Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Networks 48 (2005) 503–515
Computer Networks
www.elsevier.com/locate/comnet

# Protocols that hide user's preferences in electronic transactions

Feng Bao [a], Robert H. Deng [b,*]

[a] Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
[b] School of Information Systems, Singapore Management University, 469 Bukit Timah Road, Singapore 259756

Received 17 January 2004; received in revised form 15 September 2004; accepted 27 October 2004
Available online 20 November 2004

Responsible Editor: D. Frincke

**Abstract**

The Internet creates many new threats to personal privacy and raises some unique privacy concerns. In this paper we study the problem of how to protect users' privacy in web transactions of digital products. In particular, we introduce a system which (1) allows a user to disclose his/her identity information (such as user account or credit card number) to a web site in exchange for a digital product, but (2) prevents the web site from learning which specific product the user intends to obtain. The problem concerned here is orthogonal to the problem of anonymous transactions [M. Reed, P. Syverson, D. Goldschag, Anonymous connections and Onion Routing, IEEE Journal of Selected Areas in Communication 16 (4) (1998) 482–494; M. Reiter, A. Rubin, Crowds: anonymity for web transactions, ACM Transactions on Information System Security, 1 (1) (1998) 66–92] but commensurate with the general problem of PIR (private information retrieval) [B. Chor, O. Goldreich, E. Kushilevita, M. Sudan, Private information retrieval, in: Proceedings of 36th FOCS, 1995, pp. 41–50; B. Chor, N. Gilboa, Computational private information retrieval, in: Proceedings of 29th STOC, 1997, pp. 304–313]. Most of the existing results in PIR, however, are theoretical in nature and can not be applied in practice due to their huge communication and computational overheads. In the present paper, we introduce two practical solutions that satisfy the above two requirements and analyze their security and performance. Another issue we study in this paper is how to recover sales statistics data in our user privacy-protected system. We present a novel solution to the problem along with its security analysis.


*Keywords:* Anonymizer; Encryption; Privacy protection; On-line transaction; Digital products

## 1. Introduction

Privacy has been a sensitive issue long before the advent of the Internet. However, the Internet creates many new threats to personal privacy and

* Corresponding author. Tel.: +65 6874 7862.
*E-mail addresses:* baofeng@i2r.a-star.edu.sg (F. Bao), deng@i2r.a-star.edu.sg, robertdeng@smu.edu.sg (R.H. Deng).

1389-1286/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2004.10.010

raises some unique privacy concerns. Such concerns have been magnified in recent years due to widespread use of the world wide web and the accompanying e-commerce activities. Information sent over the Internet may pass through dozens of different computer systems on the way to its destination. Each of these systems may be capable of monitoring, capturing, and storing on-line communications. When a user surfs the web, many web sites deposit user's browsing patterns or transaction generated data on the user's hard drive which can be reused by the web sites when the user returns. Most web browsers invisibly provide web sites with information about user's computer (such as IP address, domain name, screen resolution, available plug-ins) as well as with information about the locations of other web sites a user has visited. The highly connected nature of the Internet makes it easy to automatically collect users' information from many different sources and compile a dossier about an individual—his or her likes and dislikes, shopping patterns, whereabouts and so on. Such data is a potentially valuable source of revenue for many businesses - it is useful to direct marketers as a basis for deriving targeted lists of users with similar likes; it can also be the source of abuses that may cause embarrassment for users who have accessed sensitive or controversial materials on-line.

Not only the Internet users are concerned about privacy when surfing the web, organisations are starting to take customers' on-line privacy seriously to build users' confidence on their business. Over the recent years, a number of web sites have been caught in high-profile blow-ups over how they collect and utilize user data. After on-line advertising company DoubleClick revealed its now abandoned practice to cross reference data about users' off-line purchasing behaviour with their on-line habits, the company's stock dived.

The present paper studies the technical aspect for on-line user privacy protection. A number of systems and tools have been developed to allow web user anonymity while retrieving information on the web. One type of commercial available tools is anonymizer, such as *Anonymizer* from Anonymizer.com [2] and *Freedom* from Zero-Knowledge System Inc [24]. Anonymizer is a service that submits http requests to web sites on behalf of its users. Because the request is submitted by the anonymizer agent rather than the user, the only IP address revealed to the web sites is that of the agent. However, users of this service have to trust the anonymizer to keep their IP addresses and their web activities private since they are not anonymous to the anonymizer itself.

*Onion Routing* [19] and *Crowds* [20] are two anonymity systems that do not require users to trust a single third party to maintain anonymity. *Onion Routing* is a general-purpose infrastructure for anonymous communication over a public network. It operates by dynamically building anonymous connections within a network of real-time Chaum *Mixes* [6]. A *Mix* is a store-and-forward device that accepts a number of fixed-length messages from numerous sources, performs cryptographic transformations on the messages, and then forwards the messages to the next destination in a random order. A single *Mix* makes tracking of a particular message either by specific bit-pattern, size, or ordering with respect to other messages difficult. By routing through numerous *Mixes* in the network, determining who is talking to whom is even more difficult. *Crowds* is a system for protecting users' anonymity on the web. It is named for the notion of "blending into a crowd" and operates by grouping users into a large and geographically diverse group, i.e., crowd, that collectively issues http requests on behalf of its members. In *Crowds*, web servers are unable to trace the source of a request because it is equally likely to have originated from any member of the crowd, and even collaborating crowd members can not distinguish the originator of a request from a member who is merely forwarding the request on behalf of another.

The above anonymous systems are useful for web surfing in which users have no desire or not required to be identified. Therefore, they are mostly useful when users visit free web sites and download free digital products. However, when users wish to make on-line purchasing using their credit card numbers or membership accounts, they need to provide some identifying or authenticating information. In such situations the issue of privacy protection is not user anonymity, but how to hide