Assessing the Likelihood of Privacy Policy Compliance

George O.M. Yee, Larry Korba, and Ronggong Song

Abstract Individuals interact with organizations in many different capacities (e.g. as clients, as employees). Many of these interactions require the individual to submit her personal information to the organization, which may claim compliance with privacy policy. It is important to assess this compliance quantitatively. This paper describes an approach for quantitatively assessing the likelihood that an organization will comply with privacy policy.

1 Introduction

Individuals interact with organizations in various roles that require them to submit their private information to the organization (e.g. health care patient, buyer). Given that how well an organization protects privacy is usually a matter of how well it complies with either its own privacy policy or the privacy policies of personal information owners, it is important to be able to assess this compliance quantitatively. If such assessments are publicly available, a) organizations could be challenged if their assessments are below a pre-established threshold (assuming a higher assessment is better), b) individuals could select organizations that have high compliance with which to do business, and c) organizations may be encouraged to pay more attention to protecting privacy. However, assessing an organization's actual compli-

George O.M. Yee
National Research Council Canada, Institute for Information Technology, 1200 Montreal Road, Building M-50, Ottawa, ON, Canada K1A 0R6 e-mail: george.yee@nrc.ca

Larry Korba
National Research Council Canada, Institute for Information Technology, 1200 Montreal Road, Building M-50, Ottawa, ON, Canada K1A 0R6 e-mail: larry.korba@nrc.ca

Ronggong Song
National Research Council Canada, Institute for Information Technology, 1200 Montreal Road, Building M-50, Ottawa, ON, Canada K1A 0R6 e-mail: ronggong.song@nrc.ca

ance performance may be difficult to do - the organization may be hesitant to report data needed to determine this performance, especially where the performance is bad, and even if the required data is reported, it would be difficult to ensure the reliability of the data. On the other hand, an organization's likelihood to comply with privacy policy may be more easily determined, since it could be based on what provisions it has implemented to protect privacy. In addition, an organization would welcome any opportunity to make known its investments in the protection of privacy in order to attract clients. This paper proposes a straight-forward approach for estimating the likelihood that an organization will comply with privacy policy.

Privacy refers to the ability of individuals to *control* the collection, use, retention, and distribution of information about themselves. This is the same definition as in [3] except that we also include *use*. An organization's *compliance with privacy policy* refers to the organization's use of provisions to give the protected person (PP) control over the organization's collection, use, retention, and distribution of information about the protected person, where this control is specified in the organization's privacy policy or the PP's privacy policy. An *internal violation (IV)* (or an inside attack) of privacy policy is one that is carried out by an insider of the organization (i.e. someone who has special data access privileges by virtue of the person's association with the organization, e.g. employee), whose access and use of the private information does not comply with the privacy policy. An *external violation (EV)* (or an outside attack) of privacy policy is one that is carried out by a non-insider of the organization, whose access and use of the private information does not comply with the privacy policy.

The literature appears empty of works dealing directly with estimates of an organization's likelihood to comply with privacy policy. Only works that are indirectly related were found, such as privacy impact assessment (PIA) (e.g. [6]), privacy risk analysis (e.g. [5]), and privacy audits (e.g. [2]).

2 Likelihood Estimates of Complying with Privacy Policy

A *likelihood estimate* of an organization's likelihood of complying with privacy policy is a set of numerical values that indicate the degree to which the organization will likely avoid IV and EV. The likelihood of avoiding IV and EV depends on protective provisions that the organization has in place to prevent violations. Let E denote a likelihood estimate. E will need to account for the provisions used against both IV and EV.

To account for the provisions against IV, we propose that a special PIA [6], extended to identify vulnerabilities that can lead to malicious IV, be carried out to identify IV vulnerabilities. Suppose that such an assessment identified m IV vulnerabilities and countermeasures (provisions against IV) are in place for p of these vulnerabilities. To account for provisions against EV, we propose that a special security threat analysis [5], oriented towards discovering EV vulnerabilities be carried out. Suppose that this analysis identified n security vulnerabilities and countermea-

sures (provisions against EV) are in place for q of these vulnerabilities. Then, one formulation of E is

$$\begin{aligned} E_1 &= (p+q)/(m+n), \text{ if } m+n>0, \\ &= 1, \text{ if } m+n=0. \end{aligned}$$

Let e_i account for the provisions used against IV and e_e account for the provisions used against EV. Then, another formulation of E is

$$\begin{aligned} E_2 = (e_i, e_e) &= (p/m, q/n), \text{ if } m>0,\ n>0, \\ &= (1, q/n), \text{ if } m=0,\ n>0, \\ &= (p/m, 1), \text{ if } m>0,\ n=0, \\ &= (1,1), \text{ if } m=0,\ n=0. \end{aligned}$$

Note that $0 \leq E_1, E_2 \leq 1$. In practice, the quantities E_1, e_i, e_e are expressed as percentages. E_1 has the advantage of providing a single number for ease of comparison between different organizations. A percentage threshold t for E_1 may be predetermined such that for E_1 above t, the provisions against IV and EV are deemed sufficiently likely to protect from violations. E_2 has the advantage of identifying where an organization stands in terms of its specific provisions for IV or EV. By predetermining percentage thresholds t_i and t_e for e_i and e_e respectively (thresholds above which the corresponding provisions for IV and EV are sufficiently likely to protect from privacy policy violations), E_2 defines a region in a 100 x 100 plane in which an organization's likelihood to comply with privacy policy (avoid privacy policy violations) is acceptable (shaded region in Figure 1).

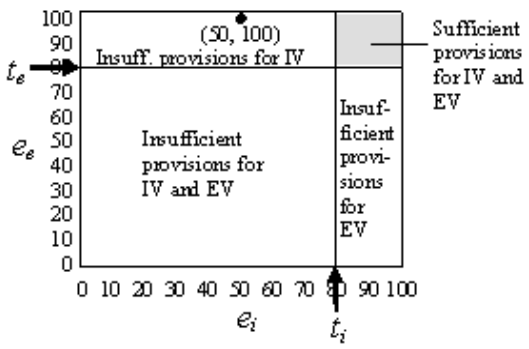

Fig. 1 Plot of E_2 = (50, 100) indicating that the corresponding organization has insufficient provisions to protect against IV.

We next give an application example. Suppose that a bank branch keeps the following personal information about its clients: name, social insurance number, home address, phone number, and financial assets. Suppose that the branch keeps this data stored within the branch itself. The branch decides to hire a privacy auditor, cer-

tified to apply the above estimation method, to estimate its likelihood of privacy policy compliance, with the intention of using the results in its advertising.

To determine values for m and p, the auditor puts together a team to do a PIA. The team analyzes where personal information originates, how it is stored, and how it is used. The PIA uncovers the following IV issues: a) there is no one accountable for private information, b) the database containing client personal data is not protected from illegal access, c) the branch's employees have been unhappy over the reduction in branch contributions to the employee pension plan, and d) the branch only does a minimal background check before hiring a new teller. The auditor is told by the branch manager that he has assigned himself to be accountable for private information in the branch's possession, and that a more thorough background check for job applicants has been initiated. However, the branch puts off any new measures to protect the database citing the fact that it already has a firewall in place. The branch also cannot do anything about the employee pension plan for the time being. Thus, $m = 4$ and $p = 2$.

To obtain values for n and q, the auditor assembles a team (with some members from the team for the PIA) to perform a threat analysis. Some examples of threats identified in this analysis are: a) the personal information flow is vulnerable to man-in-the-middle attacks (from the personal information path going into and out of the branch), b) the personal information database is vulnerable to attacks from inside and outside (via the Internet) the branch, and c) the bank tellers are vulnerable to social engineering attacks. The number of vulnerabilities n is found to be 6. Suppose that the branch has put in place countermeasures against each of these vulnerabilities, resulting in q also having the value 6. Thus,

$$E_1 = (p+q)/(m+n) = (2+6)/(4+6) = 8/10 = 4/5,$$
$$E_2 = (e_i, e_e) = (p/m, q/n) = (2/4, 6/6) = (1/2, 1).$$

Suppose that the predetermined thresholds for E_1, e_i, and e_e are t=85%, t_i=80%, and t_e=80% respectively. Then the branch has failed E_1 evaluation (since 80% for E_1 is less than the threshold of 85%). The branch has also failed E_2 evaluation (since 50% for e_i is less than the threshold of 80%). It is clear that this failure is due to the branch not providing sufficient provisions against IV (Figure 1). This branch would be motivated to improve its provisions against IV if other banks or branches of this bank are similarly evaluated and have results in the shaded area of Figure 1.

As shown by this example, E_1 provides a single number that shows whether or not the organization is likely to have sufficient provisions against IV and EV to avoid future violations. If E_1 is calculated for a number of similar organizations, the PP could easily see which organization is likely to comply with privacy policy. On the other hand, E_2 not only indicates the likelihood of an organization's future compliance, but also shows how strong the organization is in terms of its specific provisions against IV or EV. If the organization failed E_2 evaluation, it would know where it needs to make improvements in terms of provisions for IV, EV, or both.

3 Conclusions and Future Research

This work[1] has proposed estimates for evaluating the likelihood that an organization will comply with privacy policy. The estimates allow organizations to be challenged if their likelihood to comply is perceived to be inadequate. They also allow consumers to choose organization with high likelihoods of compliance.

The proposed estimates are straightforward and should be acceptable to the general public. We envision that organizations will want to publicize their estimates to show that they exceed the thresholds (which could be standardized by an international body) as for ISO 9000 [4]. This could encourage organizations to achieve higher levels of privacy policy compliance.

We suggest that the proposed approach be applied by a separate, impartial firm specialized in performing PIA and threat analysis. Application guidelines could be developed and standardized by a privacy authority, and only firms certified by the authority would be authorized to apply the approach (as done for ISO 9000 and CMMI (Capability Maturity Model Integration) [1]). This would ensure that the calculation of the estimates is done fairly and consistently across organizations.

Future research includes looking at ways to improve the accuracy of the estimates, such as incorporating the impact of past violations, as well as improving the methods for calculating the estimates, such as increasing the effectiveness of threat analysis through automation.

References

1. Carnegie Mellon Software Engineering Institute: Welcome to the CMMI Web Site. Visited April 14, 2008 at: http://www.sei.cmu.edu/cmmi/
2. Enright, K.P.: Privacy Audit Checklist. Visited May 6, 2006 at: http://cyber.law.harvard.edu/clinical/privacyaudit.html
3. Goldberg, I., Wagner, D., Brewer, E.: Privacy-Enhancing Technologies for the Internet. Proceedings, 42nd IEEE Computer Society International Conference (COMPCON'97), 103-109 (1997)
4. International Organization for Standardization: Management Standards. Visited April 16, 2008 at: http://www.iso.org/iso/management_standards.htm
5. Salter, C., Sami Saydjari, S., Schneier, B., Wallner, J.: Towards a Secure System Engineering Methodology. Proceedings of the New Security Paradigms Workshop (1998)
6. Treasury Board of Canada: The Privacy Impact Assessment Guidelines: A Framework to Manage Privacy Risk. Visited May 6, 2006, at: http://www.tbs-sct.gc.ca/pgol-pged/piatp-pfefvp/course1/mod2/mod2-5ع.asp

[1] NRC paper number: NRC 50328