Chapter 1.7
Protection of Privacy on the Web

Thomas M. Chen
Swansea University, UK

Zhi (Judy) Fu
Motorola Labs, USA

ABSTRACT

Most people are concerned about online privacy but may not be aware of the various ways that their personal information is collected during routine Web browsing. We review the types of personal information that may be collected voluntarily or involuntarily through the Web browser or disclosed by a Web server. We present a taxonomy of regulatory and technological approaches to protect privacy. All approaches to date have only been partial solutions. By its nature, the Web was designed to be an open system to facilitate data sharing, and hence Web privacy continues to be a challenging problem.

INTRODUCTION

The main appeal of the World Wide Web is convenient and instant access to a wealth of information and services. Many people will start research on a topic with a Google search. The number of Web sites has grown exponentially and reached more than 149 million in November 2007 according to Netcraft (http://news.netcraft.com/archives/web_server_survey.html).

In their search for services, users may not keep in mind that the Web is capable of collecting data as well as displaying data. The most obvious means of data collection are Web forms for registrations, logins, and messaging. These forms are voluntary disclosures of personal information that most people understand to be necessary for shopping, online banking, and other personalized services. However, users may not fully appreciate that Web sites collect information about them routinely without their consent or even notification. Web sites keep track of clients' IP (Internet protocol) addresses at a minimum and often additional information such as browser version, operating system, viewed resources, and clicked links. Moreover, this collected information may be shared among organizations in the background without the public's knowledge.

DOI: 10.4018/978-1-61350-323-2.ch1.7

Some users may have unrealistic expectations about online privacy because they ignore the fact that the Web is an open system. By design, just about anyone can easily put up a Web site and make its contents globally accessible. This means that sites should not be assumed to be trustworthy. Contrary to natural inclinations, it would be more reasonable to assume sites are untrustworthy, until a trust relationship is established (e.g., through prior experience, reputation, or third-party validation).

Web privacy is certainly not a new issue. However, if anything, concerns have escalated rather than decreased due to increasing prevalence of phishing and malware (malicious software) attacks. In phishing attacks, innocent users are lured to malicious sites designed to deceive them into revealing valuable personal information. Common types of malware include spyware, bots, and keyloggers, which can steal personal data. They can be downloaded in various ways, often without a user's awareness.

The consequences of privacy loss will be growing distrust of the Web and diminishing usage of online services. Thus, protection of privacy is an important practical problem with economic ramifications. This chapter examines regulatory and technological approaches to protect privacy on the Web. First, we survey the various threats to online privacy. Then we offer a taxonomy of approaches to provide and protect privacy.

TYPES OF PRIVATE INFORMATION

Clearly there are different types of personal information, with varying degrees of sensitivity. As shown in Figure 1, personal information on the Web might be classified into three types (Rezgui, Bouguettaya, and Eltoweissy, 2003):

- personal data such as name, address, and history;
- surfing behavior consisting of visited sites, online transactions, and searches;
- communications such as bulletin boards, messages, and feedback forms.

Personal data can be classified further into anonymous information (which can not be traceable to a specific person); personally identifiable information; or private information (Garfinkel, 2002). Information can be anonymized by "scrubbing" any identifying aspects or by aggregating multiple records into a single record. Personally identifiable information can be traced to an individual, such as name, address, e-mail address, or phone number. Although this information is

Figure 1. Types of personal information on the Web