A Survey of Privacy Policy Languages

Ponnurangam Kumaraguru
Carnegie Mellon University
ponguru@cs.cmu.edu

Lorrie Faith Cranor
Carnegie Mellon University
lorrie@cs.cmu.edu

Jorge Lobo
IBM Research, USA
jlobo@us.ibm.com

Seraphin B. Calo
IBM Research, USA
scalo@us.ibm.com

ABSTRACT

Most consumers are sensitive to privacy issues when conducting business online. Protecting information by enforcing security and privacy practices internally is a way for organizations to increase business by building trust with such consumers. They can express their privacy practices as policies in a human readable format to help consumers make informed decisions. Many privacy languages are available for representing policies, but they tend to use formats convenient to their implementations, and there is no single framework or metric to analyze and evaluate the effectiveness of these languages. In this research, we are interested in succinctly summarizing the literature available on privacy policy languages; providing an account of the features, characteristics and requirements of the languages; and, describing a comprehensive framework for analysis. We expect our results to aid implementers in choosing an existing language and to provide guidelines for building languages in the future. We expect this research to be a starting point towards developing frameworks and metrics for analyzing privacy policy languages.

1. INTRODUCTION

With the growth in Internet usage and an increase in online business, consumers expect high levels of online privacy [5]. Consumers frequently mention lack of trust as one of the reasons for not purchasing from the Internet [10]. From the organization's perspective, the need to protect consumer privacy and to comply with privacy legislation is a growing concern [9]. In particular, providing such protection is becoming an important function for the IT security management teams of many organizations. If organizations do not follow efficient privacy practices, consumers may move away from them and legal consequences may arise. Thus, the formalization of an organization's promises regarding privacy practices into privacy policies is an essential aspect in the management of customer relationships. Organizations express their internal privacy practices as statements in the privacy policies. Consumers are able to analyze the organization's stated commitment towards protecting consumers' privacy through these privacy policies.

Different types of languages are available to represent the human-readable policies in more precise and computer compatible formats. Some languages are designed to help organizations express their privacy policies in ways that are more amenable to policy enforcement, and some languages are designed to help users define their privacy preferences. These preferences can then be employed to help users make decisions. Every language has its own syntax and mechanisms for implementation. There is no standard metric available that can be used to analyze and compare these languages.

2. LANGUAGES

Privacy policy languages can help with several of the stages involved in managing privacy policies (writing, reviewing, testing, approving, issuing, combining, analyzing, modifying, withdrawing, retrieving and enforcing policy) [14]. Privacy policy languages were designed to express the privacy controls that both organizations and users want to express. Most of the privacy policy languages were designed for specific purposes with specific features and characteristics. Most of the initiatives for designing these languages have occurred in the last ten years. In 1997, the World Wide Web Consortium (W3C) began development of the Platform for Privacy Preferences (P3P) to express website privacy policies in machine-readable format [8]. A P3P Preference Exchange Language (APPEL) was also designed by W3C in 1997 to express an individual's privacy preferences, to query the data represented by P3P, and to make decisions accordingly [6], [7]. CPExchange was developed in 2000 to facilitate business-to-business communication about privacy policies [4]. Later, the industry felt the need for languages to express the internal privacy policies of the organizations themselves. With that goal IBM designed the Enterprise Privacy Authorization Language (EPAL) in 2003 [18]. During the same period a consortium of organizations joined to design the eXtensible Access Control Markup Language (XACML) [14] for expressing both privacy and security policies in a machine-readable format. There were other initiatives such as DPAL [3], and XPref [1] in 2003 and 2004.

Advances in technology and the rapid use of pervasive computing systems created a necessity for protecting context sensitive information transferred through the system (e.g., time of day and location). In 2005, the Internet Engineering Task Force (IETF) started an initiative to design Geopriv, a language that can be used to express policies to provide access on the basis of presence and location information [17].

Privacy policy languages are expected to be fairly simple and small. Therefore they have been designed as light-weight XML markup languages. These privacy policy languages are not expected to perform high-level mathematical operations or complicated flow controls.

To be included in the analysis of this research, the languages had to meet the following selection criteria: (1) the language specification should explicitly address the expression of privacy policies, because we wanted to analyze the expressiveness of privacy policy languages; and, (2) the languages should have been designed for facilitating the process of enforcement. All languages that we plan to analyze can specify privacy / security / management policies in some kind of machine-readable format.

Using the selection criteria we narrowed our analysis to the following languages (arranged in chronological order based on when development work began on them): Platform for Privacy Preferences (P3P) [8], A P3P Preference Exchange Language (APPEL) [7], Customer Profile Exchange (CPExchange) [4], Privacy Rights Markup Language (PRML) [19], XML Access Control Language (XACL) [11], Platform for Enterprise Privacy Practices (E-P3P) [2], [13], Security Assertion Markup Language (SAML) [15], Rei [12], eXtensible Access Control Markup Language (XACML) [14], Enterprise Privacy Authorization Language (EPAL) [18], X-Path Based Preference Language (XPref) [1], Declarative Privacy Authorization Language (DPAL) [3], and Geographic Location / Privacy (Geopriv) [17], [16].

3. ANALYSIS FRAMEWORK

In this section, we describe the framework that we have developed for evaluating the above privacy policy languages. Using this as the basis for the analysis of languages, we developed a framework that consists of the following attributes:

1. *Situation*: Languages have been designed and developed to address privacy management in different situations (e.g., capturing internal enterprise policies rather than user preferences) and the situation has direct influence on the characteristics of the language. We conjecture this attribute to be the most critical attribute in choosing a language.
2. *Representation*: Languages have taken many forms in representing the rules, rulesets, queries, and data. Most of the languages analyzed in this research use XML as their representation language. Some languages have adopted XML in different forms to express the features in the language. They also differ in the vocabularies they use, the basic underlying structure of the language, and in data representation. In this attribute, we plan to discuss the design features implemented in languages for representing data, rules, rulesets, and queries.
3. *Evaluation*: Languages use different techniques for making decisions based on the given rules, rulesets, queries, and data. In most of the languages the evaluation also depends on the order of the different policy components, i.e. rules and rulesets. We also discuss the error handling capabilities of the languages. In this attribute, we plan to discuss the design features of languages based on the evaluation criteria.
4. *Output Schema*: Languages produce different types of results (e.g., allow and deny) according to the evaluation of the rules, rulesets, data and queries. In this attribute, we plan to discuss the implementation of output schema in the languages.
5. *Implementation*: Languages are used in the real-world for different purposes and different deployments (e.g., type of application in which the language can be used - web or other applications). In this attribute, we plan to analyze the implementation details of languages.

4. PRELIMINARY RESULTS

We have done some preliminary analysis on the languages using the framework described in Section 3. In this section, we describe in some detail the results of the analysis of the languages with respect to one of the attributes: *situation*. We plan to analyze and write about other attributes in further publications.

For effective and efficient results, one needs to select the language that best matches the characteristics of the situation. Some common questions asked while choosing the language are: (1) which language will be helpful in representing the privacy policy of an organization in interactions with consumers; And, (2) which language is suitable for representing a user's preferences? Since these are basic and necessary questions, we suspect this attribute to be the most critical feature in choosing a language. On the basis of the situation in which the language can be used, we classify the languages described in Section 2 into the following categories:

1. *Sophisticated Access Control Languages* (SACL): SACL includes languages that were designed and developed based on Role Based Access Control (RBAC). SAC languages are mostly implemented for security policies and maintained by system administrators (e.g., XACML). In addition to representing security policies, SAC languages can also represent privacy policies.
2. *Web Privacy Policy Languages*: This category includes the languages which are helpful in representing some form of human-readable privacy policies on the Internet in machine-readable formats (e.g., P3P).
3. *Enterprise Privacy Policy Languages*: A number of languages have been designed to represent the internal policies of an enterprise, which would help the organization to perform the actions as stated in the privacy policies (e.g., EPAL). These languages are mostly used for internal purposes and they are more fine-grained than the web privacy policy languages.
4. *Context Sensitive Languages*: Since the context information can provide a personalized service, some languages were designed to represent policies that take into consideration context information. The information that is used for providing these services is very sensitive. These languages make use of the semantic web technologies for representing the policies (e.g. Geopriv).

Three of the above categories (except for enterprise privacy policy languages) can be further sub-categorized on the basis of 'whose information is being represented in the machine-readable format?' Using this information we classify the languages further into two categories:

1. *User*: This class of languages helps in representing user's privacy preferences in a machine-readable format (e.g., APPEL, XPref). Through these languages, users can express their preferences in a set of preference rules (called a ruleset), which can then be used by their user agent to make automated or semi-automated decisions regarding the acceptability of machine-readable privacy policies.
2. *Enterprise*: This class of languages helps in representing the enterprise privacy policies in a machine-readable format (e.g., XACML and EPAL).

We classified all the languages discussed in Section 2, as presented in Table 1. We can see that the enterprise category has the most entries. This was also expected because organizations

were and are in big need of enforcing privacy policies and so there have been many languages created for that purpose.

Table 1: Classification of the privacy policy languages based on the situation in which the languages can be used. Where E represents the Enterprise and U represent User category. We present the enterprise languages separately.

Sophisticated ACL	Enterprise	SAML, XACML, XACL
	User	XACML
Web	Enterprise	P3P
	User	APPEL, XPref
Enterprise		CPExchange, PRML, E-P3P, EPAL, DPAL
Context Sensitive	Enterprise	Geo-Priv, Rei
	User	Geo-Priv

5. RESEARCH PLAN

We plan to analyze the privacy policy languages described in Section 2 using all of the attributes of the framework described in Section 3. We also plan to analyze the genealogy of the languages to find out which languages have taken features from which other languages and what features have been dropped and why. In addition, we plan on surveying the various language editors, validators, and analysis tools that are available for these languages. The language editors help in expressing the privacy policy in a specific language. The validators are helpful in checking whether the syntax of the policies represented follow a particular standard independent of the implementation. For example the P3P validator checks for the existence of P3P policies on a predetermined location for any given website [16]. The validator also checks for syntactic errors in the policies. We plan on using these analyses to make recommendations for building privacy policy languages. Analysis tools help policy administrators to find potential errors (such as conflicts) during policy creation and modification.

6. CONCLUSION

This study is a first step towards a longer and more in-depth study to provide an understanding of privacy policy languages and several of their features. Unlike previous studies, in this study we plan on analyzing all major existing privacy policy languages using a framework that allows for comparison of these languages along several different dimensions.

In this research, we plan on classifying and summarizing the privacy policy languages. Although, the results from this research will not solve the problem of building efficient languages it will provide guidelines for building new languages. We believe our classification will help researchers to classify any languages designed in the future. We expect our results to aid implementers in choosing a language based on the framework discussed in the paper. The research output will bring together information from several sources discussing the capabilities and pitfalls in languages.

As the technology for representing privacy policies is growing, many open research questions and problems will be interesting to look at. They include but are not limited to: (1) an extensive study of the functionalities of the languages by expressing real-world policies in each language; (2) a study of the usability of language editors and validators; (3) conducting a field study among organizations to find out about the adoption of these privacy policy languages in business; (4) analyzing the expressiveness of privacy languages; and, (5) analyzing the differences between authorization, obligations, and delegation in languages.

7. ACKNOWLEDGMENTS

The authors thank the members of the CMU Usable Privacy and Security (CUPS) Laboratory for their feedback and criticism during discussions on the topics presented in this paper. The authors would also like to thank John Karat and Elisa Bertino for their feedback on the research presented in this paper. This research was partially funded by Carnegie Mellon CyLab and the IBM Open Collaborative Research (OCR) program.

8. REFERENCES

[1] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. An XPath-based Preference Language for P3P. In WWW '03: Proceedings of the 12th international conference on World Wide Web (New York, NY, USA, 2003), pp. 629–639.

[2] Ashley, P., Hada, S., Karjoth, G., and Schunter, M. E-P3P Privacy Policies and Privacy Authorization. In WPES '02: Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society (New York, NY, USA, 2002), pp. 103–109.

[3] Barth, A., and Mitchell, J. C. Conflict and Combination in Privacy Policy Languages. In WPES 2004: Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society (2004).

[4] Bohrer, K., and Holland, B. Customer Profile Exchange (CPExchange) Specification, Version 1.0. Tech. rep., October 20, 2000.

[5] Chellappa, R. K. Consumers' Trust in Electronic Commerce Transactions: The Role of Perceived Privacy and Perceived Security. Under review. Retrieved Sept 13, 2005, http://asura.usc.edu/~ram/rcf-papers/sec-priv.pdf.

[6] Cranor, L. F. Web Privacy with P3P. Sebastopol, CA, USA, 2002.

[7] Cranor, L., Langheinrich, M., and Marchiori, M. A P3P Preference Exchange Language 1.0 (APPEL 1.0). Tech. rep., World Wide Web Consortium, Retrieved June 12, 2005, http://www.w3.org/TR/P3P-preferences/.

[8] Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., and Reagle, J. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. http://www.w3.org/TR/P3P/.

[9] Federal Trade Commission. Privacy Online: A Report to Congress. Retrieved July 25, 2005, http://www.ftc.gov/reports/privacy3/priv-23a.pdf.

[10] Grabner, S., and Kaluscha, E. A. Empirical research in on-line trust: a review and critical assessment. Academic Press, Inc 58, 6 (2003), 783–812.

[11] Hada, S., and Kudo, M. XML Access Control Language: Provisional Authorization for XML Documents. Tech. rep., Retrieved June 21, 2005, http://www.trl.ibm.com/projects/xml/xacl/xacl-spec.html, May 2000.

[12] Kagal, L. Rei: A Policy Language for the Me-Centric Project. HP Laboratories (2002). Retrieved Aug 1, 2005, http://ebiquity.umbc.edu/_file_directory_/papers/57.pdf.

[13] Karjoth, G., Schunter, M., and Waidner, M. Platform for Enterprise Privacy Practices: Privacy-enabled Management of Customer Data. In Proceedings of the 2002, 2ndWorkshop on Privacy Enhancing Technologies, (2002), pp. 61–83.

[14] Moses, T. eXtensible Access Control Markup Language (XACML) Version 2.0. Tech. rep., Oasis, Retrieved June 17, 2005, http://xml.coverpages.org/XACMLv20CD-CoreSpec.pdf, 2004.

[15] OASIS. OASIS Specifications - SAML V 2.0. Retrieved June 26, 2005, http://www.oasis-open.org/committees/tc_home.php?wg_abbrevecurity#samlv20.

[16] P3P Validator. http://www.w3.org/P3P/validator.html. Retrieved March 23, 2007.

[17] Schulzrinne, H. A Document Format for Expressing Privacy Preferences. Tech. rep., The Internet Engineering Task Force, Retrieved June 12, 2005, http://www.ietf.org/internet-drafts/draft-ietf-geopriv-common-policy-04.txt.

[18] Schunter, M., Ashley, P., Hada, S., Karjoth, G., and Powers, C. Enterprise Privacy Authorization Language (EPAL 1.1). Tech. rep., International Business Machines Corporation (IBM), Retrieved June 17, 2005, from http://www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/index.html, 2003.

[19] Zero knowledge. Privacy Rights Markup Language (PRML) Specification. Tech. rep., Oasis, Retrieved June 17, 2005, http://www.synomos.com/html/EPML/documents/prml-spec.pdf.