Privacy Policies Examined: Fair Warning or Fair Game?

Carlos Jensen
College of Computing, GVU Center
The Georgia Institute of Technology
Atlanta, GA 30313 USA
carlosj@cc.gatech.edu

Colin Potts
College of Computing, GVU Center
The Georgia Institute of Technology
Atlanta, GA 30313 USA
potts@cc.gatech.edu

ABSTRACT

Posting privacy policies has become a popular practice with businesses as they seek to shield themselves from potential liability or regulation, as well as inform users about their privacy and rights. These policies are in many ways modeled after software license statements, and are often more legalistic than user friendly. This paper examines the current practice of privacy policies as fair warning hold up from a usability perspective, and what steps can be taken to ensure that the average user can protect their privacy online.

Keywords

Privacy Policy, Readability, Informed Consent, Fair Information Practices, Digital Divide, Vulnerable Populations, Internet Demographics

INTRODUCTION

Posting privacy policies has become a popular practice with businesses as they seek to shield themselves from potential liability or regulation, as well as inform users about their privacy and rights. An estimated 73% of websites now include a privacy policy [AEL02]. These policies are in many ways modeled after software license statements, and are often more legalistic than user friendly. How well do these policies meet the needs of users, and how can they be improved?

In a 2001 survey, 69% of respondents said they agreed with the following statement: "*I am concerned about privacy invasions* [online] *and try to take action to prevent them from happening to me*" [CM01]. Is there really cause for concern? Well, according to [AEL02], (91%) of U.S. Web sites collect personal information and 90% collect personal identifying information. Currently, privacy policies are the only mechanism for communicating privacy practices in wide scale use. For the user, they represent the only source of information on which to base decisions about participation and privacy. Therefore, it is important to examine this practice and determine whether it fulfills its purpose

The privacy policy builds on the ideas of fair warning and implicit consent. If a company posts its policy in a public place, it can assume that users have been warned and by the act of continuing to use the service, have agreed to its terms. This assumption can be made regardless of whether users have actually read the warning or not.

Businesses find this practice compelling because it requires very little effort or expense on their side. A policy must be posted and made publicly available, from there on the user is made solely responsibility for their own protection. Sites typically do not gain informed consent, but rather assume it from all visitors.

The practice of fair warning rests on three pillars [USC99]:

- Warning should be readily available to affected parties
- Affected parties should be given a clear way to voice their concerns or questions; and
- Warning should be understandable to any reasonable person making a good faith effort.

When it comes to the issue of availability, most websites featuring a privacy policy provide (at least) a link to it from their main page. Though not always prominently featured, sites which do have a privacy policy usually make it available to users, at least those users who know to look for it. This is an important issue: A privacy policy may not be sought out or consulted unless a user suspects information about them is being collected. It is unlikely that users know 91% of all US sites collect information on them [AEL02].

In general, the second principle is fulfilled since most sites provide at the minimum an email address for the webmaster. Whether this person is qualified to, or willing to answer questions about the privacy policy is debatable. But does this really matter? After all, online policies are non-negotiable. The user is presented with a set of terms and conditions, and has no leverage, or voice to negotiate new terms. While in some cases this may be an artifact of technical contraints, most often this is not the case.

We are increasingly becoming a nation of Internet users. According to a recent survey, 53.9% of the US population is now online [NTIA02]. As more of us go online, the diversity seen in our nation is also reflected online. Gone are the days when the Internet was the exclusive domain of researchers and univerisites. Now the Internet is used by people from all walks of life. For this reason, the third principle on which privacy policies are built needs to be carefully examined. We need to make sure that we are not creating a "digital literacy divide", that we are not failing vulnerable populations by allowing them to be exploited. For this reason we examine the state of literacy online, and whether privacy policies live up to the standard of being understandable to everyone.

BACKGROUND

Readability

The fair warning principle rests on the condition that "warning should be understandable to any reasonable person making a good faith effort". To investigate whether privacy policies live up to this requirement, we have analyzed the writing of 22 privacy policies. In order to do so objectively and in a manner which would allow for further analysis, we used a standardized readability measure.

There are a number of common methods to analyze readability currently in use. The most popular are summarized below:

Dale-Chall
A vocabulary-based formula normally used to assess upper elementary through secondary materials.

Flesch Reading Ease
Normally used to assess adult materials, shows scores on a scale between 0 and 100.

Flesch Grade Level
Most reliable when used with upper elementary and secondary materials.

Fry Graph
Used over a wide grade range of materials, from elementary through college and beyond.

FOG
Widely used in the health care and general insurance industries for general business publications.

Powers-Sumner- Kear
Used in assessing primary through early elementary level materials.

SMOG
Unlike any of the other formulas, SMOG predicts the grade level required for 100% comprehension.

FORCAST
Focuses on functional literacy. Used to assess non-running narrative, e.g. questionnaires, forms, tests.

Spache
A vocabulary-based formula widely used in assessing primary through fourth grade materials.

For more information see [Tef87].

Each of these methods have their own strengths and weaknesses, are based on different theoretical foundations, and have their own set of vocal proponents. Though the FOG index is widely used in the evaluation of documents in heathcare and insurance fields, we chose to use the Flesch index instead. Our choice was largely influenced by two factors: The need to classify the readability of documents into the educational grade level required of their readers, and the desire to compare our findings to previous studies [Hoc01, NTIA02, USA01].

Flesch Reading Ease Calculation

The Flesch formulas work well with upper elementary and secondary texts. The Flesch Reading Ease Score is calculated as follows [Fle49]:

$$206.835 - 84.6 * (\text{total syllables/ total words}) - 1.015 * (\text{total words/total sentences})$$

The Flesch Grade Level is calculated as follows [Fle49]:

$$(0.39 * \text{Average sentence length (in words)}) + (11.8 * \text{Average number of syllables per word}) - 15.59$$

Table 1 gives an overview of how the Flesch Reading Ease Score relates to the Flesch Grade Level [Fle49].

Table 1: Flesch Grade Level

Flesch Score	Text Difficulty	Flesch Grade Level
0-29	Very Difficult	College Graduate
30-49	Difficult	College
50-59	Fairly Difficult	10th-12th (High School)
60-69	Standard	8th to 9th Grade
70-79	Fairly Easy	7th Grade
80-89	Easy	6th Grade
90-100	Very Easy	5th Grade

Flesch Reading Ease Score and its relationship to the the Flesch grade level and complexity of texts

METHODOLOGY

To make this study as balanced as possible, we studied a set of medical, or healthcare related websites. Healtcare is an area where privacy is becoming more stringently regulated. The professional practices common to healthcare providers also leads them to go to great efforts to protect the confidentiality of their clients and patients. We therefore expect healthcare to yield more thorough privacy policies than what a random selection would.

To further aid our analysis we chose to look at the set of 23 sites examined in [AER02], a study which examined the goals of each of these policies (for a list of sites, see table 3). This goal analysis was deemed important to our study, as it could provide a way of normalizing policies in terms of their scope. One of these 23 sites, (WebRX) had to be excluded from our study because the policy was not available, leaving us with 22 policies. We examined the same version of the policies as in [AER02] in order to ensure the goal analysis was applicable to our analysis.

In this context, goals represent the abstract points that the policy maker wants to communicate to the user. Goals represent an objective and quantifiable unit of measurement for the content of these privacy policies. Two policies may express the exact same set of goals, but written in two very different styles. When doing a content analysis of the policies, this index of goals allows us to normalize based on the ammount of communication attempted in the policy.

A number of tools calculate the Flesch readability score automatically, including Microsoft Word[1], which we used to evaluate the policies. Microsoft Word also performs the calculations for the Flesch grade level, but caps the grade level at 12th grade. We were therefore unable to use Microsoft Word to evaluate the grade level and instead calculated these scores manually.

To evaluate the readability of online policies, we have to make assumptions about the reading abilities of Internet users. Given that the sites we studied provide services and information to US residents, our assumptions reflect the literacy of the adult US population. We restrict our analysis to adult readers (25 years or older). Almost all members of this reading population have completed their formal education, which means that the literacy requirements of policies and the literacy of the reading population can be fairly compared., An additional reason to exclude younger readers is that minors are protected under special legislation [FTC99] When we refer to the Internet population througout this paper, we refer to Internet users older than 25, who residie in the USA.

RESULTS

Internet Literacy

The 1992 National Adult Literacy Survey [NALS93] set the average reading ability for the adult population (adults over the age of 25) between 8th and 9th grade. 19.6% of the US population over the age of 25 have less than a high school education, and only 24.4% of the population have a bachelors degree or higher [USC02]. It is clear that there is a literacy divide in America today, but does this extend into the Internet? Literacy and education are closely linked to income and, as computers and Internet access are not universally available, we expected the Internet population to have a higher than average literacy rate.

How big is the online literacy divide? How diverse is the Internet population when it comes to education? Until the late 90's, the Internet was predominantly used by well educated men. In 1994, the Second GVU WWW User Survey [GVU94] found that 70% of respondents had a college education or higher, and an additional 19% had some college education/were still in college. Furthermore, 90% of respondents were male. Since then, Internet use has grown almost exponentially. In a 2002 survey, the National Telecomunications and Information Administration (NTIA) determined that 53.9% of the US population is now online, and that gender differences have largely disappeared [NTIA02].

The NTIA survey [NTIA02] did look at how education affected Internet use, looking at what percentage of adults in each educational category (less than high school, high school diploma, some college, bachelors degree, or beyond bachelors) use the Internet. When combined with information from the 2000 US Census [USC02] on the number of adults in each of these categories we find what percentage of the online population these groups constitute (Table 2).

Table 2: Education and Internet Use

Educational Level	General Population *		Online +	Internet Population	
Less Than High School	35,715,625	19.6%	17.0%	6,071,656	6.2%
High School Diploma / GED	52,168,981	28.6%	47.3%	24,675,928	25.1%
Some College / Associate Deg	49,864,428	27.3%	69.5%	34,655,777	35.3%
Bachelors Degree	28,317,792	15.5%	84.9%	24,041,805	24.5%
Beyond Bachelors #	10,088,500	5.5%	86.9%	8,766,907	8.9%

*2000 US Census [USC02] +2002 NTIA report [NTIA02] #US Census reports graduate studies together with technical degrees. This later category was removed to match the classification in the NTIA survey.

As we expected, education still has an effect on Internet use. Our analysis shows that the average education of the adult US Internet population is 14th grade[2], higher than high school, but not as high as a bachelors degree. This means that the Internet population, on the whole, is more literate and better educated than the general population. As the Internet population continues to grow, this average will continue to drop to more closely match that of general population.

We find that the US Internet population divides into three categories of roughly equal size; those with a high school education or less, those with some college education, and those with a bachelors degree or higher. This means that though the Internet population is currently better educated and more literate then the general populus, there is a sizable group which is vulnerable in Internet transactions requiring advanced literacy skills.

[1] http://www.microsoft.com/office/word/

[2] Average calculated with following values: All with less than high school education = 11th grade, high school = 12th grade, some college =14th grade, college = 16th grade, postgraduate = 17th grade.

Policy Readability

Our survey of 22 privacy policies found the average Flesch Reading Ease Score to be 36.9 (SD=5.1), and the average grade level required to read these policies is 13.7 (SD=1.2). For individual scores and reading levels, refer to Table 3.

Table 3: Policy Readability and Content

	Company Name	Flesch Score	Grade	Words	Goals*	Vulnerabilities *	Seal
Health Insurance	AETNA	39.4	14.2	806	10	5	N
	AFLAC	30.4	14.7	1930	2	1	N
	BCBS	40.2	15.2	638	20	7	N
	CIGNA	45.2	10.7	875	11	5	N
	EHealthInsurance	23.1	15.8	1546	15	8	Y
	Kaiser Permanente	32.0	14.3	689	5	1	N
	OnlineHealthPlan	31.9	14.2	1390	17	9	Y
Online Drugstore	CornerDrugstore	37.6	13.5	1906	24	9	Y
	DestinationRX	38.7	13.4	1925	34	18	Y
	Drugstore	38.7	13.8	1499	29	14	Y
	Eckerd	35.5	14.1	1340	15	6	N
	HealthAllies	34.5	14.4	1025	17	6	Y
	HealthCentral	41.1	13.1	1283	25	12	N
	IVillage	28.9	16.3	3382	39	18	N
	PrescriptionOnline	33.8	13.6	753	13	4	N
	PrescriptionsByMail	39.9	12.9	1082	18	7	Y
Pharmaceutical	Bayer	40.9	13.1	760	17	9	N
	Glaxo Wellcome	39.5	12.6	448	12	7	N
	Lilly (Eli)	40.4	13.6	507	7	5	N
	Novartis (Ciba)	39.7	13.5	1340	23	5	N
	Pfizer	41.1	12.1	393	7	3	N
	Pharmacia	38.7	13.3	957	18	8	N
	Average	36.9	13.7	1203.4	17.2	7.6	31.8%

** Goals and vulnerabilities taken from [AE03]. All policies dated summer 2000. Seals are from a number of different organizations, including TRUSTe, BBBOnline and WebTrust.*

On average, these policies require a reading skill equivalent to some college education. This average is somewhat lower than the average literacy level of the Internet population. Examining the individual policies we see that only one policy (that of CIGNA) scored lower than a highschool education (10.7). All other policies required more than a high school education (though Pfizer not by much). Acording to our analysis of the current Internet demographics, we find that 31.3% of the adults over 25 currently online only possess the skills to read and understand one of the 22 policies examined. Two of the policies we examined (iVillage and eHealthInsurance) effectively required the equivalent of a college education (grades 16.3 and 15.8, respectively). This means they are incomprehensible to 66.6%, a full two-thirds of all Internet users.

Our findings are in line with those found in other surveys of policies in other fields [USA01, Hoc01]. In [Hoc01], 60 financial privacy policies were examined, the average Flesch score of which was 34 (*SD*=5.8), and an average grade level of 15.6 (*SD*=1.0). In that survey, no policy scored lower than 13th grade, and 12 scored 17th grade or higher. This means that none of these policies were accessible to one-third of the online population, and 20% of the policies were only accessible to 8.9% of the population.

We examined whether there was a relationship between the length of the policy (in words) and either the Flesch score, or the grade level. Neither proved to be the case, having a correlation coefficient of 0.505 and 0.564 respectively. There also were no sign of a relationship between the number of goals expressed in the policy (a measure for content in the policy), and the Flesch score or the grade level (correlation coefficients 0.012 and 0.058 respectively). The written complexity of the policies we studied was independent of the ammount of information the policy conveyed.

Impact of Privacy Seals

The FTC encourages self-regulation when it comes to policies and their content [FTC98]. In light of this situation, privacy seal organizations, such as TRUSTe[3], BBB*online*[4] and WebTrust[5], have emerged, offering certification to sites and their policies. These seals often do not address the content of the policy, but rather the fact that the company has a policy, that this policy addresses a minimum set of issues, and that the company adheres to the stated policy. Users often mistake these seals to mean something about the level of privacy protection offered, which is not the case [BBB01].

In terms of readability, the presence of such a seal has no effect on the Flesch score, or the grade level of the policy (*t* (19) = -1.13, *p=NS*, and *t* (19) = 1.33*, p=NS* respectively). There was a marginal effect for goals (*t* (19) = 1.94, *p=0.067)*; the policies with seals addressed more issues than those without. This was to be expected, as many of the seals require that a policy address a minimum set of standard issues.

Does the fact that seals encourage policy makers to address more issues/goals in their policies make for better policies, despite the fact that it does not make for more accessible policies? In other words, are users right to assume that seals mean stronger privacy policies? [AE03] distinguishes

[3] http://www.truste.com/

[4] http://www.bbbonline.com/

[5] http://www.cpawebtrust.org/

between privacy protection goals and vulnerabilities as follows:

"Privacy protection goals are those that relate to the five FIPs" [FIP73] *"and to the desired protection of consumer privacy rights. Privacy vulnerabilities relate to existing threats to consumer privacy. In contrast to protection goals, vulnerability goals represent statements of fact or existing behavior and are often characterized by privacy invasions."*

If the assumption that privacy seals lead to policies which better protect users' privacy, we should see a higher ratio of protection goals to vulnerabilities in policies with seals. In fact, what we see is the opposite: On average, seal-carrying policies have a ratio of 0.92 vulnerabilities to each protection goal, while non-seal sites list 1.15 vulnerabilities to each protection goal (though this difference is not statistically significant ($t(19)=-1.42$, $p=NS$). Though seals may encourage more complete policies, they do not necessarily mean better protection for the user.

Policy Use

The readability of policies, and their content, only really matter if people try to access and read them. Whether policies are actually read by anyone is not something about which there is much data. In commercial circles, information on page-hits and user numbers are often guarded as trade secrets. We therefore have to rely largely on subjective user reports of what they do. Unfortunately, as is commonly the case with surveys, making comparisons between the different surveys is difficult, as wording can strongly influence responses, and definitions of the categories used by researchers are not always available.

A November 2001 survey done by Harris Interactive on behalf of the Privacy Leadership Initiative, 2,053 users were asked *"Specifically, how much time on average have you spent reading websites' privacy policies?"* 31% said they spent little or no time looking at policies, 33% glanced through, but rarely read, 33% said it depended on circumstances but that they sometimes read policies carefully, and only 3% said they read the policies carefully most of the time. [PLI01]

The Culnan-Milne survey, also from November 2001, asked 2,468 users the following question: "*How often do you read privacy notices posted by websites?*" 17% reported that they never read them, 33% reported that they rarely read them, 31% reported that they sometimes read them, 13% said they frequently read them, and 5% claimed they always read them [CM01].

A third study performed by Jupiter Research March 2002 asked 2097 users to rate the accuracy of the following question: "*Before registering, I always read the privacy statement.*" 11% strongly disagreed, 24% disagreed, 25% were neutral, 26% agreed, and 14% strongly agreed. [Jup02].

As expected, only a minority of users report reading policy statements, and more than half of Internet users reported that they rarely, if ever, read policies. Exactly how many actually do read these policies, how often, and how closely, is very difficult to determine, and no hard data is available. However, as some indication, in a June 15th 2001 article in the e-commerce Times, Forrester Research analyst Christopher Kelley reported "*less than 1 percent of the visitors to six major online travel sites during April* [2001] *actually read privacy policies*" [Reg01].

DICUSSION

Readability

We can conclude that almost half of the US Internet population does not have the reading skills to make sense of the average privacy policy in our sample. If we compare our results to other domains we find that users come out even worse, only 1/3 having the skills to interpret the average financial privacy policy [Hoc01].

What does this mean? At a very fundamental level, online privacy policies are failing to meet their purpose. They may provide notice, but fail to meet the requirement of being understandable to any reasonable person making a good faith effort. This indicates the presence of a significant digital literacy divide, at least in the context of privacy protection. As the Internet has reached a larger, more diverse audience, more and more people are left behind, and left vulnerable to abuses of their personal information.

Making policies readable is of crucial importance, because difficult language, long and confusing policies all serve to trick and confuse the user. In [BLJ01] the authors demonstate how by refrasing the same request for information, wildly different responses are given. As the authors state: "*…a study by Cyber Dialogue found that 69% of U.S. Internet users did not know they had given their consent to be included on email distribution lists. Here's how it's done: Using the right combination of question framing and default answer, an online organization can almost guarantee it will get the consent of nearly every visitor to its site*." By manipulating these simple variables they were able to bring participation up from 48.2% to 96.3% [BLJ01]. Such tactics verge on being predatory, since they aim to exploit the user's vulnerabilities, especially those with the weakest literacy skills.

Many states have readability requirements for formal documents such as insurance policies; "[f]*or example, Arkansas, Indiana, Kentucky and Ohio require a minimum score of 40 on the Flesch Reading Ease.* […] *Connecticut and Florida require a minimum of 45 on the Flesch*[…]. *Maine requires a 50*" [Hoc01]. If similar standards were legislated to apply to privacy policies, only six of the 22 policies (27.3%) would meet even the loosest requirement that applies to insurance policies (a Flesch score of 40 or higher). Only one policy (4.5%) would meet the requirements of Connecticut and Florida, and none of the policies would meet the requirements of Maine.

These requirements on insurance policy readability are intended to prevent fraud and ensure that policies meet a minimum level of fairness in information disclosure. Policies such as these are especially important when it comes to protecing vulnerable segments of the population, a group that is increasingly establishing a presence online.

Undue Burden

The surveys of how privacy policies were being read paint a grim picture; over half the respondants report never or rarely reading privacy policies. Certainly, it is to be expected that the readability issue would discourage user involvement, but is this the only factor? If we look back at the surveys examining how actively policies were used, the top two reasons given for not reviewing them are that they are too time-consuming and too hard to read (PLI; 40% and 29% of respondents respectively [PLI01]. The Culnan-Milne report does not give ratios, but lists these reasons as being the most commonly given [CM01]. Jupiter; only 31% or respondents find policies easy to read, 43% disagree [Jup02].)

The time required of a user to review policies is significant. Even if the policies were all readable by a user, the burden on that user to review every policy for every time he or she visits a site is tremendous. Yet, this is what many sites require their users to do. Clauses such as the following from the Aetna privacy policy are typical: "*Aetna Inc. may change this Statement from time to time without notice*."

In our sample 10 of the 22 sites (45.5%) used statements like these to describe their official notification mechanism. In an industry which is heavily regulated and has a longstanding tradition of strong privacy protection, only seven of the 22 healthcare organizations (31.8% of the sites) took the burden of notification and issued warnings to users that their policy had changed Most commonly, these took the form of email notifications. Five sites (22.7%) did not disclose how policy changes would be advertised, meaning that the burden to find out is given to the user. All in all, 68.2% of our sites require their users to check the policy every time they visit the site. Failure to do so is automatically interpreted as a sign of acceptance of the new terms.

Pitfalls of Policy Practices

In addition to the issues addressed so far, there are other problems associated with privacy policies and the current practice.

Implicit Consent

The current practice for collecting consent is to assume that any person using the system has agreed to the terms specified in the privacy policy and has given their consent. For this to be a fair practice, policies would have to be accessible from clearly marked "safe" areas of the websites, areas not bound by the terms of the policy and free of data collection of any form. Without such provision, the simple act of consulting the policy (requiring the user to at least access the policy page and the site's front page) means that users have already given their consent, a "Catch-22" situation. This practice violates the very essence of the concept of fair warning as well as consent.

Focus of Policies

The goal-based content analysis of these website privacy policies shows that these policies tend to focus on "*the security of the data collection process; how and what information is being collected; and contain assurances that users are given the option to decide how personal information collected about them is to be used*" [AE03]. While these are undoubtedly important issues, a survey of users' privacy concerns shows an important mismatch between these issues and what users are most concerned about. "Transfer of information to others, such as third parties; about being notified about practices before any information is collected; and how and what data are stored about a user" [EAA03]. This demonstrates the difference between what policies address and what users are looking for.

Such discrepancies are of course to be expected. The organizations responsible for these websites need to satisfy their legal requirements, and users' concerns may shift over time. It is important, however, to note that users have different information needs, and seek to satisfy them.

Lack of Transparency

In the Harris user survey, a small group (5%) gave the following as their primary justification for not reading policies: "*Do not believe the policies will protect my privacy*" [PLI01]. Although uncommon, this sentiment reflects a powerful, and possibly growing skepticism. Though connected with the issue of focus, this also addresses the issue of transparency.

Users of a website are given a policy to read, in essence a list of promises and disclosures about how their information will be used and treated. What takes place behind the scenes is hidden from the user, and so it is a daunting task to determine whether a site abides by its own policy. Companies know and users both know this. It is therefore not surprising that some users mistrust the validity of policies. The burden placed on the user to verify a site's compliance with its policy is even greater than the burden of reviewing each policy.

This is not to say that when serious policy violations are detected, investigation, punitive action, or damage to reputation necessarily follows. Some examples where negative consequences ensued include FTC vs. Toysmart [FTC00], FTC vs. Eli Lilly [FTC02a], or FTC vs. Microsoft [FTC02b]. But countless policy violations go unpunished (for instance the Microsoft FTP server leak of November 2002 [Ley02], or the 2002 Hotmail violation [Hal02]). For users to take privacy policies seriously, they need to be given verifiable guarantees, or more insight into how their information is used.

P3P Policies

Efforts to address the problems of undue burden and implicit consent have in part led to the development of P3P[6], an alternative approach to today's policies. P3P, in essence, provides a mechanism for encoding policies in a machine-readable form. This makes it possible to develop automated policy analysis tools and user agents [CLM02]; this reduces the burden that users face when checking policies by allowing for automation. In answer to the problem of implicit consent, P3P defines "safe-zones", designated areas for posting policies, areas that can be accessed without implying consent for the user. P3P also introduces other desirable features, such as binding the data collected to the policy under which it was collected. This forces providers to respect and abide by the agreement that was made at collection time, and collect new consent whenever a policy changes.

Approaches such as P3P are an important step towards making policies usable, but not perfect. Adoption has been slow, and much debate still rages on the suitability of P3P in protecing users privacy. The European Union has rejected P3P as a viable technical means for supporting their privacy laws [Epi00]. Others have argued that P3P fails to comply with baseline standards for privacy protection and is a complex/confusing protocol that hinders Internet users in protecting their privacy [Epi00]. Little evidence supports industry's claim that P3P improves user privacy, and it does not assess compliance with the five FIPs [FIP73].

Realizing Informed Consent

Ultimately, privacy is a highly subjective matter. Boundaries and sensitivities are not only culturally determined, but also highly individualized. What is acceptable and unacceptable must therefore ultimately be a decision that the affected person must make.

Awareness

The first element of informed consent must by necessity be to ensure users are actually informed of what is taking place. If users are to give consent, they need to be able to effectively evaluate the risks and benefits of any transaction. The problem is that we not only have a mismatch between what sites and users are concerned about in terms of privacy, but also a huge gap in vocabulary.

Policies refer to things such as cookies, web-bugs and ActiveX controls. Users are not concerned about technological details; they are interested in knowing what the consequences of their actions will be. It is as unreasonable to expect users to know what the risks associated with cookies are as it is to expect them to find and remember websites by their IP numbers alone. It is our job to bridge this gap, to make privacy accessible to the general public, to translate risk assessments and technological vulnerabilities into terms that users can understand.

Trust and transparency are also key elements to realizing informed consent. When selecting a provider, or deciding whether to disclose information, users must determine whether a provider will act in a manner consistent with their wishes and whether the provider will live up to their promises. In effect, users conduct a cost-benefit analysis of: the information disclosed, how sensitive it is to the user, and the (perceived) risk for misuse. This is weighted against the transaction's potential benefit. If an individual does not trust a provider to behave in an appropriate manner, no transaction will follow unless the individual is somehow forced, or the stakes are low [Kol98]. We must seek to develop certification schemes for policies, and promote transparency in applications, giving users a chance to confirm policy compliance for themselves.

History

Information disclosures often occur over time, information processing, or new features or services lead to a need for added information. To the user, this is a particularly dangerous practice. When faced with such a request for information, it is difficult for users to evaluate how it will interact with previously disclosed information, and what this aggregate may reveal about them. Sometimes this strategy is employed on purpose to confuse the user; we refer to this practice as "information creep". If users are to make an informed decision, full disclosure of already known information is needed. If the provider does not provide this, then the user should have the ability to access such information.

Finally, privacy protection begins at home. Users need to be aware of how much information they disclose about themselves (intentionally as well as unintentionally). If users are uninformed and do not protect themselves against unreasonable disclosures or terms of use, then there is little any system can do to help them. The most important thing that can be done is to try to educate users about how their practices affect them, support a process of self-reflection, and inform them about the tools at their disposal to better protect their privacy.

Control

A main deterrent to users managing their privacy is the burden it currently presents. Companies only have to define a single policy to cover all their activities, but users are expected to review the policy of every company with which they interact. Faced with such an overwhelming burden, most users quickly give up, or develop different shortcut techniques to minimize the burden. There are numerous tools that allow users to filter or warn users

[6] http://www.w3c.org/P3P

about practices, mostly based around P3P (e.g. Microsoft's Internet Explorer 6[7] and AT&T's Privacy Bird[8]).

These tools are typically based on one of two principles: linear privacy preferences or the techno-centric selection. The principle of the linear privacy preference assumes that privacy concerns can be defined on a discrete, monotonically increasing scale of risks, or compromises. While greatly simplifying the task for the users, this approach tries to force fit the users own privacy preferences and concerns into defined categories, limiting their options. The techno-centric approach asks the user to know the limitations and risks associated with different technologies, asking for their preferences to each of them.

Using scenarios, and use/misuse cases, leveraging their conceptual models, the user can be engaged using terminology that they understand. This methodology would enable us to obtain a set of performance driven requirements from users, as well as a rule-set detailing what tradeoffs and risks they are willing to accept. This rule-set may then be compared to website privacy policies to further automate the process of checking suitability and compliance.

FUTURE WORK

Our main focus as we go forward will be to look at ways to provide users with better tools for protecting their privacy. Key to this is making sure the user is informed about what risks they are exposing themselves to, what information they are disclosing, and how that information has been shared with others. We are working on a set of end-user tools to empower users to take a more active and meaningful role in protecting their own privacy online. These include awareness tools, history tools, and control tools aimed at reducing the burden placed on end users.

Other possible avenues of research include continuing our data collection, looking at how policies are actually used to complement and contrast against the survey data. We also wish to re-examine these policies to see how they have changed since this survey was done. Are policies becoming more readable, more user friendly as the market and the practice matures? It would also be interesting to compare our results to those of a random sample of sites to determine how representative our findings are.

CONCLUSIONS

The practice of privacy policies as it stands today is an ineffective way to protect users privacy online. Most users rarely consult privacy policies, and when they do they often find them unintelligible. As the Internet has moved away from being the exclusive domain of academics and researchers, more vulnerable populations have grown significantly. This trend will continue, as the differences between the Internet population and the general population continue to equalize. It will therefore be continually important to ensure that policies are clear and readable.

As the practice stands today, having a privacy policy linked to a site, or displaying a privacy seal, may have more of a negative than a positive effect on users. Because they are unlikely to follow these links and determine what the policy says, or what the seal stands for, they may make the wrong assumptions. The simple fact that a site has a policy or a seal does not mean they protect users privacy more than a site without, though users may make that asumption.

There is a clear and definitive need for tools to support and empower the user to take charge of their privacy online. Without such tools the task is proving to daunting and confusing, putting many users at risk of privacy violations. We in the HCI community are faced with a tremendous challenge in making privacy protection accessible to the end-user, and an opportunity to do something that will greatly benefit many.

ACKNOWLEDGMENTS

This work was supported by NSF ITR Grant #0113792. The authors thank Annie I. Antón, Julia B. Earp, David L. Baumer, William Stufflebeam and Qingfeng He for discussions leading to the development of these ideas.

REFERENCES

[AEL02] William F. Adkinson, Jr., Jeffrey A. Eisenach, and Thomas M. Lenard Privacy Online:A Report on the Information Practices and Policies of Commercial Web Sites http://www.pff.org/publications/privacyonlinefinalael.pdf Progress and Freedom Foundation, Washington DC. March 2002

[AER02] A.I. Antón, J.B. Earp and A.Reese. Analyzing Web Site Privacy Requirements Using a Privacy Goal Taxonomy, 10th Anniversary IEEE Joint Requirements Engineering Conference (RE'02), Essen, Germany, pp. 23-31, 9-13 September 2002.

[BBB01] BBBOnLine. Third-Party Assurance Boosts Online Purchasing: BBBOnLine Privacy, Reliability Seals Increase Consumer Confidence; Privacy Remains Public's Chief Concern (survey summary). http://www.bbbonline.org/about/press/2001/101701.asp Conducted by Greenfield Online. Arlington VA: BBBOnLine, October 17, 2001.

[BJL01] Steven Bellman , Eric J. Johnson , Gerald L. Lohse. On site: to opt-in or opt-out?: it depends on the question. Communications of the ACM February 2001 Volume 44 Issue 2

[CLM02] L. Cranor, M. Langheinrich, and M. Marchiori. A P3P Preference Exchange Language 1.0

[7] http://www.microsoft.com/windows/ie/default.asp

[8] http://www.privacybird.com/

(APPEL1.0): http://www.w3.org/TR/P3P-preferences/ W3C Working Draft, 15 April 2002.

[CM01] Mary J. Culnan and George R. Milne. The Culnan-Milne Survey on Consumers & Online Privacy Notices: Summary of Responses. http://www.ftc.gov/bcp/workshops/glb/supporting/culnan-milne.pdf Washington DC: FTC, December 2001.

[EAA03] Julia B. Earp, Annie I. Antón, Lynda Aiman-Smith, and William Stufflebeam. Crossed Signals: Internet Privacy Policies and User Concerns, Submitted to: MIS Quarterly, 31 January 2003.

[Epi00] Pretty Poor Privacy: An Assessment of P3P and Internet Privacy, http://www.epic.org/reports/prettypoorprivacy.html, Electronic Privacy Information Center, June 2000.

[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, 1973.

[Fle49] Rudolph Flesch, "The Art of Readable Writing", Macmillan Publishing, 1949

[FTC98] Privacy Online: A Report to Congress, http://www.ftc.gov/reports/privacy3/, Federal Trade Commission, June 1998.

[FTC99] Federal Trade Commision. Children's Online Privacy Protection Rule. Federal Register Vol. 64, No. 212. November 3, 1999

[FTC00] FTC v. Toysmart.com, LLC, and Toysmart.com. http://www.ftc.gov/os/2000/07/toysmartcomplaint.htm Federal Trade Commission, July 2000

[FTC02a] Eli Lilly Settles FTC Charges Concerning Security Breach, FTC Press Release, http://www.ftc.gov/opa/2002/01/elililly.htm, Federal Trade Commission, 18 Jan. 2002.

[FTC02b] In the Matter of Microsoft Corporation, http://www.ftc.gov/opa/2002/08/microsoft.htm, Federal Trade Commission, 8 August 2002.

[GVU94] GVU's Second WWW User Survey, Graphics, Visualization & Usability Center, Georgia Institute of Technology http://www.cc.gatech.edu/gvu/user_surveys/survey-09-1994/ October 1994

[Hal02] Paul Hale, "MSN coughs to adding mysteriously-appearing buttons." http://www.theinquirer.net/?article=4759 The Inquirer, 1 August 2002

[Hoc01] Mark Hochhauser. Lost in the Fine Print: Readability of Financial Privacy Notices. Privacy Rights Clearinghouse Website, July 2001. http://www.privacyrights.org/ar/GLB-Reading.htm

[Jup02] Jupiter Research, Security and Privacy Data http://www.ftc.gov/bcp/workshops/security/020520leathern.pdf Presentation to the FTC Security Workshop, May 20, 2002

[Kol98] Peter Kollock Social Dilemmas: The anatomy of cooperation. Annual Review of Sociology.1998. 24: 183-214

[Ley02] John Leyden, On the Microsoft FTP server leak. http://www.theregister.co.uk/content/55/28252.html The register, 22 November 2002.

[NALS93] Irwin S. Kirsch, Ann Jungeblut, Lynn Jenkins, and Andrew Kolstad. Adult Literacy in America: A First Look at The Results of The National Adult Literacy Survey. Washington, DC: National Center for Education Statistics, U.S. Department of Education, September 1993.

[NTIA02] National Telecommunications and Information Administration. A Nation Online: How Americans Are Expanding Their Use of the Internet http://www.ntia.doc.gov/ntiahome/dn/ Washington, D.C. February 2002

[PLI01] Privacy Leadership Initiative. Privacy Notices Miss the Mark with Consumers. http://www.ftc.gov/bcp/workshops/glb/supporting/harris%20results.pdf (survey results). Conducted by Harris Interactive. Washington DC: PLI, December 3, 2001.

[Reg01] Keith Regan. Does Anyone Read Online Privacy Policies? http://www.ecommercetimes.com/perl/story/11303.html E-Commerce Times June 15, 2001

[Tef87] C. Tefki. Readability formulas: an overview, Journal of Documentation, 43 (3): 257-269. 1987

[USA01]Your Privacy is Important to Us? A Report Card on How Bank Privacy Notices Discourage Consumers from Exercising the Right to Financial Privacy http://www.ftc.gov/bcp/workshops/glb/supporting/citaction-execsumm.pdf USAction. Presented to the FTC December 4, 2001

[USC99] Regulatory Fair Warning Act of 1999. H.R. 881 http://commdocs.house.gov/committees/judiciary/hju63853.000/hju63853_0.htm One Hundred Sixth Congress, June 29, 1999

[USC02] United State Census 2000, U.S. Summary http://www.census.gov/prod/2002pubs/c2kprof00-us.pdf United States Census Bureau, July 2002