Searchable Symmetric Encryption- A Survey

Abdul Nasir
M.Tech Scholar, R.N.Modi Engineering College, Kota

ABSTRACT
Cryptology has opened possibilities for a fully secure communication and transfer of information between parties. With time, a lot of primitives have been proposed as improvements of the earlier ones. This paper focuses on a relatively new primitive, Searchable Symmetric Encryption. It allows updations in the information while it is stored on a trusted third party due to lack of resources. Earlier considered as impossible, the proposals by Goh using secure indexes and Ostrovsky-Goldreich on oblivious RAMs failed such notions. Now, information can reliably be stored and each segment of information can be retrieved when the need arises. SSE provides practical efficiency unlike generic cryptographic primitives like Homomorphic Encryption and Multi-party computation. Being a honeypot for researchers, a lot of developments have been going on in the direction of further optimizing SSE. A state-of-art discussion followed by a brief survey of some noteworthy research works is provided in the paper.

Introduction
Cryptography aims at securing important information to be sent from one party to another through encryption such that only the receiving party can decrypt the information and use it further. The encryption and decryption processes require a key which may be shared between the parties as in the case of public/symmetric key encryption or different for both parties, i.e public for encryption and private for decryption purposes as in private/asymmetric key cryptography. Fig. 1 shows both the cryptography models through a conversation between two users Bob and Alice. Private Key encryption has been widely used seeing the complexity of handling a shared key between the two parties. When the sender cannot rely on the receiver for exchange of keys, a trusted third party comes into action with advancement in technology to allow a confidential to and fro of information between a sender and a receiver.

The standard private-key encryption and storage schemes have gained much appraisal because of their effectiveness of storing confidential documents on a third party to retrieve them back when the need arises. The need of outsourcing the information arises from the lack of resources and expertise to handle symmetric encrypted data while maintaining cost involved in the storage. However, techniques to allow searching on the stored encrypted data and retrieving only some segment of the data has been a significant area of research since long. As per DARPA, it is a matter of privacy and national security in information aggregation systems to allow searching the encrypted data [1]. Allowing search queries in the symmetric stored data can therefore be given the name “Searchable Symmetric Encryption (SSE)”.

Fig 1 (a) Symmetric/Public key encryption; (b) Asymmetric/Private key encryption
One approach to handling SSE is through secure indexes. Index can be referred to a data structure that supports keyword searches while storing documents. For searching a document with a keyword, let’s say v, a pointer is returned by the index to the documents that contain this keyword. The question arises as to how to secure an index, a solution to which is using trapdoors. A user executing a keyword search should possess a trapdoor which can be generated only through a secret key. No key, no information leakage. The very first work in the direction of using secure indexes was by Goh [2]. He proposed indexing and encrypting a document by the user followed by sending the encrypted document and secure index to the server. At the time of searching a document with keyword v, a trapdoor is generated and sent by the user to the server through which it recovers pointers to the needed documents. Fig. 2 shows the working using secured indexes.

Another full-fledged work in this direction is on oblivious RAMs by Ostrovsky and Goldreich [3,4]. Their proposed work provides a security guarantee that no information, even the access patterns of which documents contain the needed keyword, be leaked in the process. However, a limitation of the proposal is the number of logarithmic rounds involved in the interaction of a read and write. The authors worked on this issue and provided a 2-round solution but with a large square root overhead. This meant that only through weakening of the privacy guarantees can their proposal be giving efficient solutions.

Recent Developments

Cao et al [5] worked on Multiple-keyword Ranked Search over Encrypted cloud data (MRSE) after considering the large number of users and the documents in a third party like cloud. For deducing similarity, the “Coordinate Matching” multi-keyword semantic is used which focuses on as many matches as possible between the keyword query and the documents to be searched. An “Inner Product Similarity” then quantitatively evaluates the similarity. Authors give two MRSE schemes according to two threat models. The proposal when evaluated resulted in a significantly lower overhead on computation and communication.

Kamara and Papamanthou[6] designed a parallel and dynamic SSE scheme and pointed out the sequential and linear nature of the existing SSE schemes where a sequence of memory locations are accessed by the search algorithm next to the previous location. In contrast, the authors designed a sub-linear SSE scheme to be used in the multi-core environments. The search complexity is found significantly lower than the previous search schemes even with increasing number of cores. The other advantages of the scheme include a strong adaptive security against chosen-ciphertext attacks, no information leakage except what can be inferred from the previous search tokens, effective implementation in external memory and high security in a random oracle.
model. Also the proposal uses a very simple data structure of red-black trees for construction and implementation.

Popa and Zeldovich [7] proposed a multi-key SSE scheme allowing searching of a keyword in many documents encrypted with different keys. A single word search query is provided to the server with hidden document and search contents and the only information available to the server is regarding if the some word of any document matches the given keyword. The assumptions taken are the Bilinear and External Diffie-Hellman assumptions in the random oracle model. However, the two challenges encountered are the presence of no-single trusted user and the practicality of the scheme as the scheme is to be used in a real system.

Cash et al [8] in 2014 proposed the need of an SSE scheme that scales well with the increasing sizes of the data stored. Larger data sizes motivate outsourced storage; therefore the SSE should scale well to the increasing size of data. Their implementation permits encryption and secure searching on data with sizes of tens of billions record pairs. The first step in their proposal consists of the generalization and simplification step with a generic dictionary structure as an asymptotic improvement to the previous related SSE schemes. The proposal also shows making the scheme dynamic such that it can be changed effectively even after encryption. The operations that can be performed are additions and deletions in data via revocation lists.

Stefanov et al [9] also worked on achieving dynamism in search along with combating the other issues been encountered in the previous DSSE schemes—information leakage and inefficiency of the scheme in terms of complexity achieved being linear to the number of documents. The authors designed a DSSE having the best of the two worlds with significantly less information leakage than the previous schemes and a sub-linear complexity even in the worst case with a data structure having linear size. Also, unlike the other DSSE schemes, the proposal supports dynamic keywords.

Leenders [10] address the too strong security definitions made by Curtmola et al [11] regarding the index-based SSE schemes. As discussed earlier, these schemes require encrypting the index to make the transfer and search more secure. Also, Curtmola et al [11] in their paper also defined Adaptive and Non-Adaptive security, where the adversary chooses the attacks based on previous results or without any such knowledge respectively. Leender’s work [10] is in the direction of designing a simple adaptively secure SSE scheme with searching complexity linear to the amount of documents. The rest of their paper focuses on proving that no additional data confidentiality is achieved by encryption of index and hence, the security definitions defined in Curtmola et la’s paper [11] are too strong.

Yavuz and Guajardo [12] in 2015 proposed another work on Dynamic SSE (DSSE) using a bit matrix with two static hash tables. The authors claim to have designed a highly secure SSE scheme with highest privacy among all the related proposed schemes till date. The advantages of the scheme include high security, compact client storage with and without secure updates, constant update storage overhead, dynamic keyword universe, efficient and non-interactive updates, oblivious updates, parallelization and Forward privacy. The proposed scheme is found secure in the random oracle model and the practicality of the scheme is visible even with a large number of file-keyword pairs.

The earlier works were limited to keyword searches until Chase and Shen [13] proposed substring search queries in SSE. The query as reported in their paper can be outlined as “Given a data string s and a search string p, find all occurrences of p as a substring of s”. The authors first generalize two cryptographic primitives—SSE and Structured Encryption. A query able encryption scheme is then constructed for substring queries through the use of suffix trees in construction which give security comparable to the unencrypted suffix trees. With a satisfactorily lower cost and three rounds of communication, their proposal also is proved secure against malicious adversary attack. However, security guarantees are weakened and leaks information about memory access patterns.
Conclusion

In a short span of time, Searchable Symmetric Encryption has gained much popularity due to the ease of updating the stored encrypted documents which otherwise had to be completely processed, even if a segment of information was needed. Allowing search queries brings much ease to the users and saves much overhead both in terms of cost and time. Over the years, the field has attracted many researchers who have then come up with solutions to the limitations encountered in SSE. Starting from the Goh and Ostrovsky-Goldreich works, the field has seen a remarkable interest through researchers, a review of which is provided in the paper. Almost all relevant works in this field have been discussed briefly to provide a primer to this new primitive of cryptography.

REFERENCES