

A Visibility Classification Scheme for Privacy Management Requirements

Olli P. Jarvinen, Ph.Lic.

Department of Computer Science, University of Turku
Turku, Finland
olli.jarvinen@tukkk.fi

Julia B. Earp, Ph.D.

Department of Business Management, North Carolina State University
Raleigh, NC 27695-7229
919.513.1707 (Voice)
julia_earp@ncsu.edu

Annie I. Antón, Ph.D.

Department of Computer Science, North Carolina State University
Raleigh, NC 27695-7534
919.515.5764
aiananton@eos.ncsu.edu

2nd Symposium on Requirements Engineering for Information Security

ABSTRACT

Organizational privacy policies and privacy practices reflect an organization's perceived trustworthiness to those with whom it conducts business. This paper proposes a classification scheme, based upon an in-depth two-year analysis of Internet privacy policies, for examining an organization's privacy management practices within the context of its respective privacy policy. The classification scheme aids in evaluating privacy from the viewpoint of the consumer who wants their privacy protected and does not want to be misled by hidden tactics that can undermine consumer privacy. This is described in terms of protective and visible methods of managing consumer data. In this paper, we discuss a case study in which the classification scheme was employed to analyze 23 Internet health care website privacy policies.

1 INTRODUCTION

The ability for organizations to reach consumers 24 hours a day worldwide is creating new opportunities to gather large amounts of consumer information. Whenever an Internet consumer visits a website, a large amount of consumer information may easily become available to the website. The majority of data exchange between a consumer and a website is visible to the user but there are many methods in which the website can gather information without the consumer being aware. The vulnerability of an Internet consumer is significant because the Internet links companies to tens of millions of consumers around the world, offering a straightforward means to interact with other businesses and individuals at a very low cost.

We focused our study on information rich business segments in the health care industry where consumer vulnerability is exceptionally high due to the sensitive nature of information collected at these websites. This, in return, presents many challenging opportunities for healthcare related businesses. Several analyses (described in Section 3) have guided the development of the classification scheme presented in this paper. The classification scheme incorporates several perspectives that impact privacy policy content as well as privacy management practices and operational system requirements. The proposed classification scheme is effective for examining how privacy policy statements and their respective system requirements may be made apparent to a consumer without the consumer first having to read the privacy policy statement. The contribution of the classification scheme is primarily intended for software engineers, policy makers and consumer advocates.

Visible privacy practices are performed in such a way that an average Internet user is aware of data collection while accessing websites with a browser using default security and privacy settings. *Invisible privacy practices* are performed in a hidden manner that requires users to take a proactive role in learning about website privacy practices (e.g. reading the privacy policy, setting the browser's security and privacy settings, learning about cookies, etc.) The characteristics of these two perspectives are defined in Table 1. Visible and invisible privacy practices are essential trust factors for organizations that participate in online business due to the capability to easily collect data in both visible and invisible ways. Subsequently, consumers provide personal information in either a conscious or subconscious manner. Any organization embarking upon online transactions should therefore be prepared to address privacy matters in advance, clearly, and openly. A description of privacy management practices should be available to users without requiring extensive searching and reading processes. Privacy management must be evaluated from several perspectives within an organization; these perspectives primarily include legal constraints, technical measures, business rules, social norms and political norms [EAJ02]. Information technology (IT) practitioners need to be aware of the importance of different perspectives but also realize that the interplay between visible and invisible methods in protection and vulnerability settings plays a critical role in privacy management and privacy policy. The classification scheme presented in this paper addresses the visible/invisible and protection/vulnerability perspective and aims to provide a conceptual classification scheme for responsible and efficacious privacy management while also providing some basic elements and viewpoints for Internet application design.

Table 1: Characteristics of Visible vs. Invisible Privacy Management Practices

Visible	Invisible
Information voluntarily given, shared and used	Information collected, used and shared without consent
Conscious process, easy to conclude	Subconscious process, difficult for consumers to conclude
Open, choice, consent	Closed, hidden, without consumer's knowing consent
Forms, E-mails, Surveys	Cookies, Log-files, Server Files, Proxy

2 PRIVACY POLICY

A privacy policy comprehensively describes a website's information practices and is located on the site in an easily accessible location [FTC98, FTC00]. It describes the kinds of information collected by the website and the way that information is handled, stored, and used. Every organization involved in e-commerce transactions has a responsibility to adopt and implement a policy for protecting the privacy of individually identifiable information. Organizations must also consider other organizations with which they interact and take steps that foster the adoption and implementation of effective online privacy policies by those organizations as well.

Internet privacy policies are critical due to the increase in information collection from several sources and possibilities to gather and merge information in many ways. A privacy policy should directly reflect an organization's privacy rules and practices no matter what methods are employed to gather and subsequently use the data. In this paper we propose a classification scheme to aid in the design of websites to focus on features of visibility and protection, but also visibility and vulnerability. These two views reflect practices, whether protective or vulnerable, that are immediately visible to the consumer. Privacy policies and privacy practices reflect ethical views of an organization and therefore, provide an indication of perceived trustworthiness to those who conduct business with a given organization. In term, system requirements must be in compliance with these policies which are operationalized in the respective system implementations. This paper seeks to increase the IT community's understanding of privacy policy as a significant visible trust indicator for fair business practices.

3 CLASSIFICATION SCHEME FOR PRIVACY POLICIES AND PRIVACY MANAGEMENT

In this section, we introduce our privacy classification scheme that expresses four perspectives that must be considered when evaluating an organization's privacy policy, specifying system requirements and designing Internet software. These perspectives reflect the visible and hidden natures of privacy management practices, which we studied within the context of privacy protection and vulnerability. Before introducing the classification scheme, we discuss the evolution of this research to establish the context and introduce the vocabulary for the discussion in Section 3.2.

3.1 Classification Scheme Development

The classification scheme was developed in conjunction with an in-depth two-year analysis of Internet privacy policies [AE01, AER02]. It is the culmination of several phases of privacy policy and legal analyses based upon our use of the Goal-Based Requirements Analysis Method (GBRAM) [Ant96, AP98]. *Goals* are the objectives and targets of achievement for a system. In software engineering, goal-

driven approaches for requirements focus on why systems are constructed, expressing the rationale and justification for the proposed system [Lam01]. Goals are a cogent unit by which to objectively analyze and compare Internet privacy policies, enabling us to provide useful guidance to IT practitioners, policy makers, and consumers [AE01, AER02]. A discussion of goal-driven analysis is outside the scope of this paper; interested readers are referred to [Ant96, AP98, Lam01].

The classification scheme presented in this paper resulted from our two-year analysis of nearly 50 Internet privacy policies [AE01] coupled with our most recent goal-driven analyses of 16 health care website privacy policies. We have used the privacy goal taxonomy [AER02] to classify privacy goals as either privacy protection goals or privacy goal obstacles that suggest the potential and vulnerability for privacy invasions. Privacy protection goals relate to the *desired protection* of consumer privacy rights, whereas privacy goal obstacles relate to *existing threats* to consumer privacy. Privacy protection goals correspond to the five Fair Information Practice Principles [FIP73]: 1) notice / awareness, 2) choice / consent, 3) access / participation, 4) integrity / security, and 5) enforcement / redress. In contrast, privacy goal obstacles represent statements of fact that suggest the existence of vulnerabilities for privacy invasions. Privacy goal obstacles can be divided into seven classes [AER02]: 1) monitoring, 2) aggregation, 3) storage, 4) transfer, 5) collection, 6) personalization, and 7) contact.

It was apparent to our analysis team that a richer framework was needed to adequately consider privacy within a broader organizational context. Therefore, in [EAJ02] we introduced a set of five socio-technical perspectives that comprised the organizational privacy framework: legal, technical, business, contractual and social. Legal and technical perspectives offer constraints to the privacy matters of a given system. The business entities involved in online business (users, organization and third parties) have goals that are motivated by social and contractual norms that restrict the organization [EAJ02].


This paper extends that work in an effort to obtain a richer evaluation of privacy policy. We now introduce a 4-field matrix (see Table 2) that is also based upon the privacy goal taxonomy. The four fields are used to classify privacy policy goals as: visible/vulnerable, visible/protection, invisible/vulnerable or invisible/protection. It is important to distinguish between visible and invisible because they both influence Internet consumers and website companies, but each may be differently interpreted. For example visible goals may be used as trust indicators of the website to the outside world and invisible goals may unknowingly introduce additional vulnerability to the consumers. It should thus be an important focus for specifying and designing websites because consumers will value invisible practices being made visible to them. This paper advocates transforming invisible practices into visible ones as a design rule to ensure consumers are better informed and more readily able to manage their privacy.

The proposed privacy management classification scheme in Table 2 is summarized by four fields that address trust/openness, trust/protection, threat, and security. These form the focal points for this discussion. *Trust/openness* characterizes companies that clearly and openly express how they are going to process consumer information. Consumer information collected via the Web is vulnerable due to the high level of interconnectivity between website organizations and therefore, the privacy practices of such organizations should be clearly and openly stated without the consumer needing to take additional measures. *Trust/protection* refers to organizations with websites that clearly express how they are going to protect consumer privacy. Both trust/openness and trust/protection are important to enable users to make informed decisions regarding the use of their personal information. *Threat* refers to those companies that obscure (whether intentionally or unintentionally) practices that introduce vulnerabilities associated with the collection, transmission or storage of personal information. For example, a website that requires consumers to read a website's entire privacy policy each time he/she visits the site is a candidate for considering possible 'visibility' requirements. Such visibility requirements would ensure that users receive, for example, visible cues embedded in their browser that reflect the site's privacy practices. IT practitioners and security officers need to focus on technical measures necessary to provide a secure IT environment that effectively protects consumer privacy while informing consumers to ensure

sound decision making. Assessing existing policies and requirements for their position within the classification scheme aids requirements engineers as they seek ways in which to better inform website users about privacy practices and ways in which to minimize existing and potential information vulnerabilities.

Table 2: Privacy Management Classification Scheme

4-field matrix		Vulnerable		Protection	
		Trust/openness		Trust/protection	
Visible		Threat		Security	
Invisible					

Aim: 

Software engineers can benefit from the classification scheme as it will guide them during the development process of their online systems. The arrows in Table 2 indicate the target direction of website development. In other words, the most desirable kind of website is one that emphasizes consumer trust/protection by implementing visible/protection goals. The primary challenges are how to convert *threat* to *security* and *openness* to *protection*. It is not possible to convert all vulnerable goals to protection goals; therefore, at a minimum it is important to convert *threat* to *openness* and *security* to *protection*.

3.2 Case Study Classification Heuristics

Once the classification scheme introduced in Section 3.1 was developed, it was then employed in our analysis of 23 consumer-oriented healthcare website privacy policies (6 pharmaceutical companies, 7 health insurance companies, and 10 online pharmacies). The four perspectives (visible/protection, visible/vulnerable, invisible/protection, visible/protection) were used to classify 134 goals that were extracted from the 23 privacy policies. Although this case study is focused on healthcare websites, the classification scheme can be generalized to other websites since it was based on the original analysis of more than 50 Internet privacy policies [AE01]. We now briefly discuss the heuristics used to classify the privacy policy goals extracted from these policies according to our privacy management classification scheme.

During the classification process, we assumed the user's web browser maintains the default privacy/security settings. Classifying healthcare privacy policy goals involves differentiating goals according to the four perspectives of the classification scheme. Visible/protection goals are classified by analyzing each goal and asking, "*Does this goal protect user's privacy and is it visible to the user without reading the privacy policy statement?*" Consider the goal G_1 : OPT to receive emails from our company; this goal clearly protects user's privacy because the user can decide whether or not to receive emails from the website enterprise and it is visible to the user without reading the privacy policy statement. This goal is classified as a visible/protection goal. Visible/vulnerable goals are classified by asking: "*Does this goal threaten user's privacy and is it apparent to the user without reading the privacy policy statement?*" Consider the goal G_2 : USE member profile. This goal is classified as a

vulnerability goal because the user must give some personally identifiable information before he/she can continue further at the website. Gathering information is an open process and the user consciously provides information; therefore, the goal is classified as a visible goal. Invisible/protection goals are classified by asking: “Does this goal not protect user privacy and is it not apparent to the user without reading the privacy policy statement?” Consider the goal G₃: PREVENT disclosing personal identifiable information (PII) of children under 13. This goal is classified as an invisible/protection goal because it clearly protects user privacy but a user will not know this unless he or she proactively reads the privacy policy. Invisible/vulnerable goals are classified by asking: “Does this goal threaten user privacy and is it not apparent without reading the privacy policy statement?” Consider the goal G₄: SELL aggregate information. This goal clearly threatens user privacy and it is impossible for the user to be aware of this practice without reading the privacy policy statement first. Therefore, it is classified as an invisible/vulnerable goal.

Our analysis of 23 privacy policies yielded 134 goals, each of which was easily classified according to each of the four perspectives. Additionally, there were 405 total occurrences of the 134 goals. For example, the visible/protection goal G₅: NOTIFY consumer of change to privacy policy had 5 occurrences of that particular goal within our 23 privacy policies. It appeared in one of the seven health insurance website privacy policies, four of the ten drugstore website privacy policies and none of the six pharmaceutical company website privacy policies. Similarly, the invisible/vulnerable goal G₄: SELL aggregate information appeared in only one of the privacy policies; therefore, it had only one occurrence. We encountered no goals that overlapped classes. Table 3 provides an overview of the classification scheme analysis and shows the number of occurrences for goals that map to these four perspectives.

3.3 Classification Scheme Perspectives

This section discusses each of the classification scheme perspectives within the context of our analysis of Internet healthcare privacy policies.

3.3.1 Visible/Protection Perspective

The visible perspective implies that a user gives knowing consent to an Internet company to do something with information concerning him/her. It is considered to be a conscious process when the user knows they are voluntarily disclosing information and is able to prevent the disclosure if so desired. This kind of process typically occurs when a user fills out a form, sends e-mails or responds to a survey. All of these circumstances require the user to actively and consciously decide whether or not to provide the requested information. We have classified the goals that are easily observed and evaluated by the user as *visible goals*. We observed 179 occurrences of visible goals, 44% of the total goal occurrences, that included 40 (less than 30%) of the 134 distinct goals.

This perspective should define the target of a website company regarding its practices of information management because it represents an organization with open data practices that aim to protect the consumer. It should also be an important target to IT practitioners when designing and implementing Internet web solutions because it essentially defines the website’s trust factors. Considering the five protection goal classes described in section 3.1, all but the class enforcement/redress were found in this perspective in our study. This perspective got the most occurrences (126), which is 31% of the total. Each goal of this perspective saw an average of 3.7 occurrences.

3.3.2 Visible/Vulnerable Perspective

This perspective reflects to the user a direct openness of the website organization and can therefore increase user's trust. Organizations need a lot of information about their consumers to operate efficiently. If the organization shows how it is gathering, using and sharing information about a consumer, it will give necessary knowledge to the user to foster an informed opinion regarding the organization's privacy policy. The responsibility of understanding that the website receives user information is then under the user's consideration. This perspective is therefore important to notice and should be a target of an organization's website when it is not possible to convert goals to the visible/protection perspective.

According to our study, we found that four vulnerability classes fell under this perspective: transfer, collection, personalization and contact. This perspective got 53 occurrences, which is 13% of the total. This was clearly the minor perspective. The argument is more supported, when we notice that the occurrences were divided into 6 goals, which is only 4% of the total. Each of these goals got 8.8 occurrences on average.

3.3.3 Invisible/Protection Perspective

Typical to the invisible perspective is that the actions of an Internet company are impossible or difficult to observe and evaluate by a user without reading the privacy policy in advance. For example, situations where consumer information is not voluntarily provided by the consumer yet the information is collected and possibly used by the website organization. Typically this occurs when a company uses cookies to gather information about a user, tracks users' usage patterns or discloses user information without the user's consent to business partners. 70% of the goals were classified as invisible goals and they saw a total of 226 occurrences (56% of the total). When we consider that invisible goals are more common than visible goals, this verifies that consumers are vulnerable to the privacy practices of the organization.

This perspective reflects that an organization views security and privacy requirements as trivial. Typical to this perspective is that the goal supports technical security, which does not reflect directly to the user without reading the privacy policy first. Because this perspective is a protective one, it is also a promising target of website organizations. The main weakness of this perspective is a user does not get assured of security without reading the privacy policy. Three of the protection goal classes were found to belong to this perspective: enforcement/redress, access/participation and integrity/security. This perspective got 103 of the occurrences, which is over 25% of the total. Over one third of the goals were classified into this class, and every goal got 2.3 occurrences on average.

3.3.4 Invisible/Vulnerable Perspective

This perspective denotes a legitimate threat to user privacy. The threat is real because not all consumers can (or are willing to) take the time to read and understand written privacy policy statements if available. Similarly, not all Internet users are informed about privacy practices, personal data management and browser configuration. We illuminate this perspective with an example where a user begins a dialog with a website by searching for information about a particular problem or interest. Already, the website is able to collect data from the user without any visible notice to the user. This kind of gathering process is very normal for this perspective according to our analysis. The user does not know what, why and/or when information is gathered and how it is used thereof. It is important to observe that a user has not voluntarily disclosed any data for the purpose of storing it in a database of the website. The entire storing process is invisible to the user and it is, therefore, a subconscious process to the user.

According to our study, 12 of 23 websites collect technical oriented data about users. For example, goal G₆: COLLECT domain name had 7 occurrences in the 23 analyzed privacy policies. We noticed

that 2 of these 7 Web-sites shared that kind of information to 3rd parties without asking for user consent. So this kind of collecting and sharing process can happen without the user's knowledge if he/she does not carefully study the privacy policy in advance.

We found that all vulnerable goal classes belonged to this perspective. This perspective got the second most occurrences, 123, which is over 30% of the total. Occurrences were divided among 49 goals. Every goal got 2.5 occurrences in average. It is important to notice that this perspective got the most part of the goals. Almost 37% of the goals were belonged to this perspective, which makes our study focus more important. Because the most vulnerable perspective (invisible/vulnerable) is formed with so many goals, it will be addressed in the next phase of our study more extensively.

Table 3: Occurrences of Goal Perspectives in Internet Health Care Privacy Policy Analysis

Classification Scheme Perspective	Goal Class	Health Insurance	Online Pharmacies	Pharmaceutical Companies	TOTAL Occurrences
VISIBLE total:179 (44%)	Vulnerability	13	26	14	53
	Protection	18	76	32	126
INVISIBLE total: 226 (56%)	Vulnerability	23	77	23	123
	Protection	26	62	15	103
SUBTOTAL	Vulnerability	36	103	37	166
	Protection	44	138	47	229
TOTAL					405

4 SUMMARY AND FUTURE WORK

The classification scheme presented in this paper is based upon four perspectives of privacy goals: invisible/protection, invisible/vulnerability, visible/protection and visible/vulnerability. Each of these perspectives should be considered when specifying privacy requirements, which must comply with existing privacy (and perhaps even security) policies, to ensure that consumer privacy values are adequately respected and reflected in the corresponding system implementations. These perspectives, when merged with the five organizational perspectives (legal, technical, business, contractual, and social), offer a foundation for reasoning about privacy policy in various situations. For example, the process of writing an effective privacy policy can be guided by ensuring that all perspectives of both frameworks have been considered. Moreover, the process of specifying privacy-aware requirements also benefits from the availability of this classification scheme to ensure that multiple viewpoints (especially that of privacy-concerned consumers) are adequately considered. As mentioned in section 3.1, this paper encourages software engineers to transform invisible practices into visible ones; however, the details of systematic transformations have been left for future work as this paper was intended to highlight the visible/invisible concept of website characteristics through the classification scheme.

Use of the classification scheme can be effective and informative for several groups of people. From the development viewpoint, our objective is to encourage software engineers to consider ways in which to minimize the occurrence of invisible and vulnerability goals in the systems they design and implement. We have discussed one mechanism by which this may be done: by seeking to convert invisible goals to visible ones and ideally so that they are also protective. Additionally, the classification scheme provides a

means to guide communication between software engineers and policy makers as they create the privacy policy for the organization's website. Using the classification scheme as guidance, policy makers can provide visibility requirements to the developers. Together, these two groups can create a website that conveys trust to the consumer.

From the consumer viewpoint, the classification scheme provides a context for evaluating the level of trust communicated by the website to the consumer. There are several other factors that convey trust (e.g. brand name), but they are beyond the scope of this paper. However, as we begin to consider trust as viewed by the consumer, we plan to incorporate additional trust factors in our future work pertaining to the classification scheme. It is important to note that The World Wide Web Consortium has recently established the Platform for Privacy Preferences Project (P3P) as an industry standard. It will provide an automated way for users to evaluate the content of website privacy policies. P3P requires consumers to answer a set of standardized multiple-choice questions that address various aspects of Web site privacy policies. The websites implementing P3P possess a privacy policy in machine readable format and users of these sites may configure their browsers to automatically determine if a Web site's privacy policy reflects their personal needs for privacy. This is done by comparing the user's responses to the multiple choice questions with the statements in a P3P compliant policy [P3P02]. If a consumer employs the features of a P3P compliant browser, then the issue of invisible goals can be lessened for P3P compliant websites.

Health care information privacy is an important global issue due to the IT-enabled ease of international travel, communication and exchange [Klu96]. We believe that the foreseeable future will bring evolutionary improvements in technology. As IT-services (e.g. mobile and wireless technologies) are offered to the public globally, it becomes more important to evaluate the privacy policy of such IT-services and reason amongst the different cultures, laws, directives, vocabularies and other frames of reference. We have restricted our study in this paper to U.S. based websites and plan to continue the study with European based websites as part of our future work to enable us to compare privacy practices between the U.S. and Europe.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under ITR Grant #0113792 and The Fulbright Center. The authors wish to recognize Turun Kauppaseura Saatio, Emil Aaltosen Saatio, Ella and Georg Ehrnroothin Saatio, Liikesivistysrahasto, and Yrjo Jahanssonin Saatio. Additionally, the authors wish to thank Carlos Jensen, Colin Potts and William Stufflebeam for discussions leading to the development of this classification scheme.

REFERENCES

- [AE01] A.I. Antón and J.B. Earp. *A Taxonomy for Web Site Privacy Requirements*, NCSU Technical Report TR-2001-14, 18 December 2001.
- [AER02] A.I. Antón, J.B. Earp, and A. Reese. *Analyzing Web Site Privacy Requirements Using a Privacy Goal Taxonomy*, Submitted to *10th Anniversary IEEE Joint Requirements Engineering Conference (RE'02)*, February 2, 2002.
- [Ant96] A.I. Antón. *Goal-Based Requirements Analysis*, *2nd IEEE Int'l Conf. on Requirements Engineering (ICRE '96)*, Colorado, pp. 136-144, 15-18 April 1996.
- [AP98] A.I. Antón and C. Potts. *The Use of Goals to Surface Requirements for Evolving Systems*, *Int'l Conf. on Software Engineering (ICSE '98)*, Kyoto, Japan, pp. 157-166, 19-25 April 1998.
- [EAJ02] J.B. Earp, A.I. Antón, and O. Jarvinen. *A Social, Technical and Legal Framework for Privacy Management*

and Policies. Accepted to AMCIS 2002.

- [FIP73] The Code of Fair Information Practices, U.S. Dep't. of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers, and the Rights of Citizens, viii, http://www.epic.org/privacy/consumer/code_fair_info.html, 1973.
- [FTC98] *Privacy Online: A Report to Congress*, <http://www.ftc.gov/reports/privacy3/>, Federal Trade Commission, June 1998.
- [FTC00] Privacy Online: Fair Information Practices in the Electronic Marketplace. A Report to Congress. Federal Trade Commission, 2000.
- [FTC02] Eli Lilly Settles FTC Charges Concerning Security Breach, FTC Press Release, <http://www.ftc.gov/opa/2002/01/elililly.htm>, 18 Jan. 2002.
- [Klu96] Kluge, E.H.W. "Professional Ethics as Basic for Legal Control of Health Care Information", *International Journal of Bio-Medical Computing*, 43, pp. 33-37, 1996.
- [Lam01] A. van Lamsweerde. Goal-Oriented Requirements Engineering: A Guided Tour, *IEEE 5th Int'l Symp. on Requirements Engineering (RE'01)*, Toronto, Canada, pp. 249-261, 27-31 August 2001.
- [P3P02] P3P Public Overview. Accessed June 24, 2002 at <http://www.w3.org/P3P/>.