UNIVERSITÀ DEGLI STUDI DI SALERNO
Dipartimento di Informatica

Dottorato di Ricerca in Informatica
XII Ciclo - Nuova Serie

Tesi di Dottorato

The value of privacy: concerns, attitudes, behaviors online, and information protection measures

Raffaele Spinelli

Ph.D. Program Chair
Prof. Giuseppe Persiano

Supervisors
Prof. Vittorio Scarano
Dott. Delfina Malandrino

December 2013

Supervisor
Prof. Vittorio Scarano, Dip. di Informatica,
Università degli Studi di Salerno, Italy

Reviewers
Prof. Giancarlo Ruffo, Dip. di Informatica,
Università di Torino

Prof. Clemente Galdi, Dip. di Ing. elettrica e delle Tecnologie dell'Informazione
Università di Napoli

Graduate Group Chair
Prof. Giuseppe Persiano, Dip. di Informatica,
Università degli Studi di Salerno, Italy

Dean
Prof. Vincenzo Loia, Dip. di Informatica,
Università degli Studi di Salerno, Italy

Thesis submitted on
December 20th, 2013

Date of defense
April 11th, 2014

Raffaele Spinelli
ISISLab, Dip. di Informatica
Università degli Studi di Salerno
Via Giovanni Paolo II, 84084 Fisciano (Salerno), Italy
spinelli@dia.unisa.it

Abstract

Most of our lives takes place on-line. Our on-line activities, affect directly or indirectly the way other people perceive us. One have to careful decide what to expose and what not. There are a lot of personal and sensitive information that people could unintentionally disclose.
Indeed an enormous amount of data is being generated and can be disclosed by an increasing number of people on the Web, often without know who is recording what about them. The odds of being tracked without full acknowledge is growing mainly because of two reasons: the exorbitant number of company in the behavioral advertising field and a market overfilled with free services to attract users.
This thesis focus on the study of the value of privacy, as intended by people. Learning the value of privacy is of great importance. How people value their own privacy affects the way relationships among individuals are created and maintained. Not only, it have implications on how an individual relates himself with the world, it influences user behaviors and attitudes.
The mechanisms responsible for how people value their own privacy are bounded to the perception of risks and perceived level of fairness of the outside world. That component is the *awareness*. The way an individual perceives the risks around him/her, represent one of the big challenge in order to fully understand the way people value their privacy. A better understand of those mechanisms and an increased awareness will help to design and build privacy by design systems.
Increased awareness can help users to understand how and why their privacy is mined, and to become more informed about what silently happens during their navigation. *Learning from disclosure* of personal information may help to discriminate potential harmful activities from daily and regular activities that can be performed. Awareness could help people to make informed decision about privacy online, and adopt countermeasures if necessary. Protecting users on-line from privacy risks is a difficult task. Task made even more difficult by users' attitudes. Users are not fully aware of the risks of privacy leaks, even after the increasingly number of press reports about privacy leakage and personal information disclosure on the Web. They ignore that their data can be collected, aggregated and linked with ambient information for various purposes.
Anyway, even if awareness is not the only mechanism involved in evaluating privacy, it can be used to study if a privacy tool can help users to make informed decision to reduce their exposure while on the Web. To this aim, we conducted a study to analyze general perceptions, attitudes, and beliefs about privacy online, with a focus on the mutual influence with users skills.
We discovered mechanisms responsible for how a person value its own privacy: a) skills influence the perception of privacy risks b) privacy is worth the price if it is a

side effect of another well-recognized benefit.

Acknowledgement

This dissertation would not exists without the help and support of many people. I would like to thank my advisor, Prof. Vittorio Scarano and Researcher Delfina Malandrino. They helped me to better articulate my ideas and strengthened my skills as a researcher. They taught me to step back and see the bigger picture.

I am thankful to the whole ISIS research Lab: Delfina, Gennaro, Ugo, Roberto, Pina, Ilaria, Donato, Angela, Nicola, Luca, Carmine, Gigi & Andrea, Prof. Alberto Negro and Prof. Filomena De Santis. I would like to thank other people that shared life in the Lab in the past years and now work in other research facilities: Biagio, Rosario, Bernardino. A special mention for Angela and Nicola, that increased the "humanist" aspect of the Lab.

I am grateful to Balachander Krishnamurthy too, he is a unique person with an incredibly mind. Talking to him always lead to new idea and hypothesis to verify.

I have to thank the Prof. Salvatore De Pasquale of the the Department of Physics, and the colleagues that work in the NEMES Lab. Their approach improve my skills as a researcher. Especially Salvatore D'Ambrosio, who spent a lot of time with me during several experiment for other works.

There are many friends that I should thank, they made the years spent at the University unforgettable. I really enjoy those years. It is a very long list, and I will surely miss someone. The list comprises Edoardo, Gabriele, Saverio, Giancarlo, Rosa, Pietro, Anna, Barbara Farina, Pasquale di Giovanni, Vincenzo De Maio, Arcangelo, my roommates Angelo, Fabio, Antonio and many other.

A special mention goes to Biagio, who I appreciate and admire for his determination, he taught me a lot about the research process.

Gabriele deserve a special thank too, comrade of a lot of adventure. He supported me many times.

A special mention goes to my beloved one, Sabrina.

Last but absolutely not the least, comes my family who has always supported and encouraged me to grow, despite the difficulties encountered. My mother *Genoveffa*, my brother *Gerardo*, my grandparents *Rosa* and *Carmine* and my father *Carmine* that taught me to follow my passions, but did not make to see it.

Contents

List of Figures

List of Tables

1 Introduction

The discussions of privacy is a very old topic. It is a topic that get renowned each time there is a technological shift. Its fame is proportional to the impact of the technological shift. Every new technological advance of the society have an impact on the lives of every individual. They change multiple aspects of our lives, how we perceive the world, how we spent our time, even our wellness.

Most of all, technologies aims at make life easier and more *enjoyable*. But every time our lives are shaken by the introduction of new technologies, new concerns arise about the consequences of this introduction.

When photography was first introduced in newspapers, people were concerned about potential privacy violations that could happen due to the publication of photos that violate the privacy of an individual. This problem was so prominent that an attorney and a lawyer at the end of the 19th century coined one of the most well-know definition of privacy, because of the problems due to the new technology.

In the last twenty years the exponential growth of Internet, Web and computers diffusion created an environment where people can connect to each other and communicate. Companies offer services on the Internet, open online store to sell their merchandise, create innovative way to communicate with the rest of the world.

This technology have an huge impact on our lives. A good piece of our lives is spent online. We create an image of ourself, motivated by our desire to keep in touch, a sort of digital projection that depict us online and available to whoever want to know us better.

This digital picture of ourself can be less or more detailed. It can be just an email, a Web page (personal or institutional), or an Online Social Network (OSN) account, and several others. In a world so connected, social and open, an even increasing number of individual are being exposed to an enormous amount of information, online services, and new way to communicate and co-operate with each other.

With this profound shift in our lives, concerns about possible privacy violation arise again. People nowadays communicate with email, SMS, messages on OSN (Facebook, Twitter), and use several online services to search for subject of interests, read a newspaper, find a restaurant, plan a trip for holiday or work, check their savings at their bank, buy some goods, and discuss health problems. The list of such activities is never ending and includes very sensitive topic (health problems).

Given the enormous impact of this technologies on our lives, a question arise about what privacy actually mean in our days.

Various meaning of privacy exists, one of the most used, and old definition, is the one given by Samuel D. Warren and Louis D. Brandeis: "*Privacy is the right to be let alone*" [80], back to 1890. At the time, journalists started to use photography in newspapers, leading to concerns about possible privacy intrusion due to photos.

Similar to that, privacy is also defined as *the right to prevent the disclosure of personal information* [81]. Another definition for privacy is the ability of an individual to control the terms under which his/her personal information are acquired and used [22]. Privacy definitions are very sensible to context, as it can be argued from this definition too: "*Individuals have privacy to the extent that others have limited access to information about them, to the intimacies of their lives, to their thoughts or their bodies*" [68].

Recently a group of researcher [84] from the Faculty of Law from the University of Haifa, gave this definition:

> The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, property, thoughts, feelings, secrets and identity. The right to privacy gives us the ability to choose which parts in this domain can be accessed by others, and to control the extent, manner and timing of the use of those parts we choose to disclose.

The key point in all these definition is the "*ability to choose*", the capability for a person to control how their information can be accessed by other individuals and entities. Today, with all the advances in technologies, is hard to be able to control how one own personal information are disseminated and used.

The value of privacy, specially in our days, is of great importance. And its relevance has been deeply analyzed by Acquisti et al. [3]. Companies have huge interests in understanding the value of privacy and how people value their privacy. The main reason is that understand how people value their privacy, will help them to design their services in order to make their users feels comfortable while using them.

Moreover companies have interests to know how people value their privacy, because showing that the company do care about its own users' privacy, create a good image and make the company more attractive. It can result in a huge advantage respect their competitors.

The reasons to study the value of privacy are not limited to only economics ones. It has great resonance for legal communities and for policy makers. Legal experts and jurists study the problems of privacy in everyday life, what does it means in our days, and how the society should regulate privacy questions and protect citizens privacy. Moreover judiciary are making a big effort to understand the value of privacy in our time, and adapt the laws to fits the needs of the modern society.

This work is needed because there is no explicit reference to privacy protection, in the constitution of several big countries. There is no explicit reference to privacy protection or right in the American constitution, or in the European Community. Few country have a constitution that explicit talk about privacy or have specific laws about it, like Brazil, Canada and Australia. Therefore a better understand of the value of privacy would help to improve both the legislation and lawmaking process, and regulate the way private information of an individual can be accessed.

Without a proper regulation, privacy of individuals can be at risks, personal and sensitive information can be accessed by *third party* entities without consent. Privacy violation can have different nature: illegal activities as identity thefts [8, 13, 67], or

for fall under the *National Security* jurisdiction[1] or for behavioral/target advertising.

Privacy incursions can be analyzed regardless of the motivation for which they happen. Indeed, in a taxonomy of privacy violations [70] four groups of activities have been recognized as harmful for both daily life and online privacy of individuals. The first group, "*Information Collection*", includes all activities related to *surveillance*, that is: watching, listening to, or recording of an individuals' activities. The second group of activities, "*Information Processing*", involves the way information is stored, aggregated, linked and used. *Aggregation* (i.e, combination of various pieces of data about a person), *identification* (i.e., linking information about specific people) and *secondary use* (i.e., information collected for a specific purpose and maliciously used for different purposes) represent the potential harms of this group of activities. The third group of activities,"*Information Dissemination*", involves the dissemination of collected information through, as an example, the *disclosure* of both sensitive and personally identifiable information, while the final group, "*Invasion*", involves invasions into people's private affairs.

Because we cannot know how collected data about users are effectively used, we focus on the first groups of activities. One of the major reason for user tracking, and an interrupted demand for large amount of individual information, is the behavioral advertising practice. The practice increases the effectiveness of the marketer's campaigns[2]. But at the same time it also gives rise to privacy concerns since private information may be collected and centralized by a limited number of companies [18], sometimes referred as *Aggregators*. The situation has been exacerbated through the introduction of the aforementioned free popular services, such as Online Social Networks (OSN), in order to extend the ability of advertising companies to deliver targeted advertising. Users effectively pay for these free services through micro payments of ever-greater amounts of personal information.

Numerous free services came up in the last years, creating an environment where people can unconsciously disclose personal and sensitive information.

Nowadays people using these service, likely give them personal information in exchange for the service itself, for customization, or just to have their data always available, across all of their devices[3] (i.e. cloud services). Moreover privacy policy of these services are rarely[4] read and, when read, not well understood. But they do influence user, as Stutzman et al show in their work [74]. They find that users who report to have read more of a site's privacy policy tend to disclose less. An interesting study on behavior [32] show that only a small percentages of users read the privacy policy of Web Sites and, in general, most of the users are not able to reliable understand their content [10, 54], by exhibiting a little willingness to adopt privacy protective technologies [2]. The main reason is attributed to users, that find learning about privacy and reading privacy policies difficult and time consuming.

Because of this apparently apathy of users for privacy, several companies are collecting an enormous amount of data about users on the Web. On top of that, all

[1] WashingtonPost - NSA Infiltrates links to Yahoo and Google
[2] Targeted advertising is more than twice as effective as non-targeted ads [9]
[3] http://www.zdnet.com/blog/btl/how-far-do-google-drives-terms-go-in-owning-your-files/75228
[4] Less than one out of four users claims to have read Facebook's privacy policy [5].

the data gained by third party entities are used in ways that are not always known.

Numerous press report have been published about privacy scandals or misuses of users data. One of the most famous case of misuses of users' data is the *AOL Search Leak* case. In the 2006 AOL released a file for research purpose, containing 20 million search keywords used by more than 600,000 of their users. The file was made anonymous, but personally identifiable information were still available in many searches. Thank to that, it was easy to identify an individual with their associated search history, like the user 4417749[5].

The famous social network Facebook have been trough a lot of privacy problems too. One of them regards several of their popular applications (so called Apps). Apps are software made by other independent company or individuals available through Facebook. The Wall Street Journal made an astonishing discover in one of their article from the series "What They Know". They find out that several Facebook Apps sent the Facebook ID to third party companies (advertising and tracking company). Note that the transmission of this information was not bounded by privacy settings. Although it did not seems serious, it has to be stated that the Facebook ID is a string that uniquely identify a user. When the ID is available, discover the first name and last name of an individual became a trivial task. It is as simple as open a Web browser, go to Facebook and append the ID to the URL. This procedure works even if the user associated with the ID have all of his/her information set to private. It affected every profile that uses the application in question. While the ID alone do not grant any access to private information, it uniquely identify the person, allowing a third party to profile a user and associate a name, surname ant potentially other information to it (including a picture), leading to serious concerns about privacy of individuals.

At the time of the investigation, they found out that the 10 most popular Apps on Facebook, were transmitting users' ID to outside companies. Some of them even transmit personal information about user's friends.

What happen clearly show that companies that do even partial tracking, can easily associate a real person to a profile. The risks of misuses or leak of information is not to be underestimated.

Interestingly, given the enormous amount of press reports about privacy problems and the severity of the problems reported, users seems to not care about it.

Unfortunately, this apparently inconsistency does not help to clarify whenever users are really worried for their privacy and how they value it. Studying the way users value their privacy, one prominent mechanism is related to the perception of fairness of the outside world. If there is trust between an individual and a company, one is not worried to disclose personal information. Unfortunately most of the time people give trust on false basis or without enough information to make an informed decision.

Numerous studies show that people do not value their privacy, while they are still concerned about it. People are sometimes scared of being online, or open an e-commerce for privacy related concerns, but are more than pleasant to freely share personal information on social networks or e-commerce sites (as clients).

[5] http://search-id.com/user/4417749-thelma_arnold

An interesting argumentation comes from Odlyzko [58], he introduce the aforementioned contradiction on privacy, arguing that the main responsible for it relies in the benefits of price discrimination[6]. Price discrimination is a largely adopted pricing strategy, used for more than two century, that increases revenue by allowing new transactions that otherwise would not take place. The main idea is to sell the same service to different people at different price. The price is proportional to the willingness to pay of customers.

The public dislike of price discrimination provides incentive to hide it

This relationship has contributed to the increase of privacy violation and intrusion, because privacy intrusion helps to improve price discrimination without trigger raging comments from the government or associations. Profile and categorize users favor the partition of the market, helping to apply a price discrimination strategy. The so called "free" services are offered in exchange of micro-payments in personal information.

Another interesting hint on how people value their privacy come from the experiments done in numerous store[7]. Famous brand are testing technologies to track customers inside their store. All the experiments aim at obtain information on customers behaviors. They track customers by monitoring them with several technologies: a) using the Wi-Fi signal of their smartphones b) smart cameras that can tell the mood of the customers and what goods they were looking at c) special mobile App of the store. They aim at understand how much time customers spend in each sections of the store or how long they look at merchandise before buying it.

The goal is to use these information to decide how to change the store layouts, or use them to give customized coupons. It is not different from what an e-commerce site actually does in order to propose goods that a specific client could find interesting. But when the employees posted a sign telling customers that the store was tracking them, shoppers were frightened and discouraged to continue shopping.

While people seems to not care about privacy, tracking and profiling online, they do care about it when they know that someone is watching over them. Only a few manifested their desire to happily trade some of their private information for a coupon. The experiment conducted by Beresford [11] reveals that when a person is faced with a request for personal data in exchange for a cheaper price or a coupon, they do actual choice to give their data.

Unfortunately this exchange is not always transparent, therefore the importance of awareness and its influence on how people value their privacy, is not to be underestimated. An exhaustive knowledge of the extent of privacy leakage and a better understand of the mechanisms that guide people to value their privacy are the goal at which we aims.

As for the definition of privacy, the definition of privacy awareness is not well established in the literature, too. Anyway, privacy awareness of an individual should encompasses the perception of:

[6] http://en.wikipedia.org/wiki/Price_discrimination

[7] http://www.nytimes.com/2013/07/15/business/attention-shopper-stores-are-tracking-your-cell.html?_r=0

- *Who* is is tracking, receiving or collecting his/her information
- *When* information is collected
- *Which* information other entities receive, store and improperly use
- *How* pieces of information is processed, linked and aggregated to build detailed profiles
- *What* amount of 3rd party objects can undermine users' privacy on the Web

In general, users do not have perception of all the aforementioned information. Users are someway aware that their activities are largely monitored by third party entities, but they are absolutely not aware of which data is disclosed to them, and how those data may be used to derive their identity.

As a matter of fact, both awareness and full control are considered very important requirements as recommended by the Federal Trade Commission (FTC) [20], the Privacy Right Clearinghouse (PRC) [62] and other important privacy watchdog organizations.

In order to protect their privacy, users need to be, "*in primis*", educated about the existing and emerging tracking mechanisms used by third party sites and next, they have to be informed about which pieces of information is disseminated, collected and aggregated, to build consistent and truthful dossiers about them. This knowledge will contribute to increase user awareness about the dangerousness of certain activities that are increasingly airily performed on the Web, or at least, make people aware of the wide range of used tracking mechanisms and of their risks for privacy, leaving them at the same time, the full control over their own personal data and on their resultant behaviors.

1.1 Extend of privacy leakage

As stated before, given the high value of users information, in online communication, is important to give to users the concrete possibility of enhancing their privacy. Nowadays, increasing amounts of both personally identifiable information (PII) and sensitive (e.g., medical, financial and family) information continue to be leaked to third party entities [39, 41, 49].

Privacy can be undermined by third parties [21]. They intervene as uninvited guests during common online activities such as: Web searches, online shopping, online business and financial transactions, social activities and during any type of communications on the Web. However, the real risk is about the final use of these pseudo-anonymous data, that linked with personally identifiable information (i.e., phone number, credit card number, social security number and so on), may be disclosed or explicitly sold to third party entities. There exists companies (data-brokering companies) that collect, aggregate, link data and sells users profiles as their core business[8]. There have been action on the opposite side too, for example

[8] http://www.propublica.org/article/yes-companies-are-harvesting-and-selling-your-social-media-profiles. Sometimes even Internet Service Provider have sold their customer click-stream[9]

abine.com created and offers a service to remove own data across several aggregators[10].

Even pseudo-anonymous data collected and linked with PII such as email addresses and credit card number, may be sold by aggregators. The possessors of such data may use it for identity theft, social engineering attacks, online and physical stalking and so on[11] [24, 30, 42, 61]. On social network sites, individual continually share personal data, without being aware of how their information can be accessed. Liu et al [44] conduct an experiment to test if Facebook users are aware of how effective their privacy settings are. They find that almost three user out of four believe that their settings were more restrictive. They show a significant gap between the desired privacy settings and the actual privacy settings.

It would be useful to be able to measure the privacy lost. Unfortunately, privacy is usually a hard thing to measure, especially since it's hard even for users themselves to quantify. For example, photos alone are likely to have wildly varying privacy requirements, depending on who is in the photo, where it was taken, etc. Moreover, because very little is known about procedure and use of these data inside the companies, an accurate measure of what is the extent of privacy problem is not possible.

Some researcher in thsi field aimed at design mechanisms to clearly show this hidden transaction of personal information. Several solution to the problem have been proposed, but no one was adopted. In particular Riederer et al [64] propose a new mechanism called transactional privacy, that aims to reach a fully transparent transaction, where all the parties are aware of what is going on. Users control which information is released, and what is its price. Aggregators, instead, purchase these information to serve personalized ads to users. This idea, and other similar have been used in several context. Leontiadis et al [43] propose a similar idea for the advertising market on mobile devices.

Unfortunately those mechanisms are not enough attractive for companies operating in the market, that do not adopt them.

There are few points that were not investigated in depth in literature, that will be discussed here

The extent of privacy incursion in online activities Although numerous study have proved that personal and sensitive information are leaked to 3rd party sites [39, 41], there are aspects not investigated in depth.

Analysis of the online privacy threats, categorization and classification During these studies, online privacy threats are classified based on their nature, and ordered by their usage. Furthermore we design a hierarchy of the most important privacy threats analyzing the ways in which personal and sensitive information are sent to third party sites.

Study of information trading between aggregators The daisy chaining is a well-know mechanism used by aggregators to exchange information between them.

[10] http://abine.com/cnn/

[11] `http://www.priv.gc.ca/media/nr-c/2012/nr-c_120925_e.asp`

But there is no information on the extent of its usage. We studied leakages that may occur via communications among them.

Extend of privacy leakage, and aggregators capabilities In order to show how far the extent of privacy leakage is, we study if it possible to identify a user. We study if it is possible to derive its interests and browsing habits, by linking bits of personal information leaked towards different third party sites. Specifically we examined what fraction of a user's profile is available to the different aggregators.

Effectiveness of Privacy Enhancing Tools A study of the tools available to protect privacy, a comparison between them on performance, effectiveness and efficacy. We derive an ordering of the importance of existing tools according to the countermeasures they provide and their effectiveness in limiting the disclosures of important information. The order is based on the hierarchy of privacy threats.

1.2 Behaviors, perceptions and concerns

While it seems that user do not care for privacy, other statistics show that 63% of users agreed with a statement of concern for third party monitoring activities [83]. Moreover, Krishnamurthy *et al.*, in another study, emphasize how critical is the situation showing that 56% of sites analyzed (75% when considering userids) directly leak sensitive and identifiable information to third party aggregators [39]. User are worried of disclosure of personal information to third party sites without their permissions. This represents the greatest concern among them, that demand more control over the disclosure of their personal data and more effective solutions to protect themselves.

To be able to help users to make informed decision, a better understanding of how they value their privacy is needed. Two aspects have to be considered:

- Individuals have different thoughts and beliefs, with behaviors that change according to environmental and personal factors or user's orientation [25].
- Individuals are torn between the desire for privacy and the desire for personal communication with others [37].

We verify the impact of awareness on users, considering several factor. We expose users to the leakage of their personal and sensitive information, and see their reaction. One interesting question is to verify if, by seeing the leakage, they would take steps to protect themselves.

To let users face the privacy leakage of their data, we make them use a tool that shows leaked data named NoTrace [46–48]. For the experiment we recruited students from University.

We study whether their skills influence attitudes toward privacy, privacy-related behavioral intentions and actual behaviors, and whether awareness and learning from own behavior may increase the willingness of people to limit the diffusion of

personal information and protect their privacy while they are navigating the Web. The study, focus on two subgroups of students, with different skills. The students were surveyed about:

1. General understanding of privacy concerns about online activities.
2. Skills and attitudes toward privacy and behavior.
3. Willingness to limit the diffusion of their personal information

Overall, we want to evaluate whether the skilled Group would have more concerns about privacy when compared with the non-skilled Group and if the skills can influence the perception of the risks about privacy, as well as the corresponding privacy behaviors. We want also to evaluate whether awareness about information leakage and learning from online personal habits and behaviors may affect in the same way both groups.

2 Related Work

Many researcher have focused their attention on issues related to privacy online. The literature is plenty of works about privacy problem, starting from the simple analysis of potential privacy issues to extended discussions. A big effort has been done to understand the dichotomy of users apparently concerned about privacy.

Castelluccia [18] demonstrate how real are the privacy problems due to behavioral advertising. They show a simple algorithm to derive user profile and interest starting from only the customized advertising a user receive.

Several researcher have proposed mechanisms that could help to mitigate or solve the problem. Because there is no possibility to intervene in the processes that take place inside the aggregators, the only concrete way to proceed is to prevent information leakage. Every concrete solution, is based on this approach. There are client-side tools and technologies, that aims at preventing information leakage. Most of the time they are browser extension that help users to limit the diffusion of their personal and sensitive information.

This approach, deployed client-side, is obviously limited because it cannot manage what happen once an aggregator have these data. Anyway the proposed solutions are not limited to this approach. Other solutions aims at solve the problem at its core. Other studies focus on the creation of mechanisms that will not unsettle the current economic market, while improving the privacy. Some other studies, instead, aims at find mechanisms that just make the process of the economic market more transparent, so that user could make informed decision while on the Web. Anyway, as stated before, those kind of solutions are not really adopted. Till today, users can only be helped in preventing information leakage. Therefore the analysis focus on client-side tools.

To help users protect themselves, numerous studies aims at resolve, or at least to achieve a better understand of the *Privacy Paradox Problem*.

Norberg et al. [57] analyze the dichotomy of privacy, discussing advantages and disadvantages of having a legislation that protect users' privacy while the users willingly disclose personal information. Because despite protestations of users, they are the conduits of their information leakage. Rosemberg [65], among the others, state that the most effective way to control the use of users' information is to prevent it from ever coming into other hands.

Furthermore, from a policy maker's perspective, it is weird to protect users from their own chosen behaviors. Therefore the response is to give them more consciousness about what goes underground while on the Web, in such a way that they are capable of make informed decision.

2.1 Privacy leakage

All the studies that aim at quantify the extent of privacy leakage, usually relay on experiments done client-side. Most of the time the data are obtained using a Proxy/Sniffer. There are several tools available for this goal: MITMProxy[1] , Bro[2] , LiveHTTPHeader[3] , etc.

Studies on the extent of privacy leakage usually use a methodology called "surface crawling". That is a visit to the home page of the Web sites without following other links. This technique is mostly used to derive an order of the top used threats for privacy.

Indeed the UC Berkeley Center for Law and Technology[4], recently conducted a research to quantify vectors for tracking individuals on the Web. They find that the top 100 Web site dropped thousands of cookies, and that 84.7% of them were third-party cookies. They also find that Flash cookies are still used, but their use is declining among the most popular Web sites in favor of HTML5 LocalStorage.

This methodology lacks an important factor, it does not take into account the human interaction that take place on Web sites (e.g. adding items to a shopping cart, comment on an item) and did not follow links set by JavaScript code in response to users interaction. But most of all this approach do not maintain an identity across all the visited Web sites, because it does not login on any of them.

Other work do not focus on privacy threats, but on the study of concrete privacy leakage. Krishnamurthy et al. [41] during a study that lasted few years and involved more than 1,000 Web sites, made several interesting findings:

- The tracking activities have grown from 40% to 70% in few years (Oct'05 - Sep'08).
- Now only few families account for more than 75% of the tracking, due to companies acquisition.

In another study Krishnamurthy et al. [39] build a dataset composed by the top-100 Web site from 12 Alexa categories and sub-categories. They simulate user interaction by using a surface crawl, opening each one of the Web site, and login on them. They trace everything using one of the tools briefly described at the beginning of the section. Their results show that 56% of the Web sites in their dataset directly leaked pieces of private information to third party aggregators.

The major vectors of threats for privacy remains constant, their popularity vary. The only exception is the threats that come from technologies for local storage on the Web browser. In the last 5 years there was a technological shift from Flash Cookies to HTML LocalStorage. But while Flash Cookies use is decreasing in favor of HTML5 Localstorage, they are still used [53]. Furthermore there are insight of leakage presented in a study [7] where authors showed Web sites that had HTML5 local storage and HTTP cookies with matching values.

[1] http://mitmproxy.org/
[2] https://www.bro.org/
[3] http://livehttpheaders.mozdev.org/
[4] http://www.law.berkeley.edu/privacycensus.htm

Several technologies have been developed to give users the concrete possibility to protect their privacy when they are online. Most of them are stand-alone programs or extension of Web browser. The most famous one is AdBlock Plus [60] that realize the canonical filtering mechanism of blacklist and whitelist. The tool is designed to use crowd-sourcing to be useful for several purpose: blocking advertising, blocking tracking objects, enhance privacy, etc. Indeed there is a rich group of people working on lists' definition for every need. Anyway, AdBlock was design to block advertisements, and not as a tool for improving privacy online.

Instead, RequestPolicy [66] and Ghostery [28] were explicitly designed with the privacy implications of third party requests in mind. The former was meant to give the users the full control over communication between Web site and third party domain. Its mechanism is as simple as genial. For each domain, the user can chose with which other domains it can communicate (third part request) and with which not. This mechanism is powerful and help users to protect themselves, but have two main drawback: 1) It assume that a user know, or is willing to learn, which domains are malicious 2) It need a different tuning for each user. The latter focus on privacy, and give to users the control over the third party request, requests directed to a specific domain can be blocked. Instead NoScript [50] was meant to be a security tool, to prevent execution of third party (in general unwanted) JavaSript code inside the Web browser.

Other browser extension focus on specific threats, like RefControl[5] that filter out only the Referer header of HTTP request. Another example is Milk, a browser extension that alter the HTTP request/response and rewrites the HTTP cookies to strictly bind them to the first-party domains from which they were set [79].

Moreover for all these tools, there is no experiments to assert, correctness, effectiveness, or performance.[6]

Furthermore, a list of important requirements [63] that tools for privacy protection should exhibit was drawn up to better classify tools.

- **Measure privacy attitude of people**, gathering, for example preferences from observations of their behaviors. It is important to highlight that this practice should be not privacy-invasive and should involve a clear consciousness in the analyzed individuals.

- **No invasion to privacy itself** avoiding interruptions that may annoy users and and leading them to relinquish the used tools.

- **Understandable for target group** by presenting data to users in a way that they are able to handle it cognitively.

- **Tailored to the specific situations and users** by presenting information to users in a way that they are able to understand according to expertise.

[5] http://www.stardrifter.org/refcontrol/

[6] Regarding performance aspect, only a simple experiment of correctness and impact on Web sites has been presented for RequestPolicy [66].

- **Offer support, no assumption of responsibility** helping users to make informed decisions.
- **Performance and Effectiveness** in order to make the tools longer used by users since excessive delays involve an abandonment by users after first use [26].

Again, one of the key point is to give users the capabilities to make informed decision, goal that could be achieved improving privacy awareness. Most of the work reported in the literature address the problem of privacy awareness only from the point of view of the availability of privacy policies on the Web sites [36]. In addition, in spite of the concerns about risks connected to their privacy, only few users are really aware of what happens behind the scene. Regarding Facebook, for instance, users put more trust in the service provider than in average Facebook users [5]. More than half of the users (56%) believes that Facebook does not share personal information with third parties and 70% believe that Facebook does not combine information about them collected from other sources. Less than one out of four users claims to have read Facebook's privacy policy.

There are preliminary work on tool effectiveness too. In a study of 2012 [51], the authors used the FourthParty Web measurement platform[7] to study the effectiveness of 11 blocking tools. Their dataset was obtained from three consecutive crawls of the Alexa U.S. top-500. The results of the experiments is that the most effective tool in blocking tracking activities is a combination of several community-maintained blocklists. The goal was to find the most effective tool that mitigate the tracking, with no regards to personal and sensitive information leakage.

2.2 Behavior and concerns and perceptions

There are works focused on users' behaviors, concerns and perceptions of risks. Some of these works studied the behavior of users on Online Social Network, like Facebook. Most of the users think that their privacy is not at risk. They trust the companies. Acquisti et al. [5] demonstrate that only few users are really aware of what happens behind the curtains. Furthermore Facebook users put more trust in the service provider than in the average Facebook user, as already stated. More importantly, half of users believes that Facebook does not share personal information with third parties and 70% believe that Facebook does not combine information about them collected from other sources.

Most of the work in the literature, that studied users attitudes, focus on privacy issue on specific contexts. Specifically, several studies addressed the analysis of the privacy in e-commerce environments [55], the privacy of health information on social media [78], the privacy that may affect individual's purchasing decisions [75], or the privacy that may be violated by the behavioral advertising [52]. The most interesting part is the study of how different people value their own privacy and what influence them. To evaluate users perception of risks, behaviors, concerns and attitudes, a series of survey are used. Several studies aimed at analyzing if gender

[7] http://fourthparty.info/

and technical differences [27, 31, 69, 73] may influence users' behaviors and attitudes toward privacy. There are study that aims at understand if these differences can influence the willingness to disclose personal information [4, 34].

It can be claimed that users are concerned for their privacy, as many work have found [22, 29, 33, 59, 76, 77]. Other studies show that privacy concerns influence people's willingness to disclose personal information to a Web site [23, 35]. But when it comes to study the behaviors, the attitudes and actions that a user take to protect against privacy violation, the dichotomy between privacy attitudes and resultant behaviors [2, 6, 15] shows up.

Several studies show that user will more likely disclose personal information when some benefits can be obtained in return [19, 32, 71].

Recent studies in the social area show concerns among (Facebook) users, their strong negative association between privacy concerns and engagement (posting, commenting and Like-ing of content [72]), their willingness to change privacy settings [14, 45], even if, in a subsequent study [16] the authors discovered that due to the constant modifications and alterations to the policy, many users (i.e., 65.7%) are unaware of how their profiles are affected, and therefore, unaware of their personal privacy settings. Moreover, users feels not capable of correctly configure their privacy settings, and call for new tools to protect privacy [44]. Even those that read the privacy policies, do not understand them, and acts like they are reluctant when adopting privacy protective techniques and technologies [2, 17, 38] because they do not understand them.

Most of the reported studies explored differences in individuals in respect to their privacy concerns. Our goal in this work was to investigate concerns about the privacy of two groups of students from different academic areas and with different technological knowledge. We explored whether educating them about potential risks on the Web, through a direct learning from one's behavior during online activities, may involve an increasing awareness about privacy as well as an increasing willingness to reduce their degree of exposure to privacy attacks.

3 Privacy leakage: extensive study

The study of the extent of privacy leakage, has to be conducted client-side. It is the only way because we can't study the mechanisms that goes inside an aggregator. In this study, several workload were defined:

1. Workload for privacy leakage
2. Workload for performance
3. Workload for effectiveness
4. Workload for profile building
5. Workload for Awareness[1]

In all the experiments, excluded those regarding workload 2 and 5, the Fiddler proxy was used to capture the traffic and study the privacy leakage. The proxy was not used for workload 2 because we do not want to add the overhead due to the proxy. It is not present in workload 6 too, because it involves human being and we want to avoid two things:

- Make them perceive a slow connection (due to the proxy)
- Log and maintain a copy of their navigation[2]

3.1 Privacy Leakage study

This study analyze how personal and sensitive information are sent to *third party sites*, such as third party Cookies, Referer header, Web bug, third party JavaScript or advertisements. It explore the manners of leakage through which these information are leaked. The capabilities of building users' profile was tested and an extensive study of the daisy chaining was conducted. Based on the leak found, we classify the most popular threats for privacy and countermeasures to be provided by privacy preserving tools. All the tools were compared for their performance and effectiveness.

In order to study the privacy leakage, we created an alter-ego that was registered on several Web site. The Web site selected were chosen from 18 (sub)categories of Alexa, including: Health, Travel, Employment, OSN, Arts, Relationships, News, PhotoShare, Sports, Shopping, Games, Computer, Home, Kids_and_teens, Recreation, Reference, Science, and Society. For each category the first top-10 Web sites that allow users to register were selected.

[1] It was supplemented by 10 minutes of free Web navigation. See Chapter 4

[2] All the other data gathered and visualized during the test were deleted after the test ends

Every created account was filled with all the possible information that the Web site requests. For none of them was used the option to login using a third party account (like Facebook or Google account). When available, the option "Remember me" was used. The option could be used to study if private information are stored and then sent to third party sites. The information of the alter ego that were diffuse incude but are not limited to: full name, email address (required for all accounts), Date Of Birth (DOB), Social Security Number (SSN), zip code, home address, personal cellphone, school and general education information, sexual orientation, political and intellectual beliefs, general interests (music, movies, and travel). They represent the bits of private information that may be leaked towards 3rd-party sites.

A log of typical interactions between the user and the sites was created. We included actions that may uniquely identify the users from (a) search terms[3], (b) browser habits, (c) preferences about music, movie and books[4], and (d) the structure of their social networks [56]. We used the following six types of online users' interactions:

1. Account Login and Navigation. We logged in on all 180 sites and analyzed information leakage due to 3d-party cookies. We also visited 4 or 5 embedded links per page, to reflect typical navigation of a user [40].

2. Viewing/Editing Profile. To reflect the most common actions performed by users on OSN we analyzed the following actions: viewing one's own profile and editing it, viewing 5 friend's profiles, writing on the "Timeline" of 2 of them.

3. Searching the Web for Sensitive Terms. We searched using `google.com` for 20 terms in 7 sensitive categories: Health (3), Travel (5), Jobs (2), Race and Ethnicity (2), Religious beliefs (3), Philosophical and Political beliefs (4), Sexual orientation (1). For each search term we also navigated through the first 2 search result pages.

4. Popular search. We chose 10 keywords from the top Google searches in 2012 and Google Trend Web pages[5].

5. Inputting and Like-ing content. For Inputting content we analyzed the following actions: post and reply to questions on forums (2 actions), reply to dating messages (1 action), upload pictures (1 action). For Like-ing content we analyzed the following actions: "Like" on Facebook (2 actions), "Share" via Facebook (2 actions), "+1" on Google Plus (2 actions), "Share" via Google Plus (2 actions).

During the tests, every HTTP request and response were logged. At the end of the tests, the log were parsed to look for strings that match the personal and sensitive information of the alter ego. False positive were removed by hand. Every occurred leakage, were recorded with the piece of information leaked, the manner of leakage, and the third party destinations.

Categorization of the most important leakages.

To draw up an order, a weight function for every leakage is needed. The function used was a simple counting of the manner of leakage that actually leak bits. Based

[3] `http://www.nytimes.com/2006/08/09/technology/09aol.htm`
[4] `http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf`
[5] `http://www.google.com/trends/hottrends`

on the work of [39], we identified and classified the following leaked bits: Full name, Email, Country, Region, City, Zip code, Employment, Gender, Age, Political and religious beliefs.

Other leaked bit were identified there were not present in previous works: ***browser fingerprint*** information, ***Sexual orientation***, ***DOB***, ***Interests*** (Movie and Music), ***Education*** and***IP address***.

The categories of the leaked bits are based on the categorization discussed in the aforementioned work, and are: *High*, *Medium* and *Low*, that reflects their degrees of sensitivity and identifiability.

This study made several findings: the most important vehicles through which a large amount of personal and sensitive information are leaked is the Referer HTTP field, followed by Web bug.

Further, the 20% of first party sites leak personal and sensitive information to third party sites.

Specifically, for the *High* Category, health terms are leaked in 3 of the 4 sites studied. For the *Medium* Category we derived that an important bit leaked by a number of sites was the user's full name. This leakage raises concerns when this bit is combined with sensitive terms. In Job and Travel searches, 4 out of 5, and 6 out of 7 of the analyzed sites show leakage. Surprisingly Health information are leaked in almost all the visited Web sites. These information combined with user's personal information can create difficulties while seeking health insurance or lead to privacy attacks such as identity theft[6].

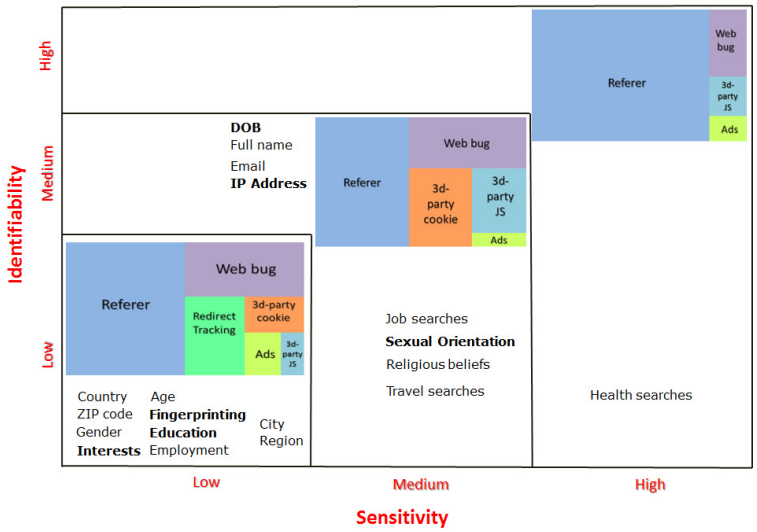


Figure 3.1: Distribution of privacy leakage vehicles across the categories.

A resume of the privacy bit leaked in show in Fig. 3.1. The privacy bits leaked are categorized in Low, Medium, and High categories. The privacy bits that were highlighted earlier, because their were not considered before, are highlighted in the

[6] `http://www.job-hunt.org/privacy.shtml`

image too. The HTTP Referer field is always the most used vehicle to track users. Only for the Low category we saw differences across all 6 manner of leakages.

3.1.1 Profile building

The goal of this section is to find if it possible for third party entities to rebuild users' profiles, and how hard it is, if possible.

To this aims, a special workload was designed (Workload 4). It was designed to perform specific actions that would mimic the behavior of a user that easily disclose personal information:

1. Logging into all 180 accounts
2. Viewing and editing all 10 profiles from the OSN category, post comments and messages, share documents with "friends"
3. Search on all 10 shopping sites from the Shopping category, add items to shopping carts (without payment), create lists, "Like" content
4. Search on all 10 Job-related sites from the Employment category, sign up for email alerts
5. Search on all 10 Health sites from the Health category, post comments
6. Search on all 10 Travel sites from the Travel category, book travel arrangements (without payment), visit Google maps site for itineraries, share with friends (via email and OSNs)
7. Reply to messages on 5 out of 10 Web sites of the Relationships category, that not required a Premium account
8. Create Photo Galleries on the `photobucket.com` Web site, upload images, add comments, share with friends, "Like" content
9. Watch videos on the `youtube.com` Web site, post comments, share with friends, "Like" content
10. Play songs on the `last.fm` Web site, post comments, share with friends, "Like" content.

The browser was instrumented with Selenium, and all the interactions were logged. No privacy preserving tools was used in this experiment to be sure to do not block any leakage that could occurs. The logs were inspected to analyze if some personal or sensitive information were being leaked to some third party entities, and what fraction of a user's profile is known by the top-10 aggregators.

The result show that a user' profile could be easily build. From the network traffic it was evident that specific bit of personal and sensitive information were being leaked to third party aggregators, more precisely the top-10 leak recipients identified in our data set are shown in Table 3.1. To analyze results the same method of Section 3.1 was used. To better identify all the information that were

currently being leaked, the set of strings to look for were enhanced to add other sensitive terms used during the test. The string being added were abut: sensitive Health terms (i.e., Pregnancy, Depression, Breast Cancer), Job terms (i.e., Analyst, Senior Analyst in New York), Travel terms (i.e., traveling from Napoli Capodichino to New York (JFK) and travel dates), music, book, and movie interests (i.e., Black Eyed Peas, Internet Traffic Measurement, and Viva l'Italia movie).

Table 3.1: Building a profile from pieces of private and sensitive information.

Aggregator	Email	IP	Country/ Region/ City	Zip Code	Gender	Age	DOB	Interests	Health/ Job Travel	Religious/ Political	Sex Orient.	Known bits [%]
doubleclick.net	–	√	√/√/√	√	√	√	√	√	√/√/√	√/–	–	81
google-analytics.com	√	–	√/√/√	√	√	–	√	√	√/√/√	√/√	√	87
scorecardresearch.com	√	–	√/–/√	√	√	√	–	√	√/√/√	√/–	–	69
adnx.com	–	–	–/√/√	√	√	√	–	–	–/–/–	√/–	–	37
yieldmanager.com	–	–	–/–/√	√	√	√	–	–	√/√/–	√/–	–	44
2o7.net	–	√	√/–/√	–	√	–	–	–	√/√/–	–/–	–	37
crwdcntrl.net	–	–	–/–/√	–	√	√	–	–	√/–/–	–/–	–	25
pubmatic.com	–	√	–/–/–	–	–	–	–	–	√/–/–	–/–	–	12
2mdn.net	–	√	√/√ /√	√	√	√	–	–	√/–/√	–/–	–	56
imrworldwide.com	–	–	√/√/√	–	√	–	–	–	√/–/–	√/–	–	37

Table 3.1 show the details for each one of the top-10 aggregators. Surprisingly Health terms were leaked to almost every aggregators. There are aggregators that can achieve a very detailed profile about users, for example Google Analytics is the top recipient of the leakages, being able to receive 87% of leaked bits.

These results prove that rebuild a users' profile is an easy and trivial task for an aggregators. They already posses all the information needed. Obviously, the use of this data are unknown but it is a fact that users profile can be easily build, with practically no big effort.

3.1.2 Daisy Chaining

The Daisy Chaining is a well known practice[7] that can increase chances of building detailed dossiers about users.

In Table 3.2 are presented information about Daisy Chaining. Column 2 shows the first party sites contacted, columns 3 and 4 the third party aggregators involved in daisy chaining while the last column shows the information leaked in that process. Daisy chaining is identified by examining the HTML body which includes an IFRAME that automatically triggers a request to the first aggregator. The aggregator's response includes a JavaScript file which triggers a request to the second aggregator. Linkage between the aggregators can be seen also via the Referer header.

3.2 Tools Comparison

In this section several tools were compared: AdBlock Plus, NoScript, Ghostery, RequestPolicy, and NoTrace. First, their functionalities and main characteristics are compared in Tables 3.3 and 3.4, respectively. Table 3.3 shows that all tools

[7] http://www.masternewmedia.org/online-advertising-management-ad-network-defaulting-and-daisy-chaining-for-ad-revenue-optimization/

Table 3.2: Leakage of private information through daisy chaining.

Count/	**Daisy Chaining**		**Bits leaked**
1st party sites	**First Aggregator**	**Second Aggregator (Family)**	
1 / www.bebo.com	bluecava.com	addvisor.net (Targus Info)	Name, Zip code
1 / www.bebo.com	bluecava.com	e.nexac.com (Datalogix)	Name, Zip code
2 / barnesandnoble.com	doubleclick.net	2mdn.net (Google)	Gender
1 / gamespot.com	doubleclick.net	2mdn.net (Google)	Gender
2 / youtube.com	doubleclick.net	googlesyndication.com (Google)	Gender
3 / www.datehookup.com	doubleclick.net	pubmatic.com (Pubmatic)	IP Address
2 / www.datehookup.com	doubleclick.net	criteo.net (Criteo)	IP Address
1 / it.bab.la	adv.adsbwm.com	bid.openx.net (openX)	Ethnicity
1 / travelocity.com	doubleclick.net	yieldmanager.com (Yahoo!)	Travel schedule
1 / www.espncricinfo.com	doubleclick.net	2mdn.net (Google)	City
1 / www.youtube.com	doubleclick.net	2mdn.net (Google)	Age, Gender
1 / www.linkedin.com	doubleclick.net	2mdn.net (Google)	Zip code, Gender

provide functionalities to filter advertisements and to block third party requests. Table 3.4 highlights the lack of awareness in almost all the compared tools.

Table 3.3: Summary of supported functionalities.

Tool	Header Removal	3rd party cookies	Flash cookies	Web bugs	HTML5 Local Storage	Opt-out cookies	3rd party requests	Ads	3rd party script execution
NoTrace	√	√	√	√	√	√	√	√	√
Ghostery	–	–	√	√	–	√	√	√	√
AdBlock Plus	–	–	–	∼[a]	–	–	√	√	√
NoScript	–	–	–	∼[a]	–	–	√	√	√
RequestPolicy	–	–	–	∼[a]	–	–	√	√	√

[a] This threat may be blocked as consequence of the application of other functionalities

Table 3.4: Summary of the main properties of the tested privacy tools.

Tool	Awareness		Crowdsourcing filtering rules	Blocking methods	Configuration properties
	Blocked URLs	Data leakage			
NoTrace	√	√	√	URL Content External	*Configuration*: Checkbox to activate/deactivate techniques *Extra Step:* Whitelist, Feedback by users, Crowdsourcing of rules
AdBlock Plus	√	–	√	URL	*Configuration:* Loading of subscription lists *Extra Step:* Adding on-the-fly new regular expressions to filter unblocked objects
Ghostery	√	–	–	URL	*Configuration:* Checkbox to activate or deactivate techniques, block/unblock *ad*-companies *Extra Step*: Selectively block a specific *ad*-company
NoScript	√	–	–	URL	*Configuration*: Checkbox to activate or deactivate techniques *Extra Step*: Whitelist, Blacklist and Embedding Objects to configure (temporarily or permanently)
RequestPolicy	√	–	–	URL	*Configuration*: Loading cross-site whitelists *Extra Step*: Add pairs of domains for which requests are allowed. Selectively enable/disable filtering on-the-fly for a Web site

All the tools were comparable in terms of functionalities: filter ads and to block third party requests. A subset of them support techniques for HTTP removal, 3d-party and Opt-out cookie blocking, HTML5 Local Storage managing and Web bug

filtering. The compared tools rely on the canonical URL-based blocking mechanism. Only one tool seems to support other mechanism, to inspect the HTTP stream to look for unwanted content, NoTrace (see Table 3.4)

The study cover the performance intended as users' perceived experience and performance (Section 3.2.1) and the effectiveness of the tested tools in terms of false positives and false negatives due to the filtering rules (Section 3.2.2).

Moreover the test done in Section 3.1 were repeated using each one of the tools described before. Several tools still leak personal and sensitive information. Specifically Noscript and RequestPolicy still leak few information while Ghostery leak more information. AdBLock Plus still have some leakage only through the Header, it still leak sensitive information while NoTrace limit more efficiently the information leakage. Plus NoTrace seems to be the only tool to offer an awareness mechanisms.

3.2.1 Performance

The workload for the performance test (Workload 2) was based on the methodology of Krishnamurthy et al. [38], and consisted of the top-100 Web sites from 15 Alexa categories (`http://www.alexa.com`). The Firefox browser was augmented by the Pagestats extension (`http://www.cs.wpi.edu/~cew/pagestats`).

The tests were done with no privacy enhancing tools installed ('NoAddons' configuration), and then with one tool installed at turn. Different Firefox profiles were created, one for each privacy protecting tools.

Response time results

One of the metric used to value the performance of privacy tools, was the mean response time when applying the filtering capabilities on the data set.

Further, the gain in terms of response time when third party objects are being removed was measured by comparing the time to retrieve an object from a page when filtering is applied, against objects' retrieve under normal conditions (i.e., the "NoAddons" configuration).

It is clear from Figure 3.2 that three tools and specifically NoTrace, AdBlok Plus and Ghostery show similar behavior (NoTrace seems to have slightly better results). But they all seems to have a greater response time when compared with NoScript and RequestPolicy (overhead of almost 600ms for both).

The main reason for these results is that NoScript and RequestPolicy are faster because of the large number of resources blocked via their filtering rules. Without a customized configuration NoScript blocks, regardless of the real danger of detected objects, *all JavaScript code*, even those that are essential to the correct behavior of the pages. RequestPolicy posses a stricter set of rules, it avoid the page break for very popular Web pages only because they are included by default in the startup whitelist. More details about this explanation could be find in Section 3.2.2.

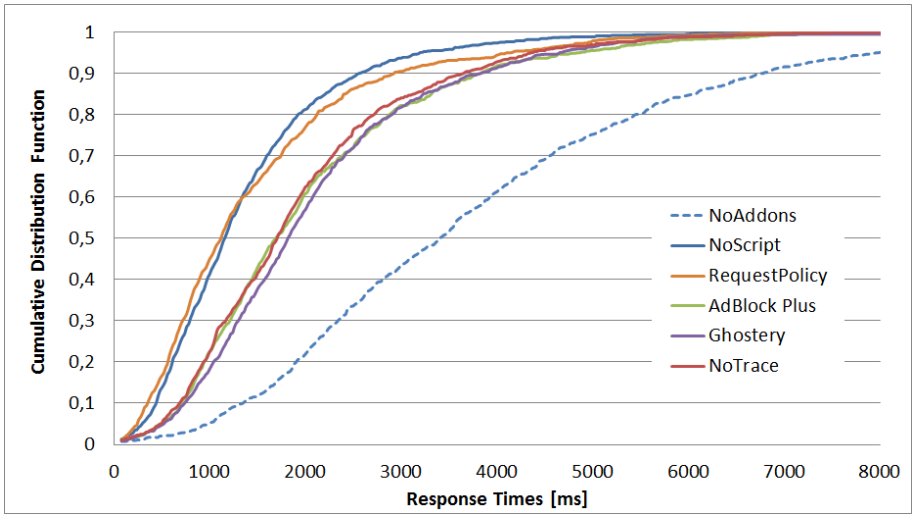


Figure 3.2: CDF of Response Time for each tools.

Browser performance results

This section explores the performance of single tools in terms of consumed memory when loading add-ons with specific techniques (i.e., ads and Web bugs filtering, 3d-party JS code execution blocking, opting-out from the tracking performed by *ad*-networks, HTTP Referer blocking). Furthermore a test of Firefox performance when multiple add-ons are installed is performed. Firefox is used with up to 8 extensions (i.e., AdBlock Plus, NoScript, Ghostery, RequestPolicy, Taco, RefControl, PrivacyChoice and TrackMeNot). The second test is necessary because most of the tools partially address the problem of privacy leak and it is already know that the Firefox browser could incur in slowdown when multiple extensions are loaded[8].

AdBlock Plus starts with the highest allocated memory since it needs to load in memory the subscription list. Ghostery shows worse performance since the resident memory at the end of the experiment was 4 times higher than the startup value. Finally, NoScript and RequestPolicy show better memory consumption values due to the high number of blocked resources.

In this case NoTrace is used as a way to provide "all in one" functionality. The results showed a higher Firefox loading time for the multiple-installations (1260ms vs 360ms for NoTrace).

Memory footprint was also analyzed, both when multiple add-ons are loaded then when single add-on is loaded. To mimic several hours of Web browsing, the MemBench script[9] was used.It is a browser benchmark that open 150 popular Web sites, one per tab, sequentially.

Memory usage was taken when Firefox starts, when all the pages were loaded (final allocated) and after all the tabs were closed (resident memory consumption). The resident memory measure was taken thank to the "about:memory" monitoring tab

[8] `http://blog.mozilla.org/addons/2010/06/14/improve-extension-startup-performance/`
[9] `http://gregor-wagner.com/tmp/mem`

of Firefox. When multiple add-ons were installed, the Firefox memory consumption were 2.8x larger than NoTrace installation.

A last test regarding the *multiple installations* memory consumption was done. The hypothesis is that more tools will imply a larger consumption of the memory. The final allocated memory for the *NoTrace single installation* was 120MB while the memory allocated by three add-ons (i.e., NoTrace, AdBlock Plus, and Ghostery) involved an increase of the value of the not released memory up to 300 MB. RequestPolicy and NoScript were not tested, since the amount of the resident and final allocated memory drastically decreased, due to the number of resources that they block and not because of better performance. The use of NoTrace alone can save on average 60% of the memory.

3.2.2 Effectiveness

The metric to measure the effectiveness of a privacy enhancing tool, was the number of correct blocked objects. In order to derive this metric, a manual analysis of each single URLs was needed, to look for false positive (FP) and false negative (FN). Because of the manual work, a different workload was used (Workload 3) that contains only 10 Web sites from the news category of *Alexa.com* This work has been done for each one of the compared tools.

In Table 3.5 are showed both FP and FN. Column 7 shows the number of FP detected when applying *intelligent filtering* (i.e., IF in Table 3.5) and without considering domains that serve their content for first party sites (i.e., NoIF).

Table 3.5: Effectiveness on popular Web sites: FP and FN.

Web Site	Web site's CDNs	AdBlock Plus		Ghostery		NoTrace		NoScript		RequestPolicy	
		FP	FN	FP	FN	FP (IF/NoIF)	FN	FP	FN	FP	FN
news.yahoo.com	yimg.com	1	10	10	3	0/10	0	10	17	9	23
edition.cnn.com	turner.com	0	34	1	12	0/3	2	21	14	1	26
weather.com	imwx.com	3	7	5	20	0/29	16	36	7	16	18
reddit.com	redditmedia.com redditstatic.com	3	3	2	8	0/2	3	5	2	24	1
my.yahoo.com	yimg.com yahoapis.com	3	5	3	9	0/2	7	2	10	4	4
bbc.co.uk/news	bbcimg.co.uk bbci.co.uk	1	8	0	11	0/14	5	36	6	106	3
foxnews.com	fncstatic.com	8	6	1	18	0/35	2	49	8	63	2
nytimes.com	nyt.com	1	11	0	12	0/7	10	48	17	63	20
huffingtonpost.com	huffpost.com	1	23	1	7	0/4	6	21	4	42	1
guardian.co.uk	guim.co.uk	5	10	4	5	0/3	8	26	5	119	1
Total		26	117	27	105	1/109	59	254	90	447	198
Recall/Precision		0.93/0.77		0.89/0.74		1.00/0.86		0.66/0.79		0.60/0.81	

While NoTrace FP are avoided, FN are still present. They can be due to:

1. First party requests for resources that are not available in the DOM
2. Objects served by CDNs of first party sites
3. 3d-party requests for resources that are not available in the DOM

The first category includes requests for Web bugs (i.e., `us.bc.yahoo.com/b`). NoTrace is not able to block them, as its technique to filter Web bugs looks at the height and weight properties of the images available in the DOM of the requested

Web page. Similar to AdBlock Plus, we allow users to add an ad-hoc filtering rule to block them.

The second category includes errors due to the inclusion of the CDN servers into the whitelist because of their role in serving needed content for the requested Web pages. *turner.com* is a well know CDN for *cnn.com*[10].

The third category includes errors due to third party requests for resources not available in the DOM. Here, the high number of errors is due to a request for a JS code that loads a certain number of both harmless scripts and malicious scripts (13 out of 16 errors are Web bugs for the `weather.com` Web site). Removing the loader will avoid all the FN but at the same time will break the page, because the harmless script is used for page formatting and other proper site's functionalities. A solution could be adopted, but it will affects loading time. It consists to intercept the request, analyze the URL, extract the harmful scripts rewrite the URL and then resubmit the modified URL.

As anticipated in Section 3.2.1 and detailed in Table 3.5, NoScript and RequestPolicy exhibit the highest number of errors. To have a better understand of the effectiveness of the tools, the number of FP and FN of the analyzed Web sites were plotted. Fig. 3.3 shows NoTrace's better behavior and the worst behavior of both NoScript and RequesPolicy with an extremely high FP. Their high error rate is due to naively blocking all third party requests, leaving users to adjust the filtering, by whitelisting URLs, or disabling filtering on a specific site when the quality appears degraded. Properly configuring the whitelist requires substantially more expertise than an average user can reasonably be expected to have.

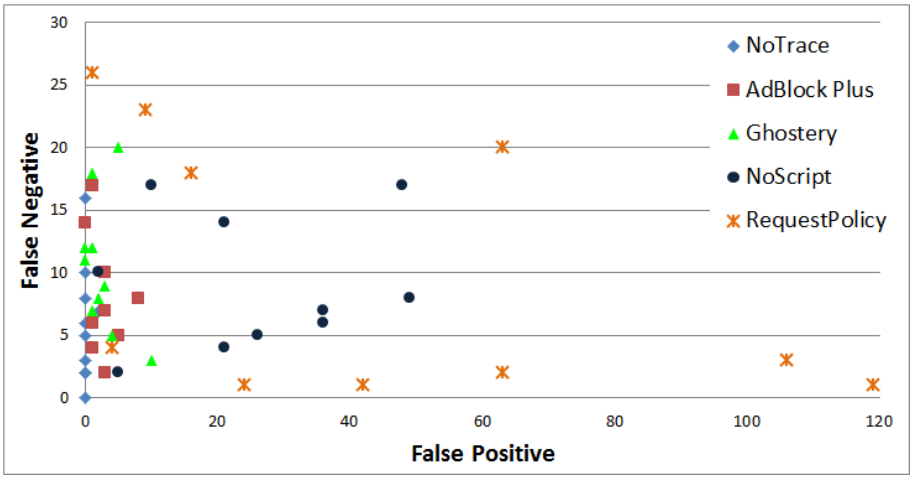


Figure 3.3: Analysis of FP and FN after blocking.

[10] `http://i.cdn.turner.com/cnn/.e/img/3.0/1px.gif`

4 Skill influence

4.1 Evaluation Study

This section aims at understand behaviors, concerns and attitudes with particular attentions to how all this characteristics are influenced by one's skills. It also want to finds if a highly customized privacy-enhancing tool can contribute to make users aware of their privacy leakage. Specifically, in this study, we tried to respond to the following questions:

1. Are there significant differences among two groups of students with regards to beliefs, attitudes and concerns about privacy? (Online privacy concerns, Section 4.2.1 for the results).

2. Experience in online activities and general technological knowledge can influence how different users "value" their privacy? (How skills influence behaviors, Section 4.2.2 for the results).

3. Learning about privacy through the help of a privacy-enhancing tool can involve more informed decisions about the countermeasures to adopt? Learning can affect in the same way different types of groups? (How privacy awareness can change behaviors, Section 4.2.3 for the results).

4.1.1 Methodology

For this experiment two samples of students were recruited from an university environment, specifically from the University of Salerno. The two samples consists respectively of 18 students from the Computer Science department and 18 students from the Cultural Heritage, Business and Law Departments. The sample of students from Computer Science Department is named *ICT* Group, the other *non-ICT*. The study was conducted at the ISIS research laboratory, at the University of Salerno. It must be emphasized that these two groups are statistically different as we will show afterwards in this Section 4.1 and 4.2.

The participation to the study was voluntary and anonymous, and students were not compensated for taking part to the interviews. Further they were recruited through email announcements to mailing lists, and word of mouth advertising. Participants were informed that all the information they provided would remain confidential. Finally, to avoid a biased sample, when recruited participants, we did not mention privacy or security, privacy risks and benefits, and we only said that we were looking for people interested in participating in an evaluation study.

Before proceed to the result of the experiment, some demographic information are shown in Table 4.1. The average age of the sample was 24, and the gender split was

almost even (i.e., 53% female and 47% male). The majority of participants spent between two and six hours on Internet per day (72% of the *non-ICT* users and 50% of *ICT* users), while only 11% of *non-ICT* users and half of *ICT* users spent on Internet more than six hours. 83% of users in the *non-ICT* Group considers themselves incompetent with online activities, while 72% of *ICT* participants consider themselves as competent/expert (χ=9.488, p-value=0.0065).

Table 4.1: Participant Demographics.

Variables		***non-ICT* Group**	***ICT* Group**	***Chi-Square* Sig. Level**
Gender	Male	28%	67%	0.0194
	Female	72%	33%	
Age	20-23 years	33%	44%	N.S.
	24-26 years	61%	50%	
	27-32 years	6%	6%	
Education	Bachelors	50%	11%	0.0113
	Masters	50%	89%	
Time spent online per day	0-2 hours	17%	0%	0.0351
	2-6 hours	72%	50%	
	6+	11%	50%	
Internet Expertise	Inexpert	83%	28%	0.0065
	Competent	17%	33%	
	Expert	0%	39%	

4.1.2 Procedure

The experiment envisioned three different phases:

1. *preliminary survey*
2. *tool testing*
3. *summary survey*

The preliminary survey consists of 32 questions in five categories, that aims to collect the following information:

- demographic information (i.e., gender, age, education level)
- information about Internet usage
- general knowledge about privacy threats on the Web
- general attitudes toward privacy
- information about awareness and general behaviors about privacy online

The first section consists of questions about demographics and skills (such as, time spent on Internet and level of expertise with online activities). The second regards knowledge about browser, and browser add-ons or extension. The third part tested participants' knowledge about some potential privacy threats on the Web (rating on a 5-point Likert scale with *strongly agree/strongly disagree* as verbal anchors). The fourth section tested participants' concerns about privacy and general attitudes toward online privacy (rating on a 5-point Likert scale). The last section was designed to assess general participants' behaviors about privacy online.

In the tool testing phase, we asked users to use a privacy enhancing tool designed to provide awareness, namely NoTrace. The participants were asked to use the evaluation tool, for a 30-minutes browsing session. Participants were free to use the tool in any modalities, they were also provided with any details about the tool and basic information on how to use it.

In order to not make participants feels uncomfortable they were not directly monitored, but a person for every need was provided in case they did not understand the instructions posed. The test was performed in an isolated environment within our research lab in order to avoid distractions due to the presence of other people. Furthermore, users were also encouraged to provide informal feedback for developers.

Before the third phase, users were asked to fill in the standard QUIS[1] and CSUQ[2] questionnaires. The aim was to provide additional information about system usability and user satisfaction when using NoTrace and differences experienced by the two tested groups. The questionnaires take less than 10 minutes. The two questionnaires were adapted to the nature of the privacy enhancing tool. For the QUIS questionnaires, th questions have a rating on a 10-point scale with appropriate anchors at each end (e.g., "Overall Reaction to the software: Terrible/Wonderful"), where small values corresponded to unsatisfactory or negative responses and large values corresponded to satisfactory results. The CSUQ questionnaires were composed by 12 questions, asking participants to indicate their agreement or disagreement through a 7-point Likert scale with *strongly agree* and *strongly disagree* as verbal anchors.

In the third and final phase, participants were asked to fill in a summary questionnaire, composed of 12 questions. The questions have a rating on a 5-point Likert scale with *strongly agree/strongly disagree* as verbal anchors. Three of them were already present in the preliminary questionnaires. They are asked in the final phase too to measure whether any change occurred in users' opinions, habits or behaviors after gaining a greater awareness about certain activities performed online by third party entities. The preliminary survey, the summary survey and the QUIS and CSUQ questionnaires are publicly available[3].

The entire experiment took between 60 and 70 minutes. Finally, for data analysis and statistics results the IBM Statistical Package for Social Sciences (SPSS, Inc.) software[4] was used.

[1] `http://oldwww.acm.org/perlman/question.cgi?form=QUIS`
[2] `http://oldwww.acm.org/perlman/question.cgi?form=CSUQ`
[3] `http://www.di.unisa.it/~delmal/research/usability/Surveys2013.pdf`
[4] `http://www-01.ibm.com/software/analytics/spss/`

4.2 Results

The main objective of the first part of the study was to determine, for both samples, beliefs and concerns about privacy, and to assess the knowledge of the threats that could undermine the privacy during online activities.

Participants were interviewed about familiarities with some privacy threats. The questions point to understand how much they are familiar with these threats. They were asked specifically if they know terms such as "Web bug", "Flash cookie", "behavioral advertising" and provide a clear definition for them. We also asked them whether they knew the risks associated with behavioral advertising.

The two groups showed differences, the *non-ICT* Group showed a small familiarity with these privacy threats, very few of them were familiar with terms such as Web bug, Flash cookie and behavioral advertising (5%, 11% and 33%, respectively). Only 28% of *non-ICT* users is aware of the risks of the behavioral advertising. Conversely, the *ICT* Group reported greater familiarity for all privacy threats. We also found statistical differences among groups, with corresponding results shown in Tale 4.2. Relation among the familiarity with privacy threats and concern about privacy online was found (question Q4 shown in Table 4.3). The analysis confirms that an increase of familiarity would obviously decrease general privacy concerns (Pearson correlation between privacy concern and familiarity with Flash cookie, behavioral advertising, and behavioral advertising risks are -0.3883, -0.3434, -0.3607, $p<.05$, respectively).

Table 4.2: Familiarity with some privacy threats.

Variables		***non-ICT*** **Group**	***ICT*** **Group**	**Chi-Square Sig. Level**
Web bug	Disagree	72%	22%	<0.0001
	Neutral	22%	0%	
	Agree	6 %	78%	
Flash Cookie	Disagree	61%	28%	0.00047
	Neutral	28%	0%	
	Agree	11%	72%	
Behavioral Advertising	Disagree	33%	0%	0.00012
	Neutral	34%	0%	
	Agree	33%	100%	
Behavioral Advertising risks	Disagree	33%	39%	0.00943
	Neutral	39%	0%	
	Agree	28%	61%	

4.2.1 Online privacy concerns

Measures that have been used when examining privacy include: concerns about privacy, attitudes toward privacy, privacy-related behavioral intentions and actual

behaviors. We look at association with self-reported concerns about privacy in everyday life (Q3), general concerns about privacy on Internet (Q4), as well as 2 questions related to the behavioral advertising (Q10 and Q11). These questions are shown in detail in Table 4.3. It has to be emphasized that the question ID used in here are the same used in the survey of the experiments.

The response of the participants show that all of them (in both groups) equally consider the privacy very important in everyday life. We obtained different result regarding privacy online. The *ICT* Group perceived privacy online with less concerns (agreement for 50% of participants). There was no observable statistical difference about privacy concerns across groups. Results are shown in Table 4.3.

Table 4.3: Participants privacy concerns. 5-Point Mean Likert scores. Groups are not statistically different according to these metrics.

ID	**Question**	**Mean**		**Agreement**	
		non-ICT	***ICT***	***non-ICT***	***ICT***
Q3	I consider important the privacy in everyday life	4.17	4.17	89%	83%
Q4	I am concerned about my privacy online	3.94	3.56	83%	50%
Q10	When I am online, I am aware that my browsing information may be collected by a third party for advertising purposes	3.28	3.94	67%	78%
Q11	I am comfortable with advertisers using my browsing history to serve me relevant ads, as long as that information cannot be tied to my name or any other personal information	2.72	2.89	39%	44%

According to participants' response, they are also concerned about tracking of their movements on the Web, performed by large business companies in order to provide them targeted advertising (Q10 and Q11 questions in Table 4.3). Although aware that their browsing history may be collected for advertising purposes, more than half of participants in both groups are uncomfortable even when their personal information cannot be tied to their browsing history. Once again, no evidence of statistical difference with regards to these concerns across groups were found.

Table 4.4: Privacy concern: typology comparison.

Group	***non-ICT*** *Group*	***ICT*** *Group*
Fundamentalists	33%	22%
Pragmatists [Personal Information Concerned]	28%	33%
Pragmatists [Behavioral Advertising Concerned]	22%	34%
Marginally Concerned	17%	11%

In order to understand if it was possible to categorize students based on their privacy concern, the k-means [2,12] algorithms was used. For this study, two questions strictly related to concerns about privacy (i.e., Q3 & Q4) and other two questions related to the behavioral advertising privacy threat (i.e., Q10 & Q11) were selected.

By clustering our set of participants, we identified, in contrast to the results by Westin [82] and Ackerman *et al.* [1], a group of *Fundamentalists* and a group of *Marginally Concerned*, while the *Pragmatists* group was further decomposed in two distinct groups whose privacy concerns focused either on the awareness of the risks of the behavioral advertising phenomenon or on the linking of users' history information with personally identifiable information. This categorization results are shown for both groups in Table 4.4.

More specifically, Fundamentalists users provided privacy-oriented responses to all the questions selected for that analysis, showing their concerns to both beliefs about privacy online and the risks of the behavioral advertising. Instead Marginally Concerned users show general concerns about privacy and a propensity to enjoy the benefits of the behavioral advertising but only if their personal information are not being collected and linked with the browser's history information. The Pragmatists users, as anticipated, were organized in two distinct groups which we called "*Personal Information Concerned*" and "*Behavioral Advertising Concerned*" because of their major concerns about personally information and the risks of the behavioral advertising, respectively.

Table 4.5: Participants privacy attitudes. 5-Point Mean Likert scores.

ID	**Question**	**Mean**		**Agreement**		**Unpaired T**
		non-ICT	***ICT***	***non-ICT***	***ICT***	**Sig. Level**
Q12	I am comfortable with the privacy I have when I use search engines	2.39	2.44	17%	11%	N.S.
Q13	It is my responsibility to protect my personal information on the Web	3.17	2.61	44%	22%	N.S.
Q14	I am aware of the tools that exist online to help me protect my privacy online	3.17	4.33	55%	100%	<0.001

Participants show to be uncomfortable with their search engines. Indeed when interviewed about the comfortability of the privacy of their search engines, our participants expressed their dissatisfaction (agreement of about 17% for the *non-ICT* Group and only of 11% for the *ICT* Group, Q12 in Table 4.5). Intuitively, this comfortability is related to concerns about privacy, as feelings of comfortability increase one would expect overall privacy concern to decrease (Pearson correlation r=-0.4497, p<.01).

Interestingly, participants stated that they are not personally responsible for protecting their online privacy (i.e., Q13, in Table 4.5), specifically 56% of users in the *non-ICT* Group and 78% of *ICT*. But none of the participants do take actions to effectively protect it (the whole *non-ICT* Group affirmed its inability to protect its personal information, p-value=0.0004). Even if more than half of all participants are

Table 4.6: Participants privacy attitudes.

ID	Question		Agreement		Chi-Square Sig. Level
			non-ICT	***ICT***	
Q23	If I have to prioritize between perfect search and perfect privacy I would choose...	Perfect Search Search ahead of Privacy Privacy ahead of Search Perfect Privacy	11% 28% 61% 0%	17% 17% 61% 5%	N.S.
Q24	If you knew for a fact that topics you search for using a search engine were saved forever, would it change your search habits?	No change Somewhat of a change Significantly change	17% 78% 5%	11% 77% 11%	N.S.

aware of tools that exist online to protect it (greater awareness for the *ICT* Group, p-value=0.0002) almost no one ever installed one. In general, users say that they care about privacy but they do not do anything about it.

When asked about their preferences about search engines and specifically when asked to chose between search quality and search privacy, most of the users express their preference for the *Privacy ahead of Search* option (rate of 61% for both groups, see Table 4.6).

A further analysis to inspect the nature of privacy concerns were conducted by analyzing the responses to the following question: "*What are your main privacy concerns online*?". By manual analyzing the response, five met-categories of concerns were identified:

1. "*Tracking my behaviors by third party entities*"
2. "*I am worried about making financial transactions online*"
3. "*Identity theft*"
4. "*Any unauthorized access to my personal information*"
5. "*Internet will never forget my personal data after their dissemination*"

All the details for both groups are presented in Table 4.7. Because some answers fall in two categories, the sum of columns is above 100%. Students from the *non-ICT* Group seems to be less concerned about being tracked by third party entities. One possible explanation is that they are aware of the presence of third party entities, but do not know which information they are collecting. Instead students from the *ICT* Group are more concerned about more general problems. The concerns about Identity theft are almost the same across the groups, there is no statistical difference.

Table 4.7: Nature of privacy concerns across the *ICT* and the *non-ICT* groups.

Meta-categories	***non-ICT* Group**	***ICT* Group**
Tracking my behavior online from third party entities	44%	56%
I am worried about making financial transactions online	11%	–
Identity theft	22%	28%
Any unauthorized access to my personal information	–	17%
Internet will never forget my personal data after their dissemination	–	11%
No answer	28%	33%

4.2.2 How skills influence behaviors

Some behaviors that could be indicative of a lack of privacy concern and that we studied include:

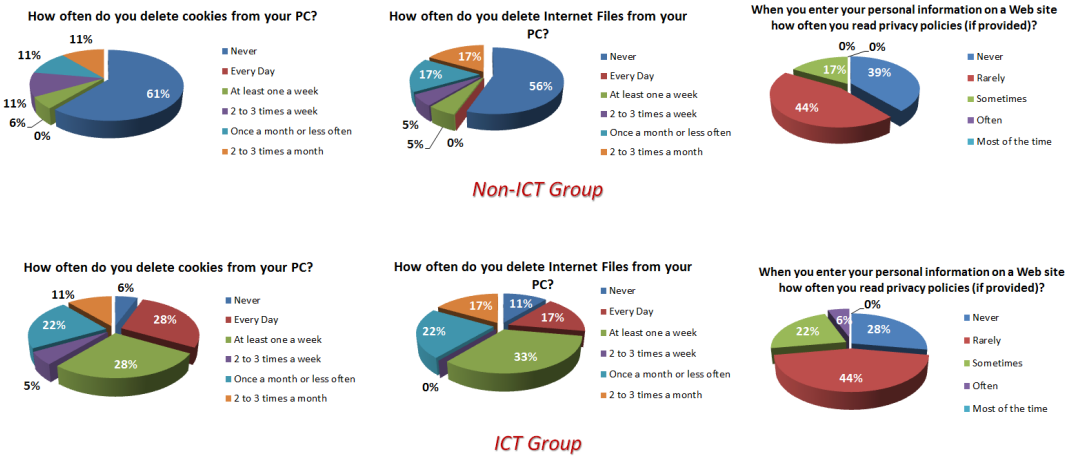
1. Increasing security settings on browsers
2. Installing tools to protect privacy
3. Deleting cookie saved on browsers
4. Deleting cache and temporary Internet files
5. Reading licence and privacy agreements.

Regarding attitudes, the participants do not change their browser settings, or install privacy tools (Q16 and Q18, Table 4.8), with differences between groups. Only 17% of *non-ICT* and 61% of *ICT* participants regularly delete cookies. There is no difference between about policy reading. More than 70% of participants in both groups rarely/never read privacy policies (see Figure 4.1 and Table 4.8).

Statistical differences between groups were found too. Specifically differences about the privacy behaviors. Significant differences across groups for "Deleting HTTP cookie" (p-value=0.0045) and "Deleting Internet files" (i.e., 0.0229). Other differences were find about "Changing browser privacy settings" and about "Installing privacy tools". There are correlation among these two actions and privacy concerns (r=-0.3371, -0.3748, $p<.05$). There was no significant difference about policy readings. Furthermore users who report to be slightly concerned about privacy online also report less engagement across the analyzed behaviors (see Figure 4.2).

From the study emerge that users of the *non-ICT* Group are less aware of the tracking activities performed by large advertising companies and aggregators. They do not know the privacy vehicles used to track users and finally, they are less aware of the tools that exist to protect the privacy. Moreover, although aware that their browsing history may be collected for advertising purposes, most of the participants in both groups (61% of *non-ICT* and 56% of *ICT* users) are still uncomfortable, even when their personal data cannot be tied to their browsing history.

Figure 4.1: Privacy behaviors: deleting HTTP cookies, deleting Temporary Internet files (cached resources), reading of Web sites' privacy policies. Results are shown for both groups.



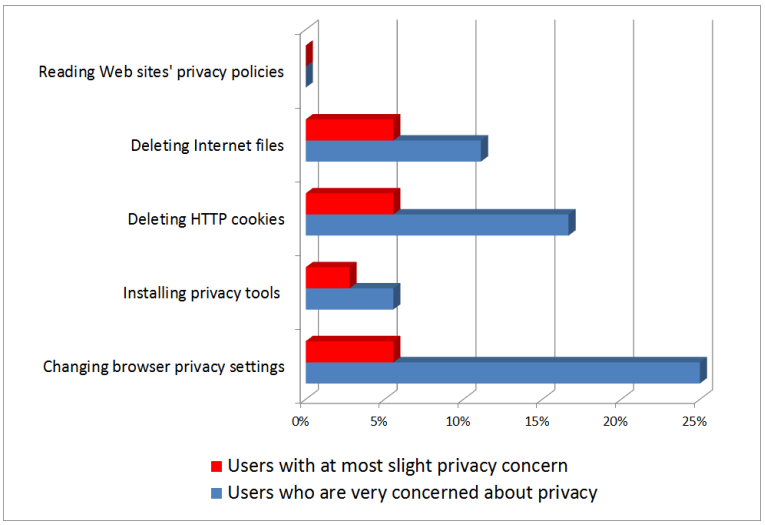


Figure 4.2: Percentages of users reporting certain behaviors, grouped by level of privacy concern.

The study showed the "lazy" attitude of *non-ICT* users, who mostly do not take actions to protect their privacy. This situation highlight a huge need for support and supporting tool for non-technical students (statistical results are shown in Table 4.8).

4.2.3 How Privacy awareness can change behaviors

Before discussing the results of the last part of the study, remind that students, in this second phase of the experiment, were asked to use NoTrace during a 30-minutes browsing session. At the end of the testing phase, we first collected information

Table 4.8: Participant privacy *actual* behaviors.

ID	Question	Agreement		Chi-Square
		non-ICT	***ICT***	**Sig. Level**
Q16	Have you ever changed your browser privacy settings?	22%	78%	0.0022
Q18	Have you ever installed a tool to protect your privacy?	0%	28%	0.0159
Q26	Are you able to protect your personal information?	0%	61%	<0.001

about users' satisfaction and perceptions about the usability of the evaluation tool. Second, we interviewed students in order to understand if NoTrace had some effect on them, in terms of increased awareness and willingness to adopt any type of strategy to improve privacy.

As we can see from Figure 4.3, on average, the posed questions were rated positively. Some lower values for the *non-ICT* users are relative to the complexity of the experimented tool and the low intuitiveness of the used terminology. In general, at the question "*Overall, It was easy to use NoTrace*" (Q27), 61% of *non-ICT* users expressed their agreement against 94% of *ICT* ones (p-value=0.0177).

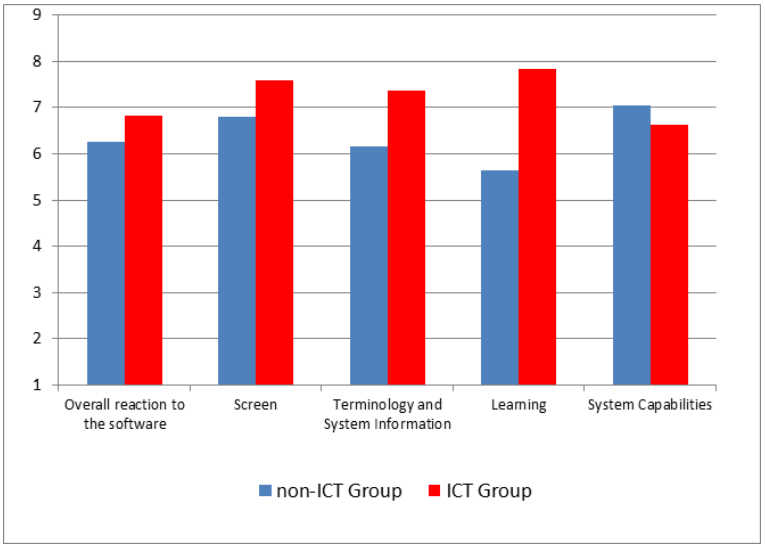


Figure 4.3: QUIS results. Comparison between *non-ICT* and *ICT* groups.

Additionally, the *ICT* Group responded with more positive results than the other Group. The same trend is still valid for the analysis of results of the second questionnaire. Specifically, as we can see from Figure 4.4, "Satisfaction" and "Clearness" are equally positively rated by both groups, while differences exists about easiness of use and learning and tool's usefulness. These results confirm that non-technological users need more support, even to understand the usefulness of any tool to protect privacy.

Finally, the reliability of the QUIS and CSUQ questionnaires were good (Cronbach's $\alpha = 0,91$ and $\alpha = 0,94$, respectively).

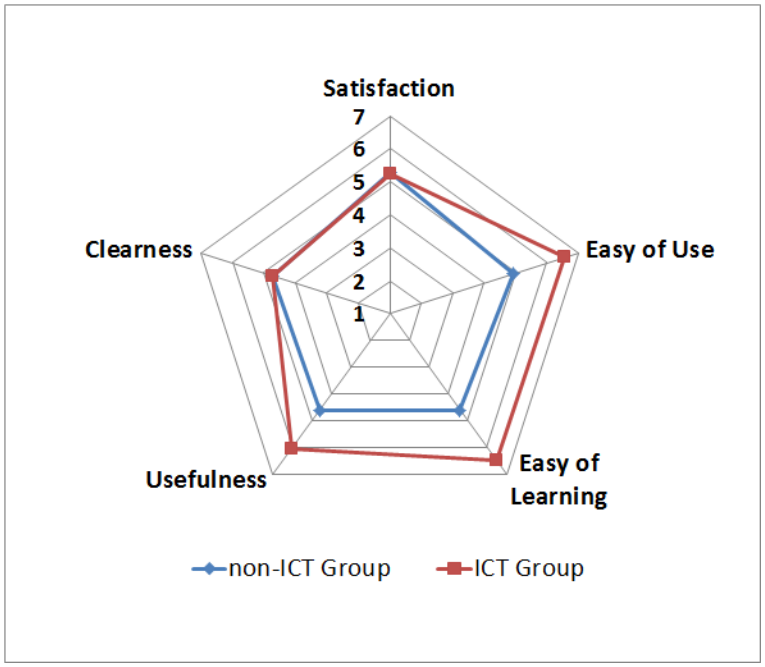


Figure 4.4: CSUQ results organized according to five metrics: Satisfaction, Easy of Use and Learning, Usefulness and Clearness.

Now we are going to analyze results about the investigation of whether a privacy-enhancing tool is able to impact on the users' perceptions, beliefs and concerns about privacy online. The question to respond is if learning about privacy through NoTrace can involve more informed decisions about the countermeasures to adopt, and if this learning affects, in the same way, both the analyzed groups.

We asked users to provide their thoughts about the concerns related to third party We sites that access to their personal data and. If after using the tool, they were willing to take actions to protect their privacy.

Results in Table 4.9 show positive ratings for both groups highlighting how the tool was able to increase user awareness about the dangers of certain activities on the Web. However, the higher positive result was for the question: "*I consider important the privacy and I want to make actions and use tools to protect it*" with agreement of 83% for *non-ICT* users and 72% of *ICT* users. Although all users rated positively all three questions shown in Table 4.9, more willingness to change their behavior online was expressed by *non-ICT* ones, even if there isn't any statistically difference.

From Figure 4.5 we can see that users who report to be slightly concerned about privacy online also report less willingness to take actions to protect themselves against privacy invasions. Another interesting result is about the high percentage of "Neutral" users (neither concerned nor unconcerned about privacy online) who reported higher willingness against the slightly concerned users, to protect their privacy online.

Table 4.9: Participants privacy *resultant* behaviors. 5-Point Mean Likert scores.

ID	Question	Mean		Agreement	
		non-ICT	***ICT***	***non-ICT***	***ICT***
Q33	I am more conscious of the leakage of my privacy on the Web and I want to change my behavior to try to protect it	4.06	3.83	72%	78%
Q34	I consider important the privacy and I want to make actions and use tools to protect it	4.11	3.94	83%	72%
Q35	Protecting my personal information means that I am willing to reduce the information I will post on the Web	3.72	3.61	67%	61%

It is worth to note that when interviewed about the possibility to reduce personal information posted on the Web, and specifically on the `Facebook.com` Web site, the agreement percentages, respect to the previous 2 questions (Q33 and Q34, Table 4.9), dropped for both groups. Even concerned about privacy, users seem to not be worried about the leakage of their personal information, when some benefits can be obtained in return. Some reasons, in fact, output of open-ended text questions, include: (a) "*I need to add accurate information to OSNs to increase my visibility*", (b) "*I need to add accurate information to OSNs to find my old friends*", (c) "*I am comfortable with how much information I share, I have a control over my personal information*".

The most important result of this part of the study is the comparative assessment between some specific questions asked before and after the tool testing phase, in order to evaluate if changes occurred in users' opinions and habits after using a tool to protect privacy. Questions are shown in Table 4.10.

Table 4.10: Comparative assessment of questions posed before and after the testing phase.

Question	***non-ICT***			***ICT***		
	Mean		**Paired T**	**Mean**		**Paired T**
	Before	**After**	**Sig. Level**	**Before**	**After**	**Sig. Level**
I consider important the privacy in everyday life	4.11	4.33	0.041	4.17	4.28	N.S.
I am concerned about my privacy online	3.94	4.44	0.024	3.61	3.89	N.S.
When I am online, I am aware that my browsing information may be collected by a third party for advertising purposes	3.22	4.22	0.011	3.94	4.44	0.0459

Figure 4.5: Percentages of users reporting final behaviors, grouped by level of privacy concern.

When students were made informed about the potential harmful activities on the Web, and have learned by the tool in which way their personal information were leaked, they expressed increased concerns about risks to their privacy, especially when their data are disclosed to third party sites.

In general, the tool allowed users to understand which personal and sensitive information are being disclosed during their online activities, in which extent and also, towards which third party sites. Support and awareness were able to involve changes in users' opinions, slightly increasing the level of concern about privacy in everyday life (mean values: 4.11 vs. 4.33 for the *non-ICT* Group and 4.17 vs. 4.28 for the *ICT* Group, Table 4.10) and mostly increasing concerns about the online privacy for the *non-ICT* Group (3.94 vs. 4.44) and the awareness about third party tracking

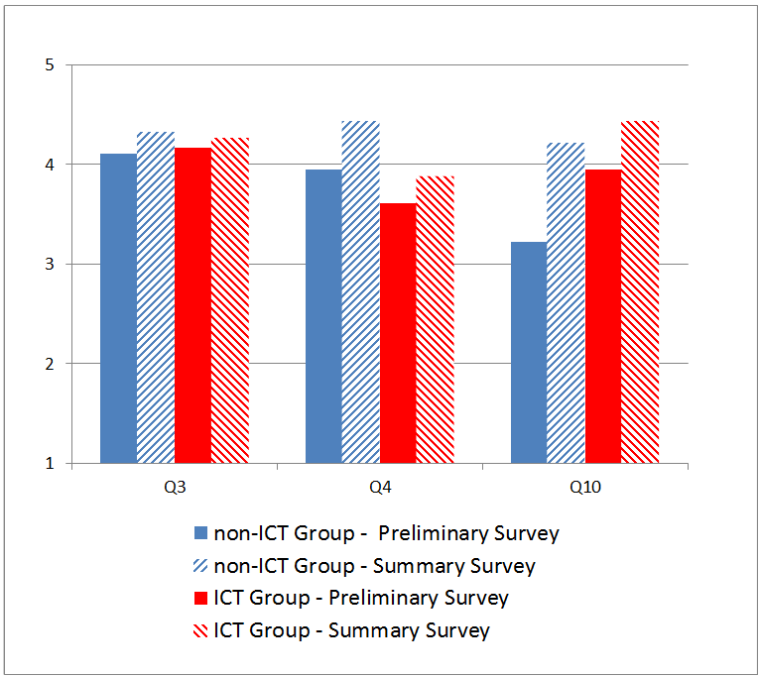


Figure 4.6: Comparison of attitudes before and after using NoTrace during the testing phase.

(for both groups). In Figure 4.6 we can see how using the tool involved greater concerns about privacy as well as a greater awareness about the tracking performed by advertising companies. Additionally, statistically significant differences for all questions exist only for the *non-ICT* users, highlighting, once again, the usefulness of NoTrace mainly for users without technological skills (see Table 4.10). For *ICT* users, instead, significant results are obtained only with regard to the awareness of the tracking performed by advertising companies. Both groups increased their awareness about the risks for their privacy related to the activities performed by large advertising companies (whose concerns were expressively disclosed in open-ended questions).

5 Conclusion

The analysis carried out showed that the extent of online privacy problems do not have to be underestimated. There are companies that actually gather an enormous amount of data about users and are able to build a detailed profile about them and link these profiles with personal information to identify an individual. Users should be aware that their data are disclosed and available to several companies, and that these companies can easily build profiles about them. Moreover the final use of these data is not fully known. These companies can voluntary on involuntary leak users' profile to other entities, without users consent.

The study, for the first time, employed a methodology that was not the surface crawling. This methodology reflects users' online real behaviors and are not limited by the surface crawling methodology. Therefore the results obtained are a precise snapshot of the current situation and not just a resume of potential privacy leakage. They reflects real privacy issues that the average user can face while online. We find that aggregators are able, at least, to easily rebuild users' profile. By using reverse engineering it was possible to show the amount of information leaked to each one of the top-10 aggregators. The profile rebuilding did not used any data mining technologies or other sophisticated techniques, it was based on simple observation of the network traffics. The most incredibly result is that one of the top-10 aggregator is able to collect 87% of a user's private information. Moreover parts of these data are exchanged between aggregators, trough the well-know practice of *Daisy Chaining*. Private bit like Name, Zip code, Gender and Travel schedule are passed from one aggregator to another in plaintext. While the number of HTTP request accounting for the daisy chaining are very low, they carry really important information.

Moreover while the capabilities of aggregtors to derive users' profile is increasing, the techniques available to users to protect themselves are not advancing at the same speed. There are a lot of tools that help to prevent users from unintentionally disclose personal and sensitive information. Anyway, almost none of them is able to make users aware of the *hidden micro-transactions* of personal and sensitive information that effectively pay for apparently free services.

Furthermore, the evaluation study about users behaviors, concerns, and attitudes supports the thesis that an increased awareness can be achieved. It has been demonstrated that one simple and effective way to increase awareness of users, is to simply show them the leaked information. It helps users to have a better perception of what goes underneath while they are on the Web, and make them capable of decide whenever to apply countermeasures. Showing to users the information that they actually leak during online activities, really increased their awareness.

While learning from one's behavior and awareness increase the users' willingness to change their behaviors online, their unwillingness to withhold information mainly

on social network sites, remains. But users are now aware that what they do online, is monitored and possible used to profile them.

The findings so far are:

- All participant students consider equally important privacy in everyday life, with no statistical differences based on skills
- That *non-ICT* students seems to be more concerned about risks to their privacy online
- Skills do influence attitudes. Indeed *non-ICT* students exhibit a little willingness to adopt privacy preserving technologies
- Both groups show the same behaviors about Web sites' policies.

Despite the increased awareness, users' need to generously provide an ever-increasing number of personal information is still prominent. It have to be emphasized that this study has some limitations. First, all participants at our study were students from an Italian academic environment. The samples were composed of users with high education levels and with an age ranging from 20 to 30 years. Therefore, the results may not necessarily be representative of the entire world population.

Future work could look into privacy attitudes and behaviors with regard to older age groups, students from other academic areas and with even more diversified technological skills. Moreover, a larger number of subjects would provide more statistically significant results.

The study prove that there are tools that help to make users aware of the risks they face during their normal online activities. More importantly, we verify that such capable tools influence users' concerns (with no difference due to skills) about the risks of behavioral advertising. It also raises general concerns about privacy, especially for *non-ICT* students.

Attitudes are largely affected by tools that increase users' privacy awareness, especially for *non-ICT* students. Usually *non-ICT* students are regretful to install and use privacy protection tools, mainly because of their low technical skills. They feels inadequate and inexperienced and do not want to spend time on a tool when they feel inexpert.

In conclusion, users expose their personal sensitive information while on the Web, mainly because of third party companies called aggregators that gather, collect and link these information to enhance the behavior advertising practice. The aforementioned aggregators can easily build profiles about users, that include their personal information and their interests. Users are fairly aware of this ecosystem, they do not know how their data can be used. To make users more aware of the online privacy risks, a tools that show what information is leaked to who was used. A better awareness was achieved thanks to the tool. A change in concerns and attitude was observed, with *non-ICT* students being more affected by the increased awareness.

This confirms the need for more support, and tools, for users who do care about privacy but are not aware of the risks they encounter during their online activities and are not able to protect themselves.

Bibliography

[1] M. S. Ackerman, L. F. Cranor, and J. Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. of the 1st ACM conference on Electronic commerce*, EC '99, pages 1–8, 1999.

[2] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. *Security Privacy, IEEE*, 3(1):26 –33, 2005.

[3] A. Acquisti, L. John, and G. Loewenstein. What is Privacy worth? In *Proceedings of the Twenty First Workshop on Information Systems and Economics (WISE)*, Dec. 2009.

[4] A. Acquisti, K. J. Leslie, and G. Loewenstein. The Impact of Relative Standards on the Propensity to Disclose. *Journal of Marketing Research*, pages 1–15, 2011.

[5] R. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Proceedings of the* 6^{th} *Workshop on Privacy Enhancing Technologies*, pages 36–58, 2006.

[6] N. F. Awad and M. S. Krishnan. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1):13–28, Mar. 2006.

[7] M. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle. Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning. Technical report, University of California, Berkeley, 2011. http://ssrn.com/abstract=1898390.

[8] M. Balduzzi, C. Platzer, T. Holz, E. Kirda, D. Balzarotti, and C. Kruegel. Abusing social networks for automated user profiling. In *Recent Advances in Intrusion Detection*, pages 422–441. Springer, 2010.

[9] H. Beales. The Value of Behavioral Targeting. `http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf`, 2010.

[10] B. Berendt, O. Günther, and S. Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 48:101–106, April 2005.

[11] A. R. Beresford, D. Kübler, and S. Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1):25–27, Apr. 2012.

[12] L. G. Berry, M. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Wiley, April, 2011 edition, 1997.

[13] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 551–560, New York, NY, USA, 2009. ACM.

[14] D. Boyd and E. Hargittai. Facebook privacy settings: Who cares. *First Monday*, 15(8), 2010.

[15] T. Buchanan, C. Paine, A. N. Joinson, and U.-D. Reips. Development of Measures of Online Privacy Concern and Protection for Use on the Internet. *Journal Of The American Society For Iinformation Science And Technology*, 58:157–165, January 2007.

[16] E. Butler, E. McCann, and J. Thomas. Privacy Setting Awareness on Facebook and Its Effect on User-Posted Content. *Human Communication*, 14(1):39–55, 2011.

[17] C. Canali, M. Colajanni, D. Malandrino, V. Scarano, and R. Spinelli. A Novel Intermediary Framework for Dynamic Edge Service Composition. *Journal of Computer Science and Technology*, 27:281–297, 2012.

[18] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran. Betrayed by your ads! In *Proceedings of the* 12^{th} *International Conference on Privacy Enhancing Technologies*, volume 7384, pages 1–17. 2012.

[19] R. K. Chellappa and R. G. Sin. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management*, 6:181–202, April 2005.

[20] F. T. Commission. Federal Trade Commission, Protecting America's Consumers. `http://www.ftc.gov/`.

[21] G. Conti. *Googling Security: How Much Does Google Know About You?* Addison-Wesley, 2008.

[22] M. J. Culnan. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly*, 17(3):pp. 341–363, 1993.

[23] T. Dinev, P. Hart, and M. R. Mullen. Internet privacy concerns and beliefs about government surveillance - An empirical investigation. *J. Strateg. Inf. Syst.*, 17(3):214–233, Sept. 2008.

[24] C. Dwyer. Privacy in the Age of Google and Facebook. *IEEE Technology and Society Magazine*, 30(3):58–63, 2011.

[25] J. B. Earp and F. C. Payton. Information Privacy in the Service Sector: An Exploratory Study of Health Care and Banking Professionals. *Journal of Organizational Computing and Electronic Commerce*, 16(2):105–122, 2006.

[26] D. F. Galletta, R. M. Henry, S. McCoy, and P. Polak. Web site delays: How tolerant are users? *JAIS*, 5(1), 2004.

[27] E. Garbarino and M. Strahilevitz. Gender differences in the perceived risk of buying online and the effects of receiving a site recommendation. *Journal of Business Research*, 57(7):768 – 775, 2004.

[28] Ghostery. `http://www.ghostery.com/`.

[29] J. Gomez, T. Pinnick, and S. Ashkan. UC Berkeley, School of Information. `http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf`, June 2009.

[30] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. Workshop on Privacy in the Electronic Society, WPES '05, pages 71–80, 2005.

[31] G. Grubbs M., Milne. Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising*, 10(2):28–45, 2010.

[32] C. Jensen, C. Potts, and C. Jensen. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203 – 227, 2005.

[33] M. Johnson, S. Egelman, and S. M. Bellovin. Facebook and privacy: it's complicated. In *Proceedings of the Eighth SOUPS*, SOUPS '12, pages 1–15, 2012.

[34] A. N. Joinson, C. Paine, T. Buchanan, and U.-D. Reips. Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*, 24:2158–2171, September 2008.

[35] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield. Privacy, Trust, and Self-Disclosure Online. *Human-Computer Interaction*, 25(1):1–24, 2010.

[36] T. Kang and L. Kagal. Enabling privacy-awareness in social networks. In *AAAI Spring Symposium: Intelligent Information Privacy Management'10*, pages 98–103, 2010.

[37] A. J. Kimmel. Ethical Issues in Behavioral Research: Basic and Applied Perspectives. 2007.

[38] B. Krishnamurthy, D. Malandrino, and C. E. Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *SOUPS, SOUPS '07*, pages 52–63, 2007.

[39] B. Krishnamurthy, K. Naryshkin, and C. E. Wills. Privacy leakage vs. protection measures: the growing disconnect. In *Web 2.0 Security and Privacy Workshop*, 2011.

[40] B. Krishnamurthy and J. Rexford. *Web protocols and practice: HTTP/1.1, Networking protocols, caching, and traffic measurement.* Addison-Wesley, 2001.

[41] B. Krishnamurthy and C. Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 541–550, 2009.

[42] V. Lawton. Privacy Commissioner of Canada. Popular websites in Canada disclosing personal information. `http://www.priv.gc.ca/media/nr-c/2012/nr-c_120925_e.asp`, 2012.

[43] I. Leontiadis, C. Efstratiou, M. Picone, and C. Mascolo. Don't kill my ads!: balancing privacy in an ad-supported mobile application market. In G. Borriello and R. K. Balan, editors, *HotMobile*, page 2. ACM, 2012.

[44] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 61–70, 2011.

[45] M. Madden and A. Smith. Reputation Management and Social Media. Technical report, Pew Internet & American Life Projects, 2010.

[46] D. Malandrino and Scarano. Privacy leakage on the Web: Diffusion and countermeasures. *Computer Networks*, (0):–, 2013.

[47] D. Malandrino and V. Scarano. Supportive, Comprehensive and Improved Privacy Protection for Web Browsing. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT)*, pages 1173–1176, 2011.

[48] D. Malandrino and R. Spinelli. NoTrace 2.3, Protect your Privacy Online. `https://addons.mozilla.org/en-us/firefox/addon/notrace/`. Development version available at `http://isis.dia.unisa.it/projects/NoTrace/`.

[49] H. Mao, X. Shuai, and A. Kapadia. Loose Tweets: An Analysis of Privacy Leaks on Twitter. WPES '11, 2011.

[50] G. Maone. NoScript. `http://noscript.net/`.

[51] J. R. Mayer and J. C. Mitchell. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, SP '12, pages 413–427.

[52] A. M. McDonald and L. F. Cranor. Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the* 9^{th} *annual ACM workshop on Privacy in the Electronic Society*, WPES '10, pages 63–72, 2010.

[53] A. M. McDonald and L. F. Cranor. A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies. Technical report, CyLab, CMUs, 2011.

[54] A. M. McDonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor. A Comparative Study of Online Privacy Policies and Formats. In *Proceedings of the* 5^{th} *SOUPS*, SOUPS '09, pages 46:1–46:1, 2009.

[55] M. J. Metzger. Communication Privacy Management in Electronic Commerce. *Journal of Computer-Mediated Communication*, 12(2):335–361, 2007.

[56] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, SP '09, pages 173–187, 2009.

[57] P. A. Norberg, D. R. Horne, and D. A. Horne. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

[58] A. Odlyzko. The unsolvable privacy problem and its implications for security technologies. *Australasian Conference on Information Security and Privacy: 8th*, Lecture Notes in Computer Science 2727, Springer:51–54, 2003.

[59] C. Paine, U.-D. Reips, S. Stieger, A. Joinson, and T. Buchanan. Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies*, 65(6):526–536, June 2007.

[60] W. Palant. AdBlock Plus. `http://adblockplus.org/`.

[61] D. Perito, C. Castelluccia, M. Kaafar, and P. Manils. How Unique and Traceable Are Usernames? In *PETS*, volume 6794, pages 1–17. 2011.

[62] Privacy Rights Clearinghouse. Empowering Consumers. Protecting Privacy. `http://www.privacyrights.org/`.

[63] S. Pötzsch. Privacy Awareness: A Means to Solve the Privacy Paradox? In *The Future of Identity in the Information Society*, volume 298 of *IFIP Advances in Information and Communication Technology*, pages 226–236. 2009.

[64] C. Riederer, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and P. Rodriguez. For sale : Your data: By : You. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, HotNets-X, pages 13:1–13:6, New York, NY, USA, 2011. ACM.

[65] A. Rosenberg. Privacy as a matter of taste and right. *Social Philosophy and Policy*, 17:68–90, 6 2000.

[66] J. Samuel and B. Zhang. RequestPolicy: Increasing Web Browsing Privacy through Control of Cross-Site Requests. In *PETS '09*, pages 128–142, 2009.

[67] B. Schneier. Risks of third-party data. *Commun. ACM*, 48(5):136, 2005.

[68] F. D. Schoeman. *Privacy and Social Freedom*. Cambridge University Press, 1992.

[69] B. K. Sheehan. An investigation of gender differences in online privacy concerns and resultant behaviors. *Journal of Interactive Marketing*, 13(4):24–38, 1999.

[70] D. J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review. GWU Law School Public Law Research Paper No. 129*, 154(3):477–553, 2006.

[71] S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. In *ACM Conference on Electronic Commerce*, pages 38–47. ACM, 2001.

[72] J. Staddon, D. Huffaker, L. Brown, and A. Sedley. Are privacy concerns a turn-off?: engagement and privacy in social networks. In *Proceedings of the Eighth SOUPS*, SOUPS '12, pages 10:1–10:13, 2012.

[73] L. S. Strickland and L. E. Hunt. Technology, security, and individual privacy: New tools, new threats, and new public perceptions: Research articles. *JASIST*, 56(3):221–234, Feb. 2005.

[74] F. Stutzman, R. Capra, and J. Thompson. Factors mediating disclosure in social network sites. *Comput. Hum. Behav.*, 27(1):590–598, Jan. 2011.

[75] J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research*, 22:254–268, June 2011.

[76] UPI. UPI Poll: Concern on health privacy. `http://www.upi.com/Topuist_News/2007/02/21/UPI-Poll-Concern-on-health-privacy/UPI-39291172098800/`.

[77] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth SOUPS, SOUPS '12*, pages 4:1–4:15.

[78] M. van der Velden and K. E. Emam. *Not all my friends need to know"*: a qualitative study of teenage patients, privacy, and social media. *Journal of the American Medical Informatics Association, JAMIA '13*, 20(1):16–24, 2013.

[79] R. J. Walls, S. S. Clark, and B. N. Levine. Functional Privacy or Why Cookies are Better with Milk. In *USENIX Workshop on Hot Topics in Security*, 2012.

[80] S. D. Warren and L. D. Brandeis. The right to privacy. *Harward Law Review*, 4(5):193–220, December 1890.

[81] A. Westin. *Privacy and Freedom*. New Jork Atheneum, New York, 1967.

[82] A. Westin. Harris-equifax consumer privacy survey. Technical report, WESTIN, A. AND HARRIS LOUIS & ASSOCIATES. Conducted for Equifax Inc, 1991.

[83] C. E. Wills and M. Zeljkovic. A Personalized Approach to Web Privacy - Awareness, Attitudes and Actions. *Information Management & Computer Security*, 19(1):53–73, 2011.

[84] O. Yael, G. Michael, D. Yaniv, Z. Ariel, T. Rom, L. Inbal, m. Arz, M. Tamar, N. Yossi, S. Yaniv, S. Saar, F. Adi, F. Maor, P. Shai, and P. Lotem. *Privacy in the Digital Environment*. 2005.