Equivalence-based Security for Querying Encrypted Databases: Theory and Application to Privacy Policy Audits

Omar Chowdhury
Purdue University
West Lafayette, Indiana
ochowdhu@purdue.edu

Deepak Garg
MPI-SWS
Germany
dg@mpi-sws.org

Limin Jia, Anupam Datta
Carnegie Mellon University
Pittsburgh, Pennsylvania
{liminjia,danupam}@cmu.edu

ABSTRACT

To reduce costs, organizations may outsource data storage and data processing to third-party clouds. This raises confidentiality concerns, since the outsourced data may have sensitive information. Although semantically secure encryption of the data prior to outsourcing alleviates these concerns, it also renders the outsourced data useless for any relational processing. Motivated by this problem, we present two database encryption schemes that reveal just enough information about structured data to support a wide-range of relational queries. Our main contribution is a definition and proof of security for the two schemes. This definition captures confidentiality offered by the schemes using a novel notion of equivalence of databases from the adversary's perspective. As a specific application, we adapt an existing algorithm for finding violations of a rich class of privacy policies to run on logs encrypted under our schemes and observe low to moderate overheads.

Categories and Subject Descriptors

H.2.0 [**DATABASE MANAGEMENT**]: General—*Security, integrity, and protection*; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy, Regulation*

Keywords

Privacy Policy Audit; HIPAA; GLBA; Querying Encrypted Databases

1. INTRODUCTION

To reduce infrastructure costs, small- and medium-sized businesses may outsource their databases and database applications to third-party clouds. However, such data is often private, so storing it in a cloud raises confidentiality concerns. Semantically secure encryption of databases prior to outsourcing alleviates confidentiality concerns, but it also makes it impossible to run any relational queries on the cloud without client interaction. Several prior research projects have investigated encryption schemes that trade-off perfect data confidentiality for the ability to run relational queries [39, 6, 21]. However, these schemes either require client-side processing [21], or require additional hardware support [6], or support a very restrictive set of queries [39]. Our long-term goal is to develop database encryption schemes that can (1) be readily deployed on commercial off-the-shelf (COTS) cloud infrastructure without any special hardware or any kernel modifications, (2) support a broad range of (non-update) relational queries on the encrypted database without interaction with the client, (3) be implemented with low or moderate overhead, and (4) provide provable end-to-end security and a precise characterization of what information encryption leaks in exchange for supporting a given set of queries. Both in objective and in method, our goal is similar to that of CryptDB [34], which attains properties (1)–(3), but not (4).

As a step towards our goal, in this paper, we design two database encryption schemes, Eunomia$^{\mathsf{DET}}$ and Eunomia$^{\mathsf{KH}}$, with properties (1)–(4). Our design is guided by, and partly specific to, a single application, namely, audit of data-use logs for violations of privacy policies. This application represents a real-world problem. Organizations are subject to privacy legislation. For example, in the US, the healthcare and finance industry must handle client data in accordance with the federal acts HIPAA [1] and GLBA [2] respectively. To remain compliant with privacy legislation, organizations record logs of privacy-relevant day-to-day operations such as data access/use and employee role changes, and audit these logs for violations of privacy policies, either routinely or on a case-by-case basis. Logs can be fairly large and are often organized in commodity databases. Audit consists of a sequence of policy-guided queries. Audit is computationally expensive but highly parallelizable, so there is significant merit in outsourcing the storage of logs and the execution of audit algorithms to third-party clouds.

Security. We characterize formally what information about an encrypted log (database) our schemes may leak to an adversary with direct access to the encrypted store (modeling a completely adversarial cloud). We prove that by looking at a log encrypted with either of our schemes, an adversary can learn (with non-negligible probability) only that the plaintext log lies within a certain, precisely defined *equivalence class of logs*. This class of logs characterizes the uncertainty of the adversary and, therefore, the confidentiality of the

CCS'15, October 12–16, 2015, Denver, Colorado, USA.

ACM 978-1-4503-3832-5/15/10.
http://dx.doi.org/10.1145/2810103.2813638.