The 6th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2016)

HIPAA-Compliant Privacy Policy Language for e-Health Applications

Youna Junga,⁎ and Minsoo Kim⁎

⁎Department of Computer and Information Sciences, Virginia Military Institute
425 Mallory Hall, Lexington, Virginia 24450, United States

Abstract

Many e-health applications collect patient’s health data and track how they are used by patients to enable and validate their effectiveness. Although e-health applications allow people to access healthcare services in easy and convenient way at the reduced cost, the lack of reliable and effective methods of privacy protection makes people hesitate to use e-health applications, and in turn, it becomes the biggest obstacle to the growth of e-Health applications. To overcome the drawback, in this paper, we first address the lack of consideration of health-related data on existing privacy policy languages and propose the HIPAA profile for existing languages, which contains the Health data schema and extensions to HIPAA-friendly policy languages. By using the HIPAA profile, e-health providers are able to specify HIPAA-compliant privacy policies and patients can express their privacy preferences on not only general usage and user data but also health-related data in detail. For better understanding, we present example policies for e-health applications and patients using the proposed profile.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Program Chairs

Keywords: e-health application; online monitoring; privacy; policy language; HIPAA

1. Introduction

E-health is an emerging area at the intersection of medical informatics, public health, business, and information technologies. By leveraging information technologies, e-health applications provide highly available, user-friendly, and personalized services at reduced cost, regardless of time and place. Currently, many e-health applications are used for diverse purposes, such as online education, healthcare research, healthcare data collection, medical interventions, and health promotion, and they are getting ingrained into the everyday life of people. To improve the effectiveness of online intervention and/or offer personalized healthcare services, some applications collect detailed, and often identifiable, health data of patients. However, collecting identifiable data on e-health applications become a privacy issue due to the sensitivity of data that e-health applications often deal with. Indiscriminate data collection and/or monitoring may result in serious privacy loss, for examples, patients’ private health data may be used for unwanted
purposes or shared with unknown people\cite{5,6,7}. In e-health applications, even generic usage data, such as login frequency into an online treatment application, can reveal a patient’s medical status.

To protect user privacy in e-health applications, the U.S. federal Health Insurance Portability and Accountability Act (HIPAA)\cite{8} stipulates that the first party of healthcare services must not disclose protected health information to other service entities (HIPAA 164.105.(a)(ii)) with only a few exceptions (HIPAA 164.512). A patient needs to check whether a privacy policy of the e-health application is fully compliant with HIPAA or not, before giving consent to disclose his/her health data. Even if a patient examines an e-health application’s policies, the application may behave different from a Service Level Agreement (SLA) that is mutually agreed between a patient and an e-health application. For example, private data can be released regardless of patients’ wishes, if a healthcare provider embeds monitoring code and/or third-party data-collecting ads in his/her applications. Although this is an obvious violation of HIPAA rules, there are no solutions that systematically detect the application’s fraud and prevent user data from undesirable use and disclosure.

To address the privacy issues on e-health applications, we preliminarily proposed the Privacy-Preserving online Monitoring (PPoM) framework\cite{9} that allows e-health applications to conduct trustworthy user monitoring and enable patients to use e-health applications without concern for loss of privacy. By using the PPoM framework, patients are able to verify user/usage data being monitored through user-friendly interface of web browser and strictly enforce their privacy policies on the client side by controlling outgoing messages sent from users’ browsers. However, the performance of the PPoM framework strongly depends on the accuracy and precision of privacy policies. Currently, a patient specifies his/her privacy policies using APPEL\cite{10} or XPref\cite{11}, while healthcare service providers use P3P\cite{12}. As general-purpose privacy policy languages, existing policy languages including P3P, APPEL, and XPref focus on generic user/usage data to be used for a variety of online applications and do not give careful consideration to health data. It is therefore impossible for both patients and e-health providers to precisely specify their privacy policies about health-related and/or HIPAA-related data, and in turn, it lowers the performance of the PPoM framework. To address the lack of consideration of health data in existing privacy policy languages, in this paper, we propose the HIPAA profile which allows an e-health provider to specify a HIPAA-compliant privacy policy and enables a patient to specify his/her privacy preference on health data in detail.

The rest of this paper is organized as follows. In Section 2, we introduce our preliminary work and identify its limitations, and in Section 3, explore existing privacy policy languages and discuss the shortcomings of existing languages. In Section 4, we identify the requirements for privacy policy languages for e-health applications and then propose the HIPAA profile. A use case is presented in Section 5, and we conclude our work in Section 6.

2. Preliminary work

As mentioned above, it is important to monitor patients’ health without a violation of privacy in e-health applications. Towards this goal, we proposed the Privacy-Preserving online Monitoring (PPoM) framework that rigorously protects user privacy by referring user policies written in APPEL or XPref and enforcing them on user side during online monitoring\cite{9}. The PPoM framework consists of three components: the PPoM Service, the PPoM Browsers, and the PPoM Tool (PPoMT). The overall architecture is shown in Fig. 1.

- **PPoM Service** – It gathers only authorized data that users allow to monitor. By specifying privacy policies, patients can determine which data can be monitored. User policies will be then enforced by the PPoM Service that selectively collects data based on user policies. Unlike the existing monitoring services where user data are collected based on an application’s policies and the policies are enforced by the application itself, the PPoM Service provides a way to enforce user policies during monitoring in a systematic manner rather than simply providing a written agreement.

- **PPoM Browser** – Even if a user is exposed to untrustworthy e-health applications that conduct indiscriminate monitoring in violation of HIPAA and a mutually agreed policy, user privacy must be protected. Towards this end, the PPoM Browser presents all data being monitored and protects user privacy on the user side by blocking outgoing messages which contain data a user does not want to disclose based on a user’s policies.

- **PPoM Tool (PPoMT)** – Although patient monitoring is essential, it is difficult for healthcare providers to develop monitoring-enabled applications and privacy policies due to lack of professional IT knowledge. The PPoM Tool (PPoMT) enables non-IT health professionals to specify privacy policies for their healthcare applications through user-friendly interfaces and helps them to convert their existing applications into monitoring-enabled applications.

To use the PPoM framework, first, a provider needs to upload the source code or enter the URL(s) of his/her application to the PPoMT. Then, he/she is required to select objects to be monitored and specify corresponding privacy policies through the interfaces generated by the In-page Selector. The Privacy Policy Generator then creates the application’s policies by analyzing selected monitoring data and policies, while the Application Converter updates source code by inserting monitoring code generated by the Monitoring Code Generator into the original source code. The application policies and the updated source code must be deployed in an application server.
By using the PPoM browser, a patient can use e-health applications without privacy concern. Whenever a patient enters a url of an application, his/her PPoM browser compares user policies and application policies. If they match, the application server sends PPoM-enabled pages which privacy-aware monitoring code is embed in. As a patient interacts with the application, the PPoM browser displays all data being monitored to enable users to verify privacy protection during monitoring. The privacy-aware monitoring code inserted in webpages collects only authorized user data according to user policies. The PPoM browser will block outgoing messages that violate the patient’s privacy policies.

3. Privacy policy languages

In the PPoM framework, both users and service providers must specify privacy preference using existing privacy policy languages, such as P3P, APPEL, and XPref. Those languages allow people to describe what kinds of user data might be monitored, what those data are used for, who those data will be shared with, and how user data are maintained. In this section, we evaluate the suitability of existing policy languages for e-health applications.

The Platform for Privacy Preferences (P3P)\(^5\) allows online applications to declare their privacy policies about data types to be collected (Data), usage of collected data (Purpose), consumers of user data (Recipient), permanence (Retention) and accessibility of collected data (Access), and dispute resolution procedure (Disputes). To specify data types that may be collected, P3P provides four pre-defined data schemas: Dynamic, User, Third Party, and Business. The Dynamic data schema is used to specify data that do not have fixed values. It includes several data types; a user’s clickstream, http to refer information carried by the HTTP protocol, clientevents to collect event data that a user triggers such as mouse movement, cookies to know if a HTTP cookie is set or retrieved by an application, searchtext for text that a user types in a search engine page, interactionrecord to have a user’s transaction history, and miscdata. The User data schema is used to generally specify a user and it includes several child data types such as name, bday, gender, home-info, business-info, and login. The Third party data schema is used to provide third party information and its data types are identical to those of the User data schema. The Business data schema features a subset of the User data relevant for describing legal policy entities. If a user and an application need to specify other data in addition to base data schemas, then they should create their own data schema.

A P3P Preference Exchange Language (APPEL)\(^6\) is a user-centric policy language that enables users to express their privacy preferences. As a complementary language to P3P, it is used by (software) user agents to make automated/semi-automated decisions regarding the acceptability of P3P policies of applications. In APPEL, a user’s policies are expressed in a set of preference rules (Ruleset). A rule (RULE) consists of a policy (p:P3P:POLICY) and a behavior (behavior). When a user tries to access an e-health application having its own P3P policies, then a user agent examines the P3P policies before proceeding further. If a policy is matched, then the user agent takes an activity such as request, limited, or block as described in the behavior attribute. Although APPEL enables users to express their privacy preference, it has not been widely accepted due to its complex syntax, error-prone semantics, and weak expression power\(^7,8\). To overcome APPEL’s drawbacks, XPref\(^9\) was proposed. It keeps APPEL’s rule heads including RULESET and RULE but replaces rule bodies represented by P3P Policies with a condition attribute expressed by XPath\(^10\). As a result, XPref is able to remove the ambiguity and complexity in APPEL’s matching patterns and enhance its expression power. However, XPref policies are also complementary to P3P policies. This implies that user preferences written in APPEL or XPref are based on the P3P data schemas. As a general-purpose policy language, P3P have focused only on generic data, rather than on specific health-related data. However, both patients and healthcare providers need to specify privacy policies regarding health data and HIPAA for effective privacy protection in the PPoM framework. Currently, there is no specific policy language dealing with health data and HIPAA. Lack of consideration for health and HIPAA data becomes a serious obstacle to widespread adoption of existing policy languages in e-health applications.
4. HIPAA profile for P3P

To address the limitations of existing policy languages, we propose the HIPAA profile that contains the Health data schema and encompasses the definitions and rules in HIPAA. The HIPAA profile enables e-health providers to specify HIPAA-compliant policies and also enables patients to easily verify compliance with HIPAA.

4.1. Health data schema

In most of the privacy policy languages, P3P’s data schemas are primarily used. The base data schemas include only four schemas; Dynamic (dynamic user activities), User (user profile), Third Party (third-party information), and Business (business information). Although a user’s health data are among the most essential data, a health data schema has not yet been proposed. If users and/or e-health applications hence want to specify their privacy policies on health data, they should define a custom schema before specifying a policy.

![Image of the Health data schema for P3P, APPEL, and XPref](image)

To avoid having inconsistent schemas across different users and applications, a standard of health data schema is needed. Towards this goal, we propose the Health data schema that contains health vocabularies which can be widely acceptable in a variety of e-health applications. The Health data schema aims to describe a patient’s health status and it has sixteen child elements as shown in Fig.2: height, weight, hearing-acuity, visual-acuity, blood-type, blood-pressure, blood-sugar-level, cholesterol-level, disabilities, allergies, lab-tests, medication, disease-history, family-medical-history, immunization-history, and healthcare-providers. The sixteen elements are selected through analysis of more than one hundred of patient charts/records that are currently used in hospitals such as the University of North Carolina Cancer Center\(^5\) and the Primary Health Medical Group\(^6\), government agency such as the U.S. Department of Health and Human Services\(^7\) and the U.S. Centers for Disease Control and Prevention\(^8\), and medical schools such as the University of California San Francisco Medical Center\(^9\). To resolve the problems of synonymy, we refine synonyms using the vocabularies in the U.S. National Library of Medicine\(^10\). Some of child elements can be extended by importing well-known public health-related ontologies, such as Disability ontology\(^11\), Disease ontology\(^12\), Drug ontology\(^13\), Laboratory Test ontology\(^14\), and Vaccine ontology\(^15\).

4.2. HIPAA-abiding privacy policy

As we mentioned above, it is critical for e-health providers to specify HIPAA-compliant privacy policies since HIPAA regulations are one of the most stringent rules for privacy protection against indiscriminate disclosure of health data. Therefore, a privacy policy language to be used in e-health applications must consider terminology and rules in HIPAA. Towards this end, we extend P3P to a HIPAA-friendly privacy policy language by modifying it as follows. A P3P policy is described as an element of STATEMENT (represented as <STATEMENT>) of a <POLICY>. Each <STATEMENT> consists of six child elements: <CONSEQUENCE>, <NON-IDENTIFIABLE>, <PURPOSE>, <RECIPIENT>, <RETENTION>, and <DATA-GROUP>. The proposed modifications are as follows and the summary of all modification is presented in Table 1.

- **<DEIDENTIFIED>** – If an e-health application collects only de-identified protected health information (De-identified PHI), the applications’ privacy policies are required to be specified as <DEIDENTIFIED> as a child element of a <STATEMENT>. De-identified PHI through the Safe Harbor method (See HIPAA 164.514(b)(2)) must not include the following identifiers of a patient or of relatives, employers, or household members of the patient (See HIPAA 164.502): Names, Geographic locations smaller than a State, Dates except year, Telephone numbers, Fax numbers, Email, Social security numbers, Medical record
numbers, Health plan beneficiary numbers, Account numbers, Certificate/license numbers, Vehicle identifiers, Device identifiers, Web URL, IP addresses, Biometric identifiers, Face photographic, and all other identifying numbers.

- **<PURPOSE>** – In P3P, twelve purposes are pre-defined as follows: <current>, <admin>, <develop>, <tailoring>, <pseudo-analysis>, <pseudo-decision>, <telemarketing>, <contact>, <historical>, <individual-analysis>, <individual-decision>, and <other-purpose>. In addition to those purposes, we add four purposes to represent the general purposes of e-health applications, which described in HIPAA: <public-health>, <research>, <healthcare-operation>, and <healthcare-reference>. Note that an e-health application may use or disclose a limited data set if the application receives a data use agreement with the limited data set recipient, only for the following purposes: <public-health>, <research>, and <healthcare-operation> (See HIPAA 164.514).

- **<RECIPIENT>** – It represents recipients of the collected data and there exist six types of recipients: <ours>, <delivery>, <same>, <other-recipient>, <unrelated>, and <public>. To represent a patient as a recipient of his/her PHI, we add <user> to the existing set of recipients. Note that, in general, a users’ access right to collected user data is represented in a <ACCESS> of a <POLICY>. HIPAA states that a patient has a right of access to inspect and/or obtain a copy of his/her PHI with few exceptions (See HIPAA 164.524). To represent a patient’s access rights regulated by HIPAA, we add <HIPAA-compliant-access/> to the value set of <ACCESS>.

- **<RETENTION>** – It indicates a retention policy that is applied to the referenced data. Currently, five retention types are defined: <no-retention>, <stated-purpose>, <legal-requirement>, <business-practices>, and <indefinitely>. However, HIPAA clearly specifies the maximum retention period of PHI (for example, an e-health application must comply with respect to the PHI of a deceased patient for a period of 50 years following the death of the patient, See HIPAA 164.502), and requires healthcare providers to document expiry date and/or expiry events (See HIPAA 164.508). Accordingly, we add <HIPAA-compliant-retention/> to the existing set of retention types and add two optional attributes, expiry-date and expiry-event, to <RETENTION>.

- **<DATA>** – It describes a data type to be collected. It has a child element <CATEGORIES> and currently seventeen categories are defined. If an e-health application collects health-related data, it must select <health/> as a category of the data to be monitored. To refer the Health data schema proposed, a value of a ref attribute of a <DATA> must start with "#health".

<table>
<thead>
<tr>
<th>Table 1. The proposed P3P associated with HIPAA regulations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proposed P3P Extension</strong></td>
</tr>
<tr>
<td><strong>&lt;ACCESS&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>&lt;DEIDENTIFIED&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>&lt;PURPOSE&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>&lt;RECIPIENT&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>&lt;RETENTION&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>&lt;DATA&gt;</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
5. Use cases

Unlike existing privacy policy languages, the HIPAA profile enables both e-health applications and patients to specify privacy preference on health-related data in detail. For example, an online obese patient tracker can specify a HIPAA-compliant policy as shown in Fig. 3. The policy indicates that the tracker collects weight and family medical history of obese patients (Fig.3.a) for public health and health care operations of a clinic which owns the patient tracker (Fig.3.b). The collected health data will be disclosed to only the first party covered entity, the obese patient clinic in this example (Fig.3.c). According to HIPAA, the clinic can retain the data (Fig.3.d) and each patient has a right to access his/her PHI (Fig.3.e). An example XPref policy for a patient that uses the HIPAA profile is shown in Fig. 4. The user policy indicates that a patient allows the first party clinic (Fig.3.f) to use his/her PHI with the purpose of its health care operations (Fig.3.g) for a HIPAA-regulated retention period (Fig.3.h) but does not want to disclose his/her family history to any party (Fig.3.i). If an administrator or a patient of an e-health application does not have enough knowledge and skill to describe their own privacy policy, they may need to use the PPoMT, shown in Fig. 1. By selecting health and usage data to be collected or to be protected through the user-friendly interfaces of PPoMT, they are able to generate their own privacy policies.

Fig. 3. An example of an e-health application policy using the proposed privacy policy language

Fig. 4. An example of a user policy using the proposed privacy policy language

6. Conclusions

There exists a most urgent need for a new privacy policy language that allows e-health applications to specify HIPAA-compliant policies and enables patients to specify privacy preferences on health data in detail. To address the requirement, in this paper, we proposed the HIPAA profile to be used in combination with existing privacy policy languages. The proposed work has the potential to make significant technical advances in user privacy on e-health applications by enabling monitoring services and user browsers to satisfy users’ privacy preferences. It enables e-health applications to obtain non-identifiable user data as well as identifiable data that users agree to disclose without concern about privacy loss and violation of HIPAA regulations. In turn, it can promote the use of e-
health services. Furthermore, it may bring positive societal impacts by contributing to further growth of online healthcare markets. By expanding the acceptance and range of e-health application participants, we can improve public health and reduce healthcare costs by replacing expensive face-to-face consultations or treatments with online healthcare services that are private and secure.

References

10. A P3P Preference Exchange Language (APEEL) 1.0. https://www.w3.org/TR/P3P-preferences/ [retrieved: May, 2016].