Searchable encryption

1. **What did the others, techniques and methodologies used, evaluation methods, results**

Over the years, the problem of encrypted search has become an important problem in information security and cryptography. This is because of the fact that, with the ever increasing amounts of data, searching has become the primary to access it. For the same reason, the storage of this data is being delegated / outsourced to third parties (i.e. cloud services providers) which raises security issues with regards to the privacy and confidentiality aspects as there is no guarantee that these third parties are to be completely trusted.

In trying to address the aforementioned issues, the problem of encrypted search has become of interest to many sub-fields in computer science (e.g., databases, security, cryptography) but also to industry and to governments, both of which are even more sensitive to the privacy of their information. For that reason, many businesses are skeptical to using cloud services, as it is necessary to authorize access to an external administrators. The data managed by a business includes information that must not be made public outside the company.

In order to avoid information leakage, there are various methods to encrypt the data entrusted to outsiders. By using common cryptographic technologies, the processing of encrypted data is not possible, i.e., it is only possible to store the encrypted data in its current state.

The Hitachi group have been researching a cryptographic technology, called “Private Information Processing”, which allows the encrypted data to be processed, developing techniques such as *“searchable encryption”*.

Also, Mitsubishi Electric has developed a searchable encryption platform software that will improve the security of cloud services where we store sensitive information in an encrypted manner and retrieve it without decrypting. This platform supports both data encryption and access control for users who perform keyword searches in the cloud.
Key features:

- Robust encryption technology - the platform uses an encryption algorithm which enables documents to be encrypted in a way that only the users of designated groups can perform keyword searches.
- Large data searches, e.g., 1-3 seconds, or less - it proposes a new index generation software for searchable encryption which grants authorized users to build a ticket, which is used to generate an index from the encrypted data for accelerated searches. It has also implemented a new automatic parallelization system which results in a searches that require 1-3 seconds for 100000 pieces of data.

There are six different ways to search on encrypted data, each based on the following cryptographic primitives:

- **property-preserving encryption** (PPE): Provides fast searching but leaks quite a bit of information to the server.
- **functional encryption** (FE): Slightly slower search but more secure and leaks less information tot the server. There are multiple FE schemes achieving various properties such as attribute-based encryption (ABE), identity-based encryption (IBE), hidden vector encryption, predicate encryption, etc.
- **fully-homomorphic encryption:** this form of encryption allows computations over encrypted data without the need to decrypt it first, and as such it is ideally suited for performing information retrievals
- **searchable symmetric encryption:** the main technique used for searchable encryption, it consists of slightly modified versions of classic symmetric cryptosistems
- **oblivious RAMs**
- **secure two-party computation**

Searchable symmetric key encryption.

Private-key storage outsourcing allows clients with either limited resources or limited expertise to store and distribute large amounts of symmetrically encrypted data at low cost. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. To address this, several techniques have been proposed for provisioning symmetric encryption with search capabilities the resulting construct is typically called symmetric searchable encryption.

Let say we have a list of words L = (w_1, w_2, ... ,w_n).
In such a scheme, the user has 3 keys, named a,b,c. The idea is to encrypt each word w_i and obtain C_i.

To encrypt w_i, we encrypt this word with the "a" key and we obtain E_a(w_i) = L_i||R_i. E is a deterministic algorithm.

Then we use a PRG(i,key_b) = S_i.

PRF_c(L_i) = k_i.

After, we calculate T_i = S_i || F_{k_i}(S_i).

And then C_i = E_a(w_i) XOR T_i

If we want to search for the word w_i, we will send to the server E_a(w_i) and K_i.

Server will XOR CT_i with E_a(w_i) and then it will get S_i || V*. if F_{k_i} == V* then true, else false;

We can easily search using boolean operators and phrases. We can also search using regular expressions, but we need to do a lot of queries.

Difficulties:

Work is linear.

Handle variable length words:

Problems at padding (it cannot be random because the user needs to know the padding for the search).

We can put before the word its length but w_i can't reveal the length to the server because it can do some statistical attacks. It can be searched in binary way, but we don't want that because of the complexity.

2. **Important names in the field, research teams**

- Dawn Xiaodong Song, David Wagner and Adrian Perrig from Berkeley
- Oded Goldreich and Rafail Ostrovsky
- Andrew Yao
- Dan Boneh and Matthew Franklin
- Stanford Applied Cryptography Group
- Microsoft Research, Cryptography Group
- Hitachi Research,

3. **Related Articles and books**

http://www.cs.berkeley.edu/~dawnsong/papers/se.pdf : paper about Practical Techniques for Searches on Encrypted Data (first SSKE scheme)

-Proved that SSKE is a PRG (security related to that of PRF and PRG used in construction)

4. **Relevant links**

http://www.slideshare.net/hnx/slides-33732614 : Searching in privacy slides.

-Motivation

-Speak about SSKE

https://crypto.stanford.edu/~eujin/papers/secureindex/2003nov-encsearch.pdf : How to search on encrypted data

-SSKE (Searchable symetric key encryption)

-SPKE (searchable public key encryption) using bilinear maps

-Secure Indexes for Searching Efficiently on Encrypted Compressed Data

http://www.mytechblog.in/wp-content/uploads/2015/01/fuzzy-keyword-search-over-encrypted-data-in-cloud-computing-www.MyTechBlog.in-Ankit-Wasankar.pptx : fuzzy keyword search over encrypted data

-The Hitachi Group

http://www.hitachi.com/rd/portal/contents/story/searchable_encryption/index.html

-Mitsubishi Electric

http://www.mitsubishielectric.com/news/2013/0703.html

http://outsourcedbits.org/category/encrypted-search-2/ : Seny Kamara's blog

-Introduction (Part 1)

-Deterministic Encryption (Part 2)

-Functional Encryption (Part 3)
-Oblivious RAMs (Part 4)
-Searchable Symmetric Encryption (Part 5)

http://www.ciphercloud.com/blog/cloud-data-encryption-easy/ : Cloud date encryption is Easy

Risk assessment

- One of the main risk in searchable encryption is finding the balance between performance and security