a regular medium of financial transactions.

IV. CONCLUSION

In this paper we studied and anlayzed various Privacy enhancing technologies and techniques available today for protecting user's online privacy. The underlying technology for each privacy solution was studied and the applicability of the solution in various real life scenarios was considered. It is concluded that no single privacy solution can fit into all real life situations and handle all types of Privacy threats. For most of the real life situations often we would need to employ two, three or more solutions simultaneously in a supplementary manner to provide a comprehensive privacy solution for handling all types of privacy threats. Further the effectiveness and benefits of each privacy technology in handling various privacy threats was examined. It was observed that most of these technologies worked successfully against casual attackers and online hackers but cannot provide complete protection against the strongest adversaries equipped with vast computing power and resources. The limitations and drawbacks of each privacy solution was analyzed and noted down. It is being concluded that every privacy solution or technology is suited best to certain scenarios and is effective against certain types of privacy attacks. For the best possible all round privacy solution we need to take into account various scenarios and often combine two or more of these technologies in a supplementary manner. The current offering of anonymity tools is encouraging, but still a lot needs to be done and achieved in this direction and we would like to see what the next generation of tools can accomplish.

REFERENCES

[1] Brian Kim, Chris Laas, Shelly O'Gilvie, Alexander Yip, "Anonymity tools for the Internet", May 2001

[2] Rolf Oppliger, "Privacy enhancing Technologies for the World wide web", *Elsevier computer Communications*, March 2005

[3] Wikipedia - identity management - http://en.wikipedia.org/wiki/ Identity_management

[4] CSO online - http://www.csoonline.com/article/2120384/identity-management/the-abcs-of-identity-management.html

[5] Vanja Senicar, Borka Jerman-Blazic, Tomaz Klobucar, "Privacy enhancing technologies – approaches and Development", *Elsevier – Computer standards and Interfaces,* Jan 2003

[6] Rakesh Agrawal, Ramakrishnan Srikant, "Privacy preserving Data mining", *Proc. 2000 ACM SIGMOD Int'l Conf. Management of Data*, ACM Press, 2000, pp. 439-450.

[7] Roberto J. Bayardo and Ramakrishnan Srikant, "Technological Solutions for Protecting Privacy", Web Technologies, Sept 2003

[8] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, Yirong Xu, "Hippocratic Databases", Very Large Data Bases Conference, August 2002

[9] Platform for Privacy Preferences (P3P) Project – "Enabling smarter Privacy Tools for the Web", http://www.w3.org/P3P

[10] Microsoft TechNet - https://technet.microsoft.com/enus/ library/cc739294(v=ws.10).aspx

[11] Ian Goldberg, David Wagner, Eric Brewer, "Privacy-enhancing technologies for the Internet", University of California, Berkeley, 1997

[12] Wikipedia – Bitcoin - http://en.wikipedia.org/wiki/Bitcoin

Kriti Arora is working as Assistant Professor in Shyam Lal college, University of Delhi. She did B.Sc. Physics(H) from University of Delhi and then did MCA from Thapar University Patiala. Prior to shifting to academics, she had a rich experience of working in IT industry for 10 years in Mainframe technology domain. She is interested in Network and Cloud Security research and has written a couple of research papers on these topics earlier.