NTNU – Trondheim
Norwegian University of
Science and Technology

Exploring CryptDB:
A Practical HE Scheme for SQL Queries

Eirik Klevstad

Submission date: October 2015
Responsible professor: Colin Boyd, ITEM
Supervisor: Chris Carr, ITEM

Norwegian University of Science and Technology
Department of Telematics

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of

the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Preface

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

Chapter 1
Introduction

1.1 Problem

Cloud services are becoming larger and more complex. Users want their content available in the cloud, and easily access to it. Companies such as Apple, [flere] are looking into health information and how your personal information can be integrated in their services. Medical research facilities stores tremendous amounts of personal data, and currently looking into how to share their research material across facilities and borders. [1] Along with these type of sensitive data, follows great responsibility and security measures.

- Homomorphic encryption to the rescue. Even if the developer of a new system that needs to store sensitive information follows the guidelines of best-practice of cryptographic storage, he will still face problems.

The idea of a homomorphic encryption (HE) scheme is to enable the system to query and compute on encrypted data (ciphertexts) without the need of decrypting it first.

1.2 Background

1.3 Scope

Background, what is the problem, scope and objective(s) of your work. It may also be relevant to state what you do not address.

If the scope/goal of your project has changed compared to what you submitted in your project description; redefine it and explain why it has changed.

Chapter 2
Homomorphic Encryption

Homomorphic encryption

* Hva det er for noe * Hvordan det fungerer * Forskning på området * Hvorfor det er viktig

Chapter 3
Overview of CryptDB

3.1 CryptDB

3.1.1 System Architecture

* Create a figure of the system architecture *

Explain the client application Explain the use of the proxy server Explain key storage Explain the use of the database server

3.1.2 SQL-Aware Encryption

CryptDB uses an encryption scheme called "SQL-aware encryption" or "onion encryption". As with onions, data items stored using CryptDB are encrypted multiple times using different types, or layers, of encryption. Each security layer allows different types of computation to be performed on the ciphertext.

Må skrive en del mer på introen her.

Random (RND) is the highest security level in CryptDB and provides the maximum security found in encryption scheme. RND does, however, not allow any computation to be done on the encrypted data. In other terms, this level is a natural choice for sensitive data that are only meant to be read. It uses a strong block cipher such as Blowfish or Advanced Encryption Standard (AES) in Cipher-Block-Chaining (CBC) mode and a random initialization vector (IV) to ensure the block cipher to be probabilistic.

A block cipher being probabilistic means that when encrypting a message M_1, the result is an encrypted message C_1. When the block cipher encrypts that same message one more time, the resulting ciphertext is an encrypted message C_2 such that $C_1 \neq C_2$.

Given two encryptions of the same plaintext $c_1 = E_k(m_1)$ and $c_2 = E_k(m_1)$, the resulting ciphertexts are c_1 and c_2 such that $c_1 \neq c_2$.

Where RND provides no computation to be done on the encrypted data, the next layer does. Deterministic (DET) is an encryption scheme enabling the application to perform standard SQL operations such as equality checks, distinct, group by and count. By allowing these sorts of computation, the application leaks information to an adversary. In particular, it leaks which ciphertexts that decrypts to the same plaintext value. Following the previous example; if the scheme encrypted the message M_1 two times, the resulting ciphertexts C_1 and C_2 are equal. DET is a deterministic scheme which, to be used correctly, should be a Pseudo-Random-Permutation (PRP). In order to cope with leaking prefix equality, the authors have designed their own version of the CMC mode [2].

Order-Preserving Encryption (OPE)

Homomorphic Encryption (HOM)

Join (Join and OPE-Join)

Word Search (SEARCH)

3.1.3 Adjusting the encryption level based on the query

References

[1] Integrated Data for Analysis, Anonymization, and Sharing kernel description. https://idash.ucsd.edu/. Accessed: 2015-10-04.

[2] Popa, R. A., Redfield, C. M. S., Zeldovich, N., and Balakrishnan, H. Cryptdb: Protecting confidentiality with encrypted query processing. In *Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles* (New York, NY, USA, 2011), SOSP '11, ACM, pp. 85–100.