NET PRIVACY

Claudio Agostino Ardagna, Chiara Braghin and Marco Cremonini
University of Milan, Italy

Abstract

In recent years, large-scale computer networks have become an essential aspect of our daily computing environment: we often rely on a global information infrastructure for e-business activities such as home banking, ATM transactions, or shopping online. One of the main scientific and technological challenges in this setting has been to provide security to individuals that operate in possibly untrusted and unknown environments. However, beside threats directly related with computer intrusions, epidemic diffusion of malwares, and plain frauds conducted online, a more subtle although increasing erosion of individuals' privacy has progressed and multiplied. Such an escalating violation of privacy has some direct harmful consequences—for example, identity thefts have spread in recent years—and negative effects on the general perception of insecurity that many individuals now experience when dealing with online services.
Nevertheless, protecting personal privacy from the many parties—business, government, social, or even criminal—which look over the value that personal information have, is an ancient concern of modern society, now increased by the features of the digital infrastructure.
In this chapter, we address the privacy issues in the digital society from different points of view, investigating:

i) the different aspects that the notion of privacy covers and the debate that the intricate essence of privacy has stimulated;
ii) the most common privacy threats and the possible economic aspects that may influence the way privacy is (and especially is not, at the current status) managed in most of the firms;
iii) the efforts, in the Computer Science community, to face privacy threats, especially in the context of mobile and database system;
iv) the network-based technologies currently available to provide anonymity when communicating over a public network.

PRIVACY IN THE DIGITAL SOCIETY

The Origins, the Debate

Privacy in today digital society is one of the most debated and controversial topics. Many different opinions about what privacy actually is and how it could be preserved have been expressed, but still no clear cut can be made to set the border that cannot be trespassed if privacy has to be safeguarded.

As it often happens when a debate heats up, the extremes speak louder. About privacy, the extremes are those that advocate the ban of the disclosure of whatever personal information, or those that just say that all personal information are already out there, therefore privacy is just dead. Supporters of the wide deployment and usage of anonymizing technologies are perhaps the best representatives of one extreme. The Chief Executive Officer of Sun Microsystems, Scott McNealy, with his "Get over it" has gained large notoriety for championing the other extreme opinion (Sprenger 1999).
However, these are just the extremes, in reality net privacy is a fluid concept that such radical positions cannot fully contain. It is a fact that even those supporting full anonymity recognize that there are several limitations to its adoption, either technical or functional. On the other side, even the most skeptics cannot avoid to deal with privacy issues, either because of laws and norms, or because of common sense. Sun Microsystems, for example, is actually supporting privacy protection and is a member of the Online Privacy Alliance, an industry coalition that fosters the protection of individuals' privacy online.

Looking at the origins of the concept of privacy, Aristotle's distinction between the public sphere of politics and the private sphere of the family is often considered as the root. Much later, the philosophical and anthropological debate around these two spheres of an individual's life evolved. John Stuart Mill in his essay, "On Liberty", introduced the distinction between the realm of governmental authority as opposed to the realm of self-regulation. Anthropologists like Margaret Mead have demonstrated how the need of privacy is innate in different cultures that protect it through concealment, seclusion or by restricting access to secret ceremonies.

More pragmatically, back to 1898, the concept of privacy was expressed by U.S. Supreme Court Justice Brandeis, which defined privacy as "The right to be let alone" (Warren & Brandeis 1890). This straightforward definition represented for decades the reference of any normative and operational privacy consideration and derivate issues and, before the advent of the digital society, a realistically enforceable ultimate goal. The Net has changed the landscape, because the very concept of being let alone while interconnected becomes fuzzy and fluid.

In 1948, privacy has gained the status of fundamental right of any individual, being explicitly mentioned in the United Nations Universal Declaration of Human Rights (Article 12): "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks." (United Nations 1948). However, although privacy has been recognized as a fundamental right of each individual, the Universal Declaration of Human Rights does not explicitly define what privacy is, simply relating it to possible interferences or attacks.

About the digital society, less rigorously but otherwise effectively in practical terms, in July 1993, The New Yorker published a brilliant cartoon by Peter Steiner that since then has been cited and reproduced dozen of times to refer to the supposed intrinsic level of privacy—here in the sense of anonymity or hiding personal traits—that can be achieved by carrying out social relations over the Internet. That famous cartoon goes with one dog that types on a computer keyboard and says to the other one: "On the Internet, no one knows you're a dog." (Steiner 1993). The Internet, at least at the very beginning of its history, was not perceived as threatening individuals' privacy, rather it was seen as increasing it, sometimes too much, since it could easily let everyone to disguise in the course of personal relationships. Today that belief may look naïve with the rise of threats to individual privacy that have accompanied the diffusion of the digital society. Nevertheless, there is still truth in that cartoon because, whereas privacy is much weaker on the Net than in the real space, even the possibility to conceal the own identity and personal traits is technically easier. Both aspects concur and should be considered.

Yet an unambiguous definition of the concept of privacy has not still been produced, as well as an assessment of its actual value and scope. It is however clear that with the term "privacy" we refer to a fundamental right and an innate feeling of every individual, not to a vague and mysterious entity. An attempt to give a precise definition at least to some terms that are strictly related to (and often used in place of) the notion of privacy can be found in (Pfitzmann & Waidner 1986; Pfitzmann & Köhntopp 2001), where the differences between *anonymity*, *unobservability* and *unlinkability* is pointed out. In the digital society scenario, anonymity is defined as the state of not being identifiable, unobservability as the state of being indistinguishable and unlinkability as the impossibility of correlating two or more actions/items/pieces of information. Privacy, however defined and valued, is a tangible state of life that must be attainable both in the physical and in the digital society.

The reason why in the two realms—the physical and the digital one—privacy behaves differently has been widely debated too, and many of the critical factors that make a difference in the two realms have been spelled out clearly. However, while it is clear that information technology and the Internet amplify threats to privacy, they also permit to develop safeguards and to mitigate risks.

Lessig in his book "Free Culture" (Lessig, 2003) provided an excellent explanation of the difference between privacy in the physical and in the digital world: "The highly inefficient architecture of real space means we all enjoy a fairly robust amount of privacy. That privacy is guaranteed to us by friction. Not by law […] and in many places, not by norms […] but instead, by the costs that friction imposes on anyone who would want to spy. […] Enter the Internet, where the cost of tracking browsing in particular has

become quite tiny. […] The friction has disappeared, and hence any “privacy” protected by the friction disappears, too."

Thus, privacy can be seen as the friction that reduces the spread of personal information, that makes more difficult and economically not convenient to gain the access to. The merit of this definition is to put privacy into a relative perspective, which excludes the extremes that advocate no friction at all or so much friction to stop the flow of information. It also reconciles privacy with security, being both aimed at setting an acceptable level of protection while allowing the development of the digital society and economy, rather than focusing on an ideal state of perfect security and privacy.

Even in an historic perspective, the analogy with friction has sense. The natural path of evolution of a technology is first to push for its spreading and best efficiency. When the technology matures, other requirements come to the surface and gain importance with respect to the mere efficiency and functionalities. Here, those frictions that have been eliminated because just a waste of efficiency, acquire new meaning and become the way to satisfy the new requirements, either in terms of safety, security or even privacy. It is a sign that a technology has matured but not yet found a good balance between old and new requirements when non functional aspects such as security or privacy become critical because not well managed and integrated.

Privacy Threats

Threats to individuals' privacy have become publicly appalling since July 2003 when the California Security Breach Notification Law (California 2002) went into effect. This law was the first one to force state government agencies, companies and nonprofit organizations that conduct business in California, to notify California customers if personally identifiable information (PII) stored unencrypted in digital archives was, or is reasonably believed to have been, acquired by an unauthorized person.

The premise for this law was the rise of "identity theft", which is the conventional expression that has been used to refer to the illicit impersonification carried out by fraudsters that use PII of other people to complete electronic transactions and purchases. The California Security Breach Notification Law lists, as PII: Social security number, driver's license number, California Identification Card number, account number, credit or debit card number, security code, access code, or passwords that would permit access to an individual's financial account (California 2002). By requiring by law the immediate notification to the PII owners, the aim is to avoid direct consequences such as financial losses and derivate consequences such as the burden to restore individual's own credit history. Starting from January 1, 2008, California's innovative data security breach notification law also applies to medical information and health insurance data.

This law, beside the benefits to customers, has been the trigger to similar laws in the U.S.A. —today, the majority of U.S. states has one—and has permitted the flourish of regular statistics about privacy breaches, once almost absent. Privacy threats and analyses are now widely debated and research focused on privacy problems has become one of the most important. Figure 1 shows a chart produced by plotting data collected by Attrition.org Data Loss Archive and Database (Attrition 2008), one of the most complete references for privacy breaches and data losses.

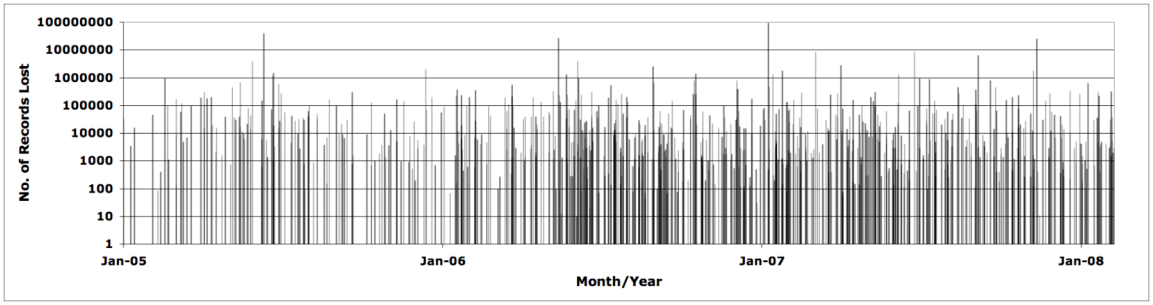

Figure 1: Privacy breaches from the Attrition.org *Data Loss Archive and Database* up to March 2008 (*X-axis: Years 2005-2008; Y-axis (logarithmic): PII records lost*)

Looking at the data series, some breaches are strikingly large. Etiolated.org maintains some statistics based on Attrition.org's database: in 2007, about 94 million records have been hacked at TJX stores in USA; confidential details of 25 million child have been lost by HM Revenue and Customs, UK; the Dai Nippon Printing Company in Tokyo lost more than 8 millions records; data about 8.5 million people stored by a subsidiary of Fidelity National Information Services were stolen and sold for illegal usage by a former employee. A similar trail path has been reported in previous years as well. In 2006, personal data of about 26.5 million U.S. military veterans was stolen from the residence of a Department of Veterans Affairs data analyst who improperly took the material home. In 2005, CardSystems Solutions—a credit-card processing company managing accounts for Visa, MasterCard and American Express—exposed 40 million debit and credit-card accounts in a cyber break-in. In 2004, an employee of America Online Inc. stole 92 million e-mail addresses and sold them to spammers. Still recently, on March 2008, Hannaford Bros. supermarket chain announced that, due to a security breach, about 4.2 millions customer credit and debit card numbers were stolen (Etiolated.org 2008).

Whereas these incidents are the most notable ever occurred, the phenomenon is distributed over the whole spectrum of breach sizes (see Figure 1). Hundreds of privacy breaches are reported in the order of few thousands records lost and all categories of

organizations are affected, from public agencies, universities, banks to financial institutions, manufacturing and retail companies, etc.

The survey "Enterprise@Risk: 2007 Privacy & Data Protection" conducted by Deloitte & Touche and Ponemon Institute (Deloitte 2007) provides another piece of data about the incidence of privacy breaches. Among the survey's respondents, over 85% reported at least one breach and about 63% reported multiple breaches requiring notification during the same time period. Breaches involving over 1000 records were reported by 33.9% of respondents, of those, almost 10% suffered data losses of more than 25,000 records. Astonishingly, about 21% of respondents were not able to estimate the record loss. The picture that results is that of a pervasive management problem with regard to PII and their protection, which causes a continuous leakage of chunks of data and few dramatic breakdowns when huge archives go lost or theft.

It is interesting to analyze the root causes for such breaches and the type of information involved. One source of information is the "Educational Security Incidents (ESI) Year In Review – 2007" (Dodge 2008) by Adam Dodge. This survey lists all breaches that occurred worldwide during 2007 at colleges and universities around the world.
For what concerns the causes of breaches, the results over a total of 139 incidents are:

- 38% are due to *unauthorized disclosure*;
- 28% to *theft* (disks, laptops);
- 22% to *penetration*/*hacking*;
- 9% to *loss* of data.

Therefore, incidents to be accounted to mismanagement by employees (unauthorized disclosure and loss) accounts for 47%, while criminal activity (penetration/hacking and theft) accounts for 40%.
With respect to the type of information exposed during these breaches, the result is that:

- *PII* have been exposed in 42% of incidents;
- *Social Security Numbers* in 34%;
- *educational* information in 11%;
- *financial* information in 7%;
- *medical* information in 5%;
- *login accounts* in 2%.

Again, rather than direct economic consequences or illicit usage of computer facilities, such breaches represents threats to individuals' privacy.

Privacy Rights ClearingHouse is another organization that provides excellent data and statistics about privacy breaches. Among others, it is particularly remarkable their analysis of root causes for different sectors, namely the private sector, the public sector (military included), higher education and medical centers (Privacy Rights ClearingHouse 2007). Table 1 reports their findings for year 2006.

Table 1: Root causes of data breaches – Year 2006
(*Source: Privacy Rights ClearingHouse*)

	Private Sector (126 incidents)	Public Sector (inc. military) (114 incidents)	Higher Education (52 incidents)	Medical Centers (30 incidents)
Outside Hackers	15%	13%	40%	3%
Insider Malfeasance	10%	5%	2%	20%
Human/Software Incompetence	20%	44%	21%	20%
Theft (non-laptop)	15%	17%	17%	17%
Laptop Theft	40%	21%	20%	40%

Comparing these results with the previous statistics, the "Educational Security Incidents (ESI) Year In Review – 2007, breaches caused by hackers in universities look remarkably different. Privacy Rights ClearingHouse estimates as largely prevalent the external criminal activity (hackers and theft), which accounts for 77%, with respect to internal problems, which accounts for 19%, while in the previous study the two classes were closer with a prevalence of internal problems.

Hasan and Yurcik (Hasan 2006) analyzed data about privacy breaches occurred in 2005 and 2006 by fusing datasets maintained by Attrition.org and Privacy Rights ClearingHouse. The overall result partially clarifies the discrepancy that results from the previous two analyses. In particular, it emerges that considering the number of privacy breaches, education institutions are the most exposed, accounting for 35% of the total, followed by companies with 25% and state-level public agencies, medical centers and banks all close to 10%. However, by considering personal records lost by sectors,

companies lead the score with 35.5%, followed by federal agencies with 29.5%, medical centers with 16% and banks with 11.6%. Education institutions record lost total for just 2.7% of the whole. Therefore, while universities are victimized by a huge numbers of external attacks that causes a continuous leakage of PII, companies and federal agencies are those that have suffered or provoked ruinous losses of enormous archives of PII. For these sectors, the impact of external Internet attacks has been matched or even exceeded by internal frauds or misconduct. The case of consumer data broker ChoicePoint, Inc. is perhaps the one that got the most publicity as an example of bad management practices that led to a huge privacy incident (Scalet 2005). In 2006, the Federal Trade Commission charged that ChoicePoint violated the Fair Credit Reporting Act (FCRA) by furnishing consumer reports—credit histories—to subscribers who did not have a permissible purpose to obtain them, and by failing to maintain reasonable procedures to verify both their identities and how they intended to use the information (FTC 2006).

The fact that threats due to hacking have been overhyped with respect to others is an opinion shared by many in the security community. In fact, it appears that considering root causes of privacy breaches, physical thefts—of laptops, disks and portable memories—and bad management practices—sloppiness, incompetence and scarce allocation of resources—need to be considered at least as serious as hacking. This is confirmed by the survey "Enterprise@Risk: 2007 Privacy & Data Protection" conducted by Deloitte & Touche and Ponemon Institute (Deloitte 2007), which concludes that most enterprise privacy programs are just in the early or middle stage of the maturity cycle. Requirements imposed by laws and regulations have the highest rates of implementation, while operational processes, risk assessment and training programs are less adopted. In addition, a minority of organization seems able to implement measurable controls, a deficiency that makes privacy management intrinsically feeble. Training programs dedicated to privacy, security and risk management look as the weakest spot. Respondents report that training on privacy and security is offered just annually (about 28%), just once (about 36.5%) or never (about 11%). Risk management is never the subject of training for almost 28% or respondents. With such figures, it is no surprise if internal negligence due to unfamiliarity with privacy problems or insufficient resources is such a relevant root cause for privacy breaches.

The Choicepoint incident is paradigmatic of another important aspect that has been considered for analyzing privacy issues. The breach involved 163.000 records and it was carried out with the explicit intention of unauthorized parties to catch those records. However, actually just in 800 cases (about 0.5%) that breach leads to identity theft, a severe offense suffered by Choicepoint customers. Some analysts have questioned the actual value of privacy, this conducts us to discuss an important strand of research about economic aspects of privacy.

ECONOMICS OF PRIVACY

The existence of strong economic factors that influence the way privacy is managed, breached or even traded off has been recognized since long (Hirshleifer 1971; Posner 1981). However, it was with the expansion of the online economy, in the 1990s and 2000s, that privacy and economy become more and more entangled. Many studies have been produced to investigate, from different perspectives and approaches, the relation between the two. A comprehensive survey of works that analyzed the economic aspects of privacy can be found in (Hui et al. 2006a).

Two issues, among the many, have gained most of the attention: assessing the value of privacy and examining to which extent privacy and business can coexist or are inevitably conflicting one with the other. For both issues the debate is still open and no ultimate conclusion has been reached yet.

The Value of Privacy

For the analysts, the problem of estimating the value of privacy has been the more puzzling one since years. On the one hand, people assign high value to their privacy when asked, on the other hand, privacy is more and more eroded and given away for small rewards. Several empirical studies have tested individuals' behavior when confronted with the decision to trade off privacy for some rewards or incentives, and when confronted with the decision to pay for protecting their personal information. The approaches to these studies vary from investigating the actual economic factors that determine people choices, to the psychological motivation and perception of risk or safety.

Syverson (Syverson 2003), then Shostack and Syverson (Shostack & Syverson 2004), analyzed the apparently irrational behavior of people which claim to highly value privacy then, in practice, they are keen to release sensible personal information for small rewards. The usual conclusion is that people are not actually able to assess the value of privacy, or that they are either irrational or unaware about the risks they are taking. While there are evidences that risks are often miscalculated or just unknown by most people, there are also some valid reasons to justify such a paradoxical behavior. In particular, the analysis points to the cost of examining and understanding privacy policies and practices, which often makes privacy a complex topic to manage. Another observation regards the cost of protecting privacy, which is often inaccurately allocated. Better reallocation would also provide government and business with incentives to increase rather than decrease protection of individual privacy.

One study that dates back to 1999 by Culnan and Armstrong (Culnan & Armstrong 1999) investigated how firms that demonstrate to adopt fair procedures and ethical behavior can mitigate consumers' concerns about privacy. Their finding was that consumers that perceive that the collection of personal information is ruled by fair procedures are more willing to release their data for marketing use. This supported the hypothesis that most privacy concerns are motivated by an unclear or distrustful stance towards privacy protection that often firms exhibit.

In 2007, Tsai et al. (Tsai et al. 2007) published a research that addresses much the same issue. The effect of privacy concerns on online purchasing decision has been tested and the results are again that the role of incomplete information on privacy-relevant decisions is essential. Consumers are sensitive to the way privacy is managed and to what extent a merchant is trustful. However, in another study, Grosslack and Acquisti (Grosslack & Acquisti 2007) found that individuals almost always choose to sell their personal information when offered with small compensation rather than keep it confidential.

Hann, Lee, Hui and Png have carried out a more analytic work in two studies about online information privacy. This strand of research (Hann et al. 2002) estimated how much privacy is worth for individuals and how economic incentives, such as monetary rewards and future convenience, could influence such values. Their main findings are that individuals do not esteem privacy as an absolute value, rather they are available to trade off it for economic benefits, and that improper access and secondary use of personal information are the most important classes of privacy violation. In the second work (Hann et al. 2007), the authors considered firms that tried to mitigate privacy concerns by offering privacy policies regarding the handling and use of personal information, and by offering benefits such as financial gains or convenience. These strategies have been analyzed in the context of the information processing theory of motivation, which consider how people form expectations and make decisions about what behavior to choose. Again, whether a firm may only offer partially complete privacy protection or some benefits, economic rewards and convenience have been found to be strong motivators for increasing the individuals' willingness to disclose personal information.

Therefore, most works seems to converge to the same conclusion: whether individuals react negatively when incomplete or distrustful information about privacy are presented, even a modest monetary reward is often sufficient for disclosing one's personal information.

Privacy and Business

The relation between privacy and business has been examined from several angles by considering which incentives could be effective for integrating privacy with business processes and, instead, which disincentives make business motivations to prevail over privacy.

Froomkin (Froomkin 2000) analyzed what he called "privacy-destroying technologies" developed by governments and businesses. Examples of such technologies are collection of transactional data, automated surveillance in public places, biometric technologies, tracking mobile devices and positioning systems. To further aggravate the impact on privacy of each one of these technologies, their combination and integration result in a cumulative and reinforcing effect. On this premise, Froomkin introduces the role that legal responses may play to limit this apparently unavoidable "death of privacy".

Odlyzko (Odlyzko 2003a, 2003b, 2004, 2007) is a leading author that holds a pessimistic view of the future of privacy, by calling "unsolvable" the problem of granting privacy because of price discrimination pressures on the market. His argument is based on the observation that the market as a whole, especially Internet-based markets, has strong incentives to price discriminate, i.e., charging varying prices when there are no cost justifications for the differences. This practice, which has its roots long before the advent of the Internet and the modern economy—one of the most illustrative examples are the 19th century railroad pricing practices—provides relevant economic benefits to the vendors and, from a mere economic viewpoint, to the efficiency of the economy. In general, charging different prices to different segments of the customer base permits to complete transactions that would not take place otherwise. On the other hand, the public has often contrasted plain price discrimination practices since they perceive them as unfair. For this reason, many less evident price discrimination practices are in place today, among which, bundling is one of the most recurrent. Privacy of actual and prospective customers is threatened by such economic pressures towards price discrimination because the more the customer base can be segmented—and thus known with greatest details—the better efficiency is achieved for vendors. The Internet-based market has provided new boost to such practices and to the acquisition of personal information and knowledge of customer's habits.

Empirical studies seem to confirm such pessimistic views. A first review about the largest privately held companies listed in the Forbes Private 50 (Peslak 2005), and a second study about firms listed in the Fortune 500 (Schwaig et al. 2006) demonstrate a poor state of privacy policies adopted in such firms. In general, privately held companies are most likely to lack privacy policies than public companies and are more reluctant to publicly disclose their procedures relative to fair information practices. Even the larger

set of the Fortune 500 firms exhibited a large majority of firms that are just mildly addressing privacy concerns.

More pragmatically, some analyses have pointed out that given the current privacy concerns, an explicitly fair management of customers' privacy may become a positive competitive factor (Brown & Muchira 2004). Similarly, Hui et al (Hui et al. 2006b) have identified seven types of benefits that Internet businesses can provide to consumers in exchange of their personal information.

PRIVACY-ENHANCING TECHNOLOGIES

The technical improvements of Web and of location technologies have fostered the development of online applications that use private information of users (including physical position of individuals) to offer enhanced services. The increasing amount of personal data available and the decreasing cost of data storage and processing make it technically possible and economically justifiable to gather and analyze large amount of data. Also, information technology gives organizations the power to manage and disclose personal information of users without restrictions. In this context, users are much more concerned about their privacy, and privacy has been recognized as one of the main reasons that prevent users from using the Internet for accessing online services. Today global networked infrastructure requires the ability for parties to communicate in a secure environment while at the same time preserving their privacy. Support for digital identities and definition of privacy-enhanced protocols and techniques for their management and exchange become then fundamental requirements. A number of useful privacy enhancing technologies (PETs) have been developed for dealing with privacy issues and previous works on privacy protection have focused on a wide variety of topics (Ardagna et al 2005; Chandramouli 2005; Cranor 2002; Karjoth & Schunter 2002; Thuraisingham 2005; Youssef et al. 2005). In the following of this section, we discuss the privacy protection problem in three different contexts. We start by describing languages for the specification of access control policies and privacy preferences. We then describe the problem of data privacy protection giving a brief description of some solutions. Finally, we analyze the problem of protecting privacy in mobile and pervasive environments.

Languages for Access Control and Privacy Preferences

Access control systems have been introduced in the past for regulating and protecting access to resources and data owned by parties. However, the importance gained by privacy requirements has brought to the definition of access control models enriched

with the ability of supporting privacy requirements. These enhanced access control models encompass two aspects: to guarantee the desired level of privacy of information exchanged between different parties by controlling the access to services/resources; and to control all secondary uses of information disclosed for the purpose of access control enforcement.

In this context, many languages for access control policies and privacy preferences specification have been defined, among which eXtensible Access Control Markup Language (XACML) (XACML 2005), Platform for Privacy Preferences Project (P3P) (Cranor 2002; W3C 2002b) and Enterprise Privacy Authorization Language (EPAL) (Ashley et al. 2002; Ashley et al. 2003) stand out.

The *eXtensible Access Control Markup Language (XACML)* (XACML 2005), which is the result of a standardization effort by OASIS, proposes a XML-based language to express and interchange access control policies. It is not specifically designed for managing privacy, but it represents a relevant innovation in the field of access control policies and has been used as the basis for following privacy-aware authorization languages. Main features of XACML are: *i) policy combination*, a method for combining policies on the same resource independently specified by different entities; *ii) combining algorithms*, different algorithms representing ways of combining multiple decisions into a single decision; *iii) attribute-based restrictions*, the definition of policies based on properties associated with subjects and resources rather than their identities; *iv) multiple subjects*, the definition of more than one subject relevant to a decision request; *v) policy distribution*, policies can be defined by different parties and enforced at different enforcement points; *vi) implementation independence*, an abstraction layer that isolates the policy-writer from the implementation details; *vii) obligations* (Bettini et al. 2002), a method for specifying the actions that must be fulfilled in conjunction with the policy enforcement.

Platform for Privacy Preferences Project (P3P) (Cranor 2002; W3C 2002b) is a W3C (World Wide Web Consortium) project aimed at protecting the privacy of users by addressing their need of assessing that the privacy practices adopted by a server provider comply with her privacy requirements. P3P provides a XML-based language and a mechanism for ensuring that users can be informed about privacy policies of the server before the release of personal information. Therefore, P3P allows Web sites to declare their privacy practices in a standard and machine-readable XML format known as P3P policy. A P3P policy contains the specification of the data it protects, the data recipients allowed to access the private data, consequences of data release, purposes of data collection, data retention policy, and dispute resolution mechanisms. Supporting privacy preferences and policies in Web-based transactions allows users to automatically understand and match server practices against their privacy preferences. Thus, users need

not read the privacy policies at every site they interact with, but they are always aware of the server practices in data handling. In summary, the goal of P3P is twofold: it allows Web sites to state their data-collection practices in a standardized, machine-readable way and it provides to users a solution to understand what data will be collected and how those data will be used.
The corresponding language that would allow users to specify their preferences as a set of preference-rules is called *A P3P Preference Exchange Language* (APPEL) (W3C 2002a). APPEL can be used by users agents to reach automated or semi-automated decisions regarding the acceptability of privacy policies from P3P enabled Web sites. Unfortunately, as stated in (Agrawal et al. 2003), interactions between P3P and APPEL had shown that users can explicitly specify just what is unacceptable in a policy, while the APPEL syntax is cumbersome and error prone for users.

Finally, *Enterprise Privacy Authorization Language* (EPAL) (Ashley et al. 2002; Ashley et al. 2003) is another XML-based language for specifying and enforcing enterprise-based privacy policies. EPAL is specifically designed to enable organizations to translate their privacy policies into IT control statements and to enforce policies that may be declared and communicated according to P3P specifications.

In this scenario, the need of access control frameworks that integrate policy evaluation and privacy functionalities arose. A first attempt to provide a uniform framework for regulating information release over the Web has been presented by Bonatti and Samarati (Bonatti & Samarati 2002). Afterwards, a solution that introduced a privacy-aware access control framework was defined by Ardagna et al. (Ardagna et al. 2008). This framework allows the integration, evaluation and enforcement of policies regulating access to service/data and release of personal identifiable information, respectively, and provides a mechanism to define constraints on the secondary use of personal data for the protection of users privacy. In particular, the following different types of privacy policies have been specified:

- *Access control policies.* They govern access/release of data/services managed by the party (as in traditional access control).
- *Release policies.* They govern release of properties/credentials/personal identifiable information (PII) of the party and specify under which conditions they can be released.
- *Data handling policies.* They define how personal information will be (or should be) dealt with at the receiving parties (Ardagna et al. 2006b).

An important feature of this framework is to support requests for certified data, issued and signed by trusted authorities, and uncertified data, signed by the owner itself. Il also allows defining conditions that can be satisfied by means of zero-knowledge proof

(Camenisch & Lysyanskaya 2001; Camenisch & Van Herreweghen 2002) and based on physical position of the users (Ardagna et al. 2006a). In the context of the Privacy and Identity Management for Europe (PRIME) (PRIME 2004), a European project whose goal is the development of privacy-aware solutions for enforcing security, an implementation of the privacy-aware access control framework has been provided (Ardagna et al. 2008). Such a prototype is part of a general architecture aimed at providing a full privacy-aware identity management solution, and integrates traditional access control mechanisms with release and data handling policies management and evaluation.

Data Privacy Protection

The concept of anonymity has been introduced first in the context of relational database to avoid linking between published data and users' identity. Usually, to protect anonymity of users, data holders encrypt or remove explicit identifiers such as, name and social security number (SSN). However, data de-identification does not provide full anonymity. Released data in fact can be linked to other publicly available information to re-identify users and to infer data that should not be available to the recipients. For instance, a set of anonymized data could contain attributes that almost uniquely identify a user, such as for instance race, date of birth, and ZIP code. Table 2 shows an example where the anonymous medical data contained in a table are linked with the census data to re-identify users. It is easy to see that in Table 2(a), there is a unique tuple with a **male** born on **03/30/1938** and living in the area with ZIP code **10249**. As a consequence, if this combination of attributes is unique also in the census data in Table 2(b), John Doe is identified, revealing that he suffers of obesity.

Table 2: User re-identification

Anonymous Medical Data

SSN	Name	DateofBirth	Sex	ZIP	Marital Status	Disease
		09/11/1984	M	10249	Married	HIV
		09/01/1978	M	10242	Single	HIV
		01/06/1959	F	10242	Married	Obesity
		01/23/1954	M	10249	Single	Hypertension
		03/15/1953	F	10212	Divorced	Hypertension
		03/30/1938	**M**	**10249**	**Single**	**Obesity**
		09/18/1935	F	10212	Divorced	Obesity
		03/15/1933	F	10252	Divorced	HIV

(b)

Census Data

SSN	Name	Address	City	DateofBirth	ZIP	…
…	…	…	…	…	…	…
…	**John Doe**	**…**	**New York**	**03/30/1938**	**10249**	**…**
…	…	…	…	…	…	…

(a)

If in the past the limited interconnectivity and the limited computational power represented a form of protection against inference processes over large amount of data, today with the advent of Internet, such an assumption does not hold anymore. Information technology in fact gives organizations the power to gather and manage vast amounts of personal information.

To address the problem of protecting anonymity while releasing microdata, the concept of *k*-anonymity has been defined. *K*-anonymity means that the observed data cannot be related to less than *k* respondents (Samarati 2001). Key to achieve *k*-anonymity is the identification of a *quasi-identifier*, which is the set of attributes in a dataset that can be linked with external information to re-identify the data owner. It follows that for each release of data, every combination of values of the quasi-identifier must be indistinctly matched to at least *k* tuples.
Two approaches to achieve *k*-anonymity have been adopted: *generalization* and *suppression*. Both approaches share the important feature that the truthfulness of the information is preserved, i.e., no false information are released.
More in details, the *generalization* process generalizes some of the values stored in the table. For instance, considering the ZIP code attribute in Table 2 and supposing for simplicity that it represents a quasi-identifier, the ZIP code can be generalized by dropping, at each step of generalization, the least significant digit. As another example, the date of birth can be generalized by first removing the day, then the month, and eventually by generalizing the year.
On the contrary, the suppression process removes some tuples from the table. Again, considering Table 2, the ZIP codes, and a *k*-anonymity requirement for k=2, it is clear that all tuples already satisfy the k=2 requirement except for the last one. In this case to preserve the k=2, the last tuple could be just suppressed.

Research on *k*-anonimity has been particularly rich in recent years. Samarati (Samarati 2001) presented an algorithm based on generalization hierarchies and suppression that calculates the minimal generalization. The algorithm relies on a binary search on the domain generalization hierarchy to avoid an exhaustive visit of the whole generalization space. Bayardo and Agrawal (Bayardo & Agrawal 2005) developed an optimal bottom-up algorithm that starts from a fully generalized table (with all tuples equal) and then specializes the dataset into a minimal *k*-anonymous table. LeFevre et al. (LeFevre et al. 2005) are the authors of *Incognito*, a framework for providing *k*-minimal generalization.

Their algorithm is based on a bottom-up aggregation along dimensional hierarchies and a-priori aggregate computation. The same authors (LeFevre et al. 2006) introduced also *Mondrian k*-anonymity, which models the tuples as points in *d*-dimensional spaces and applies a generalization process that consists in finding the minimal multidimensional partitioning that satisfy the *k* preference.

Although the advantage of *k*-anonymity for protecting respondents' privacy, some weaknesses have been demonstrated. Machanavajjhala et al. (Machanavajjhala et al. 2006) identified two successful attacks to *k*-anonymous table: *i) homogeneity attack* and *ii) background knowledge attack.* To explain the *homogeneity attack*, suppose that a *k*-anonymous table contains a single sensitive attribute. Suppose also that all tuples with a given quasi-identifier value have the same value for that sensitive attribute too. As a consequence, if the attacker knows the quasi-identifier value of a respondent is able to learn the value of the sensitive attribute associated to the respondent. For instance, consider the 2-anonymous table showed in Table 3 and assume that an attacker knows that `Alice` is born on **1966** and lives in the **10212** ZIP code area. Since all tuples with quasi-identifier <1966,F,10212> suffers of anorexia, the attacker can infer that `Alice` suffers of anorexia. Focusing on the *background knowledge attack*, the attacker exploits some a-priori knowledge to infer some personal information. For instance, suppose that an attacker knows that `Bob` has quasi-identifier <1984,M,10249> and that `Bob` is overweighed. In this case, from Table 3, the attacker can infer that `Bob` suffers of HIV.

Table 3: An example of 2-anonymous table

YearofBirth	Sex	ZIP	Disease
1984	M	10249	HIV
1984	M	10249	anorexia
1984	M	10249	HIV
1966	F	10212	anorexia
1966	F	10212	anorexia
…	…	…	…

To neutralize these attacks, the concept of *l*-diversity has been introduced (Machanavajjhala et al. 2006). In particular, a cluster of tuples with the same quasi-identifier is said to be *l*-diverse if it contains at least *l* different values for the sensitive attribute (disease in the example in Table 3). If a *k*-anonymous table is *l*-diverse, the homogeneity attack is ineffective, since each block of tuples has at least *l*>=2 distinct values for the sensitive attribute. Also, the background knowledge attack becomes more complex as *l* increases.

Although *l*-diversity protects data against attribute disclosure, it leaves space to more sophisticated attacks based on the distribution of values inside clusters of tuples with the same quasi-identifier (Li et al. 2007). To prevent this kind of attacks the *t*-closeness requirement has been defined. In particular, a cluster of tuples with the same quasi-identifier is said to satisfy *t*-closeness if the distance between the probabilistic distribution of the sensitive attribute in the cluster and the one in the original table is lower than *t*. A table satisfies *t*-closeness if all its cluster satisfy *t*-closeness.
In the next section, where the problem of location privacy protection is analyzed, we also discuss how the location privacy protection problem has adapted the *k*-anonymity principle to a pervasive and distributed scenario, where users move on the field carrying a mobile device.

Privacy for Mobile Environments

The widespread diffusion of mobile devices and the accuracy and reliability achieved by positioning techniques make available a great amount of location information of users. Such information has been used for developing novel location-based services. However, if on the one side such a pervasive environment provides many advantages and useful services to the users, on the other side, privacy concerns arise since users could be the target of fraudulent location-based attacks. The most pessimistic have even predicted that the unrestricted and unregulated availability of location technologies and information could lead to a "Big Brother" society dominated by total surveillance of individuals.

The concept of *location privacy* can be defined as the right of individuals to decide how, when, and for which purposes their location information could be released to other parties. The lack of location privacy protection could be exploited by adversaries to perform different attacks (Duckham & Kulik 2006):

- *unsolicited advertising,* when the location of a user could be exploited, without her consent, to provide advertisements of products and services available nearby the user position;
- *physical attacks or harassment*, when the location of a user could allow criminals to carry physical assaults to specific individuals;
- *users profiling*, when the location of a user could be used to infer other sensitive information, such as state of health, personal habits, or professional duties, by correlating visited places or paths;
- *denial of service*, when the location of a user could motivate an access denial to services under some circumstances.

A further complicating factor is that location privacy can assume several meanings and introduce different requirements depending on the scenario in which the users are moving and on the services the users are interacting with. The following categories of location privacy can then be identified:

- *Identity privacy* protects the identities of the users associated with or inferable from location information. To this purpose, protection techniques aim at minimizing the disclosure of data that can let an attacker infers a user identity. Identity privacy is suitable in application contexts that do not require the identification of the users for providing a service.
- *Position privacy* protects the position information of individual users, by perturbing corresponding information and decreasing the accuracy of location information. Position privacy is suitable for environments where users' identities are required for a successful service provisioning. A technique that most solutions exploit, either explicitly or implicitly, consists in reducing the accuracy by scaling a location to a coarser granularity (e.g., from meters to hundreds of meters, from a city block to the whole town, and so on).
- *Path privacy* protects the privacy of information associated with individuals movements, such as the path followed while travelling or walking in a urban area. Several location-based services (e.g., personal navigation systems) could be exploited to subvert path privacy or to illicitly track users.

Since location privacy definition and requirements differ depending on the scenario, no single technique is able to address the requirements of all the location privacy categories. Therefore, in the past, research community focusing on providing solutions for the protection of location privacy of users has defined techniques that can be divided in three main classes: *anonymity-based*, *obfuscation-based*, and *policy-based* techniques. These classes of techniques are partially overlapped in scope and could be potentially suitable to cover requirements coming from one or more of the categories of location privacy. It is easy to see that anonymity-based and obfuscation-based techniques can be considered as dual categories. Anonymity-based techniques have been primarily defined to protect identity privacy and are not suitable for protecting position privacy, while obfuscation-based techniques are well suited for position protection and not appropriate for identity protection. Anonymity-based and obfuscation-based techniques could also be exploited for protecting path privacy. Policy-based techniques are in general suitable for all the location privacy categories, although they are often difficult to understand and manage for end users.

Among the class of techniques just introduced, current research on location privacy has mainly focused on supporting anonymity and partial identities. Beresford and Stajano

(Beresford & Stajano 2003, 2004) proposed a method, called *Mix zones*, which uses an anonymity service based on an infrastructure that delays and reorders messages from subscribers. Within a mix zone (i.e., an area where a user cannot be tracked), a user is anonymous in the sense that the identities of all users coexisting in the same zone are mixed and become indiscernible. Other works are based on the concept of *k*-anonymity. Bettini et al. (Bettini et al. 2005) designed a framework able to evaluate the risk of sensitive location-based information dissemination. Their proposal puts forward the idea that the geo-localized history of the requests submitted by a user can be considered as a quasi-identifier that can be used to discover sensitive information about the user. Gruteser and Grunwald (Gruteser & Grunwald 2003) developed a middleware architecture and an adaptive algorithm to adjust location information resolution, in spatial or temporal dimensions, to comply with users' anonymity requirements. To this purpose, the authors introduced the concepts of spatial cloaking. Spatial cloaking guarantees the *k*-anonymity by enlarging the area where a user is located to an area containing *k* indistinguishable users. Gedik and Liu (Gedik & Liu 2008) described another *k*-anonymity model aimed at protecting location privacy against various privacy threats. In their proposal, each user is able to define the minimum level of anonymity and the maximum acceptable temporal and spatial resolution for her location measurement. Mokbel et al. (Mokbel et al. 2006) designed a framework, named *Casper*, aimed at enhancing traditional location-based servers and query processors with anonymous services, which satisfies both *k*-anonimity and spatial user preferences in term of the smallest location area that tcan be released. Ghinita et al. (Ghinita et al. 2007) proposed *PRIVE'*, a decentralized architecture for preserving query anonymization, which is based on the definition of *k*-anonymous areas obtained exploiting the Hilbert space-filling curve. Finally, anonymity has been exploited to protect the path privacy of the users (Gruteser et al. 2004; Gruteser & Liu 2004; Ho & Gruteser 2005). Although interesting, these solutions are still at an early stage of development.

Differently, when the users' identity is required for location-based service provision, obfuscation-based techniques has been deployed. The first work providing an obfuscation-based technique for protecting location privacy is by Duckham and Kulik (Duckham & Kulik 2005). In particular, their framework provides a mechanism for balancing the individual needs for high-quality information services and for location privacy. The idea is to degrade location information quality by adding *n* fake positions to the real user position. Ardagna et al. (Ardagna et al. 2007) defined different obfuscation-based techniques aimed at preserving location privacy by artificially perturbing location information. These techniques degrade the location information accuracy by: *i)* enlarging the radius of the measured location, *ii)* reducing the radius, *iii)* shifting the centre. In addition, a metric called *relevance* is used to evaluate the level of location privacy and

balancing it with the accuracy needed for the provision of reliable location-based services.
Finally, policy-based techniques are based on the notion of privacy policies and are suitable for all the categories of location privacy. In particular, privacy policies define restrictions that must be enforced when location of users is used by or released to external parties. The IETF Geopriv working group (Geopriv 2006) addresses privacy and security issues related to the disclosure of location information over the Internet. The main goal is to define an environment supporting both location information and policy data.

NETWORK ANONYMITY

The wide diffusion of Internet for many daily activities has enormously increased the interest in security and privacy issues. In particular, in such a distributed environment, privacy should imply also anonymity: a person shopping online may not want her visits to be tracked, the sending of email should keep the identities of the sender and the recipient hidden from observers, etc. That is, when surfing the Web, users not only want to keep secret the information they exchange, but also the fact that they are exchanging information and with whom. Such a problem has to do with traffic analysis and it requires ad hoc solutions. Traffic analysis is the process of intercepting and examining messages in order to deduce information from patterns in communication. It can be performed even when the messages are encrypted and cannot be decrypted. In general, the greater the number of messages observed, or even intercepted and stored, the more can be inferred from the traffic. It cannot be solved just by encrypting the header of a packet, or the payload: in the first case, the packet could still be tracked as it moves through the network; the second case is ineffective as well since it would still be possible to identify who is talking to whom.
In the following of this section, we first describe the onion routing protocol (Goldschlag et al. 1996, 1999; Reed et al. 1999), one of the better-known approaches that is not application-oriented. Then, we provide an overview of other techniques for assuring anonymity and privacy over networks. The key approaches we discuss are MIX network (Chaum 1981; Berthold et al. 2000), Crowds system (Reiter & Rubin 1999) and Freedom network (Boucher 2000).

Onion Routing

Onion routing is intended to provide real-time bi-directional *anonymous connections* resistant both to eavesdropping and traffic analysis in a way transparent to applications.

That is, if Alice and Bob communicate over a public network by means of onion routing, they are guaranteed that the content of the message remains confidential and no external observer or internal node is able to infer they are communicating.

Onion routing works beneath the application layer, replacing socket connections with anonymous connections and without requiring any change to proxy-aware Internet services or applications. It was originally implemented on Sun Solaris 2.4 in 1997, including proxies for Web browsing (HTTP), remote logins (rlogin), email (SMTP) and file transfer (FTP). *Tor* (Dingledine et al. 2004), generation 2 onion routing implementation, runs on most common operating systems. It consists of a fixed infrastructure of onion routers, where each router has a longstanding socket connection to a set of neighboring ones. Only few routers, called *onion router proxies*, know the whole infrastructure topology. In onion routing, instead of making socket connections directly to responding machine, initiating applications make a socket connection to an onion routing proxy that builds an anonymous connection through several other onion routers to the destination. In this way, the onion routing network allows the connection between the initiator and responder to remain anonymous. Although the protocol is called onion routing, the routing that occurs during the anonymous connection is at the application layer of the protocol stack, not at the IP layer. However, the underlying IP network determines the route that data actually travels between individual onion routers.

Given the onion router infrastructure, the onion routing protocol works in three phases:

1. anonymous connection *setup*;
2. *communication* through the anonymous connection;
3. anonymous connection *destruction*.

During the first phase, the initiator application, instead of connecting directly with the destination machine, opens a socket connection with an onion routing proxy (which may reside in the same machine, in a remote machine, or in a firewall machine). The proxy first establishes a path to the destination in the onion router infrastructure, then sends an *onion* to the first router of the path. The onion is a layered data structure where each layer of the onion (public-key encrypted) is intended for a particular onion router and contains: *i)* the identity of the next onion router in the path to be followed by the anonymous connection; *ii)* the expiration time of the onion; *iii)* a key seed to be used to generate the keys to encode the data sent through the anonymous connection in both directions. The onion is sent through the path established by the proxy: an onion router that receives an onion peels off its layer, identifies the next hop, records on a table the key seed, the expiration time and the identifiers of incoming and outgoing connections and the keys that are to be applied, pads the onion and sends it to the next onion router. Since the most internal layer contains the name of the destination machine, the last router

of the path will act as destination proxy and open a socket connection with the destination machine. Note that only the intended onion router is able to peel off the layer intended to it. In this way, each intermediate onion router knows (and can communicate with) only the previous and the next hop router. Moreover, it is not capable to understand the content of the following layers of the onion. The router, and any external observer, cannot know a priori the length of the path since the onion size is kept constant by the fact that each intermediate router is obliged to add padding to the onion corresponding to the fixed size layer that it removed.
Figure 2 shows an onion for an anonymous connection following route *WXYZ*, where the router infrastructure is as depicted in Figure 3, with *W* the onion router proxy.

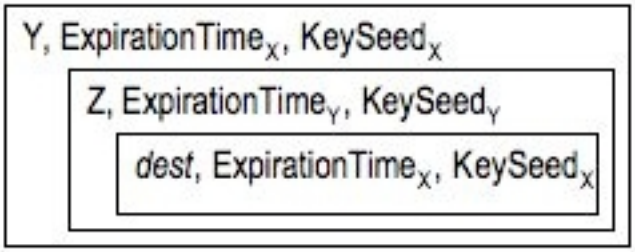

Figure 2: Onion Routing Network Infrastructure

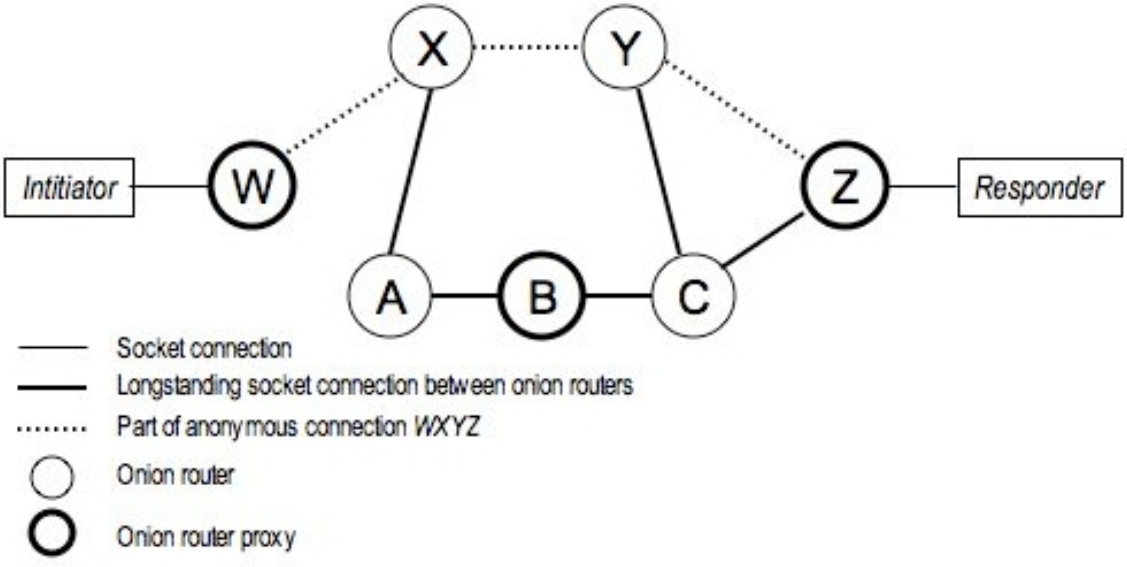

Figure 3: Onion Routing Network Infrastructure

Once the anonymous connection is established, data can be sent in both directions. The onion proxy receives data from the initiator application, breaks it into fixed size packets, and adds a layer of encryption for each onion router in the path using the keys specified in the onion. As data packets travel through the anonymous connection, each intermediate onion router removes one layer of encryption. The last router in the path sends the plaintext to the destination through the socket connection that was opened during the setup phase. This encryption layering occurs in the reverse order when data is sent backward from the destination machine to the initiator application. In this case, the initiator proxy, which knows both the keys and the path, will decrypt each layer and send the plaintext to the application using its socket connection with the application. As for the onion, data passed along the anonymous connection appears different to each intermediate router and external observer, so it cannot be tracked. Moreover, compromised onion routers cannot cooperate to correlate the data stream they see.
When the initiator application decides to close the socket connection with the proxy, the proxy sends a destroy message along the anonymous connection and each router removes the entry of the table relative to that connection.

There are several advantages in the onion routing protocol. First, the most trusted element of the onion routing infrastructure is the initiator proxy, which knows the network topology and decides the path used by the anonymous connection. If the proxy is moved in the initiator machine, the trusted part is under the full control of the initiator. Second, the total cryptographic overhead is the same as for link encryption but, whereas in link encryption one corrupted router is enough to disclose all the data, in onion routing routers cannot cooperate to correlate the little they know and disclose the information. Third, since an onion has an expiration time, replay attacks are not possible. Finally, if anonymity is also desired, then all identifying information must be additionally removed from the data stream before being sent over the anonymous connection. However, onion routing is not completely invulnerable to traffic analysis attacks: if a huge amount of messages between routers is recorded and usage patterns analyzed, it would be possible to make a close guess about the routing, i.e., also about the initiator and the responder. Moreover, the topology of the onion router infrastructure must be static and known a priori at least by one onion router proxy, which make the protocol little adaptive to node/router failures.

Tor (Dingledine et al. 2004), generation 2 onion routing, addresses some of the limitations highlighted above, providing a reasonable trade-off between anonymity, usability and efficiency. In particular, it provides perfect forward secrecy and it does not requires a proxy for each supported application protocol.

Anonymity Services

There are some other approaches that offer some possibilities for providing anonymity and privacy, but they are still vulnerable to some type of attacks. For instance, many of these approaches are designed for World Wide Web access only: being protocol-specific these approaches may require further development to be used with other applications or Internet services, depending on the communication protocols used in those systems.

David Chaum (Chaum 1981; Berthold et al. 2000) introduced *MIX-networks* in 1981, in order to enable unobservable communication between users of the Internet. Mixes are intermediate nodes that may reorder, delay and pad incoming messages in order to complicate traffic analysis. A MIX node stores a certain number of incoming messages that it receives, and sends them to the next mix node in a random order. Thus, messages are modified and reordered in such a way that it is nearly impossible to correlate an incoming message with an outgoing message. Messages are sent through a series of MIX nodes and encrypted with MIXes' keys. If participants exclusively use MIXes for sending messages to each other, their communication relations will be unobservable - even if the attacker records all network connections. Without additional information, also the receiver does not have any clue about the identity of the message's sender. As in onion routing, each MIX node knows only the previous and next node in a received message's route. Hence, unless the route only goes through a single node, compromising a MIX node does not enable an attacker to violate neither the sender nor the recipient privacy. MIX networks are not really efficient since a MIX needs to receive a large group of messages before forwarding them, thus delaying network traffic. However, onion routing has many analogies with this approach and an onion router can be seen as a real-time Chaum MIX.

Reiter and Rubin in (Reiter & Rubin 1999) proposed an alternative to MIXes, *Crowds*, a system to make only browsing anonymous, hiding from Web servers and other parties information about either the user or what information she retrieves. This is obtained by preventing a Web server from learning any information linked to the user, such as the IP address or domain name, the page that referred the user to its site or the user's computing platform. The approach is based on the idea of "blending into a crowd", i.e., hiding one's actions within the actions of many others. Before making any request, a user joins a crowd of other users. Then, when the user submits a request, it is forwarded to the final destination with probability p and to some other member of the crowd with probability $1-p$. When the request is eventually submitted the end server cannot identify its true initiator. Even crowd members cannot identify the initiator of the request, since the initiator is indistinguishable from a member of the crowd that simply passed on a request from another.

Freedom network (Boucher 2000) is an overlay network that runs on top of Internet, i.e., on top of the application layer. The network is composed of a set of nodes called anonymous Internet proxies, which run on top of the existing infrastructure. As for onion routing and MIX networks, the Freedom network is used to setup a communication channel between the initiator and the responder, but it uses different techniques to encrypt the messages sent along the channel.

CONCLUSION

In this chapter we discussed net privacy from different viewpoints, from historical to technological. The very nature of the concept of privacy requires such an enlarged perspective since it often appears indefinite, being constrained into the trade off between the undeniable need of protecting personal information and the evident utility, in many contexts, of the availability of the same information. The digital society and the global interconnected infrastructure eased the access and the spreading of personal information, therefore, developing technical means and defining norms and fair usage procedures for privacy protection is now more demanding than in the past.
Economic aspects have been introduced since they are likely to strongly influence the way privacy is actually managed and protected. In this area, research has provided useful insights about the incentive and disincentives towards a better privacy.
Then we presented some of the more advanced solutions that research has developed to date, either for anonymising stored data, hiding sensitive information in artificially inaccurate clusters, and introducing third parties and middlewares in charge of managing online transactions and services in a privacy-aware fashion. Location privacy is a topic that has gained importance in recent years with the advent of mobile devices and that is worth a specific consideration.
Furthermore, the important issue of anonymity over the net has been investigated. To let individuals surfing the Web, accessing online services, and interacting with remote parties in an anonymous way has been the goal of many efforts since years. Some important technologies and tools are available and are gaining popularity.
To conclude, whereas privacy over the net and in the digital society looks not in good shape, the augmented sensibility of individuals to its erosion, the many scientific and technological efforts to introduce novel solutions, and a better knowledge of the problem with the help of fresh data contribute to stimulate the need of a better protection and fairer usage of personal information. For this reason, it is likely that net privacy will remain an important topic in the years to come and more innovations towards a better management of privacy issues will emerge.

REFERENCES

Agrawal, R, Kiernan, J, Srikant, R & Xu, Y 2003, 'An XPath based preference language for P3P', *Proceedings of the 12th International World Wide Web Conference*, Budapest, Hungary, pp. 629-639.

Ardagna, CA, Cremonini, M, Damiani, E, De Capitani di Vimercati, S & Samarati, P 2007, 'Location privacy protection through obfuscation-based techniques', *Proceedings of the 21st Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, Redondo Beach, CA, pp. 47-60.

Ardagna, CA, Cremonini, M, Damiani, E, De Capitani di Vimercati, S & Samarati, P 2006, 'Supporting location-based conditions in access control policies', *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS'06)*, Taipei, pp. 212-222.

Ardagna, CA, Cremonini, M, De Capitani di Vimercati, S & Samarati, P 2008, 'A privacy-aware access control system', *Journal of Computer Security (JCS)*, vol.16, no. 4, pp. 369–392.

Ardagna, CA, Damiani, E, De Capitani di Vimercati, S & Samarati, P 2005, 'Towards Privacy-Enhanced Authorization Policies and Languages', *Proceedings of the 19th IFIP WG11.3 Working Conference on Data and Application Security*, Storrs, CT, pp. 16-27.

Ardagna, CA, De Capitani di Vimercati, S & Samarati, P 2006, 'Enhancing user privacy through data handling policies', *Proceedings of the 20th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, Sophia Antipolis, France, pp. 224-236.

Ashley, P, Hada, S, Karjoth, G, Powers, C & Schunter, M 2003, 'Enterprise privacy authorization language (epal 1.1)', <http://www.zurich.ibm.com/security/enterprise-privacy/epal>.

Ashley, P, Hada, S, Karjoth, G & Schunter, M 2002, 'E-P3P privacy policies and privacy authorization', *Proceedings of the ACM Workshop on Privacy in the Electronic Society (WPES 2002)*, Washington, DC, pp. 103-109.

Attrition.org 2008, 'Data Loss Archive and Database (DLDOS)', <http://attrition.org/dataloss/>.

Bayardo, RJ & Agrawal, R 2005, 'Data privacy through optimal k-anonymization', *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, pp. 217–228.

Beresford, AR & Stajano, F 2003, 'Location privacy in pervasive computing', *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46–55.

Beresford, AR & Stajano, F 2004, 'Mix zones: User privacy in location-aware services', *Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops (PERCOMW04)*, Orlando, FL, pp. 127-131.

Berthold, O, Federrath, H & Kopsell, S 2000, 'Web MIXes: A System for Anonymous and Unobservable Internet Access'. In H.Federrath, editor, *Anonymity 2000*, Volume 2009 of Lecture Notes in Computer Science, Springer-Verlag, pp. 115-129.

Bettini, C, Wang, XS & Jajodia, S 2005, 'Protecting privacy against location-based personal identification', *Proceedings of the 2nd VLDB Workshop on Secure Data Management (SDM'05)*, Trondheim, Norway, pp. 185-199.

Bettini, C, Jajodia, S, Wang, XS & Wijesekera, D 2002, 'Provisions and obligations in policy management and security applications', *Proceedings of 28th Conference Very Large Data Bases (VLDB'02)*, Hong Kong, pp. 502-513.

Bonatti, P & Samarati, P 2002, 'A unified framework for regulating access and information release on the web', *Journal of Computer Security*, vol. 10, no. 3, pp. 241–272.

Boucher, P, Shostack, A & Goldberg, I 2000, 'Freedom Systems 2.0 Architecture', <http://www.freedom.net/info/whitepapers/Freedom_System_2_Architecture.pdf>.

Brown, M & Muchira, R 2004, 'Investigating the Relationship between Internet Privacy Concerns and Online Purchase Behavior', *Journal Electron. Commerce Res*, vol. 5, no. 1, pp. 62-70.

California Security Breach Notification Law 2002, Bill Number: SB 1386, February, <http://info.sen.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.html>.

Camenisch, J, & Lysyanskaya, A 2001, 'An efficient system for non-transferable anonymous credentials with optional anonymity revocation', *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques (EUROCRYPT 2001)*, Innsbruck, Austria, pp. 93-118.

Camenisch, J, & Van Herreweghen, E 2002, 'Design and implementation of the idemix anonymous credential system', *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*, Washington, DC, pp. 21-30.

Chandramouli, R 2005, 'Privacy protection of enterprise information through inference analysis', *Proceedings of IEEE 6th International Workshop on Policies for Distributed Systems and Networks (POLICY 2005)*, Stockholm, Sweden, pp. 47-56.

Chaum 1981, 'Untraceable Electronic Mail, Return Address, and Digital Pseudonyms', *Communications of the ACM*, vol. 24, no.2, pp. 84-88.

Cranor, LF 2002, *Web Privacy with P3P*, O'Reilly & Associates.

Culnan, M & Armstrong, P 1999, 'Information privacy concerns, procedural fairness, and impersonal trust: an empirical evidence', Organization Science, Vol. 10 No.1, pp.104-15.

Deloitte & Touche LLP and Ponemon Institute LLC 2007, 'Enterprise@Risk: 2007 Privacy & Data Protection Survey', <http://www.deloitte.com/dtt/cda/doc/content/us_risk_s%26P_2007%20Privacy10Dec2007final.pdf>.

Dingledine R, Mathewson N & Syverson P 2004, 'Tor: The Second-Generation Onion Router', *Proceedings of the 13th USENIX Security Symposium*, San Diego, CA.

Dodge, A 2008, 'Educational Security Incidents (ESI) Year In Review – 2007', < http://www.adamdodge.com/esi/yir_2006 >.

Duckham, M & Kulik L 2005, 'A formal model of obfuscation and negotiation for location privacy', *Proceedings of the 3rd International Conference on Pervasive Computing (PERVASIVE 2005)*, Munich, Germany, pp. 152-170.

Duckham, M & Kulik L 2006, 'Location privacy and location-aware computing', *Dynamic & Mobile GIS: Investigating Change in Space and Time*, pp. 34–51. Taylor & Francis.

eXtensible Access Control Markup Language (XACML) Version 2.0, February 2005, <http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf>.

Etiolated.org 2008, Shedding light on who's doing what with your private information. <http://etiolated.org/>.

Federal Trade Commission (FTC) 2006, 'ChoicePoint Settles Data Security Breach Charges; to Pay $10 Million in Civil Penalties, $5 Million for Consumer Redress', <http://www.ftc.gov/opa/2006/01/choicepoint.shtm>.

Froomkin, AM 2000, 'The Death of Privacy?', 52 Stanford Law Review, pp. 1461-1469.

Gedik, B. & Liu, L 2008, 'Protecting location privacy with personalized k-anonymity: Architecture and algorithms', *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18.

Geographic Location/Privacy (geopriv), September 2006, <http://www.ietf.org/html.charters/geopriv-charter.html>.

Ghinita, G, Kalnis, P & Skiadopoulos, S 2007, 'Privè: Anonymous location-based queries in distributed mobile systems', *Proceedings of the International World Wide Web Conference (WWW 2007)*, Banff, Canada, pp. 371-380.

Goldschlag, D, Reed, M, & Syverson, P 1999, 'Hiding Routing Information', In R.Anderson, editor, *Information Hiding: First International Workshop*, Volume 1174 of Lecture Notes in Computer Science, Springer-Verlag, pp. 137-150.

Goldschlag, D, Reed, M, & Syverson, P 1999, 'Onion Routing for Anonymous and Private Internet Connections', *Communication of the ACM*, vol. 42, no.2, pp. 39-41.

Grossklags, J & Acquisti, A 2007, 'When 25 Cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information', *Proceedings of Workshop on the Economics of Information Security (WEIS)*, Pittsburgh, PA.

Gruteser, M, Bredin, J & Grunwald, D 2004, 'Path privacy in location-aware computing', *Proceedings of the Second International Conference on Mobile Systems, Application and Services (MobiSys2004)*, Boston, MA.

Gruteser, M & Grunwald, D 2003, 'Anonymous usage of location-based services through spatial and temporal cloaking', *Proceedings of the 1st International Conference on Mobile Systems, Applications, and Services (MobiSys)*, San Francisco, CA, pp. 31-42.

Gruteser, M & Liu, X 2004, 'Protecting privacy in continuous location-tracking applications', *IEEE Security & Privacy Magazine*, vol. 2, no. 2, pp. 28–34.

Hann, Il-H, Hui, KL, Lee, TS & Png IPL 2007, 'Analyzing Online Information Privacy Concerns: An Information Processing Theory Approach', *Journal of Management Information Systems*, vol. 24 no. 2, pp. 13-42.

Hann, Il-H, Hui, KL, Lee, TS & Png IPL 2002, 'Online Information Privacy: Measuring the Cost-Benefit Trade-off', *Proceedings, 23rd International Conference on Information Systems*, Barcelona, Spain.

Hasan, R & Yurcik, W 2006, 'Beyond Media Hype: Empirical Analysis of Disclosed Privacy Breaches 2005-2006 and a DataSet/Database Foundation for Future Work', *Proceedings of Workshop on the Economics of Securing the Information Infrastructure*, Washington, DC.

Hirshleifer, J 1971, 'The Private and Social Value of Information and the Reward to Inventive Activity', *American Economic Review*, vol. 61, pp. 561–574.

Ho, B & Gruteser, M 2005, 'Protecting location privacy through path confusion', *Proceedings of IEEE/CreateNet International Conference on Security and Privacy for Emerging Areas in Communication Networks (SecureComm)*, Athens, Greece, pp. 194-205.

Hui, KL & Png IPL 2006, 'Economics of Privacy', in Terrence Hendershott, Ed., *Handbooks in Information Systems, vol. 1*, Amsterdam, Elsevier, pp. 471-497.

Hui, KL, Tan, BCY & Goh, CY 2006, 'Online information disclosure: Motivators and measurements', *ACM Transaction on Internet Technologies*, vol. 6 no. 4, pp. 415-441.

Karjoth, G & Schunter, M 2002, 'Privacy policy model for enterprises', *Proceedings of the 15th IEEE Computer Security Foundations Workshop*, Cape Breton, Nova Scotia, Canada, pp. 271-281.

LeFevre, K, DeWitt, DJ & Ramakrishnan, R 2005, 'Incognito: Efficient full-domain k-anonymity', *Proceedings of the 24th ACM SIGMOD International Conference on Management of Data*, Baltimore, MD, pp. 49–60.

LeFevre, K, DeWitt, DJ & Ramakrishnan, R 2006, 'Mondrian multidimensional k-anonymity', *Proceedings of the 22nd International Conference on Data Engineering (ICDE'06)*, Atlanta, GA.

Lessig, L. 2003, Free Culture, Penguin Group (USA), <http://www.free-culture.cc/>.

Li, N, Li, T, & Venkatasubramanian, S 2007, 't-closeness: Privacy beyond k-anonymity and l-diversity', *Proceedings of the 23nd International Conference on Data Engineering*, Istanbul, Turkey, pp. 106-115.

Machanavajjhala, A, Gehrke, J, Kifer, D & Venkitasubramaniam, M 2006, '*l*-diversity: Privacy beyond k-anonymity', *Proceedings of the International Conference on Data Engineering (ICDE'06)*, Atlanta, GA.

Mokbel, MF, Chow, C-Y & Aref, WG 2006, 'The new Casper: Query processing for location services without compromising privacy', *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006)*, Seoul, South Korea, pp. 763-774.

Odlyzko, AM 2003, 'Privacy, economics, and price discrimination on the Internet', *Proceedings of the Fifth International Conference on Electronic Commerce (ICEC2003)*, N. Sadeh, ed., ACM, pp. 355-366.

Odlyzko, AM 2007, 'Privacy and the clandestine evolution of ecommerce', *Proceedings of the Ninth International Conference on Electronic Commerce (ICEC2007)*, ACM.

Odlyzko, AM 2004, 'The evolution of price discrimination in transportation and its implications for the Internet', *Review of Network Economics*, vol. 3, no. 3, pp. 323-346.

Odlyzko, AM 2003, 'The unsolvable privacy problem and its implications for security technologies', *Proceedings of the 8th Australasian Conference on Information Security and Privacy (ACISP 2003)*, R. Safavi-Naini and J. Seberry, eds., Lecture Notes in Computer Science 2727, Springer, pp. 51-54.

Peslak, AR 2005, 'Privacy policies of the largest privately held companies: a review and analysis of the Forbes private 50', *Proceedings of the ACM SIGMIS CPR Conference on Computer Personnel Research*, Atlanta, GA.

Pfitzmann, A & Köhntopp M 2001, 'Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology', *Book chapter of Designing Privacy Enhancing Technologies*, Springer Berlin, pp. 1-9.

Pfitzmann, A & Waidner, M 1986, 'Networks without user observability – design options', *Proceedings of workshop on the theory and application of cryptographic techniques on Advances in cryptology (EuroCrypt'85)*, vol. 219 LNCS Springer, Linz, Austria, pp. 245-253.

Posner, RA 1981, 'The economics of privacy', *American Economic Review*, vol. 71, no. 2, pp. 405- 409.

Privacy and Identity Management for Europe (PRIME) 2004-2008, <http://www.prime-project.eu.org/>.

Privacy Rights ClearingHouse 2007, 'Chronology of Data Breaches 2006: Analysis'. <http://www.privacyrights.org/ar/DataBreaches2006-Analysis.htm>.

Reed, M, Syverson, P & Goldschlag, D 1998, 'Anonymous Connections and Onion Routing', *IEEE Journal on Selected Areas in Communications*, vol.16, no.4, pp. 482-494.

Reiter, M & Rubin, A 1999, 'Anonymous Web Transactions with Crowds', *Communications of the ACM*, vol. 42, no. 2, pp. 32-48.

Samarati, P 2001, 'Protecting respondents' identities in microdata release', *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027.

Scalet, SD 2005, 'The Five Most Shocking Things About the ChoicePoint Data Security Breach', CSO online, <http://www.csoonline.com/article/220340>.

Schwaig, KS, Kane, GC & Storey, VC 2006, 'Compliance to the fair information practices: how are the fortune 500 handling online privacy disclosures?', *Inf. Manage.*, vol. 43, no. 7, pp. 805-820.

Shostack, A & Syverson, P 2004, 'What Price Privacy? (and why identity theft is about neither identity nor theft)', *In Economics of Information Security, Chapter 11*, Kluwer Academic Publishers.

Sprenger, P 1999, 'Sun on Privacy: 'Get Over It'', *WIRED*, <http://www.wired.com/politics/law/news/1999/01/17538>.

Steiner, P 1993, On the Internet, nobody knows you're a dog, Cartoonbank, *The New Yorker*, <http://www.cartoonbank.com/item/22230>.

Syverson, P 2003, 'The Paradoxical Value of Privacy', *Proceedings of the 2nd Annual Workshop on Economics and Information Security (WEIS 2003*), College Park, MD.

Thuraisingham, B 2005, 'Privacy constraint processing in a privacy-enhanced database management system', *Data & Knowledge Engineering*, vol. 55, no. 2, pp. 159–188.

Tsai J, Egelman S, Cranor L & Acquisti A 2007, 'The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study'. *Workshop on Economics and Information Security (WEIS 2007),* Pittsburgh, PA.

Youssef, M, Atluri, V & Adam NR 2005, 'Preserving mobile customer privacy: An access control system for moving objects and customer profiles', *Proceedings of the 6th International Conference on Mobile Data Management (MDM 2005)*, Ayia Napa, Cyprus, pp. 67-76.

United Nations 1948, Universal Declaration of Human Rights, <http://www.un.org/Overview/rights.html>.

Warren, SD & Brandais L D 1890, 'The Right to Privacy', Harvard Law Review, Vol. IV, No. 5.

World Wide Web Consortium (W3C) 2002, 'A P3P Preference Exchange Language 1.0 (APPEL1.0)', <http://www.w3.org/TR/P3P-preferences/>.

World Wide Web Consortium (W3C) 2002, 'Platform for privacy preferences (P3P) project', <http://www.w3.org/TR/P3P/>.