DataShare: A Decentralized Privacy-Preserving Search Engine for Investigative Journalists

Kasra EdalatNejad[1], Wouter Lueks[1], Julien Martin, Soline Ledésert[2], Anne Lhôte[2], Bruno Thomas[2], Laurent Girod[1], and Carmela Troncoso[1]

[1] École polytechnique fédérale de Lausanne, Switzerland
{kasra.edalat,wouter.lueks,carmela.troncoso}@epfl.ch
[2] International Consortium of Investigative Journalists

Abstract. Investigative journalists collect large numbers of digital documents during their investigations. Many of these documents contain sensitive information. Revealing possession of such documents could, therefore, endanger reporters, their stories, and their sources. As a result, even though these documents could greatly benefit other journalists' work, many documents are used only for single, local, investigations.
We present DataShare, a decentralized and privacy-preserving global search system that enables journalists worldwide to find documents via a dedicated network of peers. This work stems from the need of the International Consortium of Investigative Journalists (ICIJ) to extend their local document search on a discovery platform DataShare to a global setting.
DataShare combines well-known anonymous credentials and anonymous communication primitives with a novel multi-set private set intersection protocol (MS-PSI) into a *decentralized peer-to-peer private document search engine*. MS-PSI enables efficient search in many collections at a time, while not leaking more than state-of-the-art PSI protocols would. By significantly reducing the computation and communication cost of performing intersections, MS-PSI enables DataShare to scale to thousands of users and millions of documents.
Format: We propose a lightning talk and/or poster to introduce the challenges faced by investigative journalists, and how DataShare solves them.

Keywords: Privacy-preserving search · Private set intersection · Decentralization.