Chapter 7.6
Estimating the Privacy Protection Capability of a Web Service Provider[1]

George O.M. Yee
Institute for Information Technology, National Research Council, Canada

ABSTRACT

The growth of the Internet has been accompanied by the growth of Web services (e.g., e-commerce, e-health, etc.), leading to important provisions put in place to protect the privacy of Web service users. However, it is also important to be able to estimate the privacy protection capability of a Web service provider. Such estimates would benefit both users and providers. Users would benefit from being able to choose (assuming that such estimates were made public) the service that has the greatest ability to protect their privacy (this would in turn encourage Web service providers to pay more attention to privacy). Web service providers would benefit by being able to adjust their provisions for protecting privacy until certain target capability levels of privacy protection are reached. This article presents an approach for estimating the privacy protection capability of a Web service provider and illustrates the approach with an example.

INTRODUCTION

This work considers Web services to be: a) Web-based services that employ Extensible Markup Language (XML), Web service Definition Language (WSDL), Simple Object Access Protocol (SOAP), and Universal Description, Discovery, and Integration (UDDI) in a Service-Oriented Architecture (SOA) (O'Neill, Hallam-Baker, MacCann, Shema, Simon, Watters, et al., 2003); and b) existing and previous generations of Web-based applications that involve Web browsers interacting with Web servers that do not employ XML, WSDL, SOAP, or UDDI. This work applies to all Web services described above.

Numerous Web services targeting consumers have accompanied the rapid growth of the

Internet. For example, Web services are available for banking, shopping, learning, healthcare, and government online. However, most of these services require a consumer's personal information in one form or another, leading to concerns over privacy. For Web services to be successful, privacy must be protected. Various approaches have been used to protect personal information, including data anonymization (Iyengar, 2002; Kobsa & Schreck, 2003) and pseudonym technology (Song, Korba, & Yee, 2006). Approaches for privacy protection that are in the research stage include: treating privacy protection as an access problem and then bringing the tools of access control to bear for privacy control (Adams & Barbieri, 2006); treating privacy protection as a privacy rights management problem using the techniques of digital rights management (Kenny & Korba, 2002); and considering privacy protection as a privacy policy compliance problem, verifying compliance with secure logs (Yee & Korba, 2004).

It is also important to estimate the privacy protection capability of a Web service provider. Suppose such estimates for similar Web services A, B, and C are made available to consumers. This leads to the following benefits. If the consumer has to choose one service from among A, B, and C, then the estimates can help the consumer decide which service to select (probably the service that has the highest capability for privacy protection). In addition, the fact that consumers have access to these estimates may encourage service providers to pay more attention to protecting consumer privacy and result in higher levels of consumer trust and acceptance of Web services. Alternatively, Web service providers can use such estimates to implement services that meet predefined goals of privacy protection. Predefined levels of the estimates could be expressed as quality-of-service requirements. The estimates could then be evaluated for incremental versions of a service until the predefined levels are achieved.

The objectives of this article are to a) define estimates of the privacy protection capability of a Web service provider, b) show how the estimates can be calculated, and c) illustrate the calculation of the estimates using a Web service example.

This article extends the work of Yee (2006) by: a) improving the practicality of the approach by refocusing on estimating privacy protection capability rather than measuring how well privacy is protected; b) updating the definition of the estimates; c) updating the method for calculating the estimates; d) updating and extending the application example; e) enlarging the related works section; f) adding an evaluation section; and g) improving the clarity of the writing in all sections.

The rest of this article is organized as follows. Section "Estimates of Privacy Protection Capability" introduces the privacy protection model and defines the estimates. "Calculation of the Estimates" shows how to calculate the estimates. The section called "Application Example" illustrates the calculation of the estimates. A discussion of related work then follows. "Evaluation of Approach" discusses the strengths and weaknesses of the approach. Finally, the article ends with conclusions and directions for future research.

ESTIMATES OF PRIVACY PROTECTION CAPABILITY

Privacy

In order to define estimates of a Web service provider's capability to protect consumer privacy, it is necessary first to examine the nature of personal privacy. As defined by Goldberg, Wagner, and Brewer (1997), privacy refers to the ability of individuals to *control* the collection, retention, and distribution of information about themselves. This leads to the following definitions for this work.

DEFINITION 1: *Privacy refers to the ability of individuals to control the collection, use,*