# *Overview of CryptDB*

**Gouasmia Zakaria – Wuhan university of technology**

# CryptDB: Protecting Confidentiality with Encrypted Query Processing

## 1. ABSTRACT

Online applications are vulnerable to theft of sensitive information because adversaries can exploit software bugs to gain access to private data, and because curious or malicious administrators may capture and leak data. CryptDB is a system that provides practical and provable confidentiality in the face of these attacks for applications backed by SQL databases. It works by executing SQL queries over encrypted data using a collection of efficient SQL-aware encryption schemes. CryptDB can also chain encryption keys to user passwords, so that a data item can be decrypted only by using the password of one of the users with access to that data. As a result, a database administrator never gets access to decrypted data, and even if all servers are compromised, an adversary cannot decrypt the data of any user who is not logged in. An analysis of a trace of 126 million SQL queries from a production MySQL server shows that CryptDB can support operations over encrypted data for 99.5% of the 128,840 columns seen in the trace. Other evaluation shows that CryptDB has low overhead, reducing throughput by 14.5% for phpBB, a web forum application, and by 26% for queries from TPCC, compared to unmodified MySQL. Chaining encryption keys to user passwords requires 11–13 unique schema annotations to secure more than 20 sensitive fields and 2–7 lines of source code changes for three multi-user web applications.

**General Terms**: Security, design.

## 2. INTRODUCTION

In the face of snooping Database Administrators (DBAs) and compromise from attackers, confidentiality in Database Management Systems (DBMS) should still remain. This paper comprehensively inquires, and tests a new way of securing databases in presence of the two threats mentioned above. This new technique is called CryptDB, a system which acts as a proxy to secure the communication between the database server, and the applications server. The paper mainly deals with confidentiality of information flowing between application and database servers, and not any form of information security like integrity and availability. CryptDB receives queries from the application server, secures them and sends them to the database server. Then, it will receive encrypted data from the database, decrypts it and sends to application server to be sent to the requester. In a nutshell, CryptDB enables the BDMS to run SQL queries on encrypted database data as it could do on plaintext. This is done because by principle, curious DBAs, attackers, DBMS and system infrastructure are the entrusted fellows. The assumption made is that the application server and the database server are different and a proxy can intercept their communication. CryptDB can be implemented on a range of DBMS such as MySQL and Postgres.

## Some Confidentiality Breaches from this Year

| |
|---|
| **Apple** *12 million device records, hacked from FBI agent's computer* |
| **New York State Electric & Gas Co 1.8** *million billings via unauthorized access* |
| **Global Payments, Inc.** *1.5 million payment-card numbers hacked* |
| **Utah Dept. of Technology Services** *780,000 health files, East European hackers* |
| **53 universities worldwide (ex. John Hopkins, Harvard, Princeton, Stanford, Cornell, Ohio State, New York)** *Hacked by Team GhostShell (Oct 8)* |
| **Northwest Florida State College 279,000** *personal records hacked (Oct 10)* |
| **Intel, Citibank, DreamHost, Healing Touch Day Spa, Ruby's Diner, Masons of California, American Third Position, Boca Ski Club, Sailboat Owners Inc, US Navy, EPA, NASA, PBS, LA County Police Canine Association, NH Dept of Corrections, Verisign, High Tech Crime Solutions, CA Dept of Justice, Computer and Technology Crime High-Tech Response Team (CATCH) , ...** |

## Some Website Confidentiality Breaches from this Year

| |
|---|
| **LinkedIn.com, Mountain View, California**<br><br>6.4 million accounts had their encrypted passwords stolen and posted online |
| **Gamigo, Hamburg, Germany**<br><br>3 million US accounts hacked (8,243,809 million worldwide) |
| **Yahoo! Voices, Sunnyvale, California**<br><br>453,492 users had their plain-text passwords stolen via hacker using SQL injection. |
| **Form spring, San Francisco, California**<br><br>420,000 user accounts on development server exposed |
| **Nvidia, Santa Clara, California**<br><br>400,000 developer forum accounts exposed |
| **PlaySpan, Foster City, California.**<br><br>100,000 User IDs, encrypted passwords, and email addresses exposed. (Oct 10) |

**www.privacyrights.org/data-breach/**

**www.networkworld.com/slideshow/52525**

## 3. CryptDB Principals and Design Techniques

According to the paper, CryptDB is designed to address the weaknesses of already current solutions which are either too slow or do not provide the necessary confidentiality.

CryptDB adds a proxy server and some other components to the typical structure of database backed applications, which consists of a DBMS server and a separate application server, as shown in the figure below:

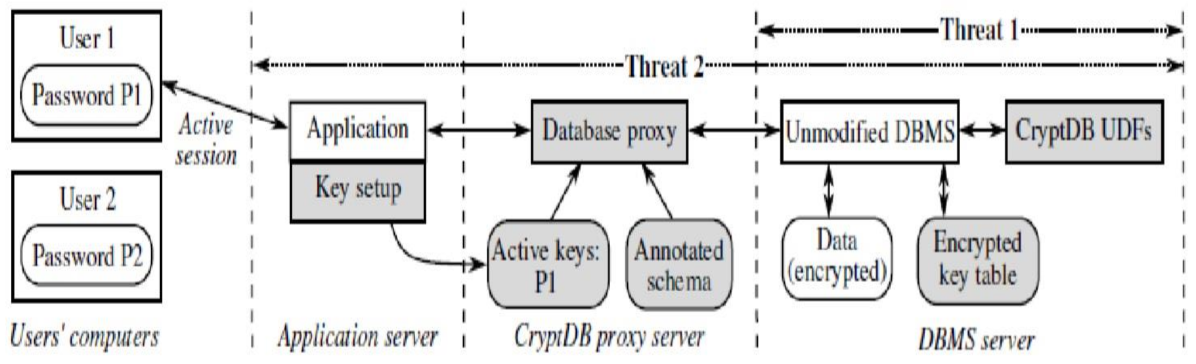

***Figure*** *Error! Bookmark not defined.* ***CryptDB Principals Design***

There are three approaches that CryptDB uses to solve the problems of the current approaches. The SQL-aware encryption approach uses the fact that SQL queries have a well-known structure consisting of operators such as order comparisons, equality check and aggregates like sum and table joins. CryptDB then uses cryptographic methods for joins to transform the queries to a form that can enable the DBMS to run them on encrypted data. On the other hand**, the adjustable query-based encryption** solves the problem seen in the first techniques where certain cryptographic schemes leak more data than required. However, because they are still needed, unions of encryption are needed to carefully adjust the queries to minimize data leakage.

The third approach, which is seen to protect users that are not logged in to a system, is to **chain the cryptographic keys to user passwords** to enable data decryption to users with access privileges.

## 4. CryptDB (Overview)

- **Key ideas:**

➢ Introduce a db proxy which executes queries over encrypted data

➢ “Onions of encryption” adjust the encryption scheme depending on the run time SQL query so that most secure encryption is used which still allows efficient query evaluation

➢ Chain encryption keys to the online user’s password so that a user’s private data can only be decrypted if that user is on line.

- **Component System:**

**Application server**

➢ runs the application code and issues DBMS queries on behalf of one or more users

➢ is modified so that it provides the db proxy with encryption keys

**DBMS server**

➢ all its data is encrypted (including table and column names)

➢ cluelessly processes encrypted data as if it were unencrypted

➢ has user defined functions (UDFs) installed to allow it to compute on ciphertexts for certain operations

➢ has some auxiliary tables (ex. Encrypted keys) used by the db proxy

**Database proxy**

- Encrypts queries received from application and sends them to DBMS
- Changes some query operators if necessary, while preserving query's semantics
- Decrypts DBMS returned results and send them to the application
- Stores master key(s) and an annotated version of application schema (which it uses to check access rights, and to keep track of current encryption level, or onion layer, on *each column*)
- Decides the key = f(user, query) to be used for encrypting/decrypting each data item

## 5. Processing a query:

**1.** db proxy intercepts application’s query and rewrites it by anonymizing table and column names & encrypting constants with the key of the encryption scheme best suited for the operation and the user

**2**. db proxy checks if the DBMS needs to adjust encryption level before executing the query

- if yes → issue an UPDATE query that invokes a UDF to adjust the encryption level layer of the appropriate columns

**3**. db proxy sends the encrypted query to the DBMS server

**4.** DBMS executes query using standard SQL (invoking UDFs for aggregation or keyword searches) and returns the encrypted results

**5**. db proxy intercepts and decrypts results, and sends them to the application

## 6. Queries over Encrypted Data

To secure data the normal database schema is changed in order to hide any relations that can be read from the database, and then stored in the CryptDB proxy. In fact, table and column names are encrypted. Column encryption depends on the data held by that column, and the type of queries to be ran by the DBMS.

Depending the type of data a column stores, there are six methods of encryption. Random (RND) produces a ciphertext from a column name using a randomly generated initial Vector (IV). Due to the randomness, RND provides a powerful encryption and is suitable when handling sensitive data. Unfortunately, it does allow running of queries that require computation e.g. MAX, SUM and ORDER BY. The second way of encryption is the Deterministic (DET). It provides a weaker security because of the leakage caused by producing the same ciphertext for on same text. DET is a pseudo-random permutation. Order-preserving encryption (OPE), as the name suggests, it preserves the order of ciphertext to remain as they were in plaintext. For example, for any key k, if $x < y$, then $OPEK(x) < OPEK(y)$. OPE is comparatively weaker that DET because it reveal order. The fourth way of encryption is Homomorphic (HOM), which is handy for any data that requires computation. With it, complex mathematical computations are possible as they could be done plaintext. However, the authors of this paper never implemented this test in their experiment. Another way of encryption is (JOIN and OPE-JOIN), which is used to join columns to hide the correlation between cross-column. This is because different DET keys are used. Joins are done for equality and order checks. Lastly, word checks (SEARCH) is a method used to search encrypted words. This method is used in queries with SQL operations such as LIKE. SEARCH is a secure as Random because it does not allow the DBMS to see whether some of keyword repeats in many rows.

## 7. Implementing Layered Encryption

Each type is best suited for a different SQL operator. Listed from strong to weak writ cryptographic strength.

- Seven different cryptographic systems are available to be cascaded into onions (RND, HOM, SEARCH, DET, (OPE-)JOIN, OPE)
- Strong systems leak less data from a compromised DBMS and are used as the onion's outer layer.
- Inner onion layers are progressively weaker systems and are only accessed as required by the query.

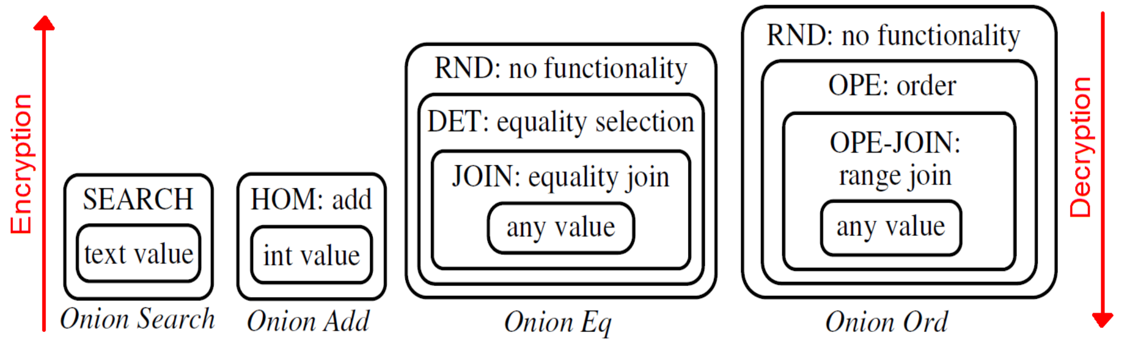

***Figure 1 Layered encryption / Decryption protocol***

The encrypted query now reaches the DBMS with encryption keys. With a few User Defined Functions (UDFs), it runs successfully, and the data that it returns to the proxy is decrypted and sent to the application. It is worth noting that CryptDB use layered encryption to provide data secrecy, as shown in figure below. This varies and this is because of the types of queries and requirements on data access.

## 8. Queries over Encrypted Data

To secure data the normal database schema is changed in order to hide any relations that can be read from the database, and then stored in the CryptDB proxy. In fact, table and column names are encrypted. Column encryption depends on the data held by that column, and the type of queries to be ran by the DBMS.

Depending the type of data a column stores, there are six methods of encryption.

Random (RND) produces a ciphertext from a column name using a randomly generated initial Vector (IV). Due to the randomness, RND provides a powerful encryption and is suitable when handling sensitive data. Unfortunately, it does allow running of queries that require computation e.g. MAX, SUM and ORDER BY. The second way of encryption is the Deterministic (DET). It provides a weaker security because of the leakage caused by producing the same ciphertext for on same text. DET is a pseudo-random permutation. Order-preserving encryption (OPE), as the name suggests, it preserves the order of ciphertext to remain as they were in plaintext. For example, for any key k, if $x < y$, then OPEK(x) < OPEK(y). OPE is comparatively weaker that DET because it reveal order. The fourth way of encryption is Homomorphic (HOM), which is handy for any data that requires computation. With it, complex mathematical computations are possible as they could be done plaintext. However, the authors of this paper never implemented this test in their experiment. Another way of encryption is (JOIN and OPE-JOIN), which is used to join columns to hide the correlation between cross-column. This is because different DET keys are used. Joins are done for equality and order checks. Lastly, word checks (SEARCH) is a method used to search encrypted words. This method is used in queries with SQL operations such as LIKE.  SEARCH is a secure as Random because it does not allow the DBMS to see whether some of keyword repeats in many rows.

## 9. Cryptographic System Details

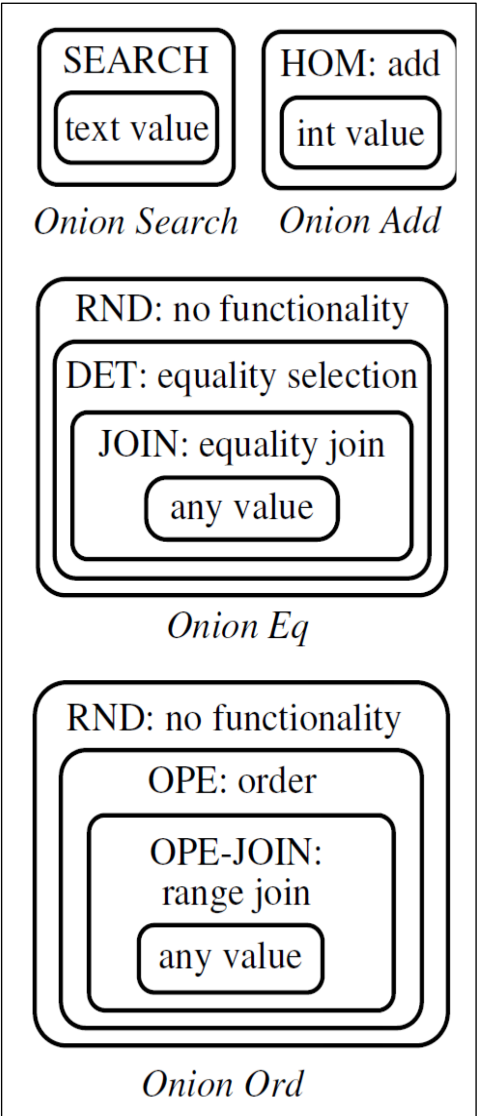

***Figure 2 Cryptographic methods***

**Random (RND):** same plaintext produces different ciphertexts

- no leaks (strong), but no efficient computation
- used for SELECTs with no predicates

**Deterministic (DET):** same plaintext produces the same ciphertext

- leaks the number of occurrences (i.e. histogram)
- allows equality checks (SELECTs with equality predicates, equality joins, GROUP BY, COUNT, DISTINCT

**Order-preserving encryption (OPE):**

plaintext1 < plaintext2 ←→ ciphertext1 < ciphertext2

- leaks order (weak)
- only used when demanded by user
- allows range queries, ORDER BY, MIN, MAX, SORT
-

**Homomorphic encryption (HOM):**

$$^{E}\mathrm{KeyHOM}^{(\mathrm{plaintext1})} * {}^{E}\mathrm{KeyHOM}^{(\mathrm{plaintext2})}$$

$$= {}^{E}\mathrm{KeyHOM}^{(\mathrm{plaintext1} + \mathrm{plaintext2})}$$

→ allows SUM, +, AVG when proxy invokes UDF

## 10.Multiple Principals

With the real checks that were used to evaluate CryptDB, with phpBB for example, it is quite open. In case the proxy and the infrastructure are untrusted, CryptDB can capture the application's access policy. This is where application developers have poser to annotate database schema in CryptDB to specify which principal has access to a certain subset of data items. CryptDB provides three straightforward steps that an application developer can use to annotate schema: specify the principle (such as users, groups, or messages) with the PRINCTYP annotation; specify columns with sensitive data and which principles will have access to them with ENC_FOR annotation; decide how to delegate a principle's rights to another with SPEAK_FOR annotation.

Each principal is associated with a randomly chosen key. If principal B speaks for principal A, then principal A's key is encrypted using principal B's key, which allows principal B to gain access to not only principal A's key but also his data. Each sensitive field is encrypted with the key of the principal in the ENC_FOR annotation. CryptDB encrypts the sensitive field into onions and onion keys are derived from a principal's key. The key of each principal is a combination of a symmetric key and a public-private key pair. CryptDB generally uses the symmetric key to encrypt any data and keys of other principals that are accessible to a certain principal. If a user is not online, CryptDB encrypts the data user his/her public key instead. When encrypting data in a query or decrypting data from a result, CryptDB follows key chains starting from user password until it obtains the desired keys.

## 11. Results

***This paper reveals that CryptDB has few application and theoretical limitations;***

- Logged in users' data is at high risk.
- How to secure the cryptographic keys seems to be an overhead to the whole system.
- The computation involved in of encrypting a query becomes intensive, as shown in figure below.
- A single encryption method is not sufficient. Thus, combing them becomes an overhead.
- In some stages, CryptDB leaks data and over time, and this will allow the attackers to study the layout, which will finally enable them eavesdrop on user's data.

***On the other hand, the standpoints of CryptDB that this paper digs into include;***

- Use of high standards of encryption
- A large number of query types are including because of use of a variety of encryption methods and query adjustments.
- It is fast. This is from the results of the real tests that were carried out on phpBB, HotCRP and grad-apply.
- The layered encryption can surely provide a complex technique that makes it easy to provide different data sets to different users.
- The ability to allow application developers to affect control by use annotated principles.

## 12. Conclusion

The paper that is presented here is complete, and with logical sentiments. The theory is rich and the authors proofs that what they have done has never been done before. This is in line with their literature review. The problem that the paper addresses have been derived in a logical and the method of inquiry is amazing.  Tests results have been presented substantially and statistically, and meaningful. What is encouraging is that the authors have remained in the scope that they had promised the reader. Database security is a nightmare in real world and the authors’ inquiry is relevant and has added to the database knowledge.

## 13. References

**[1] -** Raluca, A. P., Catherine, M. S., Nickolai, Z., & Hari, B. (2012). CryptDB: Protecting Confidentiality with Encrypted Query Processing. International Journal of Computer Applications, 45 (8), 84-99.

**[2] -** R. Agrawal,J.Kiernan,R.Srikant,andY.Xu. Order preserving encryption for numeric data. In Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, Paris, France, June 2004

**[3] -**  F. Bao, R. H. Deng, X. Ding, and Y. Yang. Private query on encrypted data in multi-user settings. In Proceedings of the 4th International Conference on Information Security Practice and Experience, Sydney, Australia, April 2008.

**[4] -** A. Boldyreva, N. Chenette, Y. Lee, and A. O’Neill. Orderpreserving symmetric encryption. In Proceedings of the 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT), Cologne, Germany, April 2009.

**[5] -** D.BonehandB.Waters. Conjunctive,subset,andrangequeries on encrypted data. In Proceedings of the 4th Conference on Theory of Cryptography, 2007.