DISS. ETH NO. 22450

Encrypting Databases in the Cloud Threats and Solutions

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

TAHMINEH SANAMRAD

M.Sc. in Computer Science, ETH Zurich

born on 14.01.1983

citizen of
Iran

accepted on the recommendation of

Prof. Dr. Donald Kossmann (ETH Zürich, Switzerland), examiner
Prof. Dr. Gustavo Alonso (ETH Zürich, Switzerland), co-examiner
Prof. Dr. Srdjan Capkun (ETH Zürich, Switzerland), co-examiner
Prof. Dr. Johannes Gehrke (Cornell University, NY, USA), co-examiner

2014

Abstract

With the growing importance of cloud computing, database encryption has become a critical technology to protect data against honest-but-curious attackers. Our goal is to encrypt the data in such a way that it remains protected against powerful attackers and at the same time achieve good performance by processing queries in the cloud without decrypting the data. *Order-Preserving Encryption (OPE)* is one of the most attractive techniques for database encryption since it allows the execution of range and rank queries on encrypted data. On the other hand, people are reluctant to use OPE-based techniques in practice because of their vulnerability against attackers with knowledge of the domain and its frequency distribution.

This dissertation makes three important contributions. First, it formalizes a set of real-world attacker scenarios on encrypted databases, namely *domain attack*, *frequency attack* and *query log attack*. Query log attack refers to the inference of secrets by observing the (encrypted) queries submitted to the encrypted database. To this end, a number of encryption techniques have been developed and studied in literature. Unfortunately, most of these schemes have ignored an important threat called query log attack. Second, based on this formalization, it shows how these attacks impact the security of an important class of encryption techniques, namely OPE. Third, it explores new encryption techniques called *Probabilistic Order-Preserving Encryption (Prob-OPE)* and *Randomly Partitioned Encryption (RPE)* which are proven to be resilient against the attacker scenarios mentioned previously. These encryption techniques address the need to encrypt databases in the cloud and at the same time execute complex SQL queries efficiently. Prob-OPE and RPE can be configured to meet different privacy and performance requirements. Privacy and performance experiments conducted using the TPC-H queries show that Prob-OPE and RPE make it indeed possible to achieve a higher level of privacy compared to the state of the art with low performance overheads.

Kurzfassung

Die Datenverarbeitung durch Internetdienste (engl. cloud computing) ist heute weit verbreitet da sie zuverlässig, skalierbar und überall verfügbar ist. In diesem Kontext stellen sich neue Herausforderung an die Datensicherheit mittels Verschlüsselung. Daten sollen verschlüsselt werden, so dass sie auch vor starken Angriffen geschützt sind. Gleichzeitig muss es möglich sein die Daten ohne allzu grosse Effizienzverluste zu verarbeiten. Ortnungserhaltende Verschlüsselungsverfahren (engl. Order-Preserving Encryption, kurz OPE) gehören zu den attraktiveren Ansätzen, da mit ihnen Bereichs- und Ortnungsanfragen effizient abgearbeitet werden können. Leider sind sie unteranderem anfällig für Angriffe wenn der Bereich der Datenwerte, z.B. Städtenamen, oder deren relative Häufigkeit bekannt ist.

In dieser Doktorarbeit formalisieren wir als erstes die für das OPE-Verfahren praxisrelevanten Angriffsarten wie z.B. den vorherig erwähntenn Angriff über den Datenbereich (engl. domain attack), die Häufigkeitverteilung der Werte (engl. frequency attack) sowie den Logbuchangriff, welcher auf der Analyse der (verschlüsselten) Datenbankanfragen basiert (engl. query log attack). Gerade der Logbuchangriff wurde bisher in der Literatur selten behandelt, er stellt aber ein in der Praxis durchaus denkbares Szenario dar.

Darauf aufbauend wird dann die Sicherheit des OPE-Verfahrens untersucht. Die daraus gewonnenen Einsichten erlauben es uns dann verbesserte Verfahren zu entwickeln mit signifikant erhöhter Widerstandskraft. Insbesondere präsentieren wir zum erstenmal eine probabilistische Erweiterung namens Prob-OPE (engl. Probabilistic Order-Preserving Encryption) und eine Erweiterung basierend auf einer zufälligen Unterteilung des Datenbereichs mit dem Kürzel RPE (engl. Randomly Partitioned Encryption). Beide Verfahren sind so desgined, dass sie erheblich sicherer sind ohne viel Effizienz zu verlieren. Durch eine geeignete Wahl der Parameter, zum Beispiel die Anzahl der Partitionierung beim RPE-Verfahren, kann in der Praxis graduell Sicherheit gegenüber Effizienz balanciert werden. Wir eruieren die Performance unserer neuen Verfahren basierend auf dem TPC-H Benchmark und zeigen mittels mathemtatischer Methoden wie stark sich die Datensicherheit erhöht. Es zeigt sich, dass Prob-OPE und RPE es ermöglichen die Datensicherheit zu erhöhen ohne einen grossen Performanceverlust in Kauf nehmen zu müssen.