

Chapter 1

State of the Art in Distributed Privacy–Preserving Protocols in Private Web Search

Mohib Ullah

The University of Agriculture, Peshawar, Pakistan

Arbab Waseem Abbas

The University of Agriculture, Peshawar, Pakistan

Lala Rukh

The University of Agriculture, Peshawar, Pakistan

Kamran Ullah

The University of Agriculture, Peshawar, Pakistan

Muhammad Inam Ul Haq

Khushal Khan Khattak University Karak, Karak, Pakistan

ABSTRACT

Web search engine (WSE) is an inevitable software system used by people worldwide to retrieve data from the web by using keywords called queries. WSE stores search queries to build the user's profile and provide personalized results. User search queries often hold identifiable information that could compromise the user's privacy. Preserving privacy in web searches is the primary concern of users from various backgrounds. Many techniques have been proposed to preserve a person's web search privacy with time. Some techniques preserve an individual's privacy by obfuscating a user's profile by sending fictitious queries with the original ones. Others hide their identity and preserve privacy through unlinkability. However, a distributed technique preserves privacy by providing unlinkability and obfuscation. In distributed protocols, a group of users collaborate to forward each other queries to WSE, providing unlinkability and obfuscation. This work presents a survey of distributed privacy-preserving protocols. The benefits, limitations, and evaluation parameters are detailed in this work.

DOI: 10.4018/978-1-6684-6914-9.ch001

INTRODUCTION

The world wide web (WWW) is a vast network of information and resources where users can access and search for a wide range of content, including text, images, videos, and audio (Khan & Ali, 2013; Khan, Ullah, Khan, Uddin, & Al-Yahya, 2021). Web search engines, such as Google and Bing, play a crucial role in helping users find the information they are looking for by processing large amounts of data and presenting relevant results based on their search queries. Search engines have become indispensable tools for Internet users as they allow easy and fast access to information on a global scale. Research has shown that people are becoming increasingly satisfied with the performance of search engines, but at the same time, they are also becoming increasingly concerned about their privacy (Ullah, 2020b). Using personalization algorithms by search engines to present search results and advertisements tailored to the user's interests is seen as both a strength and a weakness by different people.

On the one hand, personalized results can provide a more relevant and enjoyable experience for the user. On the other hand, it also raises privacy concerns as search engines collect and store large amounts of data about the user's activities, interests, and behaviours, which could be used for various purposes, including targeted advertising. Web search engines build a user profile based on various factors such as interests, preferences, and previous searches to provide more relevant results. This user profile can improve the accuracy of search results but also raises privacy concerns as it reveals sensitive information about the user. The user's profile contains their unique I.D., name, employer's details, location, and potentially sensitive information such as their health status, political views, religion, etc. (Cooper, 2008). As a result, users are often forced to trade off accuracy for privacy, which can result in less relevant search results (Dan & Davison, 2016). Search engines must balance the need for personalized results with user privacy protection.

A survey conducted in 2012 showed that many users were concerned about the privacy implications of web search engines recording their data and search queries (Ullah, 2020b). The query log, which records users' search activities, is a valuable resource for search engines as it helps them to provide more relevant results (Kaaniche, Masmoudi, Znina, Laurent, & Demir, 2020). However, the storage of query logs also poses a significant privacy threat, as this data can be disclosed to advertising agencies and media, potentially revealing sensitive information about users. The release of the AOL log in 2006, where twenty million queries generated by 658000 users over three months were published, is a well-known example of a privacy breach (Barbaro, Zeller, & Hansell, 2006; Wang, Liu, & Wang, 2020). This incident highlights the need for search engines to implement strong privacy policies and measures to protect user data and ensure users' privacy.

In some cases, web search engines may be required by court order to disclose individual queries as part of legal proceedings, such as divorce or civil lawsuits. These incidents have raised further concerns about the privacy of users' search data and the security of their personal information. The 2014 incident, where 80 million health records were lost by the second-largest insurance company in the U.S., highlights the need for better security measures to protect sensitive information (Mathews-Hunt, 2016; Yang, Onik, Lee, Ahmed, & Kim, 2019). These incidents have sparked a movement for privacy preservation among online community members and have led to questions about the privacy policies of the WSE. These incidents put a question mark on WSEs' policy regarding user privacy.

So far, many techniques have been proposed to protect users' privacy during their Web searches. These techniques can be classified into five major classes, i.e., distributed techniques, stand-alone methods, query scrambling, third-party infrastructure, and hybrid techniques. This chapter overviews the

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/state-of-the-art-in-distributed-privacy-preserving-protocols-in-private-web-search/322583

Related Content

A Resource Allocation Model for Desktop Clouds

Abdulelah Alwabel, Robert John Walters and Gary B. Wills (2019). *Web Services: Concepts, Methodologies, Tools, and Applications* (pp. 258-279).

www.irma-international.org/chapter/a-resource-allocation-model-for-desktop-clouds/217835

A Reservation-based Extended Transaction Protocol for Coordination of Web Services

Wenbing Zhao, Firat Kart, L.E. Moser and P.M. Melliar-Smith (2008). *International Journal of Web Services Research* (pp. 64-95).

www.irma-international.org/article/reservation-based-extended-transaction-protocol/3124

An Adaptive System for a Real-Time Matching Application

Taka Matsutsuka, Masatoshi Ogawa, Yohei Toriyama, Noriyasu Aso and Ichiro Iida (2022). *International Journal of Web Services Research* (pp. 1-22).

www.irma-international.org/article/an-adaptive-system-for-a-real-time-matching-application/299018

Parallel Computing for Mining Association Rules in Distributed P2P Networks

Huiwei Guan (2011). *E-Activity and Intelligent Web Construction: Effects of Social Design* (pp. 47-62).

www.irma-international.org/chapter/parallel-computing-mining-association-rules/53273

An Assessment of Several Taxonomies of Volunteered Geographic Information

Antony K. Cooper, Serena Coetzee and Derrick G. Kourie (2012). *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 21-36).

www.irma-international.org/chapter/assessment-several-taxonomies-volunteered-geographic/65107