Towards a Model for Risk and Consent Management of Private Health Information

Buffett, Scott; Kosa, T.A.

Publisher's version / Version de l'éditeur:

The Conference on Privacy, Security and Trust (PST2006) [Proceedings], 2006

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

Towards a Model for Risk and Consent Management of Private Health Information *

Buffett, S., Kosa, T.A.
October 2006

* published at The Conference on Privacy, Security and Trust (PST2006). Toronto, Ontario, Canada. October 31, 2006. NRC 48746.

Canada

Towards a Model for Risk and Consent Management of Private Health Information

Scott Buffett*, T. A. Kosa†

*National Research Council, IIT - E-business, 46 Dineen Drive, Fredericton NB, E3B 9W4, scott.buffett@nrc.gc.ca
†Smart Systems for Health Agency, 777 Bay Street, Suite 701, Toronto ON, M5G 2C8, tracyann.kosa@ssha.on.ca

***Abstract*— Health information custodians and network providers within the circle of care for a patient must meet certain legal obligations regarding the collection, access and disclosure of personal health information. We present a framework for consent and risk management that can be used to help manage a patient's consent for releasing personal health information, and analyze the risk involved in handling this type of data. A patient's preferences for specific privacy policies (expressed in P3P) are elicited through querying, and extra information is inferred using a Bayesian network. A risk analysis is performed to help a custodian to make informed decisions when handling personal health information. Thus the custodian and provider can help each other meet their respective legal obligations, and patients are more easily able to exercise their privacy rights.**

I. Introduction

One of the major technical challenges faced by Ontarian health care providers under the Personal Health Information Protection of Privacy Act, 2004 (PHIPA) is ensuring the privacy and security of personal information collected from their patients. PHIPA governs practices to protect personal health information (PHI), including consent, collection, use, disclosure and handling requests for access/correction to records by individuals.

Most patients deal with multiple health care workers, designated as health information custodians (HICs) under PHIPA. Each HIC may need different levels of access and use to a patient's health record in the course of providing care, all of which are associated with new legal obligations under PHIPA. As a result, there are several concerns:

1) How does a HIC meet their legal obligations surrounding indirect collection of PHI?
2) How does a patient retain control over access to their PHI? How does the primary care physician (PCP) ensure that appropriate personnel have access to the information when they need it, e.g. in an emergency situation?
3) How does a HIC ensure that patient preferences surrounding disclosure are being acted on?

Fulfillment of these legal obligations necessitates extracting specific preferences from the patient. Determining qualitative attitudes toward the handling of data, such as "I care most about my information being shared with third parties", is nontrivial enough. However, determining a given patient's quantitative valuations of actions on their PHI, which is necessary for a full risk analysis, can be a very difficult task.

A HIC grappling with the risks of non-compliance can use the services of a health information network provider (HINP) to enable electronic means to manage patient PHI.[1]. A HINP also has legal obligations under PHIPA[2] that require the recognition of risk: for example, one patient can be subject to a widespread unauthorized disclosure simply by accidental click of a email button by a HIC.

Given the risk and legal environment described above, it logically follows that the HINP bears some responsibility for tracking and monitoring consent as accepted by a HIC. A systematic consent management program can aid in the minimization, or potentially, elimination of some of these risks for the HIC and the HINP.

We propose a method based on utility elicitation that asks the patient appropriate questions when the health record is updated. These questions are determined and posed in such a way that allows for maximum information to be extracted and inferred on the patient's quantitative attitudes and valuations towards the handling of personal health information. This risk-based approach helps the HIC make an informed decision on whether or not to use a service at all, or simply in a particular method, whether there exists the potential to violate any of the patient's specifications.[3]

In this paper, we discuss these challenges and demonstrate a framework for a risk-based consent management system that can be used to circumvent these challenges. It is important to note that the proposed system is not yet implemented. However, the details given here offer more than a simple wish list of the types of features that are needed to help a HIC manage a patient's consent for releasing PHI. Instead, we provide a concise description of the types of problems that need to be solved, and describe our particular solutions that we plan to implement in order to achieve our goals.

II. Smart Systems Health Agency (SSHA)

One of a few HINP's currently operating in Ontario is SSHA, which offers a number of products and services to

[1]Personal Health Information Protection Act, 2004, s.10(4).

[2]Personal Health Information Protection Act, 2004, Regulation 329/04, s.6(3).

[3]For example, the patient may not consent to sending their PHI via email. However, this would not necessarily preclude a HIC from using a secure email service provided by a HINP to arrange appointment bookings for the same patient. A more detailed description of the risk environment can be found in section II

HICs. SSHA is an operational service agency accountable to the Ministry of Health and Long-Term Care (MOHLTC). They are tasked with improving the flow of patient information by building and deploying secure and reliable computer technologies to eventually connect all 150,000 healthcare providers[4] working at 24,000 locations in Ontario. The Agency is a part of the framework of the Ontario government's electronic health information management system.

The conceptual framework is almost fully developed; the actualization of the components is underway. The next step is alignment. Both the legislation and the electronic tools require a systematic capability to obtain, interpret and track consent from patients as they move through the interactive elements that constitute health care service in Ontario. For example, when a patient has preferences and instructions about information management[5], a HIC is legally obligated to follow those directions where possible. As a result, a HIC who chooses to utilize a HINP to facilitate the provision of health care can be (a) unaware of a given patient's preferences about the electronic handling of their PHI, and (b) fail to accurately interpret and follow these preferences. Thus, the HIC runs the risk of potentially violating the patient's consent via the use of an electronic means of handling PHI provided by a HINP.

Consider the provision of email services (SMI, or secure messaging infrastructure) to a HIC by SSHA.[6] The HIC will use it to transmit patient PHI. While SMI meets all governmental and legal requirements for the secure transmission of PHI, there would still be risks. Both the HINP and HIC would be required to own these risks, respectively, in the provision and use of the service. In this case, as the designated HINP, the network provider is obligated (from an operational business perspective if not from a directly legal requirement) to collect, track and log consent by the HIC to adopt these risks. Similar to an end-user license agreement, a HINP needs to be able to manage a consent framework in conjunction with the provision of SMI.

Either embedded with the SMI service, or as a stand alone add on, this tool can provide the opportunity to set preference and establish a customized acceptable risk level for patients and the HIC. At the front end, the HIC can walk through basic privacy principles, as expressed in policy, with the patient, and track their consent. At the back end, the HINP will log these preferences and append them by layering it on to the pre-existing functionality of the SMI service. At any given point during the use of the SMI service, the tool would notify the HIC (end user) if they were about to use the product in a way that steps beyond their pre-determined acceptable risk level (either for themselves, or a given patient).

[4]Including doctors, hospitals, pharmacies, labs, public health units and continuing care organizations.

[5]For the purposes of simplicity in this paper, the term information management will be used to reference all data interaction principles, including collection, access, use, disclosure, retention and disposal of PHI.

[6]In the provision of email service, it is technically possible for SSHA not to be a HINP (depending on a given client / environment) in which case its legal responsibilities under PHIPA may be subject to change.

III. Expressing Privacy Policies

The Platform for Privacy Preferences Project (P3P) [6] provides a standard language, both human- and machine-readable, for expressing policies for the handling of personal information. Developed by the World Wide Web Consortium, P3P enables websites to express their data-collecting practices in a standard format, indicating which data is to be collected, for what purposes, with whom it will be shared, and for how long it will be retained. P3P user agents then allow users to be informed of these website practices and automate the decision on whether or not to share the requested information based on the user's pre-defined privacy policy.

A P3P policy is represented by an XML file containing one or more *statements*. Each statement describes what data will be collected, with whom it will be shared, for how long it will be retained and for what purpose. Similarly, P3P statements can be used by the owners of private data to specify how their personal information should be handled. In this case, each statement describes the information-handling practices pertaining to an aspect of the user's private data that is deemed allowable. A data-collection practice that is inconsistent with the set of statements is said to violate the user's privacy preferences.

IV. Determining Patient Preferences

In this section, we outline a method for determining a patient's preferences for the handling of his/her PHI. These preferences are then used to 1) construct one or more P3P statements which make up the privacy preferences for the patient, and 2) determine the patient's perceived importance for each such statement.

Each statement specified dictates the patient's policy for a particular item of data. For example, one such statement could dictate that a patient's e-mail address may be collected by a HIC's administrative assistant, for the purpose of reminding the patient of a future appointment, and may not be shared with any third party. However, there may exist several statements for a particular item, since a patient may have one preference for how a doctor can handle the information, and another preference for how a hospital administrator can handle it, possibly with different preferences for with whom each of these HICs can share the data, how long it can be retained, etc. Given several different categories of data for which policies must be specified, and given these possibly many different specifications for each type of data, a patient's privacy preferences can be extremely complex. It would thus be impossible to ask the patient to explicitly specify his/her complete set of preferences.

To overcome this, we utilize a technique for eliciting preferences proposed by Buffett et al [4]. This technique performs utility elicitation [5] by posing several queries to the patient about her preference over various outcomes that could result from misuse of her information. For example, one such query could ask "Would you be more averse to the possibility of exposure of your e-mail address to the administration of a hospital that you have never visited, or to the possibility

of such an exposure of your home address?" If the patient answers "e-mail", then policies for the collection and use of e-mail address should be more restrictive. If the patient answers "home address", then policies dealing with home address, and also related items such as home phone number, should be more stringent.

As information on the patient's preferences is being obtained during the querying process, inferences on the patient's preferences are being made to increase the volume of knowledge. This is done by retaining anonymized data on other patients that have used the system, and using a Bayesian network to model statistical dependencies in the data. For example, if it is found that the patient places high importance on the confidentiality of visits to a physiotherapist, then, through the Bayesian network, we may find that it is likely to be the case that the patient may want to keep all records of sport-related injuries confidential as well, since this implication holds for a significant percentage of the population. As information is gathered through the query/inference process, a P3P policy file is built specifying the patient's privacy preferences.

V. Risk Analysis of Service Use

Any time a HIC uses a service or performs some action that has the potential to violate a patient's privacy preferences, a risk analysis is performed. This will help the HIC make an informed decision on whether to proceed by advising him/her on how likely the action is to violate the patient's preferences, and also on how severe the potential impact or consequences may be for such a violation.

Each statement s additionally has values $p(s)$ and $u(s)$ which specify the *probability* and the *utility* of s, respectively. These are explained as follows.

The probability of a statement indicates the likelihood that the patient will allow the information to be handled as specified by the statement. If the patient explicitly indicates s during the querying process, then $p(s) = 1$. If the patient explicitly indicates that certain actions are not allowed, then for the statement s specifying this, $p(s) = 0$. Probabilities for all other statements are inferred using the Bayesian network. For example, if a patient specifies a particular statement s for private information on physiotherapist visits, and it is the case that 73% of all patients who specify s also specify a statement s' for the handling of records on sports-related injuries, then this could imply that the probability of s' for this patient should be 0.73.

The utility of a statement is a measure of the patient's preference for that statement. More specifically, a patient's utility for a statement is a measure of the relative importance the patient places on the policy specified by the statement in comparison with other policy statements. While all statements that are explicitly specified by the user are considered important and will be enforced regardless of the patient's utility, this utility measure allows the system to profile the patient by learning what is important to her. This information is then used to determine the patient's expected utility for other policy statements that are inferred. The potential *impact* for a set of actions is then computed as a function of the patient's utilities for the relevant statements. This gives a measure of the severity of consequences that could occur by incorrectly assuming that the patient would agree with the statement. This can help in the assessment of risk for a given action by indicating the level of severity of a violation. For example, a HIC may decide to go ahead with an action that has a 12% likelihood of a violation with a minimal impact, but might be more hesitant to proceed with a 12% likelihood of a high impact.

As an example, consider three categories A, B and C of personal information, with operations x, y and z that can be performed on each. Thus there are 9 possible statements. Let these statements be denoted here for convenience by (A, x), (A, y), (A, z), (B, x), (B, y), (B, z), (C, x), (C, y) and (C, z). Suppose the user specifies in the querying process:

> "I will allow x to be done with A"
> "I will allow y to be done with B"
> "I will not allow z to be done with B"
> "I will only allow x to be done with C"

Then it is immediately known that the probability of allowance is 1 for (A, x), (B, y) and (C, x), and 0 for (B, z), (C, y) and (C, z). Based on these known probabilities, probabilities for other statements are inferred using the Bayesian network. Suppose we also receive the following statements from the patient regarding utilities:

> "My policy for C has top importance"
> "I feel less strongly about my policy for B being violated than my policy for C"

This could translate to the utilities $u(A, x) = u(B, y) = u(C, x) = 0$, $u(C, y) = u(C, z) = 1$ and perhaps $u(B, z) = 0.5$. Utilities for unknown statements are inferred in the same manner as probabilities. Note that $u(A, x) = u(B, y) = u(C, x) = 0$ indicates that the patient does not care about these statements. Since the information handling specified by the statements is permitted, the patient is indifferent as to whether this is actually done, and thus there is no impact.

When making the decision on whether or not to perform an action or actions using a patient's personal data, a HIC can enter several possible actions into the system for a full risk analysis. For each action, the system will return 1) the probability of a violation of the patient's privacy preferences, 2) the relative expected impact of the privacy violations, and 3) the probability of high impact. Additionally, the HIC has the opportunity to enter her own utilities for the candidate actions. For example, a doctor may ideally want to submit a report on a patient to researchers at a local university for inclusion in a medical journal submission, but would however be sufficiently (but less) satisfied with simply submitting the record to a colleague that had expressed interest in the diagnosis. Thus preference information can then be combined with the statistics on probabilities and impacts, and advice can

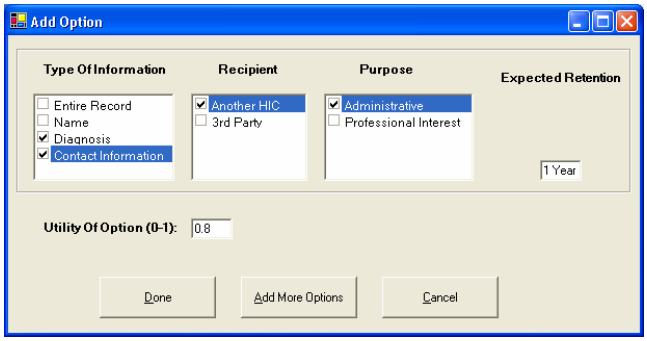

Fig. 1. Facility for adding another alternative action to the risk analysis

be given to the HIC on what is deemed to be the best course of action.

VI. System Usage

Each time a HIC uses a service that will involve the transmission of a patient's data, the service will interface with the content/risk management system to determine whether or not there may be a possible violation. An alert box will then display a warning message informing the HIC if a violation is possible. The HIC can use this information to decide whether or not to proceed, or alternatively query the system about other possibilities. When the latter is the case, the HIC will be taken to the "Add Options" screen as depicted in Figure 1. Here the HIC can construct an alternative action on the patient's data, which can be added to the analysis. For example, the HIC could choose to transmit the entire record, but perhaps specify different recipients, purposes or retention times. Or perhaps the HIC wants to keep all actions the same as the original request, but instead specify only a subset of the private data. Next the HIC indicates how preferable this option is by specifying the utility, where 0 is the worst score and 1 is the best score. Once completed, the HIC can indicate whether she is done and would like to see the results of the analysis, or that she would like to input more possible options so that a large spectrum of viable possibilities can be analyzed.

Once the user has selected the "Done" option, a risk analysis is performed and the results are displayed as in Figure 2. For each specified option, information is given on 1) the probability of violating the patient's preferences, 2) the expected impact of the violations, relative to the action with the highest impact (which by definition has an expected impact of 1), 3) the probability of high impact, which is the probability that the impact will be above some prespecified threshold (e.g 0.9), 4) the HIC utility (specified by the HIC in the Add Option screen) and 5) the overall value of the decision, which is computed by applying an adjustable formula on the other four values.

In Figure 2, the system indicates that option 2 is the best, since its probability of causing a violation was 0.46 (as compared to 0.53), its expected impact was 0.72 times that of option 1, and the probability of high impact was only 0.08 (as compared to 0.13). Option 2 was less preferred than option 1

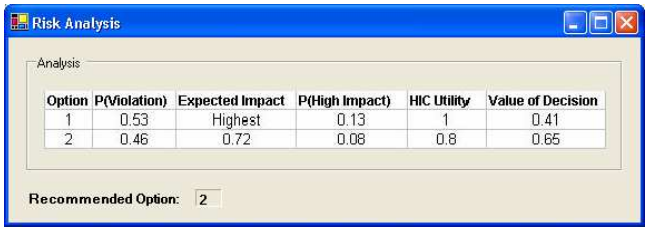

Fig. 2. Results of risk analysis

(0.8 versus 1), but that difference did not significantly detract from the overall value of the option, as it was deemed better by a score of 0.65 to 0.41.

VII. Conclusions

In this paper, we present a framework for consent and risk management that can be used to help a HIC meet their legal obligations under PHIPA by managing a patient's consent for releasing personal health information. To illustrate the need for our proposed system, we discuss the role of health information network providers, and highlight the services and associated privacy concerns offered by a specific provider, namely SSHA. We discuss a mechanism for eliciting privacy preferences from a patient, and demonstrate how these policies, both elicited and inferred by a Bayesian network, can be used to perform a risk analysis.

For future work, we plan to further develop these ideas and implement the proposed system. We also plan to investigate issues such as the role of HICs as data stewards of a patient file as set out in case law (McInerney v. MacDonald, SCR 1992), and the implications under PHIPA. Another key consideration is the work of Canada Health Infoway Inc. in delivering electronic health record (EHR) solutions. Any comprehensive consent management framework should consider the possibility of building on a future pan-Canadian architecture.

References

[1] S. S. Al-Fedaghi. How to calculate the information privacy. In *Third Annual Conference on Privacy, Security and Trust (PST05)*, 2005.

[2] S. Buffett, L. Comeau, M. W. Fleming, and B. Spencer. Monologue: A tool for negotiating exchanges of private information in e-commerce. In *Third Annual Conference on Privacy, Security and Trust (PST05)*, pages 79–88, 2005.

[3] S. Buffett, K. Jia, S. Liu, B. Spencer, and F. Wang. Negotiating exchanges of P3P-labeled information for compensation. *Computational Intelligence*, 20(4):663–677, 2004.

[4] S. Buffett, N. Scott, B. Spencer, M. M. Richter, and M. W. Fleming. Determining internet users' values for private information. In *Second Annual Conference on Privacy, Security and Trust (PST04)*, pages 79–88, 2004.

[5] U. Chajewska, D. Koller, and R. Parr. Making rational decisions using adaptive utility elicitation. In *AAAI-00*, pages 363–369, Austin, Texas, USA, 2000.

[6] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The Platform for Privacy Preferences (P3P) 1.0 Specification. http://www.w3.org/TR/P3P/, 16 April 2002. W3C Recommendation.

[7] Smart Systems for Health Agency. http://www.ssha.on.ca, 2006. Accessed: April 17, 2006.