Platform for Privacy Preferences (P3P): Current Status and Future Directions

Muyiwa Olurin
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
Oolur089@uottawa.ca

Carlisle Adams
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
Cadams@site.uottawa.ca

Luigi Logrippo
Département d'informatique et ingénierie
Université du Québec en Outaouais
Gatineau, Canada
Luigi@uqo.ca

Abstract— **Web sites usually express their privacy practices in natural language text that is often complex, informal and possibly confusing. The platform for Privacy Preference (P3P) has been proposed by W3C as a technology for expressing privacy practices of web sites in precise, machine readable language. This paper provides an account of the current status of research on P3P and proposes directions for future research, together with some possible solutions. Cloud computing (SaaS), anti-phishing, and mobile applications are some of the aspects that we consider. We claim that P3P and P3P-based techniques have considerable potential to be developed beyond their current status. The challenge is to design formalized privacy policy languages that can enable computers to process the privacy practices of web sites. In this way, many privacy issues, such as filtering web sites, combining their policies, etc., will be able to be dealt with automatically by privacy agents.**

Keywords-Privacy; P3P; Policy Language;

I. INTRODUCTION

Privacy may be defined as the right of individuals to determine for themselves when, how and to what extent information about them is communicated to others [1]. Privacy on the Internet is a growing global concern for both users and web sites. Technologies for web services and applications are becoming more elaborate for collection and analysis of user information [21]. As the amount of Personally Identifiable Information (PII) collected from users is increasing, so the privacy concerns of users also increase, as do the possibilities of fraudulent use of the information [7]. Many users want to be aware of how web sites will manage their personal data because they fear potential misuse of their PII. Users specifically worry that "web sites will sell their PII [16] data to third parties, clog their e-mails with spam, place persistent cookies on their computers or enable third parties to do so"[21]. In order to encourage users to participate in the online services and restrain web sites from data misuse, the legal and self-regulatory requirements in many countries request commercial web sites to post their privacy statements (often called privacy policies) on their web sites [4]. The privacy policies are meant to explicitly notify users about the data handling practices of web sites, and should be represented in a clear and precise manner that users can understand [4, 21]. Privacy statements posted on web sites are often lengthy, difficult to read, and complex to comprehend. Privacy statements have been shown to be time consuming to read, and are sometimes error-prone [4, 7, 8]. Users are usually not interested in reading them because they find them too legalistic or too complex. Users would be required to have at least some college education to understand their complexity and sentence structure [4]. Also, research [4, 20] has found that many privacy statements do not address data handling practices that concern users. Rather, the statements seem to have been written in order to protect the web site operators against lawsuits. In other words, these policies posted in natural languages do not reduce user privacy concerns. Users also have no easy way to verify whether the actual usage of their data is compliant with the privacy statements. In order to address the problem, The World Wide Web Consortium (W3C)[1] proposed a privacy policy specification language to make privacy policies more readable and clearer for users, while also being machine-processable. This language is called the Platform for Privacy Preferences (P3P) [3, 10].

A. Platform for Privacy Preferences

P3P is a privacy markup language that "enables web sites to encode their natural language privacy statements or privacy practices in a machine-readable XML format known as a P3P policy (sometimes called a P3P file)" [1, 9, 10]. A P3P Policy is encoded in an XML-based vocabulary and syntax and expresses the privacy policies or data handling practices of websites. A P3P Policy is a collection of statements, each of which describes the purpose, the retention, and the recipient of a piece of collected data [22]. A P3P policy consists of two main definitions (elements): the root element <POLICY>, and the body element <STATEMENT> [10]. The XML file (Policy file) that describes the privacy practices of a web site may be generated using some well-known P3P editors (e.g. JRC Policy Workbench, IBM P3P Editor). The P3P Policy of a web site can be automatically retrieved and interpreted by a P3P user agent [8]. P3P policies are transmitted and retrieved from the web server. The transport mechanism is based on the standard Hypertext Transfer Protocol / Secure (HTTP or HTTPS) [10]. Most common web browsers use these transfer protocols to communicate with the web sites when receiving and sending information over the internet. The transfer process starts with a request for the policy reference file.

[1] W3C is the international organization responsible for developing protocols and guidelines for the World Wide Web (www). http:/www.W3.org

*This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC)

B. P3P User Agent

The P3P User agent is a tool or service that translates a web site's P3P policies into a human-readable format. It also checks policies against user preferences in order to give privacy warnings. Microsoft's Internet Explorer 6 (IE6) and Netscape Navigator 7 were the first browsers to adopt the P3P functionality [7, 9]. The combination of P3P policies and user agents is a Privacy Enhancing Technology (PET) solution called a P3P solution. A P3P solution makes automated privacy policy assessment possible [7]. With the P3P solution, users can now define their privacy preferences on web browsers and choose to interact only with web sites that respect their privacy preferences [3]. While the P3P language remains one of the most widely used structured privacy policy languages on the web today [12], previous research and surveys have shown that it has achieved limited adoption by web sites and since then has remained stagnant [8, 10]. Some of the reasons why the global adoption of P3P seems slow are that "there have been few incentives driving web sites to embrace the policy format. There are no privacy regulations that require web sites to express their privacy policies in P3P format" [21]. Also, frequently P3P policies contain syntactic or semantic errors because they are not constructed correctly according to the P3P standard [8, 10], and very few web sites perform maintenance on their P3P policies because there is no law enforcing this [3, 22]. Another reason for low adoption of P3P technology is the fact that the tools do not accommodate languages other than English. In addition, P3P does not have precisely specified semantics and is not defined in such a way as to force specific policies to be attached to specific data items [22], so even when P3P is used, its precise meaning may remain unclear.

II. CURRENT STATUS OF P3P SOLUTION

P3P policies and P3P user agents have been among the research areas of web privacy and security since the early 2000s. Related research has been on recognizing and solving inconsistencies and conflicts in P3P policies or on the combination of multiple policies [12, 16, 19, 22]. Critical problems and ambiguities with P3P technology were studied by Hogben [14], and similar research, with an accent on privacy management, was done by Anton, et al. [4]. These authors pointed out some shortcomings of the P3P specification: the absence of formal semantics is one of the most crucial problems. Yu, et al. [22] have proposed a *data and purpose-centric* relational semantics for P3P in an effort to give unambiguous meaning to the syntactically different expressions of a single policy. The research reported in [4, 22] contributed a solution to the problem. However, the authors focused only on the conflicts and inconsistencies that may occur in the existing P3P language. Other research on the P3P protocol strives to improve the P3P user agent interface, in order to make the user agents more user-friendly and more effective when communicating the summary and policy warnings. Cranor et al. [11] described the Privacy Bird user interface as one of the most effective applications that can summarize privacy policies; the user agent can also be used to specify a user's privacy preferences from the web browser. [17] focused on creating a suitable user interface and clear information visualization; Cranor et al also provided a solution that summarizes the privacy settings along with graphical feedback. Arshad and Mar [5, 18] created Privacy Bird extensions for the Mozilla Firefox and Google Chrome browsers but no new technical features were added to the original Privacy Bird model. Kelley et al. [15] performed a usability study on user agents for privacy preferences; they introduced a "nutritional label design" to display the summary of P3P privacy preferences. They came up with improved P3P user agent software that can produce more meaningful results (in terms of describing the privacy practices of websites) for its users. They took into consideration the interrelationships among entities such as data, purpose, and recipient, and so on, during the interpretation of a website's P3P policy.

III. FUTURE USES AND DIRECTIONS

Until now, the characteristics and benefits of P3P solutions have been discussed based on existing applications. In this section we discuss possible future applications and the evolution of the P3P technology.

A. P3P Policies in Cloud Computing

Cloud computing has become one of the significant computing paradigm shifts in Information Technology. One can expect that in the near future, there will be need for collaborative services of different entities in the cloud. Service providers in the cloud also may want to merge their service with other providers in the same cloud to increase or maintain high productivity and customer satisfaction to end users. However, privacy is known to be a difficult issue for applications deployed in the cloud (SaaS). Presently, some service providers express privacy policies of their traditional web servers using P3P. Multiple service providers in the cloud can collaborate to achieve a single purpose for end users. They might have different privacy policies that apply to their services, so there will be need for end users to know which privacy policies should apply? The situation may be particularly confusing if the policies of the service providers are inconsistent (e.g., one says that it won't keep some information; the other says that it will).

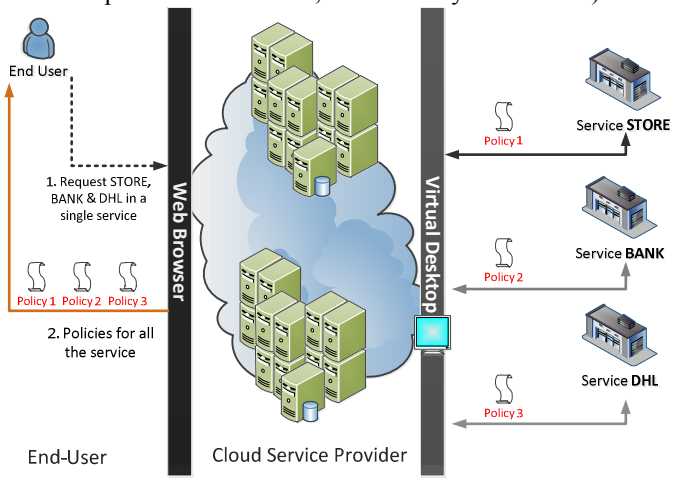

Figure 1. Privacy policies scenario in a Cloud Computing Environment

Figure 1 illustrates a scenario where service providers in the cloud (Online stores, Banks, and Postal services) may have their separate privacy policies (P3P format) declared before

migrating to the cloud. Upon joining the cloud, a user may choose to request a combination of two or more services to accomplish a particular goal. This combination request can result in policy inconsistencies, misunderstanding, and conflicts because P3P was not specified for merging policies. In order to avoid inefficiencies, the merge will have to be executed rapidly and without human intervention as much as possible. If human intervention is necessary, methods to simplify it should be developed. How to deal with these issues is a topic for future research.

B. P3P User agent on Smart Phones

Smart phones provide mobility and flexibility to users; however privacy declaration methods do not exists for smart phone applications. Usually we trust our smart phones more than our laptops because we carry them with us most of the time. As a smart phone becomes a part of our everyday life, many opportunities for privacy leaks may arise, like giving away important and personal information whenever users connect to the internet or download new applications. Smart phone applications or the mobile websites users visit often can track their everyday interactions or movements. Thus, some mobile users will want to know how the application or mobile websites will use the information available to them, e.g., will they install any tailoring cookies or tracking bugs on their smart phones? Will they allow third parties to install new applications? The third parties may not operate in the same way as the web site or application installed; they may have some entirely different privacy policy, and the fact that they have users' personal information may increase the chances of privacy compromise.
For example, users can use smart phones to assess or store their health information (web connection to the clinic) or financial information (mobile banking applications) for easy and fast connection in cases of emergency. The challenge here will be that users who are conscious of the information stored on or accessed via their phones may not want those applications or websites to take or misuse the information. This leads to the following research questions:

- How to create mobile-based privacy user agents that can communicate compact privacy policies of mobile web sites or applications to users.
- How to delegate automatic access control privileges to mobile applications and websites based on user defined privacy preferences.

C. Anti-Phishing mechanism with P3P Protocol

Fraudsters can create fake websites to lure users for the purpose of collecting their data. This type of attack on users is called phishing. Phishing attacks can steal personal identity information such as username, passwords, and credit card details from unsuspecting users by masquerading as trusted entities, such as PayPal sites. The main aim of phishers is to lead users to unsafe websites, and retrieve their personal data. Since the P3P policy on a legitimate website is freely available to the public, phishers can easily copy the P3P policies from the trusted websites and repost them on their fake websites. This can aid phishing attack schemes because P3P user agents that are trusted by the users will be misled by the P3P policy that appears to be associated with the fake site.

The P3P user agent will report a legitimate policy summary, perhaps with a (stolen) trust seal from a third party privacy seal organization. We suggest here a solution that is based on cryptographic techniques: the legitimate site should digitally sign its P3P policy using the private key that corresponds to the public key contained in a valid certificate for the site (possibly, but not necessarily, its SSL certificate). This signature can be done using the XMLSignature format. A P3P user agent, instead of simply downloading a P3P policy from a target site, would also engage in a challenge-response protocol with the site, submitting a one-time challenge (i.e., a random number that has not been used before) and receiving in return a digital signature on that challenge. If the user agent is able to verify this signature with the same public key that verifies the signature on the P3P policy, and if this public key is the one contained in the verified public key certificate issued to this site from a trusted Certification Authority, then the user can be certain that the downloaded P3P policy is actually associated with the web site currently being accessed. A fake website would thus be unable to steal a P3P policy from a legitimate site to increase the probability of phishing success because the fake site does not know the real site's private key and so would be unable to construct a proper response for the user agent's one-time challenge.

D. Geographically Locating P3P Policies

Multinational companies are taking advantage of the web's capabilities (file sharing, web applications, and fast communication); they now distribute globally their products on the web for customers in different countries. However, policy guidelines for multinational companies are usually determined by the governments of the countries where the regional websites are operating. These government guidelines may affect the posted websites' privacy policies. More precisely, a branch of a multinational company in a particular country will only be bound by government guidelines or policies in that country. Thus, there is a need for Geolocating web users to know what policies should apply to their present location on the internet. The mechanism should allow multinational websites to accurately segment their global audiences and help the users to determine where (in which country) their data will be stored or if the government of the country where the website is located will have access to their personal data. For example, Amazon is a well-known multinational website located in countries like the USA (*Amazon.com*) and Canada (*Amazon.ca*). The two websites may have two different privacy policies adapted to each government's laws and guidelines. In a scenario where a user wants to purchase a product on Amazon.ca, it may happen that because of product availability, the user gets redirected to Amazon.com. These policy differences can promote research topic and questions such as:

- How will the user know the web location where his/her personal information will be stored? Will the other websites share their personal information?
- What are the inconsistencies between the various policies involved, and what are the risks associated with them?

- Most multinational companies don't bother to change their P3P policies, even if the privacy policies are modified according to the government guidelines. (e.g. Amazon.com/w3c/p3p.xml is available but Amazon.ca/w3c/p3p.xml is not). So redirected users may not get new summaries from their 'real' P3P user agents.

In the case where the multinational company stores P3P policies in a single domain, there should be a tool to determine the P3P file that will apply to the user based on different locations.

IV. CONCLUSIONS

In current web practice, privacy statements are often informal and imprecise, making it necessary to resort to human intervention in order to detect and resolve issues. The P3P standard has been developed to promote the automation of this process. After giving a summary of P3P's main characteristics and solutions, we have given examples showing that in the future the proper use of P3P and its enhancements can help the deployment and execution of privacy-sensitive services, such as cloud services, and mobile services. A P3P-based anti-phishing method was also proposed. As much as possible, P3P-based protocols for the automatic and rapid enforcement of privacy policies and for their combination, as well as for the automatic detection and resolution of conflicts, will have to be devised. Human intervention must be reduced and made as effective as possible when it is necessary. Our contribution has been to summarize and underline the potential of P3P for this purpose and to provide a summary of issues and solutions in this area, some of which have not previously been reported in the literature. Although the P3P standard has been in existence for ten years, and has been the subject of research, we feel that its potential has not yet been achieved. It is a good base from which significant evolution and application can take place, but much has to be done.

V. REFERENCES

[1] R. Agrawal, J. Kiernan, R. Srikant and Y. Xu, "An XPath-based Preference Language for P3P," in *Proceedings of the 12th international conference on World Wide Web, ACM*, Budapest, Hungary, pp. 629-639, 2003.

[2] R. Agrawal, J. Kiernan, R. Srikant and Y. Xu, "Hippocratic Databases," in *Proceedings of the 28th international conference on Very Large Data Bases, VLDB Endowment*, Hong Kong, China, pp. 143-154, 2002.

[3] A. Antón, E. Bertino, N. Li and T. Yu, "A roadmap for comprehensive online privacy policy management,"*Communication of ACM*, vol. 50, no. 7, pp. 109-116, July 2007.

[4] A. Antón, J. Earp, D. Bolchini, Q. He, C. Jensen and W. Stufflebeam, "The Lack of Clarity in Financial Privacy Policies and the Need for Standardization," *IEEE Security and Privacy ,* vol. 2, no. 2, pp. 36-45, 2004.

[5] F. Arshad, "Privacy fox - A JavaScript Based P3P Agent for Mozilla firefox", Technical Report 17-801, in School of Computer Science, Carnegie Mellon University, Pittsburgh, 2004.

[6] M. Bartel, J. Boyer, B. Fox, B. LaMacchina and E. Simon, "XML Signature Syntax and Processing (Second edition)," 10 June 2008. [Online]. Available: http://www.w3.org/TR/xmldsig-core/. [Accessed Febuary 2012].

[7] L. Cranor, S. Byers and D. Kormann, "An Analysis of P3P Deployment on Commercial, Government, and Children's Web Sites as of May 2003," Technical Report, AT&T Labs-Research, USA, 2003.

[8] L. Cranor and J. Reidenberg, "Can User agents Accurately Represent Privacy Notices," in *Proceedings of the research Conference on Commiunication, Informtion and Internet Policy (TPRC 2002)*, pp. 1-22, 2002.

[9] L. Cranor, "P3P: Making Privacy Policies more useful," *IEEE Security and Privacy,* vol. 1, no. 6, pp. 50-55, 2003.

[10] L. Cranor, B. Dobbs, S. Egelman, G. Hogen, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, M. Reagle, Schunter.M, A. Stampley and R. Wenning, "The Platform for Privacy Preferences 1.1(P3P1.1) Specification," World Wide Web Consortium, NOTE-P3P11-20061113, 13 November 2006. [Online]. Available: http://www.w3.org/TR/P3P11/. [Accessed 21 May 2011].

[11] L. Cranor, M. Arjula and P. Guduru, "Use of a P3P User Agent by early adopters," in *Proceedings of 9th ACM Workshop on Privacy in the Electronic Society*, Washington, DC, 2002.

[12] L. Dong, Y. Mu, W. Susilo, P. wang and J. Yan, "A Privacy Policy Framework for service Aggregation with P3P," *in Proceedings of 6th International Conference on Internet and Web Applications Services, Thinkmind, St. Maarten*, pp. 171-177, 2011.

[13] S. Egelman, L. Cranor and A. Chowdhury, "An Analysis of P3P-Enabled Web sites among Top-20 Search results," in *Proceedings of the 18th Internantional Conference on Electrronic commerce, ACM*, New Brunswick, pp 197-207, 2006.

[14] G. Hogben, "Suggestions for Long term changes to P3P," in *Long term future of P3P, W3C workshop*, Kiel, Germany, 2003.

[15] P. Kelly, J. Bresee, L. Cranor and R. Reeder, "A `Nutrition Label` for privacy," in *Symposium on usable privacy and security (SOUPS), ACM*, Mountain View, CA, pp 4:1-4:12, July 2009.

[16] A. Khurat, D. Gollhann and J. Abendorth, "A Formal P3P Semantics for Composite Services," in *Proceedings of the 7th VLDB conference on Secure data management, Springer-Verlag*, Singapore, pp. 113-131, 2010.

[17] J. Kolter and G. Pernul, "Generatiung user understandable Privacy Preferences," in *Proceedings of International Conference on Availability, Reliability and Security, IEEE*, Fukuoka, pp.299-306, 2009.

[18] C. Mar, "Privacy Bird for chrome," Technical Report 19-608, in School of Computer Science, Carnegie Mellon University, Pittsburgh, 2010.

[19] M. May, C. Gunter and I. Lee, "Strong and Weak Policy relations," in *Proceedings of IEEE International Symposium on policies for Distrubted Systems and Networks,IEEE Computer Society*, London, pp 33-36, 2009.

[20] A. McDonald, R. Reeder, P. Kelly and L. Cranor, "A Comparative Study of Online policies and Formats," in *Proceedings of PETS '09 Proceeding of the 9th International Symposium on Privacy Enchaning technologies, Springer-Verlag*, Seattle, WA, pp 37-55, 2009.

[21] I. Pollach, "What's wrong with online privacy policies," *Communications of the ACM,* vol. 50, no. 9, pp. 103-108, September 2007.

[22] T. Yu, N. Li and A. Anton, "A formal semantics for P3P," in *Proceedings of the 2004 workshop on Secure web service, ACM*, Fairfax, Virginia, pp 1-8, 2004.