A Taxonomy for Web Site Privacy Requirements

Annie I. Antón*, *Member, IEEE* and Julia B. Earp, *Member, IEEE*

***Abstract*--** Privacy has recently become a prominent issue in the context of electronic commerce Web sites. Increasingly, privacy policies posted on such Web sites are receiving considerable attention from the government and consumers. In this paper we present a taxonomy for Web site privacy requirements. We have used goal-mining, the extraction of 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 web sites and their manner of operation. These goals can also be used to reconstruct the implicit requirements met by the privacy policies. We present the results of our analysis of 23 Internet privacy policies for companies in three health care industries: pharmaceutical, health insurance and online drugstores.

***Index Terms*—Requirements engineering, privacy requirements, internet security and privacy, goal-driven requirements analysis.**

I. INTRODUCTION

Requirements engineering is the principled application of proven methods and tools to describe the behavior and constraints of a proposed system. As such, it arguably influences the outcome of a software project more than any other sub-discipline within software engineering [24]. There is a need to apply a systems engineering perspective in order to consider systems and their respective policy holistically [6]. One approach to policy and requirements specification [4] relies on the application of goal and scenario-driven requirements engineering methods for secure electronic commerce systems. This approach results in the specification of: privacy policies, security policies and the corresponding system requirements for these proposed systems; inspections ensure that the requirements are in compliance with relevant policies. In this paper we explain our application of these requirements engineering techniques to Internet privacy policy analysis. We also introduce a taxonomy of privacy goals which provides an effective

* (corresponding author) A. I. Antón is with the Computer Science Department, North Carolina State University, Raleigh, NC, 27695-7534 USA (e-mail: aianton@mindspring.com).

J. B. Earp is with the Business Management Department, North Carolina State University, Raleigh, NC, USA. (e-mail: Julia_Earp@ncsu.edu).

mechanism for analyzing and comparing privacy policies, system requirements and the functionality of the respective systems.

Health care privacy, as it pertains to organizational practices, holds profound implications as service delivery impacts human life, legality and social policy [18]. The transmission and dissemination of health care information in electronic form has raised numerous privacy concerns among both consumers and providers [21]. The evolving trend toward Internet supported health care services has resulted in increased information sharing among providers, pharmacies and insurers. Unfortunately, such information sharing often conflicts with consumers' desires to be shielded from unauthorized use of their personal information. According to two recent studies [5, 29], inconsistencies exist between privacy policies and the actual privacy practices of health-care related Web sites. Moreover, visitors to Web sites (health-care sites in particular) are not anonymous, even though they think they are [29]. Web sites are concerned about safeguarding consumers' privacy as evidenced by the increasing number of privacy policies posted on these sites. However, these privacy policies fall short of truly safeguarding consumers [5, 6, 29].

In order to identify the system requirements reflected in health-care privacy policies, we employed a technique, goal-mining (the extraction of pre-requirements goals from post-requirements text artifacts), to derive the privacy-related goals of various Internet health care Web sites. Our motivation is two-fold. First, we seek to develop a reusable corpus of privacy and security goals to facilitate requirements specification for electronic commerce software developers [1, 4]. Second, goals provide an excellent unit by which to objectively analyze and compare Internet privacy policies [5], providing useful guidance to practitioners during requirements analysis activities. The results of this kind of analysis are expected to provide additional benefits to policy makers and consumers by providing more objective criteria for evaluating a Web site's privacy practices.

This work was partially supported by the Computing Research Association's Distributed Mentor Project during the summers of 2000 and 2001 as well as NSF ITR Grant #0113792.

In this paper goal-driven requirements engineering [9, 10, 31] is employed in a nontraditional manner. A privacy goal taxonomy and associated goal-mining heuristics were developed during an initial analysis of traditional electronic commerce Web site privacy policies (see Table 1). The approach taken has concurrently led to the development of an integrative taxonomy and goal-mining method as well as the analysis of Internet privacy policies. The initial electronic commerce privacy policy study enabled the development of a systematic approach to privacy goal identification and refinement as well as the privacy goal taxonomy introduced in this paper. This approach was then validated in a second privacy policy analysis, as discussed in Sections III and IV. This second effort entailed the extraction of goals from 23 Web site privacy policies that span three health care industries: health insurance, online drugstores and pharmaceutical companies. The second privacy policy analysis effort enabled evaluation and refinement of the taxonomy and its associated goal-mining heuristics. Thus, the first goal-mining effort was formative, serving as the origin of the taxonomy and goal-mining heuristics. The second health care privacy policy goal-mining effort was summative. This distinction is key in that in the summative effort previously developed methods (and the taxonomy) were being validated, whereas the formative case involved the evolution of the taxonomy and heuristics simultaneously coupled with validation.

The process of applying goal-mining heuristics in conjunction with the guidance provided by the privacy goal taxonomy aid in analyzing and comparing Internet privacy policies. This analysis provides insights into the characteristics of privacy policy content as well as the software requirements for the corresponding Web-based applications. It should be noted that the health care privacy policies we examined (and which serve as the primary focus of this paper) were in force during the month of June 2001, when the second goal-mining effort was conducted. These policies and practices are expected to change, but such change is outside the purview of discussion in this paper. For the remainder of this paper, the “first” goal-mining study is referred to as the electronic commerce goal-mining study and the

"second" is referred to as the health care goal-mining study.

The protection of personal information, such as that managed by health care Web sites, is not an option but a necessity. Goal-mining and goal analysis are effective techniques for examining how Internet Web sites manage online customer data and how they convey these practices to their customers. The remainder of this paper is organized as follows. Section II provides an overview of the state of health care privacy policy, policy evaluation mechanisms and goal-based requirements analysis. Section III introduces the taxonomy of privacy goals and employs examples from electronic commerce and health care Web site privacy policies. Section IV codifies the specific heuristics that guide the goal-mining process, providing examples from our analysis of 23 health care Internet Web site privacy policies. Results of the analysis are presented in Section V. Finally, a summary and discussion of future work is provided in Section VI.

TABLE 1: ANALYZED E-COMMERCE PRIVACY POLICIES

eCommerce Industry	**Sites for which Privacy Polices Were Reviewed**
Auction Sites	Ebay Reverse Auction Sothebys
Drug Stores	Drugstore.com Eckerd Drugs Long Drugs
Grocery Stores	HomeGrocer Lowes Peapod
Internet Service Providers	AOL Earthlink Free Internet
Online Retailers	Amazon eNews ToySmart
Traditional Mail Order Catalog	Banana Republic Eddie Bauer JCrew REI
Travel Agencies	American Express Expedia Travelocity
Trust Services	BBBOnline TRUSTe Verisign

II. BACKGROUND AND RELATED WORK

This section provides an overview of the relevant work in health care privacy policy, existing privacy policy evaluation mechanisms and the role of requirements engineering in policy analysis.

A. Health Care Privacy Policy

A privacy policy is a comprehensive description of a Web site's information practices; it is located in an easily accessible place on the site [26, 27]. Every organization involved in online transactions has a responsibility to adopt and implement a policy for protecting the privacy of Personally Identifiable Information (PII). PII is information that can identify a particular individual, either by itself or in combination with other information (e.g. name, address, phone number, email address, IP address, browsing patterns, etc.). Since 1973, the Fair Information Practice (FIP) Principles [25] have served as the basis for evaluating and establishing the policies and rules that underlie most privacy laws and practices in the United States. Although organizations engaged in electronic transactions should disclose privacy policies that are based on fair information practices [25, 26, 27], several studies [17, 22] have found that Internet privacy disclosures do not always reflect fair information practices. As discussed in Section III, this contributes to the inability to classify all Web site privacy goals as simply protection goals. Instead, privacy goals are classified as either protection or vulnerability goals.

In 2000 over 17,000 different health care Web sites offered a wide range of products and services on the Internet [29] including: purchasing, provision of clinical information, professional interaction and personal health records. The privacy practices of health care related Web sites have recently received attention in the press as with the Pharmaceutical company Eli Lilly which inadvertently revealed the electronic mail addresses of patients with depression, bulimia or obsessive-compulsive disorder [40]. The company had been using their email messaging system for more than two years to routinely send reminders about taking Prozac medication or attending to related matters. Such messages were usually addressed only to individuals, until the erred message was sent to over 600 participants and included all of their e-mail addresses. The sensitive nature of these services and recent incidents such as this suggest a need for legislation regarding information exchange.

Although the Privacy Act of 1974 provides some protection for medical records that are held by federal agencies, it does not cover medical records held by private groups where most medical records are actually created and stored[1]. Moreover, numerous exceptions are contained in the act so that its overall protection is leaky at best. When introduced, the Privacy Act was heralded as a huge step forward, but it is now considered one of the most outdated privacy acts in the world [11].

The increasing utilization of the Internet for healthcare information exchange has initiated legal reform with regard to privacy protection of electronic medical records. The 1996 Health Information and Portability Accountability Act (HIPAA)[2] mandated that the U.S. Government Administration introduce regulations regarding the control of medical records. These regulations called for the inclusion of a provision for health information privacy. The Department of Health and Human Services (HHS) published the final Privacy Rule[3] that took effect on April 14, 2001, requiring health care providers and health plans to comply by April 14, 2003. The new regulations specify several procedures regarding disclosure of PII and should therefore be reflected in health care Web site privacy policies [13].

B. Privacy Policy Evaluation Mechanisms

A 1999 survey revealed that 87% of Internet users are concerned about threats to their privacy when online [16]. However, several studies have subsequently shown that Internet users are more inclined to trust a Web site if it posts a privacy policy [20, 29]. These findings appear to have produced positive results as most online companies now post some form of privacy policy on their Web site. The downside to this is that not all consumers can (or are willing to) take the time to read and understand these policies. Consequently, several privacy policy evaluation mechanisms have been introduced to assist online consumers. As discussed in this section, privacy policies are evaluated in a rather ad hoc and inconsistent manner. Current approaches include P3P [15, 36, 39] and various privacy seal programs [12].

---

[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).

1) Platform for Privacy Preference Project (P3P)

The World Wide Web Consortium is establishing the Platform for Privacy Preferences Project (P3P) [4] as an industry standard to provide an automated way for users to gain control of and manage the use of their personal information on Web sites they visit. P3P requires consumers to answer a set of standardized multiple-choice questions that address various aspects of Web site privacy policies. The sites implementing P3P possess a privacy policy in machine readable format and users of these sites may configure their browsers to automatically determine if a Web site's privacy policy reflects their personal needs for privacy. This is done by comparing the user's responses to the multiple choice questions with the statements in a P3P compliant policy.

A P3P privacy statement is specific about the data (e.g. user's name) or types of data (e.g. user's demographic data) being collected by the site. Types of data that can be specified in the statement fall into the following categories: physical contact information, online contact information, unique identifiers, purchase information, financial information, computer information, navigation and click-stream data, interactive data, demographic and socioeconomic data, content, state management mechanisms, political information, health information, preference data, location data, government-issued identifiers and other information. By informing consumers about the data being collected and allowing them to decide if it is acceptable to them, P3P addresses the first two elements of the FIPs principles, notice/awareness and choice/consent. It addresses a third FIP principle, enforcement/redress, by including a <remedies> element to specify possible remedies in case a privacy breach occurs. However, P3P does not in anyway support the access/participation or integrity/security FIP principles.

Additionally, a report by the Electronic Privacy Information Center (EPIC) [23], asserts that P3P fails to comply with baseline standards for privacy protection and that it is a complex and confusing protocol

[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/

that will hinder Internet users in protecting their privacy. As of September 2001, only forty-six sites were publicly listed as being compatible with the latest specification of P3P[5]. In addition to the slow adoption of P3P, the European Union has rejected P3P as a viable technical means for supporting their stringent privacy laws [23]. Although P3P provides a technical mechanism for ensuring that users can be informed about privacy policies before they release personal information, it does not provide a technical mechanism for ensuring sites act according to their policies. Moreover, there is little evidence to support the industry claim that P3P will improve user privacy, as it does nothing to assess compliance with the recommended Fair Information Practices.

2) Privacy Seals

Privacy seal organizations, such as TRUSTe[6], BBBonline[7] and WebTrust[8], in a sense complicate privacy policy since consumers, understandably, often trust indirect and abbreviated indicators of privacy protection rather than reading the full privacy policy.

The seal and guarantee provided by TRUSTe appear to be comforting to consumers [20]. However, most consumers are unfamiliar with what the TRUSTe privacy seal truly signifies. In reality, this privacy seal simply ensures that TRUSTe has reviewed the licensee's privacy policy for disclosure of the following uses of information by a Web site: 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. This is not particularly stringent and does not reflect a real commitment to consumer privacy, merely an openness about what degree of privacy is or is not supported. TRUSTe requires licensees to disclose their privacy practices and adhere to established privacy principles based on the fair information practices.

[5] http://www.w3.org/P3P/compliant_sites
[6] http://www.truste.com/
[7] http://www.bbbonline.com/

This is an admirable service and evidence exists that it has brought about the protection of consumer privacy in a very real way in the case of Toysmart.com [28]. However, consumers should be alarmed by the privacy policies of some Web sites displaying this supposed "commitment to customer privacy." As long as a Web site's privacy policy openly admits that customer information is sold, leased, etc., the site is eligible for a TRUSTe privacy seal. For example, some TRUSTe licensees track what Web page visitors were at prior to accessing their Web site, whereas other TRUSTe licensees sell or share their customer email lists with other companies, allowing these third parties to send email solicitations.

The BBB*OnLine* privacy seal is posted on Web sites for which the merchant has met all BBB*OnLine* Privacy Program requirements regarding notice, choice, access and security of PII collected online. These companies must post privacy policies stating what personal information is collected; how it will be used; choices consumers have in terms of use; and the policy must verify security measures taken to protect this information. These companies commit to abide by their posted privacy policies, and agree to a comprehensive independent verification by BBB*OnLine*. Similar to TRUSTe, consumers are given a false sense of security when they encounter a BBB*OnLine* seal since they do not realize that a Web site can display it regardless of whether or not a privacy policy truly protects consumer privacy.

The CPA WebTrust seal entails a more stringent privacy evaluation mechanism, as only licensed public accountants who complete special training are able to issue a WebTrust seal. Unlike other seal programs, WebTrust requires accountants to conduct an independent examination that carries the professional equivalency of a financial statement audit. A licensed CPA, Chartered Accountant, or equivalent will only award a seal to a site if it completely passes the examination. WebTrust meets the U.S. industry consensus standards for privacy established by the Online Privacy Alliance[9]. Unlike P3P, the program substantively meets the standards for the European Union and Canada as well. However, there are very

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

few Web sites that currently display the CPA WebTrust seal. This is attributed to the extremely high cost of a CPA WebTrust seal, especially since it is not mandatory and has not been proven to markedly boost site visits.

A more effective privacy evaluation mechanism would be a policy-rating tool that considers not only the presence of certain policy content, but the implications of the policy content in reference to how such practices affect consumer privacy. Enonymous.com hosted a now defunct rating system, *www.privacyratings.com* that reviewed and rated Web site privacy policies according to how the site used PII. The three specific criteria used for the ratings were (1) whether a site contacts visitors for purposes beyond the primary purpose of data collection; (2) whether a site shares, trades, or sells user data; and (3) whether sites conduct such use with explicit visitor permission. The enonymous rating system focused on the notice and choice offered to visitors about the use of their PII. These correspond to the first two principles of the FTC's five FIP principles but do not include security, access/participation, enforcement/redress or other factors such as cookie use. Enonymous.com was dissolved in 2000 due to financial reasons. Given the growing concern for online personal privacy, however, it is evident that the public is in need of a more effective privacy evaluation mechanism. Requirements engineering provides reliable and straightforward mechanisms for evaluating privacy as discussed in the following subsection.

P3P and privacy seals only compare privacy policies based upon what content is provided from their "short-lists". In contrast, goals are a much richer unit for evaluation, enabling one to compare policies based upon not only content, but upon whether or not they actually protect private information and to what degree. A recent study [29] that analyzed the privacy practices of health care Web sites focused on comparing protection practices to the FIPs, not the vulnerability practices. This highlights a major strength and difference between the approach advocated in this paper and existing approaches; that is, the

[9] The Online Privacy Alliance is a group of corporations and associations having a common goal of protecting individual privacy online (http://www.privacyalliance.org/)

evaluation of privacy vulnerabilities in addition to protections.

C. Policy from the Requirements Engineering Perspective

Although researchers in the requirements engineering community are beginning to focus on electronic commerce applications [3, 10, 37] there remains a need to apply proven requirements analysis methods and demonstrate how to best apply these methods within the context of establishing and analyzing policy. Goal analysis has been successfully applied within the context of evolving electronic commerce systems [10] as we now discuss.

1) Goals

Goals are the objectives and targets of achievement for a system. In requirements engineering, goal-driven approaches focus on why systems are constructed, expressing the rationale and justification for the proposed system [31]. Since goals are evolutionary, they provide a common language for analysts and stakeholders. 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. Furthermore, since goals are typically more stable than requirements [9], they are a beneficial source for requirements derivation. Goals are operationalized and refined into requirements and point to new, previously unconsidered scenarios [7, 10, 30, 33, 38].

2) Goal-Based Requirements Engineering

The Goal-Based Requirements Analysis Method (GBRAM) [3, 8, 9, 10] is a straightforward methodical approach to identify system and enterprise goals and 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 and techniques through the inclusion of a set of heuristics, guidelines and recurring question types. Five sets of heuristics are included: identification heuristics, classification heuristics, refinement heuristics, elaboration heuristics and conflict identification/resolution heuristics. The heuristics are useful

for identifying and analyzing specified goals and scenarios as well as for refining these goals and scenarios. The GBRAM heuristics and supporting inquiry include references to appropriate construction of scenarios and the process by which they should be discussed and analyzed. The method has been successfully applied to the analysis of electronic commerce applications [10, 13]. In this paper, we describe our use of the method to mine privacy policies for system goals and requirements and codify the domain specific heuristics for applying the GBRAM for goal-mining Internet privacy policies. In the following sections we introduce our privacy goal taxonomy and describe the goal-mining process.

III. TAXONOMY OF PRIVACY GOALS

During the summer of 2000, we engaged in the first of two *goal-mining* exercises in which we evaluated 24 Internet privacy policies from 8 non-regulated electronic commerce industries (see Table 1). The identified goals are useful for discovering implicit internal conflicts within privacy policies and conflicts with the corresponding Web sites and their manner of operation. Additionally, these goals can be used to reconstruct the implicit requirements met by the privacy policies and to reason about expected privacy policy content for different types of Web sites (e.g online drugstores, pharmaceuticals and health insurance). This information can assist software developers in specifying requirements that address common Web site privacy goals.

Privacy experts who had viewed data from the electronic commerce goal-mining study initially suggested that all goals expressed in a Web site's privacy policy should support the Code for Fair Information Practices [25]. However, the goals derived from 24 Internet electronic commerce privacy policies proved challenging to classify in this simple manner. We attempted to classify the derived privacy goals according to the Fair Information Practices (notice / awareness; choice / consent; access / participation; integrity / security; and enforcement / redress). Unfortunately, it became abundantly clear that it was impossible to "force-fit" all the derived goals into the five FIPs. We thus analyzed the

remaining unclassified goals to determine what was different about those goals than the goals that supported the fair information practices. Careful examination revealed that the remaining goals did not exemplify privacy protection practices; instead, they reflected practices that introduce vulnerabilities into a site's ability to protect personal information. This led to the creation of a 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 privacy policies.

In the privacy goal taxonomy, privacy goals are broadly classified as either privacy protection or privacy vulnerability goals. *Privacy protection goals* are those that relate to the five FIP principles and which express the desired protection of consumer privacy rights. *Privacy vulnerability goals* are those related to existing threats to consumer privacy and represent statements of fact or existing behavior that may be characterized as privacy invasions. The five kinds of *privacy protection goals* are defined in Table 2 and the five kinds of *privacy vulnerability goals* are defined in Table 3. The following sub-sections provide concrete examples of privacy protection goals and privacy vulnerability goals. Eventually, in software development, these goals are operationalized into system requirements and checked for compliance with the respective policies [4]. Our preliminary analysis showed that the practices of several Web sites do not actually comply with the goals extracted from their privacy policies, as discussed in Section V.

A. Privacy Protection Goals

Since privacy protection goals suggest those properties to be satisfied in a system, the protection goals are subdivided into the categories of the Fair Information Practice Principles [26]. This provides an effective framework for privacy protection goals as discussed below.

TABLE 2: 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

1) Notice and Awareness

The principle of notice and awareness 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 the organization's privacy policy. In the electronic commerce goal-mining study, a number of the notice and awareness goals directly referred 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 that consumers are aware of changes to a privacy policy. The first approach is illustrated by the goal, G_{103}: `NOTIFY customer of changes to privacy policy`, which obligates the electronic commerce company to notify the Web site's users of changes to the policy; for example by sending an email message to all registered users. The second approach is illustrated by the goal, G_{104}: `POST changes to privacy policy on web site`, which places the responsibility for learning of changes on the users of the Web site, who presumably must revisit the site and read its policy carefully and 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 Web sites' posted privacy policy. The following six aspects of notice and awareness are recognized in [26] 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 is suggestive of the kinds of privacy requirements that Web-based applications should satisfy. In the examined electronic commerce privacy policies, few goals related to the identity of the organization collecting the data; the examined privacy 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 PII. Several sites returned cookies to a domain name having no obvious connection with the organization to which the site appeared to belong. The general use of information is typically addressed. 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 sometimes are described as "partner" or "member" sites. Other policies provide some form of 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 the case of goal G_{63} (`LIMIT disclosure of prescription information/PII to patient or authorized representative`).

The "voluntary provision" category of the Fair Information Practice Principles overlaps with the "Choice/Consent" principle; the taxonomy introduced in this paper classifies all goals pertaining to

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

2) Choice and Consent

The principle of choice and consent ensures 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. The collection of personal information in itself can be an invasion of privacy, 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 of choice/consent goals include: G_{14}: `OPT-IN to receive information and promotions` and G_{16}: `OPT-OUT from new use of PII in future`.

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 the removal of information about an individual from a company's databases, is an example of an access/participation goal and is incorporated into the privacy goal taxonomy.

4) Integrity/Security

The principle of integrity and security addresses the practice of ensuring that data is both accurate and secure. The following aspects of integrity/security are recognized in [26] 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 personally identifiable information about themselves. Therefore, the goal taxonomy does not classify the provision of consumer access to data as an integrity/security goal. Instead, the integrity/security 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` are examples of goals that 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 use of passwords (e.g. G_{61}: `USE password for customer accounts`); and the storage of data on secure servers or computers (e.g. G_{114}: `STORE credit card info securely (encrypted, separate DB)`).

5) Enforcement/Redress

There must be a mechanism in place to enforce privacy, otherwise a policy is merely suggestive, rather than prescriptive [26]. 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. However, 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` is an example of a self-regulation goal whereas goal G_{44}:

`DISCIPLINE employee who violates privacy policy` exemplifies private remedies taken by a company to enforce their privacy policy.

B. Privacy Vulnerability Goals

Vulnerability goals are classified according to the manner in which they violate consumer privacy. There are several kinds of insidious privacy invasions; monitoring, aggregation, storage, and transfer of information. 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 ones PII stored in a database and/or shared with third parties. However, in reality, most consumers are oblivious to these practices. Furthermore, the collection of such information presents the potential for grievous invasions of privacy simply due to the vulnerability presented by its existence and consequently the temptation for abuses.

TABLE 3: 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

Obvious privacy invasions are those that the consumer is acutely aware of or which they eventually become aware. Specifically, there exist three kinds of obvious privacy invasions: direct collection for

secondary purposes, personalization, and solicitation. This subsection provides examples from our goal-mining efforts to frame the discussion of vulnerability goals that represent privacy invasions. Benign privacy invasions are those for which access and use of PII is beneficial to the consumer; for example, access of/to information and collection of information for some positive outcome or goal achievement. Privacy vulnerability goals are classified according to Table 3 as discussed below.

1) Information Monitoring

Information monitoring goals refer to information that organizations may track when consumers visit their Web site. Sometimes such tracking may benefit the consumer; for example when an electronic commerce application maintains a shopping cart for customer purchases. Alternatively, tracking may benefit the organization, as is the case when used for statistical analysis or profit (e.g. via the selling of aggregated information to 3rd 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.

2) Information Aggregation

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

3) Information Storage

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

4) Information Transfer

Privacy by its very definition implies insurance that others cannot find something out. This wholly incorporates the idea that 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.

5) Information Collection

Information collection goals address what information is collected by Web sites. In the taxonomy, 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; the goal G_{37}*:* `COLLECT credit card information for billing/collect payment for services` is an example of a direct collection goal. Indirect collection occurs when a Web site collects information without the consent of visitors to their site (e.g G_{22}*:* `ALLOW 3rd parties to collect browsing and usage patterns information` and G_{32}*:* `COLLECT browser type`).

6) Information Personalization

Information personalization goals address the tailoring or customization of a Web site to a specific visitor, thus affecting the functionality or content offered to individual visitors. Personalization may be as simple as greeting the Web 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 the case of goal G_{110} (`CUSTOMIZE content to specific customer using demographic / profile data`), which may serve to personalize the site for purposes of targeted marketing.

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 of these goals exemplify ways in which customers or site visitors may be contacted.

IV. The Goal Mining Process

This section presents the goal-mining process and its associated heuristics within the context of our analysis of Internet health care privacy policies. The process of identifying high-level goals is fundamental to the requirements analysis and specification process. *Goal mining* refers to the extraction of goals from data sources (in this case, privacy policies) by the application of goal-based requirements analysis methods [9]. The extracted goals are expressed in structured natural language. Analysts begin the goal-mining process by first exploring any available information sources such as existing security and privacy policies, or requirements specifications and design documentation, to identify both strategic and tactical goals. Strategic goals are those that reflect high-level enterprise goals whereas tactical goals involve short-term goal achievement [4, 32]. These goals are documented and annotated with auxiliary information

including the responsible agents. Goals are then organized according to goal class (privacy protection or privacy vulnerability, as previously discussed) as well as according to keyword and subject (e.g. browsing patterns, personalization, cookies, etc. in Table 4).

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

Detailed techniques and heuristics for each of these operations are described in two theses [9, 19]. Once goals are identified, they are elaborated; goal elaboration entails analyzing each goal for the purpose of documenting goal obstacles, scenarios, constraints, pre-conditions, post-conditions, questions and rationale. Goal refinement consists of removing synonymous and redundant goals, resolving any inconsistencies that exist within the goal set, and operationalizing the goals into a requirements specification. The objective of this work to date has been to create a library of reusable security and privacy goals. This goal library will enable developers to operationalize those goals required by their respective systems. The availability of this library of privacy and security goals in the *SMaRT* (Scenario Management and Requirements Tool) [1] will enable requirements engineers and analysts to begin 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.

The electronic commerce goal-mining study led to the development of the privacy goal taxonomy

introduced in Section III and enabled us to codify a comprehensive set of goal-mining heuristics tailored to the analysis of privacy policies, as discussed in this section. The goal-mining process is comprised of three main activities: goal identification, classification and refinement. The heuristics to guide the goal-mining process are codified below. These heuristics are broadly applicable and are not simply relevant for analysis of privacy policies; they are useful for analyzing any documentation from which system requirements may be derived. For example, many software systems must enforce and/or comply with established security policies; goal-mining aids analysts throughout this requirements analysis process. This section provides a brief overview of some of the most useful heuristics, employing examples from the health care goal-mining study. Privacy policies for three health care sectors (see Table 5) were analyzed: health insurance, online drug stores and pharmaceuticals. The goals were analyzed according to different characteristics such as protection vs. vulnerability and subject matter (e.g. cookies, PII, browsing patterns, etc).

A. Heuristics for Identifying Goals

To identify goals, each statement in a privacy policy is analyzed by asking, *"What goal(s) does this statement or fragment exemplify?"* and/or *"What goal(s) does this statement obstruct or thwart?"* The identified goals are worded to express a state that is true, or the condition that holds true, when the goal is realized. Consider Privacy Policy #1 from the Blue Cross Blue Shield (BCBS) privacy policy:

> ***Privacy Policy #1:*** *Our cookies will never be used to track your activity on any third party web sites or to send spam, …*

By asking these goal identification questions, we identify the goals: G_{53}: `PREVENT cookies from tracking activity on other websites` and G_{54}: `PREVENT use of cookies to send spam`.

All action words are possible candidates for system goals. Goals in privacy policies may thus also be

identified by looking for useful keywords (verbs). This is an extension of previously supported techniques [2, 14, 35]. The following list of keywords are commonly found in most privacy policies: ADVISE, AGGREGATE, ALLOW, COLLECT, COMPLY, CUSTOMIZE, DISALLOW, DISCIPLINE, DISCLOSE, ENSURE, IMPROVE, KEEP, LIMIT, MAINTAIN, MONITOR, NOTIFY, OPT-IN, OPT-OUT, PREVENT, PROHIBIT, PROTECT, PROVIDE, RECOGNIZE, REMOVE, REPORT, REQUIRE, RETRIEVE, SELL, SEND, SHARE, STORE, TRACK, TRANSMIT, TRANSFER, and USE. To demonstrate the action word approach, consider the following statement from the Eckerd privacy policy:

> ***Privacy Policy #2:*** *Examples of information collected include the kind of web browser you used, the domain from which you connected to the Internet, the date and time you accessed the site, your computer's operating system, and the address of the website from which you connected to our site.*

The action word COLLECT appears in Privacy Policy #2. This action word serves as an indicator for several goals: G_{32}: COLLECT browser type, G_{33}: COLLECT domain name, G_{35}: COLLECT operating system, G_{25}: COLLECT date and time site was accessed, and G_{28}: COLLECT address of preceding website. Goals are thus also identified using inquiry-driven [34] and traditional action word location techniques.

Although not detailed in this paper, additional heuristics suggest synonymous words that may be expressed using one of the previously listed goal keywords. For example, consider Privacy Policy #3, taken from the Bayer privacy policy.

> ***Privacy Policy #3:*** *We use the information from cookies to provide services better tailored to our users' needs and we never save passwords or credit card information in cookies.*

In this privacy policy, the term "tailor" is clearly synonymous with "customize". Using the heuristics, which guide the identification and mapping of synonymous words to the list of acceptable keywords, we express the goal G_{112}: CUSTOMIZE experience at our site using cookies. This goal,

although expressed differently on different sites, appeared in 10 of the 23 analyzed health care privacy policies.

B. Heuristics for Classifying Goals

Classification of goals involves differentiating goals according to goal class (e.g. protection vs. vulnerability) and subject matter.

Protection goals are classified by analyzing each goal and asking, *"Does this goal protect one's private information?" and/or "Does this goal support one of the five fair information practices?"* Whereas, vulnerability goals are classified by considering each goal and asking *"Does this goal potentially compromise the privacy and/or security of one's private information?"* and/or *"Does this goal conflict with any of the five fair information practices?"* Consider Privacy Policy #1, which yielded the goal G_{54}: `PREVENT use of cookies to send spam`, this goal clearly seeks to protect one's privacy and is thus classified as a privacy protection goal. In contrast, the HealthCentral goal, G_{22}: `ALLOW 3rd parties to collect browsing and usage patterns information`, is a privacy vulnerability goal.

Table 4 provides an overview of the subject matter analysis. The 13 subject matters studied are listed in the left most column of the table. This part of the 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 the subject matter of a particular policy since one would clearly not expect certain subjects to appear in every Internet privacy policy. Both privacy protection and vulnerability goals were observed within each of the subject matter categories. This analysis is discussed in more detail in Section V. The table details additional data about the identified goals, according to subject matter, such as the number of functional, operational, synonymous, redundant and final goals; this refinement process is discussed below.

C. Goal Refinement Heuristics

Organization of goals entails eliminating redundancies and reconciling synonymous goals. Goals are considered synonymous if their intended states are equivalent or if they mean the same thing to different stakeholders who simply express the goal using different terminology. It is up to the analyst to identify these instances. For example, the goals <TRACK pages on our site using cookies> and <TRACK usage patterns using cookies> are synonymous and can be reconciled as one goal that encompasses the spirit and scope of both. The analyst can choose either of the two goal names; however, all related essential information must be maintained. In the case of these two goals, they were further merged with another goal: <TRACK usage patterns using aggregate data>. The previous two goals were merged with the latter as follows: G_{96}: TRACK usage patterns (using aggregate data or cookies). Thus, if the same goal appears more than once, all but one of the goals should be eliminated. In Table 4, merged goals are represented by the number that appears within parentheses, following the number of synonymous goals. Redundancies and synonymous goals are most easily identified after the goals have been organized according to subject matter. Table 4 shows the number of goals that were deemed synonymous or redundant in the analysis of health care privacy policies. When reducing the number of goals, the *Browsing Patterns/Site Usage, PII/HI Usage, Contacting Customer and Security/Access* goal subjects experienced the greatest reduction rate. This indicates a tendency for Web site privacy policies to over-explain these particular practices using redundant/synonymous goals or statements.

The "Total" and "% Reduction" columns in Table 4 characterize the evolution of the goal set, showing the growth and refinement of the goal set throughout the goal-mining process. The raw data initially contained 263 goals, mined from the 23 privacy policies; upon completion of the goal refinement activities, the goal set had been reduced to 171 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 functionality) or operational (these goals represent business rules or operational procedures). The goal <AGGREGATE survey results> is an example of a functional goal; the goal <REVIEW web security weekly> is an example of an operational goal.

Our privacy goal taxonomy and goal-mining heuristics were validated throughout the course of our Internet health care privacy policy analysis. The following section details our observations and findings.

V. Observations and Discussion

This study had several objectives, to: (1) create a taxonomy for classifying privacy goals for subsequent operationalization into system requirements; (2) develop a set of reusable privacy and security goals for electronic commerce software developers; and (3) use those goals to analyze and compare Internet privacy policies to reason about the corresponding requirements for these systems. This section provides an analysis of the goal data and other relevant observations.

A. Data Analysis of Protection and Vulnerability Goals

The five FIPs oblige Web sites to address notice/awareness, access/participation, choice/consent, integrity/security and enforcement/redress in their site privacy policies. However, as noted in Section III, it is challenging for Web sites to completely disclose their information-related practices (see definition of privacy policy in Section II) by simply addressing the FIPs. Therefore, during our goal-mining effort policies tended to contain both privacy protection and vulnerability goals (see Table 5). Prior to our data analysis, we set forth several tentative assumptions in order to draw out and test their logical or empirical consequences [5]. 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 as true. When comparing the number of protection goals to the number of vulnerability goals for each Web site, 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 Web site was observed to be, on average, greater than the number of vulnerability goals in that Web site. This was the case with 15 of the 23 examined health care Web site privacy policies.

TABLE 5: NUMBER OF PRIVACY PROTECTION AND PRIVACY VULNERABILITY GOALS IDENTIFIED IN 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

It is interesting to note that in 8 of the examined privacy policies we observed the number of protection goals for a given Web site to be equal to or fewer than the number of vulnerability goals in that Web site; for example, AETNA's privacy policy stated five vulnerability goals and five protection goals. This finding is noteworthy (and possibly even alarming) for consumers who hope that a health care Web site would focus more on expressing how they protect their customers' personal information, but that is not the case. Having an equal number of vulnerability goals demonstrates that Web sites continue to introduce risk to its customers. In contrast, Web sites with a greater number of protection goals demonstrate that they are making an effort to minimize risk to its customers.

B. Qualitative Observations

An examination of the protection / vulnerability goal types within the subject matter goal classes helped

surface apparently missing goals. In the sample, none of the sites had protection goals in the Browsing Patterns / Site Usage and IP Address categories. This implies that these sites do not deem it necessary to protect visitor browsing patterns or IP Addresses. Similarly, no Health Insurance or Pharmaceutical sites had any vulnerability goals pertaining to opting-in or opting-out. From this, one can infer that the opt-in and opt-out options at these kinds of sites are favorable to consumers and are expressed as protection goals.

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 Web site does in fact allow cookies to be sent to third parties. By setting Netscape Preferences to accept all cookies and warn before accepting a cookie, we tested those sites that specifically failed to include any mention of cookies sent to the third parties. Drugstore.com, for example, requires cookies to be enabled before a visitor may even view their home page; moreover, once cookies are enabled, this Web site sends cookies to third parties, and yet this was not expressed in their privacy policy.

Privacy vulnerability goals signal potential privacy invasions. Some invasions are insidious or covert in that they are not readily apparent to consumers, as is often the case when non-transient cookies are placed on a consumer's hard drive. This is especially true with cookie ads that provide no value or benefit to the consumer. 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 solicitations via email. Finally, some privacy invasions are benign; that is to say the consumer is a knowing and active contributor, facilitator, or participant in the exchange of PII. It should be noted that what one consumer considers a privacy invasion may be a valued feature or service to another consumer. This debate is outside the scope of this paper; however, we are currently creating a privacy values survey instrument to assess these value

differences and create a privacy values baseline.

VI. Summary and Future Work

Most Web sites today display privacy policies that describe the site's privacy related information practices. However, in spite of the many guidelines for the content and layout of these policies, privacy policy content inevitably differs from site to site. As one would expect, a site that supports electronic commerce transactions will obviously require more policy statements that focus on PII related privacy. The subject matter goals one expects to see in these sites' 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 the use of credit card information. This is one of the many reasons that privacy policies are so difficult to compare without consideration of the domain, business, and system requirements. It is also why goals and the goal taxonomy presented in Section III provide such an effective unit for measuring and comparing these policies, while reasoning about the respective system requirements.

The functionality of a company's Web site must reflect its privacy policy, else that policy is meaningless since the Web site implementation does not comply with the policies that govern it. A privacy policy describes the kinds of information collected by the Web site and the way that information is handled, stored, and used; if the Web site does not conform to its policy, the company may be subject to public outcry or legal action. In this paper, we have introduced a taxonomy for classifying privacy goals; we describe our use of a software requirements engineering technique, goal-mining, to examine privacy policies for system goals and requirements; and we codify domain specific heuristics for applying the GBRAM for goal-mining Internet privacy policies. While we emphasize privacy policy goal-mining in this paper, the techniques we have presented are generalizable to different software systems; for example, security goals may be observed in most multi-user systems. Examining and comparing privacy policies

using goals is an innovative and effective analysis method that provides useful guidance to software developers, policy makers and consumers.

Our plans for future work include the development of a privacy rating tool based on the goal analysis process and the values baseline that will be established using our, previously mentioned, privacy values survey instrument. In parallel, we plan to construct software agents that mine privacy policies for privacy and security goals as well as agents that determine the actual practices of a Web site. These software agents will consider not only the presence of certain policy content, but the implications of the policy content in reference to how practices affect consumer privacy and system requirements.

VII. Acknowledgment

The authors wish to thank Kevin Farmer, Angela Reese, Hema Srikanth and Ha To. Additionally, we thank Thomas Alspaugh, Colin Potts, Richard Smith and Gene Spafford for discussions leading to our classification of privacy protection and vulnerability goals.

VIII. References

[1] T. Alspaugh, A.I. Antón, T. Barnes and B. Mott. An Integrated Scenario Management Strategy, *IEEE Fourth International Symposium on Requirements Engineering (RE'99)*, University of Limerick, Ireland, pp. 142-149, 7-11 June 1999.

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

[3] 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*, Springer-Verlag, Volume 6, pp. 63-73, May 2001.

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

[5] A.I. Antón, J.B. Earp and A.Reese. Goal Mining to Examine Health Care Privacy Policies, Submitted to: *IEEE 2002 Symposium on Security and Privacy*, NCSU Technical Report, TR-2001-10, 6 November 2001.

[6] A.I. Antón, J.B. Earp, C. Potts and T.A. Alspaugh. The Role of Policy and Privacy Values in Requirements Engineering, *IEEE 5th International Symposium on Requirements Engineering (RE'01)*, Toronto, Canada, pp. 138-145, 27-31 August 2001.

[7] A.I. Antón, W.M. McCracken and C. Potts. Goal Decomposition and Scenario Analysis in Business Process Reengineering, *Advanced Information System Engineering: 6th International Conference, CAiSE '94 Proceedings*, Utrecht, The Netherlands, 6-10 June 1994, pp. 94-104, 1994.

[8] A.I. Antón. Goal-Based Requirements Analysis, *Second IEEE International Conference on Requirements Engineering (ICRE '96)*, Colorado Springs, Colorado, pp. 136-144, 15-18 April 1996.

[9] 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.

[10] A.I. Antón and C. Potts. The Use of Goals to Surface Requirements for Evolving Systems, *International Conference on Software Engineering (ICSE '98)*, Kyoto, Japan, pp. 157-166, 19-25 April 1998.

[11] Bartley L. Barefoot, Enacting a Heath Information Confidentiality Law: Can Congress Beat the Deadline?, 77 N.C.L. Rev. 283 (1998).

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

[13] D. Baumer, J.B. Earp and F.C. Payton. Privacy of Medical Records: IT Implications of HIPAA. *ACM Computers and Society*, 30(4), pp.40-47, December 2000.

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

[15] R. Clarke. Platform for Privacy Preferences: An Overview, *Privacy Law & Policy Reporter*, 5(2), pp. 35-39, July 1998.

[16] L.F. Cranor, J. Reagle and M.S. Ackerman. *Beyond Concern: Understanding Net Users' Attitudes About Online Privacy*, AT&T Labs-Research Technical Report TR 99.4.3, http://www.research.att.com/library/trs/TRs/99/99.4/99.43/report.htm, April 1999.

[17] M.J.Culnan, *Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission*. Washington, DC: Georgetown University, The McDonough School of Business, http://www.msb.edu/ faculty/culnanm/gippshome.html, 1999.

[18] Darr, *Ethics in Health Services Management*, Third Edition, Health Professions Press, Inc. Baltimore, MD, 1997.

[19] J.H. Dempster. *Conflict Identification*. M.S. Thesis, North Carolina State University, 2000.

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

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

[22] Surfer Beware III: Privacy Policies without Privacy Protection, December 1999 at http:www.epic.org/reports/surfer-beware3.html.

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

[24] W.J. Fabrycky and B.S. Blanchard. *Life Cycle Cost and Economic Analysis*, Prentice-Hall, 1991.

[25] *The Code of Fair Information Practices*, U.S. Department of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers, and the Rights of Citizens, viii, http://www.epic.org/privacy/consumer/ code_fair_info.html, 1973.

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

[27] Privacy Online: Fair Information Practices in the Electronic Marketplace. A Report to Congress. Federal Trade Commission, 2000.

[28] FTC Sues Failed Website, Toysmart.com, for Deceptively Offering for Sale Personal Information of Website Visitors. FTC File No. 002-3274. July 10, 2000.

[29] 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, January 2000.

[30] Jarke, M., X.T. Bui and J.M. Carroll. Scenario Management: An Interdisciplinary Approach *Requirements Engineering Journal*, Springer Verlag, 3(3-4), pp. 154-173, 1998.

[31] A. van Lamsweerde. Goal-Oriented Requirements Engineering: A Guided Tour, *IEEE 5th International Symposium on Requirements Engineering (RE'01)*, Toronto, Canada, pp. 249-261, 27-31 August 2001.

[32] Policy Framework for Interpreting Risk in eCommerce Security. CERIAS Technical Report, http://www.cerias.purdue.edu/ techreports/public/PFIRES.pdf, Purdue University, 1999.

[33] C. Potts. ScenIC: A Strategy for Inquiry-Driven Requirements Determination, Proceedings *IEEE 4th International Symposium on Requirements Engineering (RE'99)*, Limerick, Ireland, 7-11 June 1999.

[34] Potts, C., K. Takahashi, and A. Antón. Inquiry-Based Requirements Analysis, *IEEE Software*, 11(2), pp. 21-32, March 1994.

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

[36] J. Reagle and L. F. Cranor. The platform for Privacy Preferences. *Communications of the ACM*. 42(2), pp.48-55, February 1999.

[37] W.N. Robinson. Electronic brokering for assisted contracting of software applets, *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, Vol. 4 , pp. 449-458, 1997.

[38] C. Rolland, C. Souveyet and C.B. Achour. Guiding Goal Modeling Using Scenarios, *IEEE Transactions on Software Engineering*, 24(12), pp. 1055-1071, December 1998.

[39] P. Thibodeau. Companies Moving Slowly on P3P Adoption, *Computerworld*, 35(44), pp.17, October 29, 2001.

[40] C. Wilson. “Lilly reveals Prozac patients' identities.” http://www.infobeat.com/cgibin/WebObjects/IBFrontEnd.woa/wa/fullStory?article=409190643, 17 July 2001.