Order-Preserving Encryption

-state of the art + risk assessment-

I. Introduction

Order-Preserving Encryption (OPE) has been considered in literature during the last years and is a popular tool to encrypt data before outsourcing it. We want to emphasize that sorting is a very useful property to preserve on the encrypted data because it enables efficient range queries.

OPE is part of a large cryptographic area called *Searches on encrypted data*. Besides OPE, searches on encrypted data can also be performed using other cryptographic schemes such as *searchable, functional and homomorphic encryption.*

In this document we will discuss the work made in this direction, techniques and methodologies used and evaluation methods, results in the Related work section. The security of OPE has been being debated in the last years and we will describe the possible attacks in the Risk assessment section. We conclude in the last section.

II. Related work

[1] Agrawal, R., Kiernan, J., Srikant, R. and Xu, Y. *Order preserving encryption for numeric data.* In *Proceedings of the 25th International Conference on Management of data (2004), SIGMOD.*

It represents the first OPE scheme. The authors assumed that the distribution of plaintext is known which might lead to some statistical attack. Hence, they modify this distribution to mach a randomly chosen one. The paper does not provide any formal security guarantee for the scheme. The paper provides an entire section about the methods evaluation, but the most significant are those related to the time required for inserting new tuples, retrieval of tuples. They tested their scheme on DB2 IBM database and concluded that the performance is quite good and might be used in production.

[2] Boldyreva, A., Chenette, N., Lee, Y. and O' Neil, A. *Order-preserving symmetric encryption.* In *Proceedings of the 28th International Conference on Advances in Cryptolgy (2009), EUROCRYPT.*

This paper gives the first formal security guarantee of OPE and introduces the notion of *indistinguishability under ordered chosen plaintext attack* (IND-OCPA). The authors also proved that no stateless scheme can achieve this notion, in particular their scheme requires storing a key on the client. The paper uses new techniques from probability theory, it gives a suggestion for the implementation and still the authors do not recommend the practical use of their construction. Evaluation of the scheme was performed in a mathematical sense and the authors do not specify concrete practical results.

[3] Popa, R., A., Li, F.H. and Zeldovich, N. *An ideal-security protocol for order preserving encoding.* In *34th IEEE Symposium on Security and Privacy (2013), S&P.*

This paper gives the first IND-OCPA secure OPE scheme. It is stateful and requires storing information on an OPE server that they assume is ideally placed at the client side. They run a multi-round protocol which makes their scheme very inefficient due to the network delay. The size of the stored information is linear in the numbers of distinct plaintext. Also the performance is influenced by a lot of updates of the ciphertexts. The implementation was made in C++ with MySQL .

[4] Kerschbaum, F. *Frequency-Hiding Order-Preserving Encryption.* In *Proceedings of the 22nd Conference on Computer and Communications Security (2015), ACM SIGSAC*

This paper gives a new OPE scheme that achieves a strictly stronger notion of security called *indistinguishability under frequency- analyzing ordered chosen plaintext attack (*IND-FAOCPA*).*The frequencies of the plaintexts are hidden by using randomized encryption which is a new approach never used in the other previous schemes. It is stateful and ciphertext might change their initial form. The state is represented as a binary search tree which can be rebalanced to make searches more efficient. The scheme was implemented in Java 7 on a Intel-Core I5-4300 CPU with 1.9-2.5Ghz and 16 Gb RAM. Also the paper provides a security proof under IND-FAOCPA. The client storage space was reduced by different compression techniques which makes it less secure. The scheme can be favorable when higher security than deterministic OPE is desired and client storage and search time should still remain low.

III. Risk Assessment

All the schemes described above cannot be semantically secure (IND-CPA) because it is revealed the order of plaintexts. So, we are interested whether there exist other *leakages* such as plaintext bits, plaintext frequencies, relative distance between plaintexts and approximation for plaintext location. In this sense we present the most practical attacks that one should consider when implements an OPE scheme:

A) Frequency analysis: is a well-known attack that decrypts deterministic-encrypted columns given an auxiliary dataset that is well-correlated with the plaintext column (i.e. dictionary attack). The extent of the correlation needed, however, is not significant and many publicly-available datasets can be used to attack various kinds of encrypted columns with this attack

B) L_p-optimization: is a new family of attacks we introduce that decrypts deterministic-encrypted columns. The family is parameterized by the $\mathbf{L_p}$-norms and is based on combinatorial optimization techniques.

C) Sorting attack: is an attack that decrypts OPE-encrypted columns. This attack is very simple but, as we show, very powerful in practice. It is applicable to columns that are dense in the sense that every element of the message space appears in the encrypted column. While this may seem like a relatively strong assumption, we show that it holds for many real-world datasets.

D) Cumulative attack: is a new attack we introduce that de-crypts OPE-encrypted columns. This attack is applicable even to low-density columns and also makes use of combinatorial optimization techniques.

IV. Conclusions

We spent some time for researching the OPE domain which is still at the beginning and no one should use existing constructions in production. There is a lot of work that need to be done in order to make the OPE practical.

Our project will consist of implementing [4] because it is a very recent work and as far as we have seen there no exists an implementation on the public repositories. This work will make us to better understand the OPE concept and might represent a good starting point for our future research.