

The Lack of Clarity in Financial Privacy Policies and the Need for Standardization

Annie I. Antón¹, Julia B. Earp², Davide Bolchini³, Qingfeng He¹, Carlos Jensen⁴, William Stufflebeam¹

¹College of Engineering, North Carolina State University, Raleigh, NC 27695-8207, USA (919.515.5764)

²College of Management, North Carolina State University, Raleigh, NC 27695-7229, USA (919.513.1707)

³Faculty of Communication Sciences, University of Lugano, Lugano TI 6900, Switzerland (+41.91.912.47.13)

⁴College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280, USA

¹{aiananton,qhe2,whstuffle}@eos.ncsu.edu ²Julia_Earp@ncsu.edu ³davide.bolchini@lu.unisi.ch ⁴carlosj@cc.gatech.edu

Abstract

This paper discusses the lack of clarity of 40 online privacy statements from nine financial institutions that are covered by the Gramm-Leach-Bliley Act (GLBA), which states that policies must be “clear and conspicuous.” The study is novel in that it uses two complimentary approaches to analyze the clarity of policies: goal-driven requirements engineering and readability analysis of privacy policy statements based on proven metrics. Our findings show that compliance with the GLBA “clear and conspicuous” requirement by the analyzed policies is at best questionable, and demonstrate that most policies require a reading skill considerably higher than the Internet population’s average literacy level.

1 Introduction

The Federal Trade Commission (FTC) states that a privacy policy is a comprehensive description of a domain’s information practices, located in one place on a website, and may be reached by clicking on an icon or hyperlink [FTC98]. More specifically, privacy policies inform consumers about how organizations collect and use their personal information and theoretically serve as a basis for consumer browsing and transaction decisions. Each policy differs greatly because of the lack of standardization across different industries and organizations. This lack of standardization for expressing business practices affecting privacy presents a daunting learning curve for those wanting to compare different organizations’ policies before deciding in which organization to entrust their personal information.

The Progress and Freedom Foundation (PFF) recently surveyed a random sample of highly-visited websites and found that 83% of those websites posted a privacy policy [PFF02], showing a significant increase from the 1990’s, when only 14% provided any notice regarding information privacy practices [FTC98]. Overall, the PFF study shows that commercial website privacy practices and policies are improving in two major ways: they are claiming they are collecting less information from consumers, and websites policies are increasingly reflecting fair information practices (see [FIP73]).

Although many organizations are now posting online privacy policies, these organizations must realize that simply posting a privacy policy on their website does not guarantee compliance with existing legislation. To date, privacy protection law in the U.S. includes coverage for healthcare data (the Health Information Portability and Accountability Act, HIPAA), information obtained from and/or about children (the Children’s Online Privacy Protection Act, COPPA) and financial data (the Gramm-Leach-Bliley Act, GLBA). The GLBA contains the most extensive financial privacy legislation in U.S. history [FC99] and serves as the focus of this paper. It was enacted in 1999 and became effective on July 1, 2001. 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 institution¹.

Though many organizations are taking strides to improve their privacy practices, and consumers are becoming more privacy-aware, it remains a tremendous burden for users to manage their privacy. Consider for example, that organizations only have to define one policy, but users are expected to read the policy of every website with which they interact. Anecdotally, users rely on the assumption that others must have already read the website's privacy policy and identified any potential vulnerabilities. Moreover, as we discuss, these policies generally require the reading equivalency of someone with at least two years of college education to be fully comprehended. These factors contribute to most website users not attempting the challenging task of reading and understanding the policies themselves [EB03]. Additionally, existing standards and tools which are meant to help end-users manage their privacy, such as P3P (Platform for Privacy Preferences Project)² and Privacy Bird³, force-fit users' preferences and concerns into defined categories, limiting their options. To bridge the gap between website privacy policies and consumer understanding, there is a need to standardize the terms used to express organizational privacy practices thereby increasing their clarity.

We have completed an in-depth analysis of online privacy practices as expressed in documents such as privacy policies and terms of use from GLBA-covered institutions. We have evaluated the level of clarity of these privacy documents with the intent to develop a more standard vocabulary for expressing privacy statements; this can be used to increase the clarity of website privacy statements. The results of this analysis are important because they help policy makers and consumers identify practices that potentially threaten consumer privacy as well as guiding software engineers in developing systems that are aligned with their organization's privacy policies. Our findings also demonstrate that most privacy policies require a reading skill considerably higher than the Internet population's average literacy level. The social impacts of our study are addressed throughout our discussion.

2 Analyzing Financial Privacy Policies

For this study, 40 online privacy statements from nine GLBA-covered institutions⁴ were examined using goal-driven requirements engineering and text readability metrics. 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)⁵. The financial privacy policies we examined, and which serve as the focus of this paper, were in force during June of 2003.

We employed a content analysis technique, goal-mining (the extraction of pre-requirements goals from post-requirements text artifacts) [AER02], to analyze website privacy documents. Our goal-mining efforts were conducted in the spirit of Grounded Theory, in which existing phenomena are analyzed 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 analyzed [GS67]. Therefore, the work presented in this paper is not based on 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 website's privacy practices.

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

² <http://www.w3.org/P3P/>

³ <http://www.privacybird.com/>

⁴ As explained later, some institutions post multiple privacy policy documents.

⁵ These rankings are available at The Financial Services Fact Book's corresponding website <http://www.financialservicesfacts.org/>.

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.

2.1 Mining Policies

Goal mining refers to the extraction of goals from data sources (in this case, privacy policies) [AER02] by the application of goal-based requirements analysis methods [Lam01]. The extracted goals are expressed in structured natural language. We begin the goal-mining process by first exploring any privacy policies (or requirements specifications and design documentation in the case of more traditional requirements work), 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. These goals are documented and annotated with auxiliary information, including the responsible actors, in a web-based Privacy Goal Management Tool (PGMT) developed at North Carolina State University (NCSU).

To identify goals, each statement in a privacy policy is analyzed by asking, “*What goal(s) does this statement or fragment exemplify?*” and/or “*What goal(s) does this statement obstruct or thwart?*” All action words are possible candidates for goals. Goals in privacy policies are thus also identified by looking for useful keywords (verbs). The identified goals are worded to express a state that is true, or the condition that holds true, when the goal is realized. Consider Privacy Policy #1 from the Bank of America Privacy Policy:

Privacy Policy #1: *Employees are authorized to access customer information only when they need it, to provide you with accounts and services or to maintain your accounts.*

By asking the goal identification questions, we identify the goal: G_{144} : PROVIDE access to CI (Customer Information) to authorized personnel with authorized roles from Privacy Policy #1⁶. The goal’s actor is the Institution and it is an integrity/security goal according to our taxonomy as discussed in Section 2.3. This same goal was then reused during the analysis of *Privacy Policy #2* taken from the Citigroup Privacy Promise for Consumers:

Privacy Policy #2: *We will permit only authorized employees, who are trained in the proper handling of customer information, to have access to that information.*

Our analysis was supported by the PGMT that assists analysts in the goal mining, reconciliation and management processes. The PGMT maintains a goal repository for use in continuing analyses of policies and other documents from which goals can be derived. Each goal in the repository is associated with a unique ID, a description, a responsible actor, its sources and a privacy taxonomy classification. All goals are fully traceable to the policies in which they are stated and are distinguished as either policy goals (strategic goals) or scenario goals (tactical goals) in each policy. The tool greatly improved the efficiency of the privacy policy mining and analysis process over our previous two studies, which were conducted without tool support. A total of 1,032 different goals were extracted from the 40 examined policies using goal-mining identification heuristics. The PGMT was used to keep track of the number of goal occurrences in each policy (see column 6 in Table 1).

We now discuss how the goal refinement and reconciliation process aid in making privacy policies more clear.

⁶ The careful reader will note that there is no explicit mention of authorized role in Privacy Policy #1; we are interpreting the statement in the bank’s favor for purpose of this analysis, making the assumption about the definition of “need”.

Table 1: Summary of all GLBA-covered financial institution and respective privacy policies analyzed for this study.

	Policy Document	Protection Goals	Vulnerabilities	Unclassified	Total Goals	FRES	FGL
BANKS							
Bank of America	Overview	4	0	0	4	36.50	11.15
	Privacy Policy	73	55	4	132	32.20	12.99
	Online Practices	23	19	5	47	41.60	12.45
	Information Security	29	2	12	43	39.10	11.68
	Identity Theft	18	0	4	22	43.90	10.42
	Accounts & Services	4	4	5	13	41.70	11.72
	FAQ (State: NC)	63	32	3	98	43.40	11.29
	Subtotal	214	112	33	359	40.10	11.77
Citibank	Citigroup Promise	21	15	0	36	24.00	15.65
	Citi com Online Data Policy	8	21	2	31	31.30	15.38
	Citi MyAccounts Promise	19	14	1	34	31.30	14.09
	Citi MyAccounts Notice	12	14	1	27	30.10	15.54
	Citi Terms of Use	4	11	2	17	21.60	17.03
	Subtotal	64	75	6	145	28.50	15.47
Wachovia	Privacy Statement	60	28	10	98	29.70	13.32
	Internet Privacy	20	40	4	64	35.20	13.91
	Privacy Statement FAQ	44	25	7	76	30.20	13.25
	Fraud Prevention	57	1	6	64	39.40	11.76
	Security Statement	25	1	0	26	35.90	13.07
	Online Banking & Billpay	20	2	10	32	42.00	13.04
		Subtotal	226	97	37	360	35.00
INSURANCE COMPANIES							
Allstate	Privacy Statement	55	23	11	89	41.10	11.68
	Terms of Use	9	12	8	29	31.40	15.86
	Subtotal	64	35	19	118	38.10	12.89
American Int'l Group	Privacy Policy	8	9	1	18	27.00	17.24
	Conditions of Use	1	15	4	20	25.20	16.90
	Subtotal	9	24	5	38	26.10	16.97
State Farm	Privacy Principles	3	2	1	6	23.10	14.93
	Privacy Policy Customers	15	23	2	40	40.80	12.46
	Privacy Policy Consumers	10	9	2	21	26.30	14.56
	Privacy and Security	8	8	0	16	37.20	13.14
	Privacy Policy for PHI	24	41	3	68	36.20	13.50
	State Privacy Rights	1	0	0	1	33.50	17.08
	Privacy Policy FAQ	70	34	8	112	48.20	10.78
	Terms of Use	11	8	11	30	31.00	13.98
	Subtotal	142	125	27	294	40.10	12.41
SECURITIES FIRMS							
Goldman Sachs	Privacy Policy	33	28	4	65	25.70	15.56
	Terms & Conditions of Use	0	4	0	4	22.50	18.72
	Subtotal	33	32	4	69	24.10	17.15
Merrill Lynch	Global Privacy Pledge	30	34	5	69	30.90	14.84
	Online Privacy Statement	11	15	3	29	37.10	12.93
	Legal Info	2	6	0	8	25.20	17.27
	Subtotal	43	55	8	106	29.50	15.55
Morgan Stanley	Privacy Pledge	5	0	1	6	28.90	14.20
	US Individual Investor PP	23	41	3	67	28.90	15.76
	Internet Security Policy	8	10	0	18	31.40	14.79
	ClientServe ISP	18	3	3	24	35.90	13.30
	Terms of Use	10	28	1	39	26.10	17.32
	Subtotal	64	82	8	154	29.50	15.70
TOTAL		859	637	147	1643	33.07	14.11

2.2 Achieving Clarity Through Reconciliation

Goal refinement in requirements engineering consists of resolving ambiguities, redundancies and inconsistencies that exist within the goal set, for eventual goal operationalization into a requirements specification. The goal refinement process is guided by heuristics. For example, goals are considered

synonymous if their intended end states are equivalent, or if they mean the same thing to different stakeholders who simply express the goal using different terminology. The analyst must identify these instances because tool support for this activity is currently not available. 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 (e.g. actor, source, etc.). 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).

We are also codifying heuristics for how to reconcile hyperonymous (broader in meaning) and hyponymous (narrower in meaning) goals that are expressed using different keywords. For example, consider G_{644} : PREVENT sharing CI with telemarketers and G_{629} : AVOID selling/sharing CI with telemarketers/other companies for marketing purposes. Two heuristics were applied to these goals during the goal refinement process. These two goals were initially deemed synonymous with respect to their keywords. The keywords AVOID and PREVENT both mean to keep from happening. However, PREVENT implies an ensurance via the existence of a physical or procedural mechanism. In the case of goals G_{629} and G_{644} , there was no mention of mechanisms to keep such sharing from happening, thus the keyword AVOID was maintained. Additionally, because goal G_{629} is hyperonymous with respect to the argument of goal G_{644} , this goal (G_{629}) was maintained. Similar distinctions are made among different keywords. For example, the keywords PROVIDE and INFORM both make something available. The distinction encoded in the PGMT is that PROVIDE refers to the provision of services or specific functionalities, whereas INFORM refers to the imparting of knowledge or information.

Each keyword in the repository has been formally defined to ensure consistent use of the keywords throughout the analysis process. The PGMT currently contains a list 57 keywords that are commonly found in Internet privacy policies (see Table 2).

Table 2: Privacy Policy 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			

Additionally, we have codified a set of rules that suggest keywords used in different organizations' policies that should be reconciled with keywords in the repository. As shown in the example above, different institutions' policies express the same practices using different terms. These differences require end-users to calibrate their understanding of different website policies, imposing a tremendous (and unfair) burden on the end-user. It is not surprising, therefore, that end-users may find it difficult to trust the practices expressed by companies in their policies. From a systems perspective, a standard vocabulary would be extremely beneficial because it would enable us to formalize the informal, ambiguous and sometimes inaccurate statements found in Internet privacy policies. A standard vocabulary would also facilitate the development of tools to help policy makers standardize their policies across institutions and help consumers more readily understand what is said in website policies. Tools employing such a

vocabulary would make privacy policies more clear to consumers, enabling them to compare privacy practices and make more informed decisions concerning to whom they entrust their NPI. The 57 formally defined keywords provide a useful, extensible vocabulary for examining privacy policies because they standardize what different policies express with different terms in a manner that can increase visibility and understanding for end-users.

After this reconciliation process was completed, the final goal set contained 910 goals extracted from financial policies. The goal repository in the tool, however, contains a total of 1,042 goals because it also includes 132 goals from our previous analysis of health care privacy policies.

2.3 Privacy Protection or Privacy Vulnerability?

In general, policies should express the ways in which they protect NPI but, according to the Fair Information Practice Principles [FIP73], institutions should also inform their customers of potential vulnerabilities that may threaten one's privacy. *Privacy protection goals* express the desired protection of consumer privacy rights, whereas *privacy vulnerabilities* describe practices that potentially threaten consumer privacy. These two dimensions, privacy protections and privacy vulnerabilities, are extensively intertwined, but are not clearly separated in website privacy policies. However, from an end-user's perspective, it is important to help end-users clearly distinguish between practices that protect one's privacy and practices that introduce potential vulnerabilities. For this reason, we have developed a privacy goal taxonomy in which privacy statements are broadly classified as either privacy protection goals or privacy vulnerabilities [AER02, AE03]. The goal taxonomy provides a framework of understanding for relevant privacy issues concerning the treatment of customer data by a financial institution. Some of these issues have already been addressed by structured approaches that may help provide an enhanced awareness of the different levels of privacy responsibilities [IAM03].

Privacy protection goals are subdivided into five categories [AER02]: notice and awareness; choice and consent; access and participation; integrity and security; and enforcement and redress. Notice and awareness goals reflect ways in which consumers should be notified and/or made aware of an organization's information practices before any information is actually collected from them. Choice and consent goals reflect ways in which an organization ensures that consumers are given options as to what personal information is collected, how it may be used and by whom. Access and participation goals reflect ways that organizations 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. Integrity and security goals reflect the ways in which an organization ensures that data is both accurate and secure. Finally, enforcement and redress goals reflect ways in which an organization enforces its policies.

Privacy vulnerabilities relate to existing threats to consumer privacy and represent statements of fact about existing behavior that may be characterized as privacy invasions. There are *obvious* privacy invasions and *insidious* privacy invasions. Obvious privacy invasions are those that consumers are acutely aware of or about which they eventually become aware. Specifically, there exist three kinds of obvious privacy invasions: direct collection for secondary purposes, personalization and solicitation. Insidious privacy invasions are those about which the end-user may not be aware. There are several kinds of insidious privacy invasions: monitoring, storage, aggregation and transfer of information. Some may argue that if a consumer opts in to being monitored, the following practices cannot possibly be insidious: having one's usage patterns or other data aggregated with that of other consumers or having one's NPI stored in a database and/or shared with third parties. However, collection of such information presents the potential for grievous invasions of privacy simply because of the vulnerability presented by its existence and consequently the potential for abuses.

We categorized all 910 goals according to these taxonomy classes by carefully considering each policy goal's actual intent. This activity enabled us to compare the number of privacy protection goals and privacy vulnerabilities in each policy. Consider the following goal: STORE credit card info securely (encrypted, separate DB). At first glance the keyword, STORE, would have led this goal to be classified as a vulnerability (information storage). However, using the taxonomy enables one to identify this goal as a protection goal (integrity/security). Some goals were not truly relevant to privacy or privacy-related functionality and were unclassified for purposes of this study (see Table 1). For example, the Wachovia goal, G_{548} : MAINTAIN efficient service, reflects a general declaration of commitment by the company that expresses neither a privacy protection goal nor vulnerability.

In our previous data analysis of healthcare privacy policies, we hypothesized that the number of protection goals in a healthcare privacy policy would be greater than the number of vulnerabilities; this hypothesis was confirmed (p-value=0.0089) [AE03]. For our current analysis, we also hypothesized that the number of protection goals in a financial privacy policy would be greater than the number of vulnerabilities; this hypothesis was also confirmed. When comparing the number of protection goals to the number of vulnerabilities in each financial policy (see Table 1), the t-test analysis revealed a statistically significant difference (p-value=0.02215) between them. The number of protection goals for a given website was observed to be, on average, greater than the number of vulnerabilities in that website. This was the case with 22 of the 40 examined financial website privacy policies. The p-value for the financial analysis is not as significant as that of the healthcare analysis. We could possibly attribute this to the presence of financial regulation and the lack of healthcare regulation at the time of our healthcare analyses. Though there were a higher number of protection goals in these policies, nearly 40% of the overall goals found expressed a privacy vulnerability.

2.4 Identifying Policy Conflicts

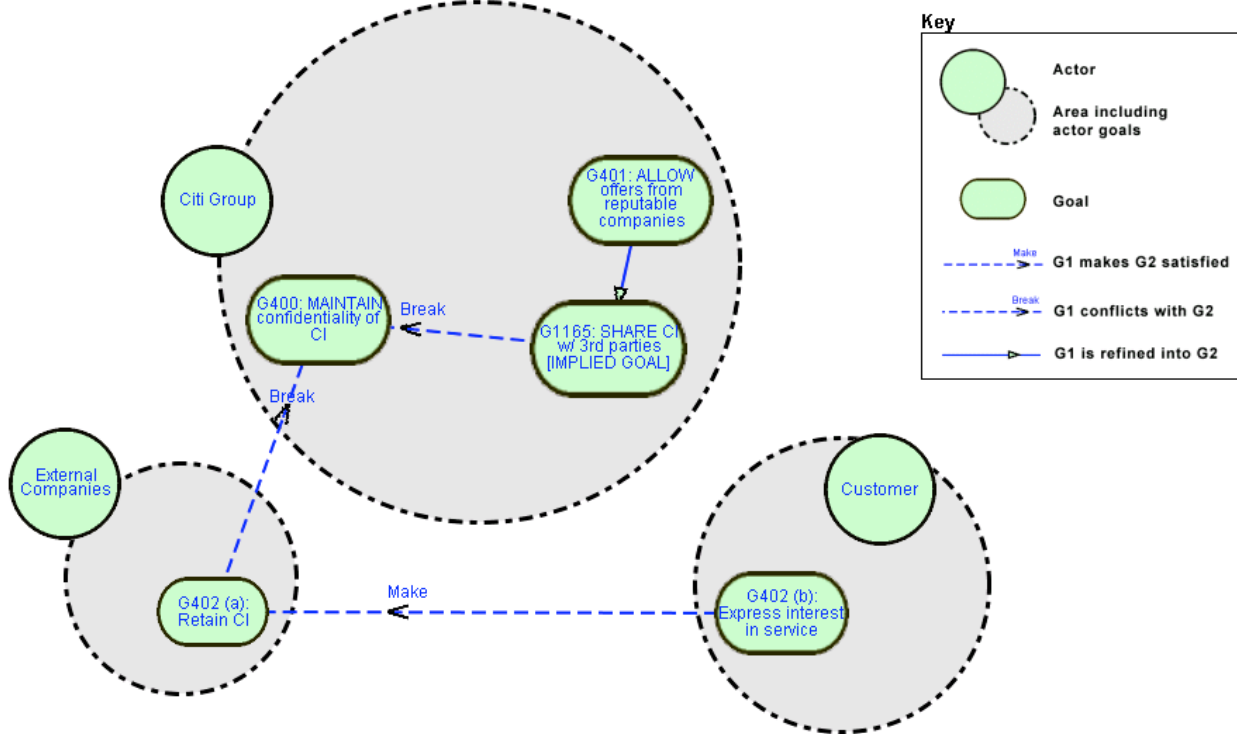
The goal-mining heuristics coupled with the taxonomy provides a basis for identifying conflicting statements within a privacy policy. To more vividly express the relationship between policy statements, we used the i^* notation, which supports the modelling of goals, their semantic relationships and the corresponding stakeholders [Yu93]. For example, the goal mining activity identified a potential conflict, shown in Figure 1, 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 1, 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.

There may be conflicts that are even more evident than the one shown in Figure 1. For example, the Bank of America Privacy Policy had two consecutive goals: G_{231} (HONOR customer privacy preferences once specified by customer) and G_{232} (ALLOW six to eight weeks for privacy preferences to be fully effective). This situation, portrayed in Figure 2, presents an interesting question: How can customer preferences be consistently honored, when there is a six to eight week period in which the institution does not make those preferences effective? This is one of many questions that

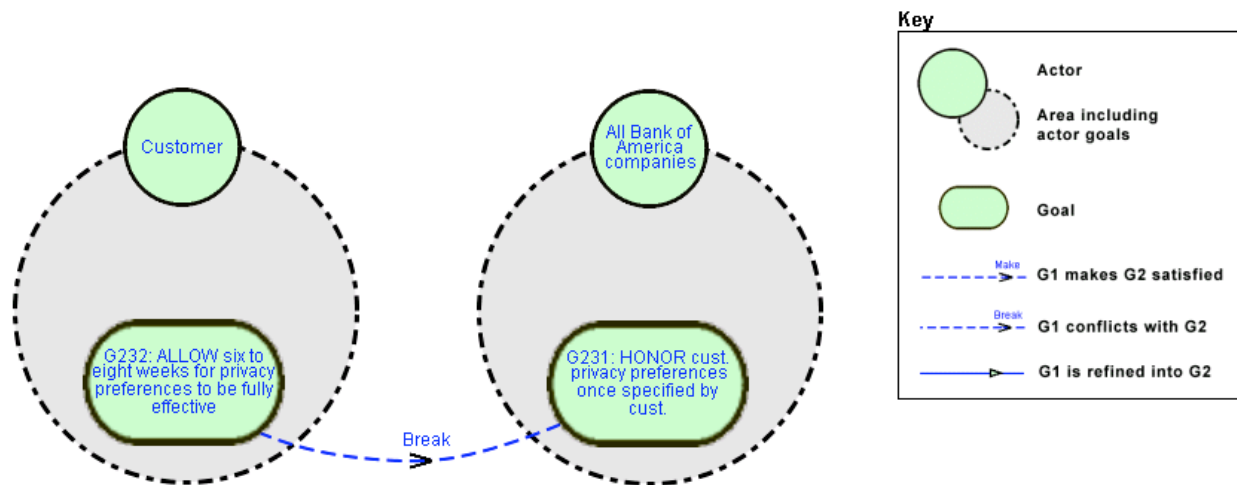
arise when considering such conflicts.

Figure 1: Potential Conflicts in Citigroup Privacy Promise modelled with i^* .



Customers should be aware that when reading strong commitments expressed by an institution in their privacy policy, there may also be other easily overlooked potential vulnerabilities that undercut those commitments. Goal-based analysis exposes these kinds of conflicts that an end-user may not discern based upon a superficial reading. Another benefit of this form of analysis is that it enables policy makers to have deeper insights into the semantic relationships that form the meaning of their current policy. Based on these insights, policy makers can devise strategies to resolve conflicts and inconsistencies, resulting in higher quality policy statements that better conform to the GLBA's requirement for clarity.

Figure 2: Potential Conflict in Bank of America Privacy Policy modelled with i^* .



3 Are Privacy Policies Readable?

GLBA-covered institutions “must provide a clear and conspicuous notice that accurately reflects [the institution’s] privacy policies and practices,”⁷ where clear and conspicuous notice is defined as “a notice that is reasonably understandable and designed to call attention to the nature and significance of the information in the notice.”⁸ This clarity requirement is critical because consumers are increasingly becoming privacy-aware and are more interested in knowing how to protect their own privacy. Therefore, it is useful to evaluate privacy policies in a manner that examines their clarity and readability.

From a practical standpoint, what constitutes a clear notice hinges on the language used in that notice, and whether it is reasonable to expect the target audience to understand it. Although the GLBA provides useful examples, the examples are not exhaustive and it is largely up to the financial institution to make a subjective judgment with regard to the clarity of their notices. A more objective measure entails considering the reading and comprehension skills of the target audience, as well as performing a readability analysis of the notices to determine whether they are clear enough to be understood.

Literacy and education are closely linked to income and, as computers and Internet access are still relatively expensive, we expect the online adult U.S. population to have a higher than average education and literacy rate. Thus, we employ the education level statistics for the adult U.S. population of Internet users rather than that of the general population (see Table 3).

Table 3: Education and Internet Use.

Educational Level	General Population (GP)		% GP Online	Internet Population	
	# People (in millions)	% of Total Population		# People (in millions)	% of Internet Population
Less Than High School	27.5	15.5	12.8	3.5	3.8
High School Diploma / GED	57.4	32.4	39.8	22.8	24.5
Some College / Associate D.	45.4	25.6	62.4	28.3	30.5
Bachelors Degree	30.6	17.7	80.8	24.7	26.6
Beyond Bachelors	16.3	9.2	83.7	13.6	14.6

Source: 2002 National Telecommunications and Information Administration report [NTI02]

We know from the 2000 U.S. Census that 15.5% of the U.S. population over the age of 25 has less than a high school education, and only 26.9% of the population has a bachelor’s degree or higher [NTI02]. Additionally, 65.6% of the U.S. population has access to a computer, 53.9% is now online, and gender differences are largely nonexistent in terms of Internet adoption [NTI02]. The average education of the adult U.S. population is 13.5 years of education⁹, whereas the Internet population has an average 14.4 years, approximately the equivalent of an Associate degree, or two years in college. It is important to note that though Internet users, on average, are more educated, 28.3% of adult U.S. Internet users have the equivalent of a high school education or less. As more Americans go online, the Internet population will more closely match the general population.

With a greater understanding of the population’s literacy level, we can now examine whether privacy notices are clear, as required by the GLBA. The most commonly used method is to employ a standardized, statistical readability metric that allows an objective evaluation and simple comparison between notices. The Flesch Reading Ease Score (FRES) [Fle49] is a metric for evaluating more complex texts and is often used both to evaluate school texts as well as legal documents. Flesch, like other metrics, gives an approximate score for the difficulty of a text. Though no metric is universally employed or accepted, the Flesch metrics have been commonly accepted benchmarks for decades.

The FRES rates texts on a 100-point scale, where higher scores signify simpler texts. This score is

⁷ See 16 C.F.R. Part 313.4(a)

⁸ See 16 C.F.R. Part 313.4(a)

⁹ Average calculated with following values: All with less than high school education = 11th grade, high school = 12th grade, some college = 14th grade, college = 16th grade, postgraduate = 17th grade.

computed by looking at the average number of syllables per word, as well as the average sentence length. Longer words and sentences are more difficult to read and therefore produce a lower FRES. The Flesch Grade Level (FGL) determines the U.S. grade-school equivalency level of a text and is also based on the average number of syllables and sentence length. The metrics can be computed as follows:

Flesch Reading Ease Score (FRES):

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

Flesch Grade Level (FGL):

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

A number of tools calculate the FRES, including Microsoft Word¹⁰, which we used to evaluate the policies discussed herein. MS Word also calculates the FGL up to the 12th grade; for more complicated texts, we calculated these scores manually using the formula above.

Our survey found the average FRES to be 33.1 (SD=6.8), and the average grade level required to read these policies is 14.1 (SD=2.1) (see Table 1). This average is lower than the average education level of the U.S. adult Internet population, but higher than that of the general population. The most difficult policy had a FGL of 18.7 (Goldman Sachs Terms & Conditions of Use), the equivalent of a postgraduate education. The most readable notice (as rated by the FGL) required the equivalent of 10.4 years of schooling to understand (Bank of America Identity Theft).

Of the 40 policies examined, 8 require the equivalent of a high school education or less (12 years), 13 require the equivalent of some college education (12-14 years), 12 require 14-16 years of schooling and 7 require the equivalent of a postgraduate education (more than 16 years). Moreover, of the nine institutions from whom these policies were drawn, six of them had at least one policy document requiring the equivalent of a postgraduate education. This means that even though the average grade level equivalent is 14.1, full disclosure, or a full understanding of what two thirds of these organizations are doing, is perhaps only available to one sixth of the adult U.S. Internet population.

The fact that 28.3% of individuals in the adult population are likely unable to understand these privacy policies challenges the principle of clear notice; none of the nine institutions examined herein live up to the intent of the GLBA clarity requirement. Bank of America comes close, though 2 of its 7 notices require more than a high school education to understand, but none require more than 13 years of schooling. Setting the bar more leniently at the equivalent of a college education (excluding 58.8% of the population), only 3 of the 9 organizations examined (Allstate, Wachovia and Bank of America), manage to give appropriate notice. In either case, it is clear that we are a long way from meeting this GLBA requirement.

We also examined the relationship between the length (in words) of the policies and their complexity. Users are often put off by lengthy documents and notices, but are these policies in fact any worse? There proved to be no correlation between the length of the policy and the FRES or the FGL (correlation coefficients 0.303 and 0.251 respectively). Similarly, there was no relationship between the number of goals expressed in the individual policies and the FRES or the FGL of those policies (correlation coefficients 0.091 and 0.103 respectively). This means that the complexity of the policies we studied was independent of the amount of information conveyed, or how verbose the policies were in conveying that information. Longer policies did not provide more information, but on average, they were no harder to read than their shorter counterparts either.

4 Conclusions

Consumer privacy concerns can pose a serious impediment to expanded growth of e-commerce and Internet usage. The process of applying our goal-mining heuristics in conjunction with our privacy

¹⁰ <http://www.microsoft.com/office/word/>

protection/vulnerability taxonomy allows both vendors and consumers to analyze and compare Internet privacy policies. This analysis provides insights into the characteristics of privacy policy content and specifically, the identification of vulnerabilities, ambiguities and conflicts. Additionally, our finding that most privacy policies require a reading skill considerably higher than the adult U.S. Internet population's average literacy level suggests an improvement that can be made by financial institutions to honor the GLBA's requirement for clarity.

Even though we examined policies from a single domain (financial institutions), it was challenging to understand what the different policies meant. This was because certain statements used distinct vocabularies despite the existence of the GLBA, which advocates standardization. This vocabulary difference required us to spend a great deal of time recalibrating our understanding during the goal refinement activities. What is perhaps most concerning is that if we, as experienced privacy policy analysts, encountered difficulties in understanding various policy statements (even with tool and methodological support), it is reasonable to expect that customers will continue to face difficulties if no changes are made. Our research shows there is need for institutions to standardize the way in which they express their privacy practices, and that it is possible to make online privacy policies more clear, benefiting both the institutions and end-users.

Additionally, our preliminary analysis of these privacy policies has revealed ambiguities and conflicts that we are now analyzing so as to codify rules for their identification and subsequent resolution. For example, some policies contain temporal statements that are simple to check for conflicts because they establish constraints with which other policy statements must comply. However, there exists a need for additional, systematic ways to identify conflicts and ambiguities within a policy as well as across an organization's various privacy documents. Our analysis thus far has been limited to identifying conflicts and ambiguities within each single privacy document. Having many policies and goals per institution requires additional analysis to understand, and such an analysis is the focus of the next phase of our research.

Privacy will continue to be important to consumers. We believe that clearly articulated, meaningful privacy policies are also important to good business practice. We are continuing to investigate methods to support the goals of all stakeholders with an interest in clear privacy practices.

Acknowledgements

This work was supported by NSF ITR Grant #0113792 and a doctoral grant from the Swiss National Fund (SNF). The authors wish to thank Thomas Alspaugh for his participation in the goal-mining process as well as Gene Spafford for his comments on this paper.

References

- [AE03] A.I. Antón and J.B. Earp. A Requirements Taxonomy to Reduce Website Privacy Vulnerabilities. To Appear: *Requirements Engineering Journal*, Springer Verlag, 2003.
- [AER02] A.I. Antón, J.B. Earp and A. Reese. Analyzing Web Site Privacy Requirements Using a Privacy Goal Taxonomy. *10th Anniversary IEEE Joint Requirements Engineering Conference (RE'02)*, Essen, Germany, 9-13 September 2002.
- [EB03] J.B. Earp and D. Baumer. Innovative Web Use to Learn about Consumer Behavior and Online Privacy. *Communications of the ACM*, v.46, n.4, April 2003.
- [FC99] L.R. Fischer and Clarke Drysen Camper. Reform Law and Privacy: A Road Map, *American Banker*, No. 19, 1999.
- [FIP73] *The Code of Fair Information Practices*, U.S. Department of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers, and the Rights of Citizens, viii, http://www.epic.org/privacy/consumer/code_fair_info.html, 1973.
- [Fle49] R. Flesch, *The Art of Readable Writing*, Macmillan Publishing, 1949,
- [FTC98] *Privacy Online: A Report to Congress*, <http://www.ftc.gov/reports/privacy3/>, Federal Trade Commission, June 1998.
- [GS67] B.C. Glaser and A.L. Strauss. *The Discovery of Grounded Theory*. Chicago: Aldine Publishing Company, 1967.

- [IAM03] L. Ishitani, V. Almeida and W. Meira. Masks: Bringing Anonymity and Personalization Together, *IEEE Security & Privacy*, 1(3), pp. 18-23, 2003.
- [Lam01] A. van Lamsweerde. □Goal-Oriented Requirements Engineering: □A Guided Tour, *IEEE 5th International Symposium on Requirements Engineering (RE'01)*, Toronto, Canada, pp. 249-261, 27-31 August 2001.
- [NTI02] National Telecommunications and Information Administration. A Nation Online: How Americans Are Expanding Their Use of the Internet <http://www.ntia.doc.gov/ntiahome/dn/> Washington, D.C. February 2002.
- [PFF02] W.F. Adkinson, J.A. Eisenach and T.M. Lenard. *Privacy online: A Report on the Information Practices and Policies of Commercial Web Sites*. Washington, DC: Progress & Freedom Foundation, 2002. Downloaded July 18, 2003: <http://www.pff.org/publications/privacyonlinefinalael.pdf>.
- [Yu93] E. Yu. Modeling Organizations for Information Systems Requirements Engineering, *IEEE 1st International Symposium on Requirements Engineering (RE'93)*, San Jose (CA), 1993.