

The Use of Goals to Extract Privacy and Security Requirements from Policy Statements

Annie I. Anton¹, Davide Bolchini², Qingfeng He¹

¹College of Engineering, North Carolina State University, Raleigh, NC 27695

²Faculty of Communication Sciences, University of Lugano, Lugano TI 6900, Switzerland

¹{aianton, qhe2}@eos.ncsu.edu ²davide.bolchini@lu.unisi.ch

Abstract

This paper addresses the use of goals to extract non-functional requirements from policy statements. Goals are important precursors to software requirements, but the process of abstracting them from security and policy policies has not been thoroughly researched. We present a summary of a goal-based approach for extracting standard security and privacy requirements from policy statements and illustrate its application to analyze 40 financial privacy policies. We present heuristics to support goal analysis, goal refinement, and the development of tool support, including the establishment of a goal repository that can be used in future goal analyses. To gain a deeper understanding of the goal set, and to identify potential conflicts and inconsistencies between goals, we used i to model semantic relationships between goals, their actors and strategic dependencies. The goal-based process will assist software engineers in the specification of system requirements that are in alignment an organization's policies.*

1. Introduction

Keeping sensitive information secure is increasingly important within the context of e-commerce and web-based applications due to concerns regarding the electronic transmission and dissemination of personally identifiable information. Privacy policies, as well as security policies, for Web-based systems are often developed as an afterthought [AE01]. The implications are significant because these systems' requirements, policies and functionality are often misaligned and/or in conflict with one another [AER02]. When requirements, policies and functionality are not aligned, consumer privacy is jeopardized and it is increasingly difficult for software developers to ensure their systems are secure and privacy-aware. A website's security and privacy policies should be reflected in the actual system requirements [AE01], and yet they are often not considered during requirements specification activities (either during the planning of new systems or feature enhancements to existing systems).

Security and privacy policies establish the rules

and guidelines by which a system (or set of interacting systems) must operate. A software requirement is an abstract statement of some behavior that a system is expected to exhibit or service that it must provide. Policies, whether explicit or implicit, are high-level requirements or meta-requirements that determine the normative context in which a system or technology may be successfully employed. Policies and requirements share some similarities and differences. They are similar in that both policies and requirements express desire or worth, rather than fact. Additionally, policies and requirements are typically expressed as statements in the optative mood [Jac95]; that is, they specify what must or ought to be done. However, policies express a broader scope than do requirements and policies are more open-ended than requirements. Perhaps the most important distinction to be made between policies and requirements is that software requirements are system-specific whereas policies are system-independent.

Obtaining requirements for software systems is a complex and challenging activity. Traditionally, security and privacy requirements have not been explicitly stated. Instead, they are often implied by the system's architectural security mechanisms [Fir03]. Security and privacy requirements are expressed in policies. However, it is especially challenging to obtain requirements from policy statements because they often refer to nebulous entities (e.g. "personal information" and "transaction history"). These requirements usually do not establish a meaningful connection between the terminology appropriate for the specific function being supported, and the more open-ended terms in which the policy is typically expressed. Goal analysis, as discussed in this paper, offers a methodical and systematic approach to extract privacy and security requirements from these often nebulous policy statements. These requirements, while challenging to identify and express are critical because they ensure the software systems they specify are in compliance with governing policies.

This paper presents a summary of a goal-based approach for extracting standard security and privacy requirements from policy statements and illustrates its application to analyze 40 financial privacy

policies, evaluating the method in the process. We present heuristics to support goal analysis, goal refinement, and the development of tool support, including the establishment of a goal repository that can be used in future goal analyses. The analysis performed was based upon the GBRAM (Goal-Based Requirements Analysis Method) [Ant97] and built upon previous studies of e-commerce and healthcare privacy policies [AER02]. In order to gain a deeper understanding of the goal set, we used *i** [Yu93] to model semantic relationships between goals, their actors and strategic dependencies. This process allows analysts to study potential conflicts within a given policy and identify possible resolutions, which will assist software engineers in the specification of system requirements that are in alignment with the organization's privacy policies.

The remainder of this paper is organized as follows. Section 2 provides an overview of the relevant work. Section 3 briefly summarizes the process of extracting privacy and security goals from policy statements. Section 4 discusses an in-depth analysis of 40 Internet privacy policies using goal-driven requirements engineering techniques. Section 5 discusses our plans for future work.

2. Background and Related Work

In requirements engineering, goals are used to model the high-level objectives of a software system's stakeholders. Goals are recognized as powerful drivers for the development process because they help relate system requirements to the business and organizational needs, enable traceability of design rationale, and help analysts identify conflicts and tradeoffs in the early stages of the software development lifecycle. In goal identification, requirements are derived from a variety of information sources and although the need to include security and privacy policies among those information sources has been recognized as important [Lic97, PFI99, RHA03, AEC03].

Much of the recent research into requirements acquisition has investigated the formal refinement of goals [Ant97, DFv93, Pot95, vDM95]. Goal refinement is intended to reduce the risk of incomplete requirements by explicit consideration of the rationale for requirements and the derivation of other requirements that are implied by this rationale. The emphasis of much of the work on goals has been on using them to derive formal specifications [DFv93, vDM95]. In contrast, this paper focuses on the actual extraction of goals from documents that have traditionally not been used by requirements engineers as a source for requirements: privacy and security policy statements. Policies have become

increasingly important because of the need to ensure that software systems are secure and privacy aware.

The KAOS framework [DFv93] states system goals as a strategy to derive complete and consistent requirements through a formal refinement process. Whereas KAOS primarily focuses on functional requirements, the *i** model [Yu93] incorporates non-functional goals to uncover quality requirements for envisioned systems. Moreover, *i** extends the spectrum of goals to be considered to those of the organization and its actors, thus enabling the identification of interdependencies and social relationships between stakeholders. The GBRAM's heuristics [Ant97] help analysts extract the actual goals and requirements of new or existing systems from available project documentation and designs so the goals may ultimately be structured into a requirements specification. The analysis of policy statements that are inherently more far reaching and broader in scope than requirements, is especially challenging as we discuss in this paper.

Tools are needed to help analysts gain insights about the actual compliance of a system's behavior with the needs and wants of the stakeholders as well as any governing policies. In the privacy domain, the GBRAM has been extensively used to analyze website privacy policies to systematically extract the requirements and goals underlying organizations' privacy practices [AER02]. This process, called "goal-mining," makes use of a privacy taxonomy [AER02] that classifies the extracted requirements as either protection goals or vulnerabilities. Protection goals express the effort declared by an organization to honor/respect its customers' privacy. Vulnerabilities reflect potential threats to customer privacy as derived from current organization practices such as information collection, storage and transfer. The taxonomy and the lessons-learned from the previous goal-mining experiences represent the methodological basis for the work presented in this paper.

In software engineering, compliance is most commonly documented via requirements traceability [DP98, Ram98]; a measure that reduces the risk of, for example, changes not propagating across lifecycle artifacts. In these cases, system documentation often remains unmodified after their initial creation, often becoming obsolete [Ram98]. An important component in policy and requirements alignment entails ensuring that a system's SRS (Software Requirements Specification), security policy and privacy policy are never obsolete by adopting an iterative risk analysis activity to ensure compliance, minimizing the risk of inconsistencies across the resulting requirements and policy artifacts. The

ultimate objective is to ensure compliance between the requirements, all policies, and the system's functionality [AEP01].

Researchers are beginning to explore complimentary ways to employ roles and goals to elicit and model security and privacy requirements to ensure compliance. Crook et al. have proposed an analytical role modeling framework to model access policies using roles in requirements engineering [CIN03]. They regard access policies as security requirements that specify restrictions of access to valuable assets. He et al. proposed a framework for modeling privacy requirements in role engineering, which employs goal-based requirements analysis techniques to analyze business processes and produce a complete specification of roles, permissions and constraints [HA03, He03].

The NFR Framework is a general approach to model and analyze nonfunctional requirements, including security requirements [CNY00]. Liu et al. employ the combination of goal-oriented and agent-oriented techniques to analyze security and privacy requirements with i^* [LYM02, LYM03]. Fontaine employs KAOS to elaborate security requirements and integrates KAOS with a policy specification framework and language called Ponder [Fon01]. Ponder is a framework and a comprehensive language for specifying and deploying security policies [Dam02]. Fontaine's approach refines certain security requirements into specific authorization rules and access control policies with help of Ponder. However, the goal-driven policy analysis approach discussed in this paper is requirements level analysis. We extract high-level security and privacy goals from existing policy statements and refine them into nonfunctional requirements.

3. Goal Driven Policy Analysis

The process of identifying high-level goals is fundamental to the requirements analysis process. The GBRAM assumes that goals have not been previously documented or explicitly elicited and that analysts must work from various sources of available information, each with its own scope of knowledge, to determine the goals of the desired system. It also supports the elaboration of goals to represent the desired system. A detailed presentation of how to apply the method from the initial identification of goals to translation of those goals into operational requirements is available in [Ant97].

This paper provides a brief overview of how the GBRAM is used to extract goals from policies,

particularly privacy policies, which are inherently different from more traditional sources used during software system design. Privacy and security policies are typically not even considered project documentation. And yet, there is an increasing awareness of the need to treat policies as important sources for requirements derivation as discussed at [RHA03]. Project documents typically adhere to standards of clarity, consistency, and conciseness (the three "C's") because they are intentionally created to be read by computer experts (e.g. software engineers, designers, coders). Project documents provide a basis for systems requirements specification. Privacy and security policies are intended for different audiences and purposes than those documents. Internet privacy policies tell website users about how their information may be used. Security and privacy documents are rarely communicated to software developers. Furthermore, the architects of those policies seldom emphasize the three "C's." As a result of the differences in audience, intended purpose, and structure, these policies are often redundant, inconsistent or ambiguous. This makes extraction of requirements from those policies much more difficult.

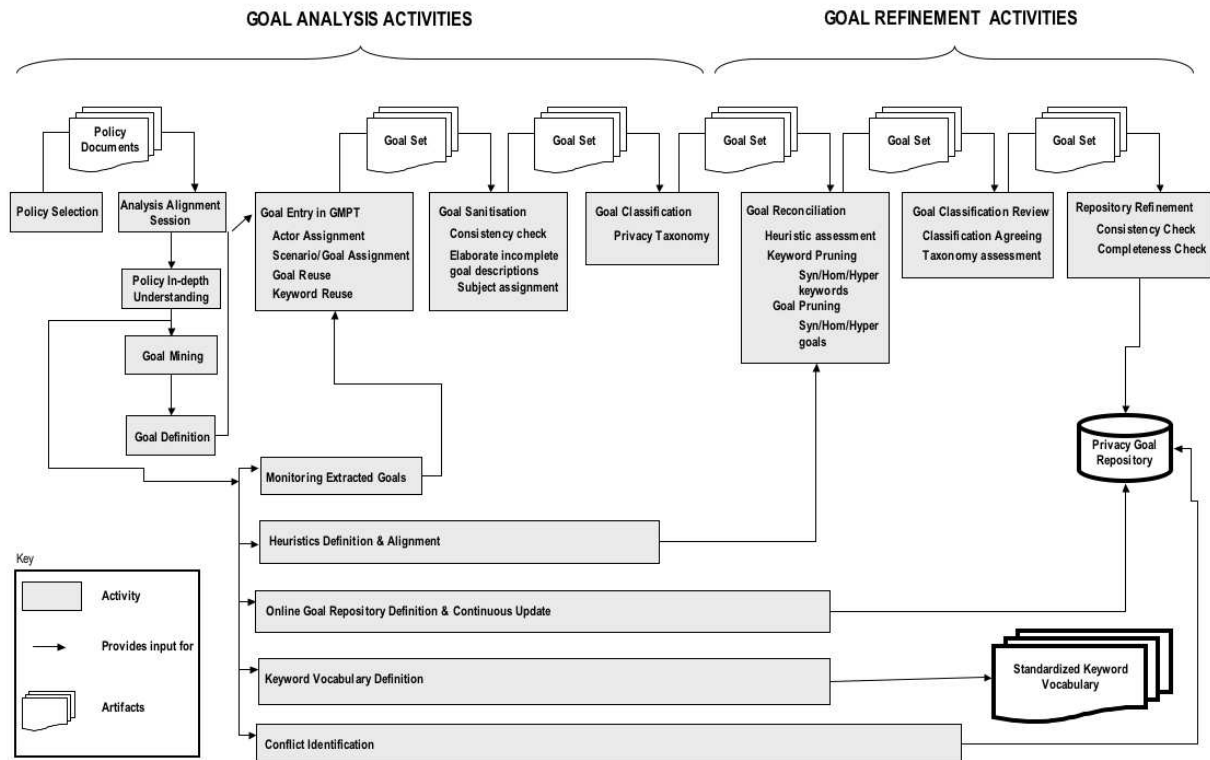
The method presented in this paper supports goal analysis and goal refinement. *Goal analysis* concerns the exploration of available information sources for goal identification followed by the organization and classification of goals. *Goal refinement* concerns the evolution of goals during goal reconciliation. In [Ant97] goals are classified as either achievement or maintenance goals. In this paper, we discuss a more meaningful goal classification, focused on privacy and security requirements, in which goals are classified as either privacy protection goals or vulnerabilities.

3.1 Goal Analysis

Figure 1 portrays the goal analysis and refinement activities that an analyst conducts when applying the GBRAM to analyze privacy and security policy statements.

The goal *analysis* activities may be summarized as follows. The analyst team simultaneously studies a given policy document during an initial "Alignment Session" (see Figure 1) to establish both a shared analysis practice as to how to apply the GBRAM and a common vocabulary for the domain of study (e.g. privacy policy or security policy). During the alignment session the following issues are addressed while examining each policy statement:

Figure 1: Goal Analysis and Refinement Process Model



- Is this statement worth considering?
- Is it in or out of scope?
- What are the boundaries of the scope for our analysis?
- Should we keep the original phrasing of the policy statement or should we rephrase it as an unambiguous goal?

Setting specific and domain-dependent analysis rules is not a trivial task. It requires a cooperative effort that benefits the team by making the subsequent analysis steps more efficient and less error-prone. Although we analyzed the policies of nine organizations (discussed in Section 4) because of our desire to develop a repository of reusable privacy and security goals, each organization in our study had multiple policies (see Table 1). Given the increased need for software engineers to derive requirements from an organization's multiple policy documents, it is imperative that all analysts on a project collectively analyze the first policy to accomplish the goals of the alignment session. In our experience, this initial session is critical in

minimizing future misinterpretations and errors in the final requirements.

The individual in-depth understanding of a set of policies would be in vain if the practical procedures for capturing analyzing relevant goals were not first clarified by the analyst team. The alignment session thus represents the starting point for an important procedure alignment activity that continues throughout the analysis process as analysts develop an in-depth understanding of each policy.

The next analysis activity, goal-mining (see Figure 1), entails extracting pre-requirements goals and their responsible agents from text artifacts [AER03]. Goals are not only identified, but need to be defined as clearly and consistently as possible throughout the analysis effort. Here, trade-offs between clarity and conciseness are hard to resolve. Consider the goal: <Prevent third parties to share PII with affiliates and subsidiaries of the institution for marketing and promotional purposes without previous written customer consent>. What should be retained and/or discarded in the goal definition? Is

“Prevent third parties to share PII with affiliates” enough? It is clear that the condition “without customer consent” completes the sense of the statement. Does it suffice to generically define <Prevent third parties to share PII without customer consent>? Should “affiliates and subsidiaries” give rise to two different goals (one concerning affiliates and the other concerning subsidiaries) or should they be generalized into one entity? This is just a small example of the intricate issues that arise when systematically defining privacy goals. Agreement and convergence to shared rules is facilitated as the analysis proceeds and the number of goals to be reused in the repository is augmented.

Monitoring is an important facet during goal analysis activities. While privacy goals are entered into a goal Privacy Goal Management Tool (PGMT), discussed in Section 4, the analyzed goals are monitored (see Figure 1) by querying the tool’s goal repository or by relying on the memory of the analysts.

3.2 Goal Refinement

During *goal refinement*, goals are classified, synonymous goals are reconciled and redundant goals are eliminated. Additionally, conflicts are identified and resolved. Keeping track of the goals in the PGMT makes it easier to identify conflicts within an institution’s policy statements. From the user’s perspective, conflicting privacy statements reflect privacy practices that are communicated unclearly or ambiguously to the user, often confusing him/her about the actual treatment of their personal data by the given institution. Organizations, in turn, risk customer mistrust based upon customer perceptions of conflicting privacy statements [AEP01]. Software engineers run the risk of introducing errors in the software systems they design when these errors are not identified prior to requirements specification [AE03]. Finally, policy managers and the organization’s staff may gather contradictory messages that may impede their tasks.

During goal analysis, conflicts mainly arise from:

- **Contradictions:** contradictions occur when two or more privacy statements cannot hold at the same time and under the same conditions. For example, consider the following two privacy policies: PP_1 : *we allow a period from 3 to 8 weeks for customer preferences to be enacted* and PP_2 : *We honor customer preferences from the moment they have been specified by customers*. These two policies are clearly in conflict. When policy statements express temporal constraints, it is easier to spot contradictions.

Table 1: GLBA covered financial institutions and respective privacy policies analyzed for this study

	Policy Document	Protection Goals	Vulnerabilities
Bank of America	Overview	4	0
	Privacy Policy	73	55
	Online Practices	23	19
	Information Security	29	2
	Identity Theft	18	0
	Accounts & Services	4	4
	FAQ (State: NC)	63	32
	Subtotal	214	112
Citibank	Citigroup Promise	21	15
	Citi com Online Data Policy	8	21
	Citi MyAccounts Promise	19	14
	Citi MyAccounts Notice	12	14
	Citi Terms of Use	4	11
	Subtotal	64	75
Wachovia	Privacy Statement	60	28
	Internet Privacy	20	40
	Privacy Statement FAQ	44	25
	Fraud Prevention	57	1
	Security Statement	25	1
	Online Banking & Billpay	20	2
	Subtotal	226	97
Allstate	Privacy Statement	55	23
	Terms of Use	9	12
		Subtotal	64
American Int'l Group	Privacy Policy	8	9
	Conditions of Use	1	15
		Subtotal	9
State Farm	Privacy Principles	3	2
	Privacy Policy Customers	15	23
	Privacy Policy Consumers	10	9
	Privacy and Security	8	8
	Privacy Policy for PHI	24	41
	State Privacy Rights	1	0
	Privacy Policy FAQ	70	34
	Terms of Use	11	8
	Subtotal	142	125
Goldman Sachs	Privacy Policy	33	28
	Terms & Conditions of Use	0	4
		Subtotal	33
Merrill Lynch	Global Privacy Pledge	30	34
	Online Privacy Statement	11	15
	Legal Info	2	6
	Subtotal	43	55
Morgan Stanley	Privacy Pledge	5	0
	US Individual Investor PP	23	41
	Internet Security Policy	8	10
	ClientServe ISP	18	3
	Terms of Use	10	28
	Subtotal	64	82
TOTAL		859	637

- **Inconsistencies:** inconsistencies occur when privacy statements differ with regard to the treatment of the same issue (e.g. treatment of a data) and may create confusion as to the organization’s actual practice. For example, consider the following two goals: G_1 : Session data are not shared for promotional purposes and G_2 : Session data are shared only for purposes aiming at enhancing the personalization of the service offers. Do “promotional purposes” include “personalized service offers”? What are the specific and

objectively verifiable conditions under which session dates are shared?

Additionally, conflicts may be explicit or implicit. Explicit conflicts are identified by simply comparing two privacy statements as they are expressed in the original privacy policies. This kind of conflict may seem easier to spot. However, in case of several, long privacy policy documents from the same institution, teamwork is essential to help track statements from different documents for deeper evaluation. In fact, teamwork increases the probability of maintaining a comprehensive recall of the analyzed statements.

Implicit conflicts are generated by statements that, on the surface, are not apparently contradictory or inconsistent but they *imply* or suggest conflicting practices. When confronting these conflicts, analysts must extract the hidden statements from those presented to surface the actual conflict (see Figure 2). To help analysts understand conflict, the *i** model is employed to set the scene of a potential conflict and to clarify the responsible stakeholders, the goals and the semantic relationships (explicit or implicit) between the goals [Yu93].

Figure 2 shows a potential between G_{400} (MAINTAIN confidentiality of CI) and G_{401} (ALLOW offers from reputable companies) in the CitiGroup Privacy Promise policy. Actually, G_{401} *per se* is not conflicting with G_{400} , but it implies G_{1165}

(SHARE CI with 3rd parties), which is strongly conflicting with G_{400} .

G_{400} is also in potential conflict with G_{402} (OBLIGATE external companies to not retain PII unless customer expresses interest in their products/services). To elucidate the conflict, also shown in Figure 2, we split the goal into two parts: G_{402a} (Retain CI) – owned by the external company – and G_{402b} (Express interest in service) – owned by the customer. The policy suggests that G_{402a} is achieved only if G_{402b} is satisfied. However, the criteria for determining ‘customer interest’ in 3rd party services are not explained in the policy (see G_{402b}). This omission leads to ambiguity. This ambiguity makes the conditions for which CI is retained (G_{402a}) by external companies unclear, thus compromising the confidentiality of CI. The ongoing monitoring of extracted goals — greatly supported by the PGMT’s functionality — enables existing goals to be reused extensively. Goal entry is not a simple mechanical process because it also entails some decisions intended to add a richer semantics to the goals being defined: *actor assignment* (a) and *scenario/policy classification* (b).

Figure 2: Potential Conflicts in Citigroup Privacy Promise modelled with *i**.

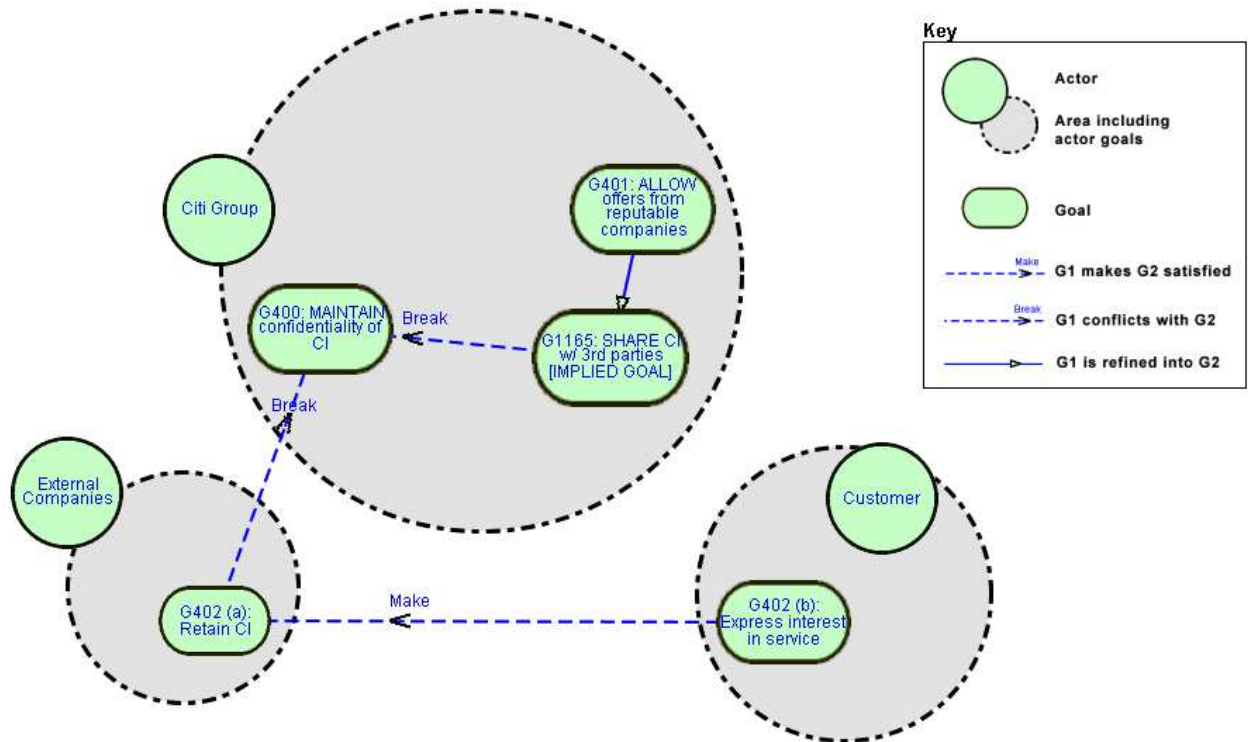


Table 1: Privacy Policy Action Keywords.

ACCESS	CONNECT	DISCLOSE	MAINTAIN	INVESTIGATE	RESERVE
AGGREGATE	CONSOLIDATE	DISPLAY	MAKE	POST	REVIEW
ALLOW	CONTACT	ENFORCE	MAXIMIZE	PREVENT	SHARE
APPLY	CONTRACT	ENSURE	MINIMIZE	PROHIBIT	SPECIFY
AVOID	CUSTOMIZE	EXCHANGE	MONITOR	PROTECT	STORE
BLOCK	DENY	HELP	NOTIFY	PROVIDE	UPDATE
CHANGE	DESTROY	HONOR	OBLIGATE	RECOMMEND	URGE
CHOOSE	DISALLOW	IMPLY	OPT-IN	REQUEST	USE
COLLECT	DISCIPLINE	INFORM	OPT-OUT	REQUIRE	VERIFY
COMPLY	DISCLAIM	LIMIT			

(a) A responsible *actor* is assigned (not invented from scratch) to each goal. Actors may be, for example, the respective policy’s institution, the website user, third parties, and subsidiaries. When defining actors, analysts seek to answer the question: “*Who is responsible for taking the action expressed in this goal?*” In this sense, actors make evident the goal’s relevance and role within its social context. For example, filtering the goal repository “by actor” (Who [actor] does what [goal]) gives a clear picture of the burden allocated to the user or the amount of responsibility assumed by the institution.

(b) Goals are also classified according to their granularity. *Policy* goals are general statements concerning high-level commitments directly reflecting an institution’s mission. For example, a recurring policy statement is “*Our institution protects the privacy of its customers by technical, organizational and procedural measures.*” Goals extracted from such policy statements are often coupled with more specific examples that partially describe mechanisms in place or detailed procedures to be followed. For example, “*Users should log-off securely after finishing the session,*” or “*state locator cookies may be activated to facilitate personalized financial offers.*” The latter two policy statements exemplify scenario goals, because they envision possible (and not necessary) usage situations or practices suggested to the user that have the advantages of being precise and the disadvantage of not being exhaustive.

The classified goals are subsequently sanitized (see Figure 1) according to the action keywords they use to express their purpose. Table 2 lists the 57 keywords currently contained in the PGMT; this list reflects the keywords that are commonly found in Internet privacy policies. Each keyword in the repository has been formally defined to ensure consistent keyword use keywords the analysis process. Sanitizing the data represents a further activity to advance the level of consistency in a goal

set, and to fill the gaps left along the mining and definition process.

A third way in which goals are classified is according to subject matter. For example, consider the goals: <Store credit card info [actor: institution]>, <Secure credit card info [actor: user]>, <Prevent sharing credit card info>, <Use SSL to encrypt transaction involving credit card info>. Classifying these goals according to their subject matter results in obtaining a body of goals concerning credit card information, facilitating the identification of inconsistencies and requirements creep in the declared practices and procedures to ensure the security and privacy of such critical information.

Finally, goals are organized according to a general privacy taxonomy [AERO2] that distinguishes the goals concerning consumer privacy protection from potential threats and vulnerabilities emerging from the policy analysis. Table 1 lists the number of privacy protection goals and vulnerabilities identified in each of the 40 policies analyzed for this study.

A final reconciliation phase is necessary to completely standardize the goal set. This entails merging and reconciling synonymous and homonymous goals as well as the keywords used to define them. Moreover, the analysis heuristics and keyword vocabulary are assessed and validated to allow useful methodological tools that support future analyses. This assessment is important because even the taxonomy and actor-subject-scenario classification show their limitations once the goal repository comprises a significant number of goals. Team reviews are important because they help ensure consistency and completeness.

4 Analyzing Financial Privacy Policies

In this section, we detail our study in which we applied the techniques discussed in Section 3 to analyze 40 online privacy statements from nine

Gramm-Leach-Bliley Act (GLBA)¹-covered institutions². The GLBA requires financial institutions — including banks, insurance companies and securities firms — to protect the security and confidentiality of nonpublic personal information (NPI) for distribution beyond the institution³.

4.1 Overview of the Study

Our sample consists of websites from three banks (Bank of America, Citibank and Wachovia), three insurance companies (Allstate, American International Group and State Farm) and three securities firms (Goldman Sachs, Merrill Lynch and Morgan Stanley). This sample was taken from a cumulative listing of the top five U.S. banks (by revenue, 2002), the top five U.S. property/casualty insurance companies (by net premiums written, 2001) and the top five U.S. securities firms (by revenue, 2001)⁴. The financial privacy policies we examined, and which serve as the focus of this paper, were in force during June of 2003. Table 1 lists the organizations' respective policies that we analyzed.

This requirements analysis effort was conducted by five analysts for approximately 30 hours a week over a period of four weeks and a sixth analyst for 1.5 days. The principle analyst led the team meetings in which all five analysts participated and other meetings were held in pairs or groups of three. These meetings were each one to three hours in duration. The meetings were primarily goal mining sessions during which goals were extracted from privacy policy statements and goal refinement and reconciliation sessions during which redundant goals were eliminated, synonymous goals were merged.

The goal-mining heuristics that guide the process of extracting goal statements from policies have previously been successfully employed to analyze nearly 50 privacy policies in two domains: general e-commerce and healthcare websites [AER02]. Herein, we discuss our third such study in which the focus was on financial institutions.

Our analysis efforts were supported by a web-based Privacy Goal Management Tool (PGMT) developed at North Carolina State University (NCSU). The tool enables analysts to document all identified goals and annotate those goals with

auxiliary information, including the responsible actors, source (origin), privacy taxonomy classification, subject classification, etc. Each goal in the repository is also associated with a unique ID and a description. An important feature of the tool is its goal repository that supports reuse by providing easy access to all previously defined goals. PGMT provides strong management; for example, goals can be searched according to flexible user-defined conditions (e.g. goal ID, keywords, taxonomy, subject, actor, and occurrences). These conditions can be combined together using AND/OR operators. All goals are fully traceable in the repository (i.e. - we can trace in which policies a goal appears) as well as grouping of related policies (i.e. - we analyzed 7 Bank of America policies, each of which can either be analyzed independently or as a group). The tool supports one goal multiple occurrences in a single policy and the recognition of policy goals (strategic goals) and scenario goals (tactical goals) in each policy. Finally, the tool supports automatic multi-user analysis results comparison. For example, each analyst classifies the goals separately and the tool can automatically check the differences in their classification results for their resolution.

4.2 Discussion

As previously mentioned, privacy and security policies are typically not considered as project documentation by software engineers examining existing documentation for requirements derivation. However, these policies are a critical source for requirements derivation because they contain the requirements that concern the security of sensitive information that is particularly vulnerable in web-based and e-commerce systems.

This study serves as preliminary validation of goal-driven process and associated tool support to extract security and privacy requirements from existing policies. Moreover, the preliminary validation of a reusable privacy goal repository and the PGMT has proven successful, and we are now extending the tool and its associated heuristics to better aid analysts in knowing when to make certain design decisions, such as when to use a particular action keyword with the aid of glossaries and heuristics. Our goal repository, which contains 1,036 privacy and security goals, can be used to analyze most financial, healthcare and e-commerce privacy policies. The reusability of the privacy and security goals has proven instrumental throughout the goal-mining and policy analysis process described in this paper. Initially, we had only 136 goals from healthcare policies [AER03]. As our analysis proceeded, we observed that more goals were

¹ Gramm-Leach-Bliley Act of 1999, 15 U.S.C. §§ 6801- 6809 (2000).

² Some institutions post multiple privacy policy documents.

³ Gramm-Leach-Bliley Act of 1999, 15 U.S.C. §§ 6801- 6809 (2000).

⁴ These rankings are available at The Financial Services Fact Book's corresponding website <http://www.financialservicesfacts.org/>.

reusable from the repository as shown in Table 3. We believe the introduction of legislation that requires organizations to explicitly state their practices and adhere to certain government enforced standards has facilitated this increase in reuse.

Table 3: Summary of goals reused during goal mining

Goal Mining	Total # of Goals	# of New Goals	# of Reused Goals	% of Reused Goals
Day 1	99	96	3	3.03%
Day 2	112	106	6	5.36%
Day 3	112	87	25	22.32%
Day 4	62	62	0	0%
Day 5	26	23	3	11.54%
Day 6	258	123	135	52.33%
Day 7	322	162	160	49.69%
Day 8	329	158	171	51.98%
Total	1320	817	503	38.11%

The goal reconciliation process helped to refine the goal set as shown in Table 4 below. Traceability to original policy statements is critical to support examination of goal context. This is particularly important when trying to resolve contradictions and inconsistencies during goal refinement. Examination of synonymous keywords for reconciliation also contributed to pruning of the final goal set.

Table 4. Goal Evolution during Goal Reconciliation

* These goals are from previous studies of healthcare and e-commerce privacy policies. ** 35 goals were eliminated during the goal-mining process mostly because they are duplicate goals.

Activity	# of keywords	# of goals
Before goal-mining	34	136*
Upon completion of goal-mining	132	1159
Before the 1 st goal reconciliation	132	1124**
After the 1 st goal reconciliation	80	1116 (8 deleted, 99 changed)
After the 2 nd goal reconciliation	74	1108 (8 deleted, 71 changed)
After the 3 rd goal reconciliation	68	1094 (14 deleted, 122 changed)
Current repository	68	1042

5. Summary and Plans for Future Work

Requirements elicitation and analysis is a challenging task. In this paper, we discussed an

effective approach for extracting security and privacy requirements from existing policy statements with goal-oriented requirements analysis techniques. Preliminary validation of this approach in analyzing financial privacy policies demonstrated that this technique and associated tool support is effective for identifying reusable security and privacy requirements from policy statements. By employing the library of reusable privacy and security goals, requirements engineers and policy makers can identify potential conflicts and inconsistencies within policy statements and between system requirements and policies to bring policies and requirements into alignment. The reusable security and privacy goal repository is also helpful for analyzing security and privacy policies in other domains. The goals can be further used to standardize Internet policy statements, making it possible to provide a better understanding of privacy policies to customers.

Our process of extracting requirements from existing documents was applied only to security and privacy requirements. We plan to further validate the method to extract other nonfunctional and functional requirements from security policies as well as privacy policies. For example, some organizations have operational procedures for certain tasks, which are documented as handbooks. We believe is possible to apply this technique to analyze these documents and extract functional requirements. This will be part of our future work.

The PGMT that was developed and utilized in this work will be integrated into our Scenario Management and Requirements Tool (SMaRT) [SAA03]. The PGMT's goal extraction, reuse, and refinement capabilities are a perfect extension to SMaRT's current capabilities, and the integration of the two will yield a powerful tool for software engineers as well as policy analysts. To validate the feasibility of goal reuse in our repository, most of which specify privacy and security related requirements in the domain of financial systems, we are currently analyzing privacy policies that govern health care systems. Ultimately, we will provide tools that broadly support reuse across various application domains.

Acknowledgements

This work was supported by NSF ITR Grant #0113792 and a Swiss National Fund (SNF) doctoral grant . The authors thank goal-miners Julie Earp, William Stufflebeam and Thomas Alspaugh, as well as Gene Spafford for his comments.

References

[AE01] A.I. Anton and J.B. Earp. Strategies for Developing Policies and Requirements for Secure Electronic

- Commerce Systems. in *E-Commerce Security and Privacy*, ed. by A.K. Ghosh, Kluwer Academic Publishers, pp. 29-46, 2001.
- [AEP01] A.I. Anton, J.B. Earp, C. Potts and T.A. Alspaugh. The Role of Policy and Privacy Values in Requirements Engineering, *IEEE 5th Int'l Symposium on RE*, pp. 138-145, 27-31 August 2001.
- [AER02] A.I. Anton, J.B. Earp and A. Reese. Analyzing Web Site Privacy Requirements Using a Privacy Goal Taxonomy, *10th Anniv. IEEE Joint RE Conference*, September 2002.
- [AEC03] A.I. Anton, J.B. Earp and R.A. Carter. Precluding Incongruous Behavior by Aligning Software Requirements with Security and Privacy Policies, To Appear: *Information and Software Technology*, Elsevier, 2003.
- [Ant97] A.I. Anton. Goal Identification and Refinement in the Specification of Software-Based Information Systems, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, 1997.
- [BR89] B. Boehm and R. Ross. Theory-W Software Project Management: Principles and Examples, *IEEE Transactions on Software Engineering*, July 1989.
- [CIL02] R. Crook, D. Ince, L. Lin, and B. Nuseibeh. Security Requirements Engineering: When Anti-requirements Hit the Fan, *10th Anniv. IEEE Joint Req'ts Eng. Conf. (RE'02)*, 9-13 September 2002.
- [CIN03] R. Crook, D. Ince, and B. Nuseibeh. Modelling Access Policies Using Roles in Requirements Engineering, To appear: *Information and Software Technology*, Elsevier, 2003.
- [CNY00] L. Chung, B.A. Nixon, E. Yu, and J. Mylopoulos. *Non-Functional Requirements in Software Engineering*. Kluwer Academic Publishers, 2000.
- [Dam02] N.C. Damianou. *A Policy Framework for Management of Distributed Systems*, PhD Thesis, Imperial College, London, 2002.
- [DFv93] A. Dardenne, S. Fickas and A. van Lamsweerde. Goal-Directed Requirements Acquisition, *Science of Computer Programming* 20 (1-2), pp. 3-50, 1993.
- [DP98] R. Dömges and K. Pohl. Adapting Traceability Environments to Project-Specific Needs, *Comm. of the ACM*, 41(12), pp. 54-62, December 1998.
- [Eas93] S. Easterbrook. Domain Modelling with Hierarchies of Alternative Viewpoints, *Intl. Symp. on Req'ts Eng. (RE '93)*, pp. 65-72, January 1993.
- [Fir03] D.G. Firesmith. Analyzing and Specifying Reusable Security Requirements, *2nd International Workshop on Requirements Engineering for High Assurance Systems (RHAS'03)*, pp. 7-11, 9 September 2003.
- [Fon01] P.-J. Fontaine. *Goal-Oriented Elaboration of Security Requirements*, Project Dissertation, Universite Catholique de Louvain, Belgium, 2001.
- [HA03] Q. He and A. I. Anton. A Framework for Modeling Privacy Requirements in Role Engineering, *Proc. of the 9th Int'l Workshop on Req'ts Eng.: Foundations for Software Quality (REFSQ 2003)*, Austria, 2003.
- [He03] Q. He. A Structured Role Engineering Process for Privacy-Aware RBAC Systems, *RE'03 Doctoral Symposium*, Monterey, CA, September 2003.
- [Jac95] M. Jackson. *Software Requirements and Specifications*. Addison-Wesley, 1995.
- [Lam01] A. van Lamsweerde. Goal-Oriented Requirements Engineering: A Guided Tour, *5th Int'l Symp. on Req'ts Eng.*, pp. 249-261, 27-31 August 2001.
- [Lic97] S. Lichtenstein. Developing Internet Security Policy for Organizations, *30th Hawaii Int'l Conf. on System Sciences*, Vol. 4, p350-357, 1997.
- [LYM02] L. Liu, E. Yu, and J. Mylopoulos. Analyzing Security Requirements as Relationships Among Strategic Actors, *2nd Symp. on Req'ts Eng. for Information Security (SREIS'02)*, Raleigh, 2002.
- [LYM03] L. Liu, E. Yu, and J. Mylopoulos. Security and Privacy Requirements Analysis within a Social Setting, *11th Intl. Req'ts Eng. Conf. (RE'03)*, Monterey, 2003.
- [Mof99] J. D. Moffett. Requirements and Policies, *Policy Workshop 1999*, HP-Laboratories, Bristol, UK, 1999.
- [NKF94] B. Nuseibeh, J. Kramer and A. Finkelstein, A Framework for Expressing the Relationships Between Multiple Views in Requirements Specification, *IEEE Trans. on Software Eng.*, 20(10), pp. 760-773, Oct. 1994.
- [PFI99] *Policy Framework for Interpreting Risk in eCommerce Security*. CERIAS Technical Report, Purdue University, 1999.
- [Pot95] Colin Potts. Using Schematic Scenarios to Understand User Needs, *DIS'95: Designing Interactive Systems*, Ann Arbor: MI, pp.247-256, 23-25 Aug 1995.
- [Ram98] B. Ramesh, Factors Influencing Requirements Traceability Practice, *Communications of the ACM*, 41(12), pp. 37-44, December 1998.
- [RHA03] *2nd International Workshop on Requirements Engineering for High Assurance Systems (RHAS'03)*. Eds: C. Heitmeyer and N. Mead. Monterey, CA, pp. 7-11, 9 September 2003.
- [SAA03] W. Stufflebeam, A.I. Anton and T.A. Alspaugh. Scenario Management and Requirements Tool, *RE'03*, Monterey, CA, September 2003.
- [vDM95] A. Van Lamsweerde, R. Darimont, and P. Massonet. Goal-Directed Elaboration of Requirements for a Meeting Scheduler: Problems and Lessons Learnt, *2nd Int. Symposium on Requirements Engineering (RE'95)*, York, UK, pp. 194-203, March 1995.
- [Yu93] E. Yu, Modeling Organizations for Information Systems Requirements Engineering, *Proc. 1st International Symposium on Requirements Engineering*, RE'93, San Jose, USA, 1993.