Analyzing Website Privacy Requirements Using a Privacy Goal Taxonomy

Annie I. Antón[1], Julia B. Earp[2], Angela Reese[3]
[1] College of Engineering, North Carolina State University, Raleigh, NC 27695-7534
[2] College of Management, North Carolina State University, Raleigh, NC 27695
[3] College of Arts and Sciences, University of Dayton, Dayton, OH 45469-0800
[1] aianton@eos.ncsu.edu [2] Julia_Earp@ncsu.edu [3] angelamreese@hotmail.com

Abstract

Privacy has recently become a prominent issue in the context of electronic commerce websites. Increasingly, privacy policies posted on such websites are receiving considerable attention from the government and consumers. We have used goal-mining, to extract pre-requirements goals from post-requirements text artifacts, as a technique for analyzing privacy policies. The identified goals are useful for analyzing implicit internal conflicts within privacy policies and conflicts with the corresponding websites and their manner of operation. These goals can be used to reconstruct the implicit requirements met by the privacy policies. This paper interrelates privacy policy and requirements for websites; it introduces a privacy goal taxonomy and reports the analysis of 23 Internet privacy policies for companies in three health care industries: pharmaceutical, health insurance and online drugstores. The evaluated taxonomy provides a valuable framework for requirements engineering practitioners, policy makers and regulatory bodies, and also benefits website users.

1 Introduction

Requirements engineering (RE) is the principled application of proven methods and tools to describe the behavior and constraints of a proposed system. Our approach to policy and requirements specification [AE01a] applies goal and scenario-driven RE methods to specify: privacy policies, security policies and the corresponding system requirements. This paper explains our application of these techniques to Internet policy analysis. We also introduce a privacy goal taxonomy that provides an effective mechanism for analyzing and comparing privacy policies, system requirements and the respective system's functionality.

Health care privacy holds profound implications since service delivery impacts human life, legality and social policy [Dar97]. Electronic health related information transmission and dissemination has raised privacy concerns among both consumers and providers [EP00]. The evolving trend toward Internet supported health care services has inevitably resulted in increased information sharing among providers, pharmacies and insurers. Unfortunately, such information sharing often conflicts with consumers' desires to be shielded from unauthorized personal information use. We employ goal-mining to derive the privacy-related goals (and system requirements) from Internet health care website privacy policies.

This study is focused on three objectives. The first was to create a taxonomy for classifying privacy goals and requirements. Second, to develop a corpus of reusable privacy and security goals for e-commerce software developers [AAB99, AE01a]. Goals are a cogent unit by which to objectively analyze and compare Internet privacy policies, enabling us to provide useful guidance to RE practitioners, policy makers, and consumers. Goal-mining with the privacy goal taxonomy is effective for examining how websites claim they manage online customer data and how they convey these practices to their customers [AEP01]. Thus, the third objective is to provide a basis for analyzing and comparing Internet privacy policies.

This paper is organized as follows. Section 2 discusses relevant related work. Section 3 introduces a privacy goal taxonomy. Section 4 briefly discusses the goal-mining process employed to analyze 23 health care website privacy policies. A discussion and plans for future work are provided in Sections 5 and 6, respectively.

2 Background and Related Work

This section discusses the relevant work in health care privacy policy and legislation, policy evaluation and the role of goal-based RE in policy analysis.

2.1 Health Care Privacy Policy and Legislation

A privacy policy comprehensively describes a website's information practices and should be easily accessible on the site [FTC98, FTC00]. Although organizations engaged in electronic transactions should disclose privacy policies that are based on the Code of Fair Information Practices (FIPs) [FIP73, FTC98, FTC00], Internet privacy disclosures do not always reflect the FIPs [Cul99]. Health-care related website privacy practices are receiving increased attention. Last year Eli Lilly unintentionally released their Prozac clients' email addresses when the company sent customers subscribing to their prescription refill reminder notification service an email message, containing all the service subscribers' email addresses, informing them that the service was being cancelled [FTC02]. Such incidents stress the need for technical measures as well as health information exchange legislation. Although the Privacy Act of 1974 provides some protection for medical records held by federal agencies, it fails to cover medical records held by private groups where most medical records are actually created and stored[1]. Moreover, the act contains numerous exceptions so that its overall protection is leaky at best.

Increased Internet utilization for health information exchange has initiated legal reform. The 1996 Health Information and Portability Accountability Act (HIPAA)[2] mandated the U.S. Government Administration to introduce medical records controls regulations. These regulations called for a provision for health information privacy. The Department of Health and Human Services' (HHS) final Privacy Rule[3] took effect on April 14, 2001 and required health care providers and plans to comply by April 14, 2003. We are clearly at a critical juncture and thus chose to focus this analysis effort within the health related information privacy domain.

2.2 Privacy Policy Evaluation Mechanisms

Privacy policies are evaluated in a rather ad hoc and inconsistent manner. Current solutions include the Platform for Privacy Preferences Project (P3P) [RC97] and various privacy seal programs [Ben99]. Internet users are concerned about threats to their privacy when online [CRA99]. Several studies have shown, however, that Internet users are more inclined to trust a website if it simply posts a privacy policy [EB02, GHS00]. Most online companies now post privacy policies on their website, but not all consumers can (or are willing to) take the time to read and understand them. Several privacy policy evaluation mechanisms do seek to assist online consumers.

The World Wide Web Consortium has established the P3P[4] as a standard to provide an automated way for users to gain control over the use of their personal information at websites they visit. P3P requires users to answer standardized multiple-choice questions that address various aspects of website privacy policies. Sites implementing P3P have a machine-readable privacy policy; users configure their browsers to automatically determine if the site's privacy policy reflects their personal privacy preferences (users' multiple choice question responses are compared with P3P compliant policy statements). Adoption has been slow; as of June 13 2002[5], only 89 sites were compatible with the P3P 1.0 specification[6]. Additionally, the European Union has rejected P3P as a viable technical means for supporting their stringent privacy laws [Epi00]. 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 FIPs.

The FTC encourages self-regulation [FTC98] but some online businesses have yet to adopt the fundamental FIPs that address consumer privacy. Privacy seal organizations, such as TRUSTe[7], BBB*online*[8] and WebTrust[9], complicate privacy policy since consumers often trust indirect and abbreviated privacy protection indicators rather than reading the full privacy policy.

The TRUSTe privacy seal simply ensures that TRUSTe has reviewed a licensee's privacy policy for the following information use disclosures: what personal information is being gathered; how the information will be used; who the information will be shared with; the choices available regarding how collected information is used; safeguards in place to protect personal information from loss, misuse, or alteration; and how individuals can update or correct inaccuracies in information collected about them. TRUSTe requires licensees to disclose their privacy practices and adhere to privacy principles based on the FIPs. The BBB*online* privacy seal similarly ensures that a privacy policy has been reviewed for disclosures akin to TRUSTe's. These mechanisms are not stringent and do not reflect a real commitment to consumer privacy, merely an openness about the degree to which privacy is supported.

A more effective privacy evaluation mechanism would consider not only the presence of certain policy content, but the policy content implications in reference to how practices affect consumer privacy. RE provides reliable and straightforward mechanisms for evaluating privacy.

2.3 Policy from the RE Perspective

There is a need to apply proven requirements analysis methods and demonstrate how to best apply these methods within the context of analyzing policy and system requirements. Goal analysis is especially well suited to support these activities. *Goals* are the objectives and targets of achievement for a system. In RE, goal-driven approaches focus on why systems are constructed, expressing the rationale and justification for the proposed system [Lam01]. Focusing on goals, instead of specific requirements, allows analysts to communicate with stakeholders using a language based on concepts with which they are both comfortable and familiar.

The Goal-Based Requirements Analysis Method (GBRAM) [Ant96, AP98, ACD01] is a straightforward methodical approach to identify strategic and tactical goals as well as requirements. It is useful for identifying and refining the goals that software systems must achieve, managing trade-offs among the goals, and converting them into operational requirements. The method suggests goal identification and refinement strategies through via heuristics, guidelines and recurring question types. In this paper, we describe the method's use to mine privacy policies for system goals and requirements. We now introduce our goal taxonomy.

[1] 5 U.S.C. 552a (1994)

[2] Health Insurance Portability and Accountability Act of 1996, 42 U.S.C.A. 1320d to d-8 (West Supp. 1998).

[3] Federal Register 59918 et seq., Department of Health and Human Services, Office of the Secretary, 45 CFR Parts 160 through 164, Standards for Privacy of Individually Identifiable Health Information, (December 28, 2000).

[4] http://www.w3.org/P3P/

[5] This date is actually misleading, since in September of 2001, 13 sites were listed. Since then, the # of sites listed has increased, but the "September 2001" date has never been updated accordingly.

[6] http://www.w3.org/P3P/compliant_sites

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

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

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

3 A Privacy Goal Taxonomy

During the summer of 2000 we applied goal-mining in a pilot study to evaluate 24 Internet Privacy Policies from 8 non-regulated e-commerce industries (e.g. Online travel agencies and online retailers). The identified goals were useful for discovering implicit internal conflicts within privacy policies and conflicts with the corresponding websites and their manner of operation. These goals can also be used to: reconstruct the implicit requirements met by privacy policies; reason about expected policy content for different website types; and aid developers in creating policies that address common goals for a given site.

Initially, privacy experts who viewed our pilot study data suggested that all goals expressed in a website's privacy policy should support the Code for FIPs [FIP73]. However, the derived privacy goals proved challenging to classify in this simple manner. We attempted to classify the goals according to the five FIPs (notice/awareness; choice/consent; access/participation; integrity/security; and enforcement/redress), but found it was impossible to "force-fit" all the derived goals into these five categories. Careful examination revealed that broader coverage was need because the remaining unclassified goals did not exemplify privacy protection practices; instead, they reflected practices that introduce vulnerabilities in a site's ability to protect personal information. Clearly, a taxonomy based solely on the FIPs is idealistic and does not represent the realistic state of Internet privacy policy. This led us to create a comprehensive taxonomy for privacy-related system goals so that consumers and system developers can more accurately compare privacy practices and reason about a site's functionality and alignment with its stated policies. Although there is no taxonomy for privacy requirements in existence, we chose to derive our taxonomy from the FIPs since it serves as the standard for all privacy recommendations in the U.S. [FTC98, FTC00, FIP73, Cul99].

The taxonomy broadly classifies privacy goals as either privacy protection or privacy vulnerability goals. *Privacy protection goals* are those that relate to the five FIPs and to the desired protection of consumer privacy rights. *Privacy vulnerability goals* 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. We now discuss these goal classes, providing concrete examples of health care privacy protection goals and vulnerability goals.

3.1 Privacy Protection Goals

Since privacy protection goals suggest those properties to be satisfied in a system, the protection goals are subdivided according to the FIP categories [FTC98] as defined in Table 1.

3.1.1 Notice and Awareness

The notice and awareness principle asserts that consumers should be notified and/or made aware of an organization's information practices before any information is actually collected from them. The mechanism by which consumers are typically made aware of such practices is a site's privacy policy. Some notice and awareness goals directly refer to the privacy policy itself. One can argue that the over-reliance on a privacy policy for such notifications places the burden and responsibility for notice and awareness on the consumer.

Two opposing approaches are evident in ensuring consumers awareness of privacy policy changes. The first approach is illustrated by goal G_{103}: `NOTIFY customer of changes to privacy policy`, which obligates the site to notify its users of changes to its policy (e.g. by sending an email message to registered users). The second approach is illustrated by goal G_{104}: `POST changes to privacy policy on website`, which places the responsibility for learning of changes on the site's users, who presumably must revisit the site and read its policy carefully on a regular basis. We found this second approach to be more common than the first one. All notice/awareness goals, however, do not revolve around a websites' posted privacy policy.

The following six aspects of notice and awareness are recognized in [FTC98] as essential and have been incorporated into the privacy goal taxonomy:

identification of the entity collecting the data;
identification of the uses to which the data will be put;
identification of any potential recipients of the data;
nature of the data collected;
means by which data is collected (if not obvious);
whether the provision of the requested data is voluntary; and
steps taken by the data collector to ensure the confidentiality, integrity and quality of the data.

This list suggests the kinds of privacy requirements that Web-based applications should satisfy. In the examined pilot study, few goals related to the organization collecting the data's identity; the examined policies either did not address this issue at all or in a few cases simply noted that their sites contained links to other sites that collected Personally Identifiable Information (PII), such as name, address, phone number, etc. Several sites returned cookies to a domain name having no obvious connection with the organization to which the site appeared to belong. General information use is typically addressed, but some privacy policies state that data collected by the site will be distributed to entities other than the one collecting the information; these entities are usually unspecified "third parties" but may be described as "partner" or "member" sites.

Other policies provide some assurance that data will not be transferred elsewhere (e.g. G_{57}: `PREVENT selling/ renting customer lists`). Most health care privacy policies address the nature of the data to be collected presumably due to the fact that these sites handle sensitive information concerning health care records. For example, medical prescriptions and diagnoses as in goal G_{63} (`LIMIT disclosure of prescription information/PII to patient or authorized representative`). The FIP principles "voluntary provision" category overlaps with the "Choice/Consent" principle; the taxonomy introduced in this paper classifies

Table 1: Privacy Protection Goal Taxonomy Goal Classifications

Protection Goal Taxonomy	Protection Goal Sub-Classifications
Notice/Awareness Goals asserting that consumers should be notified and/or made aware of an organization's information practices before any information is actually collected from them (e.g., an organization's privacy policy).	• General Notice/Awareness • Identification of the Uses to Which the Data Will be Put • Identification of Any Potential Recipients of the Data • 3rd Party Limitations • Nature of the Data Collected • Steps Taken by the Data Collector to Ensure the Confidentiality, Integrity, & Quality of the Data
Choice/Consent Goals ensuring that consumers are given the option to decide what personal information collected about them is to be used and whether it may be used for secondary purposes.	• Choice of How Data is Used • Choice of Sharing Data • Choice of What Data is Taken/Stored
Access/Participation Goals allowing or restricting access to a particular site or functionality based on whether or not the consumer provides their PII. Goals in this category address also the ability for consumers to access or correct any personally identifiable information about themselves.	• PII Provision Required • PII Provision Optional • Providing consumer access to data
Integrity/Security Goals ensuring that data is both accurate and secure. Security and accuracy comes from both the consumer and the organization collecting the PII. Goals in this category range from vague statements stating only that PII is kept securely to specific technical goals of what security protocols will be used to transfer PII over the Internet.	• Mission Statement • User-Supplied Integrity Goals • Using Anonymous PII • Destroying Untimely or Sensitive Data • Managerial Measures to Protect Against Loss and the Unauthorized Access, Destruction, Use, or Disclosure of the Data • Technical Measures to Protect Against Loss and the Unauthorized Access, Destruction, Use, or Disclosure of the Data
Enforcement/Redress Goals addressing the mechanisms that are in place to enforce privacy, otherwise a policy is merely suggestive, rather than prescriptive. Prescribe a way of working and general guidelines companies should follow. These include both self-imposed and government imposed work restrictions.	• Operational Prevention Assurance • 3rd Party Prevention Assurance • Failure of Assurance

all goals pertaining to voluntary information provision as Choice/Consent goals. The last aspect of notice and awareness concerns ensuring confidentiality, integrity and data quality; this is typically expressed by goals that impose mechanisms to ensure that consumer data and information is kept confidential and secret.

3.1.2 Choice and Consent

The choice and consent principle ensures that consumers are given the option to decide what personal information collected about them is used and whether it may be used for secondary purposes. Personal information collection in itself can be a privacy invasion, one over which consumers should have some control. Choice and consent goals are typically identified by focusing on key words, such as `OPT-IN` and `OPT-OUT`. Examples include: G_{14}: `OPT-IN to receive information and promotions` and G_{16}: `OPT-OUT from new use of PII in future`.

3.1.3 Access/Participation

The principle of access and participation asserts that consumers are able to access, correct and challenge any data about themselves; it refers to providing a means for consumers to ensure that their data is accurate and complete. Access must encompass timely and inexpensive access to data; a simple means for contesting inaccurate or incomplete data; a mechanism by which the data collector can verify the information; and the means by which corrections and/or consumer objections can be added to the data file and sent to all data recipients. The goal G_1: `ALLOW customer to modify/remove their PII`, which concerns removing information about an individual from a company's databases, is an access/participation goal.

3.1.4 Integrity/Security

The integrity and security principle addresses the practice of ensuring that data is both accurate and secure. The following integrity/security practices are recognized in [FTC98] as essential: providing consumer access to data; destroying untimely data or converting it to anonymous form; managerial measures to protect against loss and the unauthorized access, destruction, use, or disclosure of the data; and technical measures to protect against loss and the unauthorized access, destruction, use, or disclosure of the data.

The principle of providing consumer access to data overlaps with "Access/Participation"; as previously mentioned, access/participation goals address the ability for consumers to access or correct any PII about themselves. Therefore, the goal taxonomy does not classify the provision of consumer access to data as an integrity/security goal. Instead, this goal subclass focuses on protecting sensitive data via managerial or technical measures. Managerial measures address organizational procedures that limit access to data and ensure that those individuals with access do not utilize the data for unauthorized purposes. Goal G_{63}: `LIMIT disclosure of prescription information / PII to patient or authorized representative (prescribing physician)` and goal G_{80}: `DISALLOW access to PII by non-affiliated persons` address managerial measures. Technical measures to prevent unauthorized access include encryption in the transmission and storage of data (e.g. G_{60}: `PROTECT order information using SSL encryption technology`); limits on access through password use (e.g. G_{61}: `USE password for customer accounts`); and data storage on secure servers or computers (e.g. G_{114}: `STORE credit card info securely (encrypted, separate DB)`).

3.1.5 Enforcement/Redress

There must be a mechanism in place to enforce privacy, otherwise a policy is merely suggestive, rather than prescriptive. Although the FIP principles list three specific types of enforcement and redress (self-regulation, private remedies, and government enforcement), the examined privacy policies did not address each individually. Goals pertaining to self-regulation and private remedies are more common than those addressing government enforcement. Goal G_{50}: REQUIRE employees to comply with company privacy policy addresses self-regulation whereas goal G_{44}: DISCIPLINE employee who violates privacy policy exemplifies private remedies taken by a company to enforce their privacy policy.

3.2 Privacy Vulnerability Goals

Vulnerability goals are classified according to the manner in which they violate consumer privacy (see Table 2). There are several kinds of insidious privacy invasions: monitoring, aggregation, storage, and information transfer. Some may argue that if a consumer opts in to being monitored the following practices cannot possibly be insidious: having ones usage patterns or other data aggregated with that of other consumers or having one's PII stored in a database and/or shared with third parties. However, in reality, most consumers are oblivious to these practices. Moreover, such information collection presents the potential for grievous privacy invasions simply due to the vulnerability presented by its existence and consequently the temptation for abuses. Obvious privacy invasions are those that consumers are acutely aware of or of which they eventually become aware. The three kinds of obvious privacy invasions are: direct collection for secondary purposes, personalization, and solicitation. Benign privacy invasions are those for which access and PII use is beneficial to the consumer (e.g. access of/to information and information collection for some positive outcome or goal achievement).

When studying the processes involved with the vulnerability goal classes (e.g. information collection, monitoring, and aggregation) there may appear significant overlap. However, looking at each goal individually, there exists a strong differentiation between these three types of goals. It was important to be able to distinguish between these types of goals since our analysis involved examining individual goals.

Consider the following scenario: a website uses cookies to monitor a customer, then collects the customer's PII and finally aggregates the data with a larger pool of data. It is reasonable to think the corresponding goals would overlap, but in reality they are individual goals that can be categorized independently. However, the strong dependencies that exist between these goals cloud this distinction. We thus distinguish vulnerability goals as avoidance goals because these refer to things the consumer would want to avoid. A protection goal is something we want to achieve whereas a vulnerability goal is something we want the website to avoid because it denotes the existence of threats to one's privacy.

3.2.1 Information Monitoring

Information monitoring goals refer to information that organizations track when consumers visit their website. Sometimes such tracking may benefit the consumer; as when an e-commerce site maintains a shopping cart for customer purchases (information collection). Alternatively, tracking may benefit the organization when used for statistical analysis or profit (e.g. by selling aggregated information to third parties). Goal G_{25} (COLLECT date and times at which site was accessed) seems innocuous, unless someone who surfs the Web at 3 A.M. begins to receive advertisements for insomnia cures, indicating the existence of a privacy vulnerability.

3.2.2 Information Aggregation

Aggregation combines previously gathered PII data with data from other sources. Aggregation goals appear to be more prevalent in e-commerce privacy policies than in health care privacy policies. The goals pertaining to aggregated information in the examined health care privacy policies were classified as another goal type (e.g. information transfer or information collection). In contrast, e-commerce websites commonly aggregate information for various purposes, including targeted marketing (e.g. AGGREGATE purchase information by zip code) and statistical analysis of website usage (e.g. AGGREGATE statistics about user browsing patterns). This suggests that goals may be somewhat domain-specific. Although aggregation goals are included in the taxonomy, this does not imply that every privacy policy must include information aggregation goals.

3.2.3 Information Storage

Information storage goals address how and what records are stored in an organization's database. The two main reasons for information storage are: consumer use and corporate use. Storage for consumer use is intended to ease, for example, purchase transactions for the user (e.g. STORE purchase records). In contrast, storage goals pertaining to corporate use tend to operationalize and/or instantiate business rules (e.g. STORE credit card information until dispute is resolved).

3.2.4 Information Transfer

Privacy by definition implies ensurance that others cannot find something out. This wholly incorporates the idea that some information must not be transferred. These goals address the practice of allowing information to be transmitted, the reason(s) why information may be transferred, and to whom that information is transferred. Information transfer goals are among the easiest to identify due to a standard set of keywords for their identification: DISCLOSE, SELL, SHARE, and PROVIDE. Goal G_{124}: DISCLOSE collected PII when required by law is representative of one information transfer practice and goal G_{129}: SHARE PII for offers/promotions justifies the reason for which information is being transferred.

Table 2: Privacy Vulnerability Goal Taxonomy Classifications

Vulnerability Goal Taxonomy	Vulnerability Goal Sub- Classifications
Information Monitoring Goals concerning what organizations may track what consumers do on their site through means such as cookies. This could be for the consumer's benefit, like when an electronic-commerce application maintains a shopping cart for a consumer, or for the organization's benefit, be it for purely statistical use or for profit (via selling of aggregated information to 3rd parties).	• Monitoring for Services • Monitoring for Statistics • Limitation of Monitoring
Information Aggregation Aggregation combines previously gathered PII data with data from other sources.	N/A
Information Storage Goals addressing how and what records are stored in an organization's database. These goals cover a broad range, from security to monitoring and basically storage-specific.	• Storage for Customer Use • Storage for Corporate Use
Information Transfer Goals concerning any transfer of information. Privacy by its very definition means an insurance that others can not find something out. This wholly incorporates the idea that information must not be transferred. These goals address safeguards against the transfer of information, as well as to whom what information is transferred.	• Sharing PII with users • Sharing/Selling with Other Companies/Sites • Limitation of Sharing
Information Collection Goals addressing how and what information is being collected. Collection occurs when an organization collects information from a consumer either by directly requesting that they enter information, or by collecting information without their consent, such as browser information.	• Direct Collection (e.g. user provided information) • Indirect Collection (e.g. browsing patterns)
Information Personalization Goals addressing personalization as when consumers either change their PII, or when cookies are used to customize, thus affecting the functionality or content offered to them.	• Personalization by User Preference • Personalization of Site and Service • Personalization of Advertising, Offers, and Promotions
Contact These goals deal with how and for what purpose organizations contact consumers using their PII. This could be helpful, such as contacting customers to validate an email address, or annoying, such as sending out unwanted promotions based on past patterns.	• Contact for Promotions and Offers • Contact for Security and Verification • Contact Based on Preference

3.2.5 Information Collection

Information collection goals address what information websites collect. Information collection is characterized as either direct or indirect. *Direct collection* occurs when an organization directly requests visitors to enter information about themselves in a form, for example (e.g. G_{37}: COLLECT credit card information for billing/collect payment for services exemplifies a direct collection goal). *Indirect collection* occurs when a website collects information without the site visitor's consent (e.g G_{22}: ALLOW 3rd parties to collect browsing and usage patterns information and G_{32}: COLLECT browser type).

3.2.6 Information Personalization

Information personalization goals address website tailoring or customization to a specific visitor, thus affecting the functionality or content offered to individual visitors. Personalization may be as simple as greeting the site visitor by name (e.g. "Welcome, George.") as suggested by goal G_{107} (RECOGNIZE repeat customers using cookies) or may be more elaborate as in goal G_{110} (CUSTOMIZE content to specific customer using demographic / profile data), which may serve to personalize the site for targeted marketing.

3.2.7 Contact

Contact goals address how and for what purpose organizations contact visitors or others. Such contact may be helpful, as when customers are contacted to validate an email address. However, sometimes contact is perceived as annoying, such as the practice of sending out unwanted promotions based upon visitors' browsing patterns. Consider goals G_{38} (ALLOW affiliates to use PII for marketing/promotional purposes) and G_{41} (SEND email to customer); both these goals exemplify ways in which site consumers and visitors may be contacted.

4 The Goal Mining Process

The process of identifying high-level goals is fundamental to the requirements analysis and specification process. *Goal mining* refers to extracting goals from data sources (in this case, privacy policies) by applying goal-based requirements analysis techniques [Ant97]. The extracted goals are expressed in structured natural language (examples were provided in Section 3).

Our pilot study led us to develop our privacy goal taxonomy and enabled us to codify a comprehensive set of goal-mining heuristics tailored to privacy policy analysis. The goal-mining heuristics and privacy goal taxonomy were validated via their application in the examination of three kinds of health care website privacy policies. Goals were extracted from 23 privacy policies: 6 pharmaceutical companies, 7 health insurance companies, and 10 online pharmacies. Table 3 lists the analyzed health care websites privacy policies. The goals were analyzed according to different characteristics such as protection vs. vulnerability and subject matter (e.g. cookies, PII, browsing patterns, etc) as shown in Table 4.

Three main activities comprise the goal-mining process: goal identification, classification and refinement. Analysts first explore any available information sources such as existing security and privacy policies, or requirements specifications and design documentation, to identify both strategic and tactical goals. Goals are documented and then classified according to goal class (privacy protection or privacy vulnerability) as well as according to keyword and subject (e.g. browsing patterns, personalization, cookies, etc. in Table 4). Goal refinement entails removing synonymous and redundant goals, resolving any inconsistencies within the goal set, and operationalizing the goals into a requirements specification. The heuristics to guide the goal-mining

Table 3: Number of Privacy Protection and Vulnerability Goals in 23 Health Care Privacy Policies

	Company Name	Number of Protection Goals	Number of Vulnerability Goals
Health Insurance	AETNA	5	5
	AFLAC	1	1
	BCBS	13	7
	CIGNA	6	5
	EHealthInsurance	7	8
	Kaiser Permanente	4	1
	OnlineHealthPlan	8	9
Online Drugstore	CornerDrugstore	15	9
	DestinationRX	16	18
	Drugstore	15	14
	Eckerd	9	6
	HealthAllies	11	6
	HealthCentral	13	12
	IVillage	21	18
	PrescriptionOnline	9	4
	PrescriptionsByMail	11	7
	WebRX	18	7
Pharmaceutical	Bayer	8	9
	Glaxo Wellcome	5	7
	Lilly (Eli)	2	5
	Novartis (Ciba)	18	5
	Pfizer	4	3
	Pharmacia & Upjohn	10	8

process are detailed in [AE01b]; they are broadly applicable and not simply relevant for privacy and/or security policy analysis.

Table 4 summarizes the subject matter analysis. The 13 subject matters studied are listed in the table's left most column. This part of our analysis is clearly domain specific; for example, PII/HI refers to Personally Identifiable Information and Health Information (as in medical records concerning one's prescription medication, etc.). However, it is useful to reason about a particular policy's subject matter since one would clearly not expect certain subjects to appear in every Internet privacy policy. We observed both privacy protection and vulnerability goals within each subject matter category. Table 4 details additional data about the identified goals, according to subject matter, such as the number of functional, operational, synonymous, redundant and final goals. It also lists the number of health care privacy goals deemed synonymous or redundant. During goal refinement the *Browsing Patterns/Site Usage, PII/HI Usage, Contacting Customer and Security/Access* goal subjects enjoyed the greatest reduction rate. Merged goals are represented by the number that appears within parentheses, following the number of synonymous goals. The "Total" and "% Reduction" columns characterize the goal set's evolution, showing growth and refinement throughout the goal-mining process. Our raw data initially contained 263 goals, mined from the 23 privacy policies; upon goal refinement completion, the goal set had been reduced to 131 goals. Some goals were not truly relevant to privacy or privacy-related functionality. These goals were classified as either functional (meaning they support some system features) or operational (these goals represent business rules or operational procedures). The goal `<AGGREGATE survey results>` exemplifies a functional goal; the goal `<REVIEW web security weekly>` exemplifies an operational goal.

5 Observations and Discussion

This study had several objectives, to: (1) create a taxonomy for classifying privacy goals and requirements; (2) develop a set of reusable privacy and security goals for e-commerce software developers; and (3) use those goals to analyze and compare Internet privacy policies. Comparing privacy policies using goals is an innovative and effective analysis method that enables us to provide useful guidance to practitioners, policy makers and consumers.

Privacy policies are difficult to compare without considering the domain, business, and system requirements. A site that supports e-commerce transactions will obviously require more policy statements that focus on PII related privacy. The subject matter goals one expects to see in these site's policies include credit card information, PII, information transfer and storage. In contrast, sites whose primary mission is information dissemination with few transactions have little or no need to address credit card information usage. It is also why goals and the goal taxonomy presented in Section 3 provide such an effective unit for measuring and comparing these policies.

It is challenging for sites to completely disclose their information practices by simply addressing the FIPs, and policies tend to contain both privacy protection and vulnerability goals (see Table 3). This is noteworthy because recommendations for creating privacy policies in the past have suggested addressing the FIPs as a comprehensive strategy for policy creation [FTC98]. Prior to our analysis, we hypothesized that the number of protection goals in a health care privacy policy is greater than the number of vulnerability goals for that policy; this hypothesis was confirmed. When comparing the number of protection and vulnerability goals for each website, the t-test analysis revealed a statistically significant difference (p=0.0089) between them. In other words, the number of protection goals for a given website was observed to be, on average, greater than the number of vulnerability goals in that website. This was the case in 15 health care website privacy policies. It is interesting to note, however, that in 17 health care privacy policies we observed the number of protection goals to be equal to or greater than the number of vulnerability goals in a given website; and in 6 privacy policies there were more vulnerability than protection goals. Eli Lilly's privacy policy stated two protection goals and five vulnerability goals (interesting given recent events). Perhaps more alarming is the existence of 18 vulnerability goals in iVillage's policy. This finding is noteworthy for

Table 4: Subject Matter Goal Classes

Subject Matter	Total	Functional	Operational	Synonymous	Redundant	Final	% Reduction
Cookies/Bugs	14	7			1	7	50
Browsing Patterns/Site Usage	16			8 (1)		6	62.5
IP Address	4			1		3	25
Aggregate Info	12	3		1 (1)		7	41.7
Information	18			1 (1)		15	17
PII/HI	49	1		8 (2)	10	26	47
PII/HI Usage	42	1		13 (6)	8	14	67
Credit Card Info	9			1 (1)	3	4	56
Policies/Procedures	29	5	6	3		15	48
Contacting Customer	14		1	1	6	5	64
OPT In/Out	10			1		9	10
Security / Access	33	3	1	13 (1)	3	12	64
Children	13		1	2	2	8	38
TOTAL	263	20	9	53 (13)	33	131	50.2

consumers who hope that a health care website would focus more on expressing how they protect their customers' personal information. Having an equal number of vulnerability and protection goals demonstrates that websites introduce risk to its customers. In contrast, websites with a greater number of protection goals demonstrate that they are making an effort to minimize risk. The ability to easily identify these vulnerabilities aids requirements engineers in ensuring that system requirements address all potential vulnerabilities.

While some companies may not recognize an obligation to reveal their privacy-related goals in their privacy policy (e.g. a website using third-party cookies without giving appropriate notice in their policy), the taxonomy and goal-mining exposes a general lack of awareness, lack of compliance and potential for liability. We recognize that a taxonomy based upon existing privacy policies may seem suspect since some websites may not conscientiously construct their privacy policies. However, companies are beginning to take privacy policies very seriously as evidenced by the Chief Privacy Officer hiring trend and we believe there is much to be learned from analyzing existing privacy policies. Specifically, we submit that: (1) websites need to be held accountable for what they state in their privacy policies; (2) privacy policy authors need a taxonomy to better understand that their site's privacy policy implies certain assurances of functionality in the respective website; and (3) requirements engineers need a taxonomy to better understand the kinds of requirements needed to satisfy governing policies and help consider the potential vulnerability scenarios a system must address.

The fact that requirements specifications are often incomplete also applies to privacy policies. A careful analysis of selected goals revealed that one privacy policy failed to include the goal `<ALLOW third parties to use cookies>` even though the respective website did in fact allow cookies to be sent to third parties. By setting our browser preferences to warn before accepting cookies, we were able to test those sites that specifically fail to include any mention of cookies sent back to third parties. Drugstore.com requires cookies to be enabled before a visitor may even view their home page; moreover, once cookies are enabled, the website sends cookies to third parties, yet this was not expressed in their privacy policy.

Privacy vulnerability goals signal potential privacy invasions. Some invasions are insidious or covert since they are not readily apparent to consumers, as when non-transient cookies are placed on a consumer's hard drive. This is especially true when the cookie provides no additional value or benefit to the consumer, such as with cookies that offer personalization or purchase history information. Alternatively, some privacy invasions are obvious in that the consumer is aware or becomes aware of the privacy invasion, such as when a consumer begins to receive email solicitations. Finally, some privacy invasions are benign; the consumer is a knowing and active contributor, facilitator, or participant in the exchange of PII. What one consumer considers a privacy invasion may be a valued feature or service to another consumer [AEP01]. This debate is outside this paper's scope; however, we have created a privacy values survey instrument to assess these value differences and create a privacy values baseline.

6 Summary and Future Work

In this paper, we present a preliminary attempt to structure the privacy policy domain with goal taxonomies. We introduce a taxonomy for classifying privacy goals; we describe our use of a RE technique, goal-mining, to examine privacy policies for system goals and requirements. While we emphasize privacy policy analysis in this paper, the techniques are generalizable to different software systems. Eventually, in software development, goals are operationalized into system requirements and checked for compliance with the respective policies. One of our objectives is to create a library of reusable security and privacy goals; and we are well on our way to achieving this objective. The availability of this library of privacy and security goals in the *SMaRT* (Scenario Management and Requirements Tool) [AAB99] will enable requirements engineers and analysts to build security and privacy into e-commerce applications early on rather than having to add it in afterwards due to oversight or external pressures.

Examining and comparing privacy policies using goals is an innovative and effective analysis method that enables us to provide useful guidance to RE practitioners, policy makers and consumers. Our preliminary analysis showed that several websites with these goals stated in the privacy policies do not actually comply with the goals, a subject of discussion for a future paper. It is important to stress that we have not made any bold claims about this taxonomy; it will not make a site secure based on the protection/vulnerability goals. Instead, we claim that the taxonomy helps analysts evaluate a site's trust by aiding in the examination of its policies, requirements and practices.

Our plans for future work include developing a privacy rating tool based on goal analysis and the values baseline that will be established using a privacy values survey instrument. We also plan to consider the relationship between the severity of specific goal instances and the occurrence of privacy protection and vulnerability goals.

Acknowledgements

This work was supported by NSF ITR Grant #0113792 and the CRA's Distributed Mentor Project. The authors thank Kevin Farmer, Hema Srikanth and Ha To as well as Thomas Alspaugh, Colin Potts, Richard Smith and Gene Spafford for discussions leading to the privacy protection and vulnerability goal classification.

References

[AAB99] T. Alspaugh, A.I. Antón, T. Barnes and B. Mott. An Integrated Scenario Management Strategy, *4th IEEE Int'l Symp. on Req'ts Eng.*, pp. 142-149, June 1999.

[Abb83] R.J. Abbot. Program Design by Informal English Descriptions. *Communications of the ACM,* 26(11):882-894, November 1983.

[ACD01] A.I. Antón, R.A. Carter, A. Dagnino, J.H. Dempster and D.F. Siege. Deriving Goals from a Use-Case Based Requirements Specification, *Requirements Engineering Journal*, Vol. 6, pp. 63-73, May 2001.

[AE01a] A.I. Antón and J.B. Earp. Strategies for Developing Policies and Requirements for Secure Electronic Commerce Systems. in *E-Commerce Security and Privacy*, ed. by A.K. Ghosh, Kluwer Academic Publishers, pp. 29-46, 2001.

[AE01b] A.I. Antón and J.B. Earp. *A Taxonomy for Web Site Privacy Requirements*, NCSU Technical Report TR-2001-14, 18 December 2001.

[AEP01] A.I. Antón, J.B. Earp, C. Potts and T.A. Alspaugh. The Role of Policy and Privacy Values in Requirements Engineering, *5th Int'l Symp. on Requirements Engineering,* pp. 138-145, Aug. 2001.

[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.

[Ant97] A. I. Antón. *Goal Identification and Refinement in the Specification of Software-Based Information Systems*, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA, 1997.

[AP98] A.I. Antón and C. Potts. The Use of Goals to Surface Requirements for Evolving Systems, *Int'l Conf. on Software Engineering,* pp. 157-166, April 1998.

[Ben99] P. Benessi TRUSTe: An Online Privacy Seal Program. *Communications of the ACM*. 42(2), pp.56 – 59. February 1999.

[Boo91] G. Booch. *Object-Oriented Design with Applications*. Benjamin Cummings. Redwood City, California, 1991.

[CRA99] L.F. Cranor, J. Reagle and M.S. Ackerman. Beyond Concern: Understanding Net Users' Attitudes About Online Privacy, *AT&T Labs-Research Technical Report TR99.4.3,* http://www.research.att.com/library/trs/TRs/99/99.4/99.43/report.htm, 1999.

[Cul99] M.J.Culnan, Georgetown Internet Privacy Policy Survey: Report to the FTC. Georgetown Univ., The McDonough School of Business, http://www.msb.edu/faculty/culnanm/gippshome.html, 1999.

[Dar97] Darr, Ethics in Health Services Management, Health Professions Press, Inc. Baltimore, MD., 1997.

[EB02] J.B. Earp and D.Baumer. Innovative Web Use to Learn about Consumer Behavior and Online Privacy. To Appear: *Communications of the ACM*, 2002.

[EP00] J.B. Earp and F. C. Payton. Dirty Laundry: Privacy Issues for IT Professionals, *IT Professional*, 2(2), pp. 51-54, March/April 2000.

[Epi00] *Pretty Poor Privacy: An Assessment of P3P and Internet Privacy* http://www.epic.org/reports/prettypoorprivacy.html, EPIC, 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.

[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.

[GHS00] J. Goldman, Z. Hudson and R.M. Smith. Privacy Report on the Privacy Policies and Practices of Health Web Sites, Sponsored by the California HealthCare Foundation, Jan. 2000.

[Lam01] A. van Lamsweerde. Goal-Oriented Requirements Engineering: A Guided Tour, *IEEE 5th Int'l Symp. on Requirements Engineering,* pp. 249-261, August 2001.

[RBP91] J. Rumbaugh, M. Blaha, W. Premerlani. F. Eddy and W. Lorensen. *Object-Modeling and Design,* Prentice Hall, New York, NY, 1991.

[RC97] J. Reagle and L. F. Cranor. The platform for Privacy Preferences. *Communications of the ACM*. 42(2), pp.48-55, February 1997.