Requirements Eng (2004) 9: 169–185
DOI 10.1007/s00766-003-0183-z

ORIGINAL ARTICLE

Annie I. Antón · Julia B. Earp

A requirements taxonomy for reducing Web site privacy vulnerabilities

Received: 15 September 2002 / Accepted: 15 March 2003 / Published online: 5 December 2003

Abstract The increasing use of personal information on Web-based applications can result in unexpected disclosures. Consumers often have only the stated Web site policies as a guide to how their information is used, and thus on which to base their browsing and transaction decisions. However, each policy is different, and it is difficult—if not impossible—for the average user to compare and comprehend these policies. This paper presents a taxonomy of privacy requirements for Web sites. Using goal-mining, the extraction of pre-requirements goals from post-requirements text artefacts, we analysed an initial set of Internet privacy policies to develop the taxonomy. This taxonomy was then validated during a second goal extraction exercise, involving privacy policies from a range of health care related Web sites. This validation effort enabled further refinement to the taxonomy, culminating in two classes of privacy requirements: protection goals and vulnerabilities. Protection goals express the desired protection of consumer privacy rights, whereas vulnerabilities describe requirements that potentially threaten consumer privacy. The identified taxonomy categories are useful for analysing implicit internal conflicts within privacy policies, the corresponding Web sites, and their manner of operation. These categories can be used by Web site designers to reduce Web site privacy vulnerabilities and ensure that their stated and actual policies are consistent with each other. The same categories can be used by customers to evaluate and understand policies and their limitations. Additionally, the policies have potential use by third-party evaluators of site policies and conflicts.

A. I. Antón (✉)
College of Engineering, North Carolina State University,
Raleigh, NC 27695-8207, USA
E-mail: aianton@eos.ncsu.edu

J. B. Earp
College of Management, North Carolina State University,
Raleigh, NC 27695, USA

Keywords Privacy requirements · Security requirements

1 Introduction

A 1999 survey revealed that 87% of Internet users are concerned about threats to their privacy when online [1]. However, several studies have subsequently shown that Internet users are more inclined to trust a Web site if it posts a privacy policy [2, 3]. A privacy policy is a comprehensive description of a Web site's information practices, located in an easily accessible place on the site [4, 5]. These policies are typically the only information source available to consumers who are deciding whether or not to correspond with the Web site. There is a need to apply a systems engineering perspective so that Web sites and their respective policy may be treated holistically [6]. One approach to policy and requirements specification [7] relies on the application of goal and scenario-driven requirements engineering methods for secure e-commerce systems. In this paper we explain our application of these requirements engineering techniques to Internet privacy policy analysis. We also introduce a taxonomy of privacy protection goals and vulnerabilities, developed using a grounded theory approach (content analysis), that provides an effective mechanism for analysing and comparing privacy policies, system requirements and the functionality of the respective systems. The taxonomy categories can be used by Web site designers to ensure that their stated and actual policies are consistent with each other and that they reflect the values and beliefs of online consumers. The same categories can be used by customers to evaluate and understand policies and their limitations. The categories also have potential use by third-party evaluators of site policies and conflicts.

In this paper, we employ goal-driven requirements engineering [8, 9, 10, 11] in a non-traditional manner. Goals and requirements are typically extracted from traditional information sources, such as: existing requirements and design specifications; stakeholder interview transcripts; and legacy code. We employed a

content analysis technique, goal-mining (the extraction of pre-requirements goals from post-requirements text artefacts), to derive the privacy-related goals of various Internet health care Web sites. These goals enabled us to identify the system requirements reflected in health care privacy policies. Our goal-mining efforts were conducted in the spirit of grounded theory, in which existing phenomena is analysed to develop an understanding of the current state of a particular subject of interest. Grounded Theory is theory derived from data that has been systematically gathered and analysed [12]. Therefore, the work presented in this paper is not based upon a distinct preconceived theory or hypothesis that we hope to support or refute. Instead, our goal-mining effort was a scientific analysis to develop new theory. 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.

A privacy goal and vulnerabilities taxonomy and associated goal-mining heuristics were developed during an initial analysis of traditional e-commerce Web site privacy policies (see Table 1). The approach concurrently led to the development of an integrative taxonomy and goal-mining method as well as the analysis of Internet privacy policies. The initial e-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 Sects. 3 and 4. 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.

Table 1 Analysed e-commerce privacy policies (June 2000)

E-commerce 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 catalogs	Banana Republic
	Eddie Bauer
	JCrew
	REI
Travel agencies	American Express
	Expedia
	Travelocity
Trust services	BBBOnline
	TRUSTe
	Verisign

The process of applying goal-mining heuristics in conjunction with the privacy goal taxonomy aids in analysing and comparing Internet privacy policies. Such an analysis provides insights into the characteristics of privacy policy content (and specifically, the identification of vulnerabilities) as well as the software requirements for the corresponding Web-based applications. The health care privacy policies we examined (and which serve as the primary focus of this paper) were in force during June of 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 e-commerce goal-mining study and the "second" is referred to as the health care goal-mining study. 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 [13].

The remainder of this paper is organised as follows. Section 2 provides an overview of the state of health care privacy policy, policy evaluation mechanisms and goal-based requirements analysis. Section 3 introduces the taxonomy of privacy goals and vulnerabilities, employing examples from e-commerce and health care Web site privacy policies. Section 4 codifies the specific heuristics that guide the goal-mining process, providing examples from our analysis of 23 health care Internet Web site privacy policies, and presents the results of our analysis. Finally, a summary and discussion of future work is provided in Sect. 5.

2 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.

2.1 Health care privacy policy

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 unauthorised use of their personal information. According to two recent studies [3, 13], 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 [3]. 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 [3, 6, 13].

In the 1970 s, the United States Congress held hearings in which privacy advocates sought to outlaw the use of centralised computer databases by credit bureaus. These hearings lead to the recognition that organisations have certain responsibilities and individuals have certain rights in terms of information collection and use. And, since 1973, the Fair Information Practice (FIP) Principles [14] 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 organisations engaged in electronic transactions should disclose privacy policies that are based on the FIPs [4, 5, 14], several studies [15, 16] have found that Internet privacy policies do not always reflect fair information practices.

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. The increasing utilisation of electronic medical records for health care information management, in addition to online health care information exchange, has initiated legal reform in the United States (US) with regard to privacy protection of electronic medical records. The 1996 Health Information and Portability Accountability Act (HIPAA)[2] mandated that the US 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 14 April 2001, requiring health care providers and health plans to comply by 14 April 2003. The new regulations specify several procedures regarding disclosure of PII and should therefore be reflected in health care Web site privacy policies [17]. As we discuss in Sect. 4, health care Web sites are inconsistent in their treatment of PII.

[1] 5 USC 552a (1994)

[2] Health Insurance Portability and Accountability Act of 1996, 42 USCA. 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, (28 December 2000).

2.2 Privacy policy evaluation mechanisms

Consumer attitudes regarding online privacy [1, 2, 3] 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 limited manner. Current approaches include P3P [18] and various privacy seal programs [19]. Whereas these privacy policy mechanisms are implemented to benefit only consumers, the privacy taxonomy and goal-mining heuristics presented in this paper are additionally intended to assist software engineers in assessing implicit requirements met by privacy policies and identifying vulnerabilities that must be addressed in order to prevent their operationalisation into system requirements and eventual implementation.

2.2.1 Platform for privacy preference project (P3P)

The World Wide Web (WWW) 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 standardised multiple-choice questions via their Web browser 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 [20].

A P3P privacy statement specifies the data (e.g., user's name) or types of data (e.g., user's demographic data) being collected by the site, as well as the uses, recipients and retention of that data. Types of data that can be specified in the statement include: physical contact information, online contact information, unique identifiers, purchase information, financial information, computer information, navigation and click-stream data, interactive data, demographic and socio-economic data, content, state management mechanisms, political information, health information, preference data, location data, government-issued identifiers and other information.

A detailed discussion of how P3P is implemented is outside the scope of this paper. However, it is important to note that P3P does not simply consist of a set of multiple choice questions. P3P policies express the

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

privacy policy of the site that hosts the policy (server-side). The main goal of P3P is awareness, so that users may make decisions based upon the available information. P3P is limited in that no matter how well a site defines its policy in P3P, the mechanism relies on site owners' participation and honesty. Even if P3P policies are implemented and accurate, users may dismiss a warning because they typically lack an understanding of the effects of the warning upon their decision. Additionally, because most browsers do not support interactive privacy preference interaction, many possible privacy invasions are not thwarted. To support the client-side, the creators of P3P have introduced APPEL (A P3P Preference Exchange Language). APPEL allows users to specify pre-defined, yet customisable, privacy preference rules, termed rulesets, which are represented as XML documents [21]. The ruleset is then used to make automated or semi-automated decisions about whether a P3P enabled Web site's policies are acceptable to the user. A preliminary analysis reveals that P3P and APPEL currently support only one class of goals in our taxonomy: notice/awareness (introduced in Sect. 3), but this analysis is not the subject of this paper. However, we are currently mapping the taxonomy to P3P, developing needed extensions as we seek to ensure sites operate according to their policies.

A report by the Electronic Privacy Information Center (EPIC) [22], asserts that P3P fails to comply with baseline standards for privacy protection and that it is a complex and confusing protocol that will hinder Internet users in protecting their privacy. As of August 2002, 119 sites were publicly listed as being compliant with the 11 April 2002 specification of P3P (P3P 1.0).[5] However, a closer examination reveals that they do not really take full advantage of the entire specification. Whereas P3P may help standardise privacy notices, it still does not offer privacy protection [23]. P3P can support the growth of privacy choices, including anonymity and our efforts to map our privacy goal taxonomy to P3P seek to support such growth. P3P does have its limitations, however. It does not: protect the privacy of users in countries lacking privacy laws; have the ability to create public policy; nor can it demand that its specifications be followed in the marketplace [23]. More importantly, P3P cannot guarantee a company's Web site is operating according to the practices expressed in the company's privacy policy. The only way to penalise a company for failing to comply with its privacy policy is via law or membership in a self-regulatory body. Using the goal taxonomy presented in this paper, we seek to develop tools that enable software engineers to develop systems which comply with governing privacy policies. We are also investigating the use of run-time requirements monitoring [24, 25, 26] in conjunction with the taxonomy to support end-users seeking to determine whether the sites they visit are operating responsibly, according to their expressed privacy policies.

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

2.2.2 *Privacy seals*

Privacy seals are displayed on Web sites in an attempt to convey a sense of trustworthiness to site visitors. A Web site displaying a privacy seal is often considered more trustworthy than other sites because the Web site's privacy policy has been evaluated by a third party. Privacy seal organisations, such as TRUSTe,[6] BBBonline[7] and WebTrust,[8] thus indicate endorsement from a trusted organisation, but they also mislead consumers who subsequently trust indirect and abbreviated indicators of privacy protection rather than reading the full privacy policy.

The guarantee provided by TRUSTe, and other seals, appear to be comforting to consumers [2]. However, most consumers are unfamiliar with what these privacy seals truly signify. In reality, a TRUSTe 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 FIPs. 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 [27]. 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

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

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

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

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 realise 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 assesses an organisation's privacy policy and business practices for compliance. A licensed CPA, Chartered Accountant, or equivalent will only award a seal to a site if it completely passes this examination. Thus, WebTrust approved sites are certified as protecting and managing information according to their privacy policy. This differs from the TRUSTe seal, which launches an investigation into actual business practices only after a privacy breach is reported. Nevertheless, there are very 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 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, http://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, but did not include security, access/participation, enforcement/redress or other factors such as cookie use. Enonymous.com was dissolved in 2000 for 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. Although P3P has the potential to automate such a rating tool, 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 shortlists of privacy protection practices. In contrast, goals enable one to compare policies based upon not only content, but upon whether or not they actually protect private information and to what degree. A major strength and difference between the approach advocated in this paper and existing approaches; that is, the evaluation of privacy vulnerabilities in addition to protections.

2.3 Policy from the requirements engineering perspective

Although researchers in the requirements engineering community are beginning to focus on e-commerce applications [9, 28, 29] there remains a need to apply proven requirements analysis methods and demonstrate how to best apply these methods within the context of establishing and analysing policy. Goal analysis has been successfully applied within the context of evolving e-commerce systems [9] as we now discuss.

2.3.1 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 [10]. 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 [8], they are a beneficial source for requirements derivation. Goals are operationalised and refined into requirements and point to new, previously unconsidered scenarios [9, 30, 31, 32, 33].

2.3.2 Goal-based requirements engineering

The Goal-Based Requirements Analysis Method (GBRAM) [8, 9, 28, 34] 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 analysing 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 analysed. The method has been successfully applied to the analysis of e-commerce applications [9, 17].

Whereas this paper may be viewed as presenting techniques to distill natural-language goals and warnings from stated (natural-language) policies, in requirements engineering goals are in fact used as one input toward specifying a software system [8, 10, 11]. Thus, the

approach is important for software engineers in addition to company management and legal departments. It enables policy and requirements to be considered more holistically because the approach also helps specify corporate policy and business practices which must ultimately be operationalised into software requirements. In this paper, we describe our use of a 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.

3 Taxonomy of privacy goals

Content analysis is a grounded theory approach for making valid inferences from data to their context while remaining replicable [35]. It includes classifying observed data into a set of content categories where entries in a given category have similar meanings. Content analysis was employed to structure the privacy policy domain to develop a privacy taxonomy [13].

During the summer of 2000, we engaged in the first of two content analysis exercises (using goal-mining) in which we evaluated 25 Internet privacy policies from 8 non-regulated e-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. Web site designers can use these goals to ensure that their stated and actual policies are consistent with each other and they can be used by customers to evaluate and understand policies and their limitations. 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. Finally, third-party evaluators of Web site privacy policies can use the goals to identify conflicts.

The goals derived from 25 Internet e-commerce privacy policies were categorised according to common characteristics that emerged and coded into the following categories: notice/awareness, choice/consent, access/participation, integrity/security, enforcement/redress, monitoring, aggregation, storage, transfer of information, collection, personalisation and solicitation. Some of the goals exemplified privacy protection practices while others 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 and potential vulnerabilities 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 goals or privacy vulnerabilities. *Privacy protection goals* express the desired protection of consumer privacy rights. In contrast, *privacy vulnerabilities* relate to existing threats to consumer privacy and represent statements of fact or existing behaviour that may be characterised as privacy invasions. The five kinds of *privacy protection goals* are defined in Table 2 and the seven kinds of *privacy vulnerabilities* are defined in Table 3. When classifying the goals from 48 Internet privacy policies, the researchers (Antón & Earp) initially agreed upon 124 goal classifications. We initially disagreed on 8 goal classifications. The source of confusion that led to the classification discrepancies for six of the eight goals was the lack of a clear definition for "Aggregate Data". The terminology was subsequently refined and disambiguated — aggregate data refers to non-identifying statistical data often used for marketing and promotional purposes. For example, how many users are received daily, the types of services used most often and the overall demographics of the users. This data is collected in mass and therefore, does not identify anyone. The last two disagreements were quickly resolved through a very brief discussion leading to a consensus. Thus, there was an initial agreement of 94% for the goal classifications.

The following sub-sections provide concrete examples of privacy protection goals and privacy vulnerabilities for the Web-based application domain. Eventually, in software development, these goals are operationalised into system requirements and checked for compliance with the respective policies [7]. 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 Sect 4.

3.1 Privacy protection goals

Privacy protection goals suggest those properties to be satisfied in a system, and are subdivided into five categories: notice and awareness; choice and consent; access and participation; integrity and security; and enforcement and redress. These categories provide an effective framework for privacy protection goals as discussed below.

3.1.1 Notice and awareness

Notice and awareness goals reflect ways in which consumers should be notified and/or made aware of an organisation'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 organisation's privacy policy. In the e-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

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 organisation's information practices before any information is actually collected from them (e.g., an organisation'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 organisation 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 unauthorised access, destruction, use, or disclosure of the data Technical measures to protect against loss and the unauthorised 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

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_{102}: `NOTIFY customer of changes to privacy policy`, which obligates the e-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_{103}: `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 do not revolve around a Web site's posted privacy policy. In the examined e-commerce privacy policies, few goals related to the identity of the organisation 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 organisation to which the site appeared to belong. Such cookies are referred to as ''third party'' cookies and often occur when a Web site allows an outside company to place an advertisement on the Web site. This, in turn, allows the advertising company to place cookies on the unsuspecting consumer's computer. 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_{56}: `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_{62} (`LIMIT disclosure of prescription information/PII to patient or authorised representative`).

The last aspect of notice and awareness concerns ensuring confidentiality, integrity and quality of the data; this is typically expressed by goals that impose

Table 3 Privacy vulnerability classifications

Privacy Vulnerability Taxonomy	Privacy Vulnerability Sub- Classifications
Information monitoring	
Goals concerning what organisations 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 organisation'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 organisation'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 organisation 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 personalisation	
Goals addressing personalisation as when consumers either change their PII, or when cookies are used to customise, thus affecting the functionality or content offered to them.	Personalisation by user preference Personalisation of site and service Personalisation of advertising, offers, and promotions
Contact	
These goals deal with how and for what purpose organisations 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

mechanisms to ensure that consumer data and information is kept confidential and secret.

3.1.2 Choice and consent

Choice and consent goals reflect ways in which a Web site ensures that consumers are provided 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`.

3.1.3 Access/participation

Access and participation goals reflect ways that Web sites allow consumers to access, correct and challenge any data about themselves; for example, by 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. For example, the goal G_1: `ALLOW customer to modify/remove their PII`, concerns the removal of information about an individual from a company's databases, is classified as an access/participation goal.

3.1.4 Integrity/security

Integrity and security goals reflect the ways in which a Web site ensures that data is both accurate and secure. 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 organisational procedures that limit access to data and ensure that those individuals with access do not utilise the data for unauthorised purposes. Goal G_{62}: `LIMIT disclosure of prescription information/PII to patient or authorised representative (prescribing physician)` and goal G_{79}: `DISALLOW access to PII by non-aliated persons` are examples of goals that address managerial measures. Technical measures to prevent unauthorised access include encryption in the transmission and storage of data (e.g., G_{59}: `PROTECT order information using SSL encryption technology`); limits on access through use of passwords (e.g., G_{60}: `USE password for customer accounts`); and the storage of data on secure servers or computers (e.g., G_{113}: `STORE credit card info securely (encrypted, separate DB)`).

3.1.5 Enforcement/redress

Enforcement and redress goals reflect ways in which a Web site enforces its policies. Goals pertaining to self-regulation and private remedies are more common than those addressing government enforcement are. 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.

3.2 Privacy vulnerabilities

Privacy vulnerabilities reflect ways in which a Web site may 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.

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, personalisation, and solicitation. This subsection provides examples from our goal-mining efforts to frame the discussion of vulnerabilities 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 vulnerabilities are classified according to Table 3 as discussed below.

3.2.1 Information monitoring

Information monitoring vulnerabilities reflect the occurrence information tracking by organisations 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 organisation, 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.

3.2.2 Information aggregation

Aggregation refers to combining previously gathered PII data with data from other sources. It is important to note that aggregation is more prevalent in e-commerce privacy policies than in health care privacy policies. E-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 aggregation vulnerabilities are somewhat domain-specific. Although aggregation vulnerabilities are included in the taxonomy, this does not imply that every privacy policy includes one or more information aggregation vulnerability. In fact, the two examples provided here are actually taken from our e-commerce case study because no aggregation goals were expressed in the analyzed health care privacy policies.

3.2.3 Information storage

Information storage vulnerabilities reflect how and what records are stored in an organisation's database. There are two main reasons for an organisation 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 operationalise and/or instantiate business rules (e.g., `STORE credit card information until dispute is resolved`).

3.2.4 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. Information transfer vulnerabilities reflect 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 vulnerabilities are among the easiest to identify due to a standard set of keywords for their identification: DISCLOSE, SELL, SHARE, and PROVIDE. Goal G_{123}: DISCLOSE collected PII when required by law is representative of one information transfer practice and goal G_{127}: SHARE PII for oers/promotions justifies the reason for which information is being transferred.

3.2.5 Information collection

Information collection vulnerabilities reflect what information is collected by Web sites. In the taxonomy, information collection vulnerabilities are characterised as either direct or indirect. Direct collection occurs when an organisation 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 vulnerability. 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).

3.2.6 Information personalisation

Information personalisation vulnerabilities reflect the tailoring or customisation of a Web site to a specific visitor, thus affecting the functionality or content offered to individual visitors. Personalisation may be as simple as greeting the Web site visitor by name (e.g., "Welcome, George.") as suggested by goal G_{106} (RECOGNIZE repeat customers using cookies) or may be more elaborate as in the case of goal G_{109} (CUSTOMIZE content to specific customer using demographic/profile data), which may serve to personalise the site for purposes of targeted marketing.

3.2.7 Solicitation

Solicitation vulnerabilities reflect how and for what purpose organisations 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/solicitations based upon visitors' browsing patterns. Consider goals G_{38} (ALLOW aliates to use PII for marketing/promotional purposes) and G_{41} (SEND email to customer); both of these vulnerabilities exemplify ways in which customers or site visitors may be contacted.

4 Case study: health care privacy policy Requirements Analysis

This section describes the goal-mining process within the context of our health care privacy policy analysis. Section 4.1 discusses how taxonomy goals are identified, classified and refined. Section 4.2 discusses the results of our case study, involving the analysis of 24 Internet health care Web site privacy policies.

4.1 The *Goal Mining* process

This subsection presents the goal-mining process and its associated heuristics within the context of our content analysis of Internet health care privacy policies. *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 [8]. 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 [7, 36]. These goals are documented and annotated with auxiliary information including the responsible agents. Goals are then organised according to goal class (privacy protection or privacy vulnerability, as previously discussed) as well as according to keyword and subject (e.g., browsing patterns, personalisation, cookies, etc.). Once goals are identified, they are elaborated; goal elaboration entails analysing 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 operationalising the goals into a requirements specification.

The e-commerce goal-mining study led to the development of the privacy goal taxonomy introduced in Sect 3 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 analysing 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 were analysed: health insurance, online drug stores and pharmaceuticals. The goals were analysed according to different characteristics such as protection goals vs. vulnerabilities and subject matter (e.g., cookies, PII, browsing patterns, etc).

4.1.1 Heuristics for identifying goals

To identify goals, each statement in a privacy policy is analysed 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 realised. 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_{52}: `PREVENT cookies from tracking activity on other Web sites` and G_{53}: `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 [37, 38, 39]. 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 Web site 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 Web site`. Goals are thus also identified using inquiry-driven [40] 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_{111}: `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.

4.1.2 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 analysing each goal and asking, *''Does this goal protect one*'s private information?'' Whereas, vulnerabilities are classified by considering each goal and asking *''Does this goal potentially compromise the privacy and/or security of one*'s private information?'' Consider Privacy Policy #1, which yielded the goal G_{53}: `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.

Redundancies and synonymous goals are most easily identified after the goals have been organised according to subject matter (Table 4). The 13 subject matters studied are listed in the left most column of the table. Merged goals are represented by the number that appears within parentheses, following the number of synonymous goals. The ''Total'' and ''% Reduction'' columns characterise the evolution of the goal set as discussed below.

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 goals and vulnerabilities were observed within each of the subject matter categories. This analysis is discussed in more detail in Sect. 4.2. Table 4 details additional data about the identified goals, according to subject matter, such as the number of functional, operational, synonymous,

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

redundant and final goals; this refinement process is discussed below.

4.1.3 Goal refinement heuristics

Organisation 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_{95}: 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.

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.

As previously mentioned, the "Total" and "% Reduction" columns in Table 4 characterise 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.

4.2 Observations and discussion

This study had several objectives, to: (1) create a taxonomy for classifying privacy goals for subsequent operationalisation into system requirements; (2) develop a set of reusable privacy and security goals for e-commerce software developers; and (3) use those goals to analyse 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.

4.2.1 Data analysis of protection goals and vulnerabilities

Before our data analysis of the health care privacy policies, we set forth several tentative assumptions in order to draw out and test their logical or empirical consequences [13]. We hypothesised that the number of protection goals in a health care privacy policy is greater than the number of vulnerabilities for that policy; this hypothesis was confirmed as true. When comparing the number of protection goals to the number of vulnerabilities for each Web site (see Table 5), 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 vulnerabilities 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 vulnerabilities identified in health care privacy Policies

	Company name	Number of protection goals	Number of vulnerabilities
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 important to note that we are not advocating a simple count of the number of goals in each category because it can hardly be called conclusive; an additional analysis is required. This analysis involves using the taxonomy to guide the careful consideration of a policy goal's actual intent. Some goals may be clearly distinguished as either protection goals or vulnerabilities; for example G_{53}: `PREVENT use of cookies to send spam` is obviously a protection goal and G_{38}: `USE PII for marketing & promotional purposes` is obviously a vulnerability. However, there are cases in which this distinction is not obvious without careful analysis of the intent according to the taxonomy's goal class definitions (see Tables 2 and 3). Consider the following goal: G_{113}: `STORE credit card info securely (encrypted, separate DB)`. Simply focusing on the keyword, `STORE`, without consideration of the taxonomy (which would provide an understanding of the goal's true intent) would have led this goal to be classified as a vulnerability (information storage). In reality, this goal is a protection goal (integrity/security) and the taxonomy enables an analyst to classify it as such. Goals that are classified as protection goals should be operationalised into a system's requirements; they express desired behaviours. In contrast, goals that are classified as vulnerabilities require additional analysis and possible design decisions to eliminate the given vulnerability. For example, had goal G_{113} not referenced security, then it would have indicated the need for refining the policy and/or corresponding requirements [7]. An additional benefit of the taxonomy is thus realised by analysts in search of a stopping criteria. The existence of vulnerabilities clearly indicates the need for further goal refinement because the ultimate objective is to ensure a trusted and secure system as represented by a set of privacy protection goals [41, 42]. Thus, the existence of vulnerabilities suggests an incomplete requirements analysis effort.

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 expressed vulnerabilities in that Web site; for example, AETNA's privacy policy stated five vulnerabilities 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 vulnerabilities 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 minimise risk to its customers. Table 6 provides examples of health care privacy policy goals organised according to their goal class.

4.2.2 Qualitative observations

An examination of the protection goal and vulnerability types within the subject matter goal classes helped surface 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 vulnerabilities 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 favourable 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 browser 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 sets cookies on behalf of third party sites, and yet this was not expressed in their privacy policy.

Privacy vulnerabilities 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.

Table 6 Example privacy protection goals and privacy vulnerabilities

Privacy Protection Goals	Privacy Vulnerabilities
Notice/Awareness	**Information Monitoring**
G_{99}: NOTIFY customer of intended use of PII	G_{92}: ALLOW 3rd parties to use cookies
G_{101}: NOTIFY customer when PII is removed	G_{93}: ALLOW 3rd parties to use Web bugs
G_{103}: NOTIFY customer of changes to privacy policy	G_{96}: CONNECT IP address w/ PII
G_{104}: POST changes to privacy policy on website	G_{25}: COLLECT date and times at which site was accessed
Choice/Consent	**Information Aggregation***
G_4: PREVENT using friend's email w/o consent	G_{134}: AGGREGATE purchase information by zip code
G_5: DISCLOSE HI at request of patient	G_{135}: AGGREGATE statistics about user browsing patterns
G_7: PREVENT licensing/trading/renting/selling PII w/o consent	
G_{10}: ALLOW customer to change preferences (opt-out)	**Information Storage**
G_{14}: OPT-IN to receive information and promotions	G_{112}: STORE friend info
G_{16}: OPT-OUT from new use of PII in future	G_{136}: STORE PII
Access/Participation	**Information Transfer**
G_3: REMOVE PII of children under 13	G_{128}: SHARE PII w/ aliates
G_1: ALLOW customer to modify/remove their PII	G_{131}: ALLOW links to other sites who's privacy policy is dierent
	G_{132}: SELL all customer info as business asset (in event of buy out)
	G_{124}: DISCLOSE collected PII when required by law
	G_{129}: SHARE PII for oers/promotions
Integrity/Security	**Information Collection**
G_{53}: PREVENT use of cookies to send spam	G_{22}: ALLOW 3rd parties to collect browsing and usage patterns information
G_{52}: PREVENT cookies from tracking activity of other websites	G_{27}: ALLOW 3rd parties to collect IP address
G_{55}: PREVENT sharing aggregate data	G_{37}: COLLECT credit card info for billing/collect payment for services
G_{75}: PREVENT saving credit card/password info using cookies	G_{22}: ALLOW 3rd parties to collect browsing and usage patterns information
G_{78}: PREVENT sending unsolicited email	G_{32}: COLLECT browser type
G_{57}: PREVENT selling/renting customer lists	G_{37}: COLLECT credit card information for billing/collect payment for services
G_{63}: LIMIT disclosure of prescription information/PII to patient or authorised representative	
G_{80}: DISALLOW access to PII by non-aliated persons	
G_{60}: PROTECT order information using SSL encryption technology	
G_{61}: USE password for customer accounts	
G_{114}: STORE credit card info securely (encrypted, separate DB)	
Enforcement/Redress	**Information Personalisation**
G_{42}: DISALLOW use of collected PII from aecting HI coverage by our company	G_{111}: CUSTOMIZE experience at our site using cookies
G_{43}: COMPLY with internationally recognised standards of privacy protection	G_{107}: RETRIEVE customer info using cookies
G_{50}: REQUIRE employees to comply with company privacy policy	G_{107}: RECOGNIZE repeat customers using cookies
G_{44}: DISCIPLINE employee who violates privacy policy	G_{110}: CUSTOMIZE content to specific customer using demographic / profile data
	Contact
	G_{38}: USE PII for marketing & promotional purposes
	G_{40}: SEND one-time email to friend
	G_{38}:ALLOW aliates to use PII for marketing/promotional purposes
	G_{41}: SEND email to customer

*Aggregation goals in this table were identified in the e-commerce case study.

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 have created a privacy values survey instrument to assess these value differences and create a privacy values baseline.

5 Summary and future work

Most Web sites display a privacy policy that describes 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 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 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 taxonomy presented in Sect. 3 provide such an effective unit for measuring and comparing these policies, while reasoning about the respective system requirements. Our study 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 [7 42, 43]. The third objective is to provide a basis for analysing and comparing Internet privacy policies and the corresponding system requirements which they govern.

Goal-mining with the privacy goal taxonomy is effective for examining how Web sites claim they manage online customer data and how they convey these practices to their customers [6]. Requirements engineers need to better understand the kinds of requirements needed to satisfy governing policies and help consider the potential vulnerability scenarios a system must address. The taxonomy offers the ability to systematically identify privacy vulnerabilities that should be refined into protection requirements to better ensure that system requirements reflect a trusted system.

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. In this paper, we have introduced a taxonomy for classifying privacy protection goals and vulnerabilities; we describe our use of a software requirements engineering content analysis 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 emphasise privacy policy goal-mining in this paper, the techniques we have presented are generalisable 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.

The taxonomy presented in this paper is admittedly biased towards the status quo for Web sites in the United States since the sample of privacy policies were primarily from US-based organisations . The five principles that the privacy protection goals reflect, have been the focus of US industry, the US Department of Commerce and the Federal Trade Commission (FTC), but these principles are certainly less restricting than the OECD Guidelines or the European Union Directive for the protection of personal data and privacy. The OECD guidelines are more complete than the FIPs as they address the following aspects of data protection: Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation, and Accountability [44]. The European Union Directive is more comprehensive as it addresses all five FIPs[9] in addition to several other privacy elements including one vulnerability class, information transfer.[10] This emphasises the extent to which the privacy goal taxonomy addresses privacy goals, as it includes a thorough coverage of vulnerability attributes in addition to standard protection principles.

Our plans for future work include empirical investigations in which we examine the usefulness of the taxonomy to those who specify the requirements for new Web-based systems. To date, anecdotal evidence suggests that the taxonomy is useful for ensuring consistency and better requirements coverage during requirements specification [42]. Our plans also 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. This survey has been administered to 1005 Internet users in an effort to establish a privacy values baseline for correlation with the taxonomy presented in this paper. The detailed data analysis results and survey instrument design are forthcoming. However, our preliminary analysis shows that consumers are most concerned with (in order): information collection, information personalisation, notice/awareness, and information transfer. In contrast, privacy policies emphasise (in order) choice/consent, information collection, integrity/security, and information transfer. We are also investigating the use of run-time requirements monitoring to support end-users seeking to determine whether the sites they visit are operating in compliance with their privacy policies. In parallel, we are mapping our privacy goal taxonomy to P3P to establish

[9] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (). Articles 6,10 and 11 address notice/awareness; Article 7 addresses choice/consent; Article 12 addresses access/participation; Articles 16 and 17 address integrity/security; Articles 22, 23 and 23 address enforcement/redress.

[10] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (). Articles 25 and 26 address information transfer

an automatable representation of the taxonomy. Finally, the taxonomy goals will be used to populate the SMaRT (Scenario Management and Requirements Tool) to support the specification of privacy related requirements by software developers, providing additional validation.

Acknowledgments This work was supported by NSF ITR Grant #0113792 and the CRA's Distributed Mentor Project. The authors wish to thank Shane Smith, 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 goals and vulnerabilities.

References

1. Cranor LF, Reagle J, Ackerman MS (1999) 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
2. Earp JB, Baumer D (2003) Innovative Web use to learn about consumer behavior and online privacy. Commun ACM 46(4):81–83
3. Goldman J, Hudson Z, Smith RM (2000) Privacy report on the privacy policies and practices of health Websites, Sponsored by the California HealthCare Foundation
4. Federal Trade Commission (1998) Privacy online: a report to congress.http://www.ftc.gov/reports/privacy3/
5. Federal Trade Commission (2000) Privacy online: fair information practices in the electronic marketplace. A report to congress
6. Antón AI, Earp JB, Potts C, Alspaugh TA (2001) 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. Antón AI, Earp JB (2001) Strategies for developing policies and requirements for secure electronic commerce systems. In: Anup K (ed) E-commerce security and privacy. Kluwer, Glosh, pp 29–46 CHECK STYLE
8. Antón AI (1997) Goal identification and refinement in the specification of software-based information systems. Dissertation, Georgia Institute of Technology, Atlanta, GA
9. Antón AI, Potts C (1998) 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
10. van Lamsweerde A (2001) 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
11. Mylopoulos J, Chung L, Liao S, Wang H, Yu E (2001) Exploring alternatives during requirements analysis. IEEE Softw 18(1):92 –96
12. Glaser BC, Strauss AL (1967) The discovery of grounded theory. Aldine, Chicago
13. Antón AI, Earp JB Reese A (2002) Analyzing Web site privacy requirements using a privacy goal taxonomy. 10th Anniversary IEEE Joint Requirements Engineering Conference (RE'02). Essen, Germany, pp 23–31, 9–13 September 2002
14. The code of fair information practices (1973) 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
15. Culnan MJ (1999) Georgetown Internet privacy policy survey: report to the federal trade commission. The McDonough School of Business , Georgetown University, Washington, DC,http://www.msb.edu/faculty/culnanm/gippshome.html
16. Electronic Privacy Information Center (1999) Surfer beware III: privacy policies without privacy protection.http://www.epic.org/reports/surfer-beware3.html
17. Baumer D, Earp JB and Payton FC (2000) Privacy of medical records: IT implications of HIPAA. ACM Comput Soc 30(4):40–47
18. Reagle J, Cranor LF (1999) The platform for privacy preferences. Commun ACM 42(2):48–55
19. Benessi P (1999) TRUSTe: An online privacy seal program. Commun ACM 42(2):56 – 59
20. P3P Public Overview. http://www.w3.org/P3P/, cited 24 June 2002
21. Cranor L, Langheinrich M, and Marchiori M (2002) A P3P preference exchange language 1.0 (APPEL1.0): W3C working draft.http://www.w3.org/TR/P3P-preferences/, cited 15 April 2002
22. Electronic Privacy Information Center (2000) Pretty poor privacy: an assessment of P3P and Internet privacy.http://www.epic.org/reports/prettypoorprivacy.html
23. Mulligan D, Schwartz A, Cavoukian A, Gurski M (2000) P3P and privacy: an update for the privacy community.http://www.cdt.org/privacy/pet/p3pprivacy.shtml, cited 28 March 2000
24. Cohen D, Feather MS, Narayanaswamy K, Fickas SS (1997) Automatic monitoring of software requirements. International Conference on Software Engineering, pp 602 –603
25. Fickas S, Feather MS (1995) Requirements monitoring in dynamic environments. Second IEEE International Symposium on Requirements Engineering, pp 140 –147
26. Feather MS, Fickas S, van Lamsweerde A, Ponsard C (1998) Reconciling system requirements and runtime behaviour. Ninth International Workshop on Software Specification and Design, pp 50 –59
27. FTC sues failed Website, Toysmart.com, for deceptively offering for sale personal information of Website visitors. FTC File No. 002–3274. 10 July 2000
28. Antón AI, Carter RA, Dagnino A, Dempster JH, Siege DH (2001) Deriving goals from a use-case based requirements specification. Req Eng (6):63–73
29. Robinson WN (1997) Electronic brokering for assisted contracting of software applets. Proceedings of the Thirtieth Hawaii International Conference on System Sciences, vol. 4, pp 449–458
30. Antón AI, McCracken WM, Potts C (1994) Goal decomposition and scenario analysis in business process reengineering. Advanced Information System Engineering: 6th International Conference, CAiSE '94 Proceedings, Utrecht, The Netherlands, pp 94–104, 6–10 June 1994
31. Jarke M, Bui XT, Carroll JM (1998) Scenario management: an interdisciplinary approach. Req Eng 3(3/4):154–173
32. Potts C (1999) ScenIC: A strategy for inquiry-driven requirements determination. Proceedings IEEE 4th International Symposium on Requirements Engineering (RE'99), Limerick, Ireland, 7–11 June 1999
33. Rolland C, Souveyet C, Achour CB (1998) Guiding goal modeling using scenarios. IEEE Trans Softw Eng 24(12):1055–1071
34. Antón AI (1996) Goal-based requirements analysis. Second IEEE International Conference on Requirements Engineering (ICRE '96), Colorado Springs, Colorado, pp 136–144, 15–18 April 1996
35. Krippendorff K (1980) Content analysis: an introduction to its methodology, vol. 5. Sage, Newbury Park, CA
36. Policy framework for interpreting risk in eCommerce security. CERIAS Technical Report (1999), Purdue University, http://www.cerias.purdue.edu/techreports/public/PFIRES.pdf
37. Abbot RJ (1983) Program design by informal english descriptions. Commun ACM 26(11):882–894
38. Booch G (1991) Object-oriented design with applications. Benjamin Cummings, Redwood City, CA
39. Rumbaugh J, Blaha M, Premerlani W, Eddy F, Lorensen W (1991) Object-modeling and design. Prentice Hall, New York
40. Potts C, Takahashi K, Antón AI (1994) Inquiry-based requirements analysis. IEEE Softw 11(2):21–32
41. Jarvinen O, Earp J, Antón AI (2002) A visibility classification scheme for privacy management requirements. 2nd Symposium

on Requirements Engineering for Information Security, Raleigh, NC, 17–18 October 2002
42. Antón AI, Earp JB, Carter RA (2003) Precluding incongruous behavior by aligning software requirements with security and privacy policies. Inf Softw Technol 45(14):967–977
43. Alspaugh T, Antón AI, Barnes T, Mott B (1999) 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
44. CDT (2000) CDT's guide to online privacy: privacy basics: the OECD guidelines.http://www.cdt.org/privacy/guide/basic/oecdguidelines.html, cited 6 August 2002