Contents

1 Introduction 1

2 Methods 1

3 Results 3
   3.1 OXT Extension to Full Security 3
   3.2 MC-OXT Extension to Full Security 4

4 Discussion 5
1 Introduction

Verifiable search on encrypted files is motivated by the recent surge toward cloud storage. A weak client with storage limitation can delegate the burden of data processing to a server. This process allows the client to efficiently search through a database without investing much computational power. This problem is commonly referred to as Verifiable Keyword Encrypted Search (VKES). The aim of this project is to develop efficient protocols to verifiably perform searches over outsourced encrypted database with full security guarantee.

In the literature, a variant form of VKES problem called Searchable Symmetric Encryption (SSE) has been considered. The SSE protocols are efficient on massive data sets and can handle expressive queries such as Boolean search. However, the SSE protocols are limited because for their inability to verify the correctness of outsourced computations. We hope to improve the current SSE solutions to have full security akin to those provided by VKES protocols. Adding this new feature would allow a client to efficiently verify whether the server faithfully executed a search query.

2 Methods

**Verifiable Sets using VKES Protocol:** For a given set $S$, a client stores $S$ with the server along with some authentication information $Auth_S$. Later the client queries the server for an arbitrary keyword $w$; the server returns the resulting output in the form $(w \in S$ or $w \notin S)$ together with a unique authentication tag $auth_w$. The client then verifies the answer using the authentication information. The authentication tag prevents the server from returning false results. This primitive guarantee, with negligible probability, that a malicious server is incapable of forging a false answer that passes the verification test. For the implementation of the above protocol consult [2, 6].
THE OXT SSE PROTOCOL proposed in [1], works as follows. The database owner defines a pseudo-random function \( f \), which maps keyword space into integer space.\(^1\) The database owner creates an inverted index for all documents \( \{D_1, \ldots, D_n\} \) and stores them in a table \( T \). For each keyword \( w_i \), the \( i^{th} \) row of \( T \) contains a vector of indices \( (i_1, \ldots, i_k) \) such that \( w_i \in D_{i_\ell} \) for all set of documents containing the keyword \( w_i \).\(^2\) The database owner also creates another data structure called \( XSet \). The elements of \( XSet \), called cross-tags, are computed as follows: for every keyword \( w_j \) in document \( D_i \), the value \( H(i, w_j) \) is stored in \( XSet \).\(^3\) Finally, the database owner outsources the encrypted data structures \( T \) and \( XSet \) to the cloud.

During the searching phase, let’s suppose that a client wants to search for the following conjunction \( (w_1 \land \ldots \land w_m) \) of keywords. First the client identifies the primary term \( w_1 \) and queries the server for \( f(w_1) \); this primary search is performed in \( T \) by the server and is used as filter for subsequent searches. This gives the client the list \( (i_1, \ldots, i_k) \) of the indices of the documents containing the primary term \( w_1 \). The client is then able to compute all the cross-tags \( H(i_\ell, w_2), \ldots, H(i_\ell, w_m) \) through a shared computation and the server checks membership of the cross-tags in \( XSet \).

THE MC-OXT SSE PROTOCOL proposed in [3] is an extension of OXT protocol to a multi-client setting. Here, the database owner outsources encrypted data to the server and issues search tokens to clients. The clients learn information about the database through queries they are authorized to learn. To avoid unauthorized searches, the MC-OXT protocol uses shared computations between the server and the clients. The implementation of this protocol uses Diffie-Hellman blinded exponentiation; the blinding factors are used to avoid more leakages to the server.

\(^1\) A pseudo-random function \( f \) is an efficient computable function that looks random to polynomial time adversary.
\(^2\) For simplicity, we used \( i^{th} \) row to represented a pseudo-random function \( f(w_i) \) of the keyword \( w_i \).
\(^3\) \( H \) is a hashing function which is analogous to a pseudo-random function.
The OXT and MC-OXT SSE protocols have well-defined leakages. They allow some harmless information to leak to either the client or the server. This relaxation on leakage helps improve the efficiency of these protocols and its practical applicability. In the OXT protocol, there are two leakages to the server; the server learns the indices of the documents that contain the conjunction and learns the cardinality of the number of matching document from the primary search term. Whereas in the MC-OXT protocol, in addition to the leakage to the server, there are also leakages to the client.¹

3 Results

We focused on two SSE protocols: the OXT and the MC-OXT protocols. We successfully extended the above SSE protocols to full VKES security. In essence, we appended verification steps to each of the protocols. The verification steps are disparate in both protocols because the allowable leakages are different.

3.1 OXT Extension to Full Security

To add verification to OXT, we proceed as follows. First, we eliminated the Diffie-Hellman exponentiation in the protocol and used fast asymmetric cryptosystem for most of the computations involved. At the preprocessing phase, there are three modifications: The database owner creates a master list of all the keywords in every documents i.e. \( w_i = \{ w \in D_i \} \), he then (i) stores a verifiable set \( S \) of the master list; (ii) also stores the set \( XSet \) as a verifiable set; and (iii) authenticate the individual entries of the inverted index stored in table \( T \). These encrypted metadata is then

¹The client in the MC-OXT is different from the database owner.
outsourced to the server; after the setup, a client is then allowed to issue search queries.

Our modification in the search protocol is as follows. To search for a conjunction of keywords \((w_1 \land \cdots \land w_n)\), the client first chooses a primary term \(w_1\) and queries for \(w_1\) in the set \(S\). The server answers with an authenticated result. If the primary term does not appear in \(S\) then the search is terminated since no document contains all the conjunction terms. On the other hand, if the primary term appears in \(S\) then the search proceeds as in the original OXT, except that the client can now (a) verify that the vector of indices returned by the server is correct since it is authenticated and (b) can verify if the cross-tag appears in the \(XSet\) or not since the \(XSet\) is stored as a verifiable set.

The resultant protocol maintains the same leakage as in the OXT; the server learns if a search was successful or unsuccessful. The server also learns if the client repeats the same query multiple times. Although, resultant protocol features more steps; the steps does not grow with the size of the document. The verification steps actually take logarithmic in time and communication. Moreover, the use of fast asymmetric cryptosystem will severely improve its efficiency. Most importantly, in the new protocol, the server will answer any query with a proof of correctness of the result. The client would accept or reject the result depending on whether the verification evaluates true or not.

### 3.2 MC-OXT Extension to Full Security

In the MC-OXT, we did not eliminate the Diffie-Hellman exponentiation because in this setting, exponentiation is needed for the shared computation between the client and the server. Instead during the database setup, in addition to our modification to OXT, for each keyword \(w_i\) the database owner stores a unique group element with authentication. He then stores the set \(T\) and \(XSet\) as a publicly verifiable sets. The publicly verifiable models in [4, 5] provide efficient method for

---

5The primary term is a term, from the conjunction, that appears least in the database
verifying outsourced computations.

Our modification for the token generation and search phase is as follows. The database owner also provides extra information to the client, in addition to what was specified in the original protocol. This new information is in the form of a group element and it is considered as a harmless leakage to the client. In order to minimize data leakage to the party issuing the queries, the protocol in [3] requires that the cross-tags are computed by the server. We also added a verification step to ensure that the cross-tags were computed correctly by the server. The computation of the cross-tags are verified through a novel application of zero-knowledge proofs.\(^6\)

4 Discussion

A possible proof of security for our modified protocols might utilize simulation and reduction arguments to complexity theory. The full security analysis for the zero knowledge protocol can be approached as follows. Construct a simulator that generates a fake conversation indistinguishable from the real conversation. We can show that an intruder learns nothing from the real conversation that he couldn't have learned from running the simulator. In other words, any knowledge obtainable through computation in polynomial time is a knowledge already known.

**New Skills and Knowledge Gained:** This research has equipped me with numerous skills in cryptographic research. I gained insight on how to analyze the computational complexity of an algorithm. I learned the basic foundation necessary to construct efficient algorithms. Data structure is very important in building efficient algorithm; it can improve the efficiency of a search algorithm from linear in the size of the document to a constant. One particular data structure we

\(^6\)A zero-knowledge proof is an efficient method use to prove the validity of that a given statement.
used often was the hash table, which is uses a hashing function to store and access elements of the table.

The interdisciplinary nature of the program was outstanding. More importantly the peer to peer collaboration was essential in the research process. I obtained a team collaboration skills, which is a necessary skill require for excellent research. My team, comprising of a faculty mentor and a fellow graduate student, exposed me to diversity of knowledge in research. Our unique academic backgrounds and intellectual capabilities were quite useful in problem solving; I gained several insights from the team. Indeed, my theoretical background in mathematics was helpful in during the project. I learned from this experience that interdisciplinary collaboration is indispensable in solving research problems.

My participation in this research has allowed me to consult various literatures relating to verifiable encrypted search. I enjoyed reading papers mostly dealing with searchable symmetric encryption. As a result of my exposure to various literatures, I am more aware of the current state of knowledge in the area of cloud computing. In addition, I was also exposed to different techniques used in the analysis of the security of a protocol. One particular proving technique that I enjoy the most was the notion of zero knowledge proofs. Collectively, I believe that these newly obtained knowledge and skills would be useful during the tenure of my PhD studies. Moreover, it would also allow me to collaborate in interdisciplinary research projects.

RELEVANCE TO THE DHS MISSION: This research is relevant to the DHS mission of securing the cyber space. Secured encrypted search over a outsourced database is essential for privacy. Our research has many pragmatic applications in the communication systems. One specific application is in the area intelligence sharing between two or more disparate groups. Two parties can share intelligence efficiently without leaking unauthorized information in the process. For
example, if the FBI has been authorized to search the CIA database, encrypted search, will ensure
the privacy of CIA’s data while still allowing FBI to search for target activities. In order words, the
searching party should only gain knowledge of the database through authorized queries.

Computation outsourcing is important when dealing with big data. Accessing huge data
efficiently relies on the ability to delegate the search work to the server known as and cloud computing. Since there are increasing trend of such delegation, developing an efficient means
for users to verify the result of the server without investing too much computational power is
deemed essential. Verifiable search over an encrypted database addresses verifiability concerns and
guarantees full security to the searching party. A search protocol that ensures security for both the
data user and the data, will help obviate some of the problem that are prevalent in cyber world and
thus further advance the DHS mission.

Impact of Experience: My participation in this summer team program has presented me
with greater understanding of the DHS mission in the area of cybersecurity. Having successfully
completed this program, I now have a better understanding of various research opportunities in
cryptography. Specifically, this program presented me with a broader view of various applicability
of cryptographic research. I have developed a desire to continue working on similar projects
and hope to publish in an academic journal. Moreover, I plan to engage in other interdisciplinary
projects and hope to pursue a career in research and perhaps obtain a job in academia or industrial
laboratory.

Through this research, I acquired not just the necessary knowledge and skills in research
but also the confidence in applying mathematical theories in real-life situations especially in the
area of security system. The competence and expertise that I obtained through this research will
enable me to effect change in the current understanding of communication and interoperability.
This interdisciplinary collaboration has equipped me with the skills that will be useful during my doctoral study.

Ultimately, my appointment was a rewarding experience not just academically but also professionally. I obtained knowledge that will advance both my academic and professional goals. These experiences have inspired me to work harder and broaden my horizon. My participation in the DHS summer team program challenged me to work hard and drove me closer to my aspirations.
References


