Volume XXIII 2020
ISSUE no.1
MBNA Publishing House Constanta 2020

Scientific Bulletin of Naval Academy

SBNA PAPER • OPEN ACCESS

Multi-level access using searchable symmetric encryption with applicability for earth sciences

To cite this article: Marius Iulian Mihailescu, Stefania Loredana Nita and Ciprian Racuciu, Scientific Bulletin of Naval Academy, Vol. XXIII 2020, pg.213-220.

Available online at www.anmb.ro

ISSN: 2392-8956; ISSN-L: 1454-864X

doi: 10.21279/1454-864X-20-I1-029

Multi-level Access using Searchable Symmetric Encryption with Applicability for Earth Sciences

Marius Iulian Mihailescu[1], Stefania Loredana Nita[2], Ciprian Racuciu[3]

[1]R&D Department, Dapyx Solution, Bucharest, Romania
[2]Computer Science Department, University of Bucharest, Romania
[3]Computer Science Department, University of Bucharest, Romania

E-mail: marius.mihailescu@hotmail.com

Abstract. Accessing the files remotely offers a rentable solution for file storage. If the files are sensitive these should be encrypted before they will be outsourced for assuring confidentiality. These being said, the searching process becomes a real challenging problem. Most of the schemes consider only the scenarios where the users can search entirely over the encrypted files or data. In practice, the sensitive data are classified using an access control policy and different users should have different rights. The current paper will examine and propose a scheme of a multi-level access control policy (MlACP) based on searchable symmetric encryption (SSE) for the software infrastructures that ensure the interoperability of the software applications that access and work with documents and files. The proof of MlACP is The accessibility of the documents is defined by an access control policy designed to take into consideration the applicability in a multitude of IT infrastructures', such as maritime, environmental protection and ecology, physics analysis and statistics software applications.

1. Introduction

Every day, the distributed systems (e.g., cloud computing and big data) for different multi-disciplinary software applications (e.g., meteorology – earth sciences [16, 17, 18, 19, 20]) becomes more and more relevant. A variety of distributed systems have been launched in the last few years, and several educational institutions have switched to the cloud and digitized the courses. During the last few years, many distributed systems have been introduced, and many research institutions have turned to the cloud and digitized their activity. Searchable encryption may be used to ensure the protection of the data. In this way, all forms of data can be stored in encrypted format in the cloud, and users can get certain materials or documents from the server that satisfy a search requirement (based on keywords).

Searchable encryption gives the possibility for the search process to be done directly over encrypted data, without needing to retrieve it from the cloud server. Another benefit is that the search and decryption can only be performed by users who are allowed to work with that particular data.

Searchable encryption (SE) is one of the most powerful encryption techniques, which gives the possibility to the user to search for keywords over encrypted documents. It is very important to understand which are the participants into the system and to draw a line between them, especially when access control policies are being implemented into such a system. The participants can be categorized as the data user, who owns a set of documents $Doc_{set} = \{Doc_1, \dots, Doc_2\}$, putting the system into the state that is necessary to generate the keys, to assure their encryption and store them within a cloud server; the data user can submit

search queries on the cloud server; the cloud server, which stores the encrypted documents and invokes the search algorithm (see Section 4).

A standard searchable encryption technique contains the following algorithms **Error! Reference source not found.**:

- $Key_{generation}(\lambda) \rightarrow (Public_{key}, Private_{key})$: to generate the key, the security parameter (λ) is required, without it, the generation of the key will not be possible. The security parameter λ will help us to build a pair formed out of a public key and a private key is generated, ($Public_{key}$) respectively ($Private_{key}$);
- $E(Document_i, Public_{key}) \rightarrow Enc(C_i)$: the output of the algorithm will consist in the encrypted document $Enc(C_i)$. The algorithm output is based on the encryption function E which has two parameters, the public key $Public_{key}$ and a document $Document_i$;
- $Build_{Index}(Document_i, keyword, Public_{key}) \rightarrow Index$: build index algorithm has as input the following parameters, the document D_i, the $keyword$associated with the document and the public key $Public_{key}$. The output is represented by an index structure that is based on the association between the documents and the keywords;
- $Trapdoor(keyword, Private_{key}) \rightarrow trapdoor_{keyword}$: trapdoor algorithm has two parameters as input, the keyword-based on which the search is made and the secret key. The output is a trapdoor value $trapdoor_{keyword}$;
- $Search(trapdoor_{keyword}, Public_{key}, Index) \rightarrow Enc(C)$: The search algorithm has as input the following parameters, trapdoor value, and the public key. The output is represented by the encrypted documents $Enc(C) = \{C_{il}, \dots, C_{iw}\}$ in association with their keyword;
- $Decryption(C, Private_{key}) \rightarrow Dec(Enc(C))$: decryption algorithm has as input parameter the C of encrypted documents and the secret key $Private_{key}$. The output is represented by a set $Documents_{set} = \{D_{i1}, \dots, D_{iw}\} \subset Search$ of decrypted documents.

The general goal of our work. The general goal of our paper is to provide a first practical attempt in combining searchable encryption and access control policy in a real distributed system, in such a way that the main servers of the data center to be able to provide access to the files based on the access level of each user.

The paper structure. The workpaper is structured in five sections, as follows:

- *Section 1. Introduction.* The section gives a comprehensive and quick overview of the importance of searchable encryption and why it should be treated with serious importance, especially that not so many implementations are not yet available.
- *Section 2. State-of-the-art*. The section will present a short state-of-the-art of the most important contributions of searchable encryption and how searchable encryption was born and which are the challenges raised by the concept itself when we want to implement it.
- *Section 3. The Multi-Level Searchable Symmetric Encryption Scheme.* The section covers will give the basic notions that are necessary for the reader to follow to understand the proposed scheme. The section will provide a complex analysis of searchable encryption.
- *Section 4. 4. The model and workflow.* The section contains the explanations of the full idea that we have designed and proposed. To explain how searchable encryption works in a real distributed environment, we have chosen a multi-disciplinary field, such as Earth Sciences: Meteorology and Weather Forecasting Stations from a country or region.
- *Section 5. Conclusions.* The section will provide shortly how the results were achieved and introduce a couple of new future research directions that we will want to focus on and the readers are welcomed to participate in our research as well and joint works are welcomed as well.

Our contributions. Our contributions are brought as follows: (1) a multi-level symmetric searchable encryption scheme; (2) proposing an access policy for the users of a weather station; (3) proposing a theoretical framework by presenting the main mathematical background of the algorithms used in a

searchable encryption scheme; (4) a practical framework on how the algorithms from (3) could be implemented in a real distributed system using as an example the case of a weather forecasting system for the weather stations from a country or region.

2. State-of-the-art

One question that we have asked ourselves above is the link between searchable encryption and access control. In [1], Nils Löken has the answer to the question and has shown how searchable encryption and access control can be combined.

Starting with the work of Song et al. [22], the authors provide a very interesting classification of different searchable encryption flavors.

Searchable Symmetric Encryption (SSE). Curtmola et al. [21] has brought significant contributions to SSE, providing for beginning access only to single users. In [15, 6, 5, 18] the non-equivalent security notion from [12, 7, 10, 9] are proved and demonstrated. To achieve multi-user SSE re-encryption some proposals were presented in [26] by using re-encryption and broadcast encryption [10]. In [20] we can see that SSE has been combined with oblivious RAM. In [13] we have an interesting solution for retrieval of private information. In [21] examined the blind storage to provide a limitation for what the servers or data owners are being able to learn from participating in the search.

Public key encryption with keyword search (PEKS). In [2] we can see searchable encryption using settings for multiple data creators and one single recipient [11], which is quite useful when software applications dedicated for different fields, such as earth sciences (e.g. meteorology), physics or military, are developed with respect for accessing classified documents or documents which are dedicated only for certain user groups.

Several schemes for multiple recipients and access control have been proposed, and we can observe how searchable encryption and access control were separated and relying on third parties with the capability to filter the search results [14] or designing searching queries [5, 16]. The literature also contains contributions for attribute-based encryption with keyword search [9, 5].

3. The Multi-level Searchable Symmetric Encryption Scheme

Below, there is a complex example of a searchable encryption scheme and how it is structured. Our chosen example is formed from 6 algorithms, from which we have four probabilistic algorithms (*KeyGen*, *BuildIndex*, *AddUser*, and *RevokeUser*) and two deterministic algorithms (*Query* and *Search)*.

The scheme is structured as follows:

1. $(SecretKey_{Owner}, Server_{key}, PublicParam) \leftarrow KeyGeneration(1^{\lambda}, Policy, Server)$. The algorithm is invoked by the owner. This is a *probabilistic algorithm* that is invoked by the owner of the data $Owner$ which will take the security parameter λ, policy $Policy$ and the identity of the server $Server$, and based on these parameters he will *output* the owner's secret key $SecretKey_{owner}$, a server key $Server_{key}$ and the public parameters $PublicParam$.

2. $Index \leftarrow BuildingIndex(D^{desc}, SecretKey_{Owner}, PublicParam)$. This represents a *probabilistic algorithm* that is invoked by the owner. It will take the description of the data set D^{desc} and the secret key of the owner ($SecretKey_{Owner}$) and it will *output* an index $Index$.

3. $UserSecret_{key} \leftarrow AddingUser(user, \lambda(user), SecretKey_{Owner}, PublicParam)$. This is a *probabilistic algorithm* that is invoked by the owner $Owner$ to enroll a new user within the e-learning platform system. The algorithm will take the new identity of the user and level of access of the user, and the owner's $Owner$ key, and it *output* the secret for the new user.

4. $QueryToken_{word,\lambda(u)} \leftarrow Quering(word, UserSecret_{key})$. This is a *deterministic algorithm* that is invoked by the user which has the proper clearance $\lambda(user)$ to generate a search query. The

algorithm will take as an input a keyword $word \in \Delta$ (where Δ represents a dictionary of keywords) and the user's secret key and it will output the query token $QueryToken_{word,\lambda(user)}$.

5. $Results_{\omega,\lambda(u)} \leftarrow Searching(QueryToken_{word,\lambda(uuser)}, Index, Server_{key})$. *Deterministic algorithm.* The algorithm is run by $Server$ to search the index for a specific set of data items that have in their structure a keyword associated, $word$. Based on the search query and the index, it will return the results of the search as $Results_{word,\lambda(user_i)}$, including a set of identifiers of the data items $d_j \in D_{\omega,\lambda(u)}$ which contains $word$ that has to satisfy the property $\lambda(d_j) \leq \lambda(user)$, where $\lambda(user_i)$ represents the access level of the user which is submitted to the search query, or a failure symbol φ.

6. $(SecretKey_{Owner}) \leftarrow RevokingUser(user, SecretKey_{Owner}, PublicParam)$. The *probabilistic algorithm* is run by the owner $Owner$ to revoke a specific user from the system. It will take the user's id, the secret keys of the data owner and server, and it will *output* the new keys for the owner and server.

Our proposed scheme is correct if for all $k \in \mathbb{N}$, for all $SecretKey_{owner}, Server_{key}$ outputted by $KeyGeneration(1^{\lambda}, Policy)$, for all D^{desc}, for all $Index$ that is outputted by $BuildingIndex(D^{desc}, SecretKey_{Owner})$, for all $word \in \Delta$, for all $user \in U$ for all $UserSecret_{key}$ outputted by $AddingUser(SecretKey_{Owner}, user, \lambda(user), PublicParam)$, $Search(Index, QueryToken_{word,\lambda(u)}) = D_{word,\lambda(user)}$.

4. The model and workflow

Our current contribution starts from the idea that the users of a weather station should use a multi-level access policy based on searchable symmetric encryption to access documents stored encrypted on an untrusted server.

The workflow of the scheme as depicted below in Figure 1, has the following flow and structure:

- *User*: The user will interact with the multi-level searchable symmetric scheme when he wants to search for a document ($Query()$) based on a keyword that is assigned to those documents. For example, if the user wants to receive the documents that are characterized (or contain) the $keyword$ "weekly", the server will invoke *Searching()* algorithm and it will return a set of data items $Results_{word,\lambda(user_i)}$ that contains "weekly".

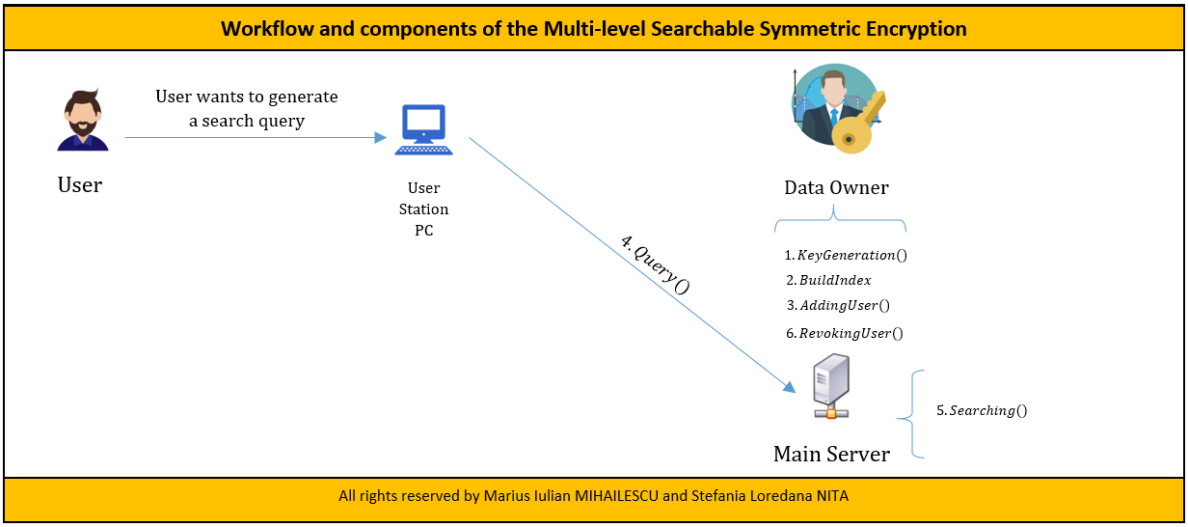

Figure 1. Workflow and components of the Multi-Level Searchable Symmetric Encryption

In Figure 2 we can see how a user will interact with the weather station PC. It is very important to understand that each station has two types of users: one weather chief station and $1 \dots n$ weather meteorologist technicians. This is very necessary for implementing the access policy.

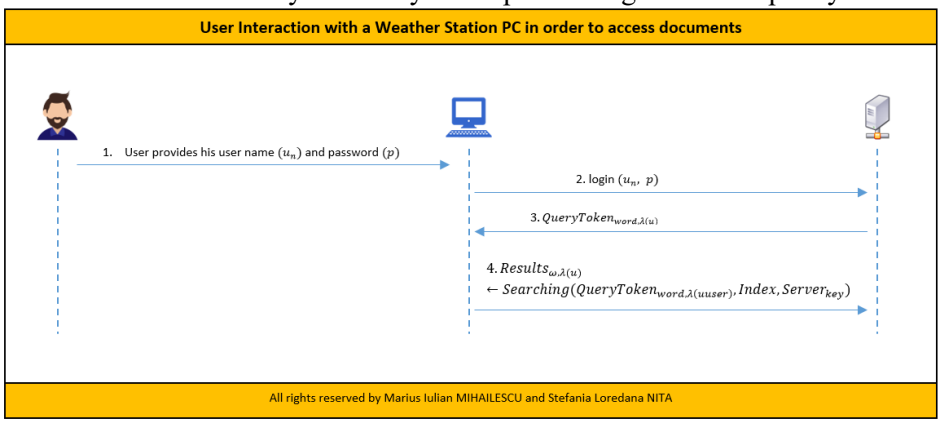

Figure 2. How a user interacts with a weather station PC to access documents

- *Data Owner*. The data owner is represented by the main data servers or regional servers. The data owner doesn't need to be looked at as a person. It can be a service as in our case which is responsible for the following operations:
  - Initializing the multi-level searchable symmetric encryption scheme ($KeyGeneration$).
  - Building the indexes for a data set of items ($BuildingIndexes$).
  - Adding, modifying, deleting, or updating users ($AddingUser$).
  - Revoking user ($RevokingUser$).
- *Main Server*. The main server contains a service that will take the query from the user and it will execute it accordingly on the server.

In Figure 3 we can observe how the weather stations are connected to a local server $\boldsymbol{L_{S_1}}$ which is represented by a county node $\boldsymbol{C_{N_{S_1}}}$. Multiple county nodes form a region $\boldsymbol{R_{C_1}}$.Multiple regional servers are connected to a big farm – data center (see Figure 4), which is the main point of controlling and collecting the weather data and observations from the weather stations from all over the country. Also we will add support for big data and obtaining recommendations [7].

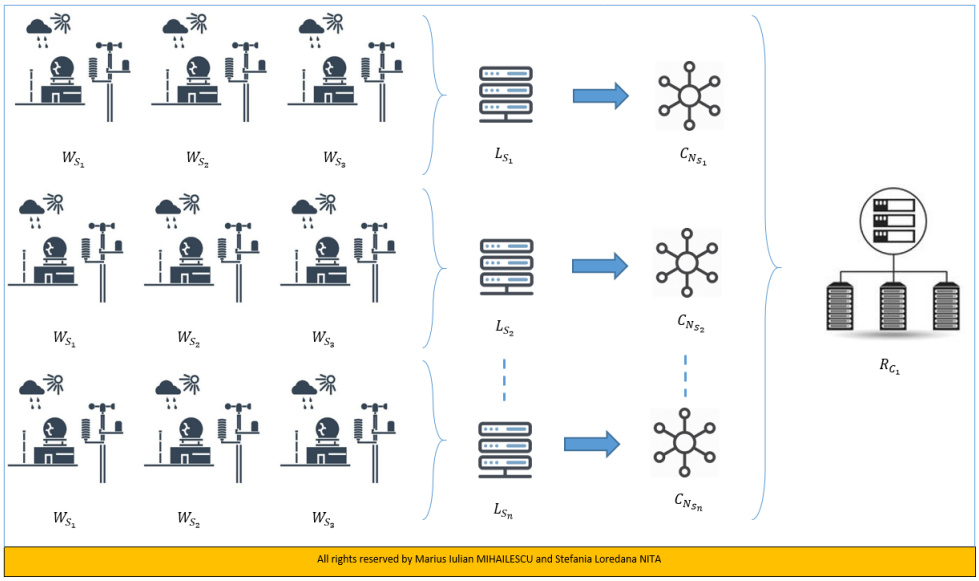

Figure 3. Interconnection of the weather station with their local server and county nodes

The local servers L_{S_n} are the first level where the data are collected. The county national servers $C_{N_{S_n}}$ take the data from the local servers and distribute them further to regional county servers R_{C_n} . From the regional country servers, the data are sent to the data-center.

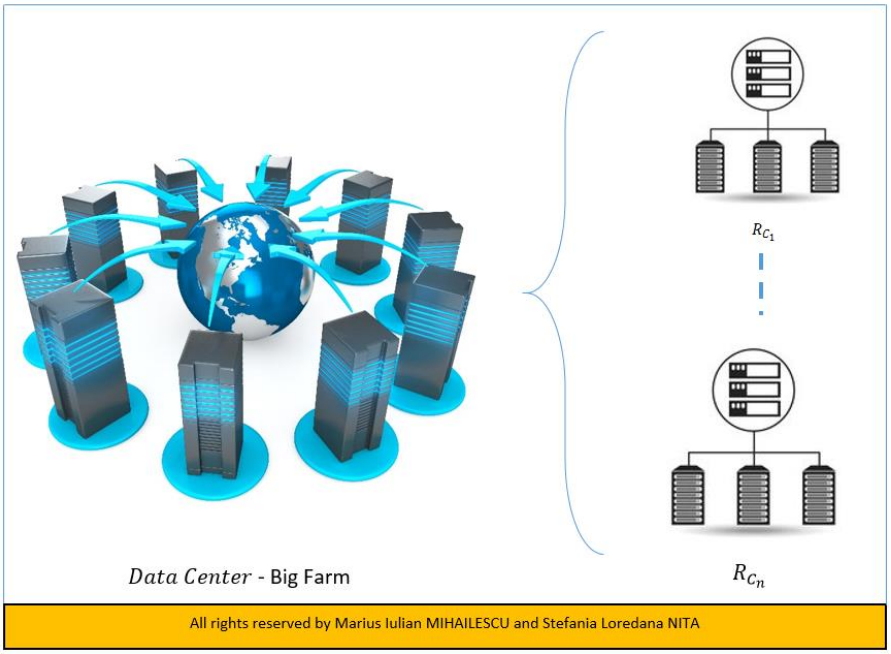

Figure 4. Regional nodes communicating with the main data center

5. Conclusions

We have managed to examine and demonstrate how a multi-level searchable symmetric encryption scheme can be applied in a real distributed environment. Each contribution has raised a set of difficulties and challenges, as follows:

- During examining the theoretical background of the most important searchable encryption schemes proposed in the last 2-3 years, we concluded that in practice, the hardware resources that need to be allocated are modest and there is no necessity to invest in very expensive network equipment. Theoretical, we have experienced interesting challenges in understanding and incorporating the algorithms in the infrastructure, in order to obtain maximum reliability, efficiency and time execution of the computations and the algorithm implementations.
- When we have proposed the access policy for the users of a weather station PC, the challenges raised were in determining exactly what documents each type of user needs to access. The policy was implemented and the results were positive without having any security issues.
- The proposed theoretical framework (see Section 3) represents the main mathematical algorithms that were implemented within the model presented in Section 4.
- The practical framework (see Section 4) has been implemented in a real distributed system using as an example the case of a weather forecasting system for the weather stations from a country or region. Our model proves to be a successful one especially due to the lack of implementations of searchable encryption primitive.

Future research directions. As future research directions, we have proposed several directions and we want to expand the model and to implement in many as possible distributed systems for different multi-disciplinary fields (physics, biology, etc). We want to compare the results and to provide better solutions

(security analysis, timely execution, developers effort, etc) for searchable encryption implementations. Also big data support will be added and providing analytics [7].

References

[1]. Nils Löken. 2017. Searchable Encryption with Access Control. In Proceedings of the 12th International Conference on Availability, Reliability, and Security (ARES '17). Association for Computing Machinery, New York, NY, USA, Article 24, 1–6. DOI: https://doi.org/10.1145/3098954.3098987.

[2]. Wang, S., Zhang, X., & Zhang, Y. 2016. Efficiently Multi-User Searchable Encryption Scheme with Attribute Revocation and Grant for Cloud Storage. PloS one, 11(11), e0167157. DOI: https://doi.org/10.1371/journal.pone.0167157

[1] James Alderman, Keith M. Martin, and Sarah Louise Renwick. 2017. Multi-level Access in Searchable Symmetric Encryption. IACR Cryptology ePrint Archive (2017), 211.

[2] Christoph Bosch, Pieter H. Hartel, Willem Jonker, and Andreas Peter. 2014. A Survey of Provably Secure Searchable Encryption. ACM Comput. Surv. 47, 2 (2014), 18:1--18:51.

[3] Hirano, Takato & Kawai, Yutaka & Koseki, Yoshihiro. (2018). Efficient Trapdoor Generation from Multiple Hashing in Searchable Symmetric Encryption: 14th International Conference, ISPEC 2018, Tokyo, Japan, September 25-27, 2018, Proceedings. 10.1007/978-3-319-99807-7_10.

[4] Guan, Wenhao & Wang, Yunling & Wang, Jianfeng & Fu, Xiaotong. (2018). Verifiable memory leakage-resilient dynamic searchable encryption. Journal of High-Speed Networks. 24. 201-217. 10.3233/JHS-180591.

[5] Alderman, James & Martin, Keith & Renwick, Sarah. (2017). Multi-level Access in Searchable Symmetric Encryption. 35-52. 10.1007/978-3-319-70278-0_3.

[6] Tarik Moataz and Abdullatif Shikfa. 2013. Boolean symmetric searchable encryption. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security (ASIA CCS '13). Association for Computing Machinery, New York, NY, USA, 265–276. DOI: https://doi.org/10.1145/2484313.2484347.

[7] N. S. Loredana, "On recommendation systems applied in big data," *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, Ploiesti, 2016, pp. 1-6.

[8] Johannes Blömer and Nils Löken. 2018. Cloud Architectures for Searchable Encryption. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES 2018). Association for Computing Machinery, New York, NY, USA, Article 25, 1–10. DOI:https://doi.org/10.1145/3230833.3230853

[9] David Cash, Paul Grubbs, Jason Perry, and Thomas Ristenpart. 2015. Leakage-Abuse Attacks Against Searchable Encryption. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). Association for Computing Machinery, New York, NY, USA, 668–679. DOI:https://doi.org/10.1145/2810103.2813700

[10] Ioannis Demertzis, Rajdeep Talapatra, and Charalampos Papamanthou. 2018. Efficient searchable encryption through compression. Proc. VLDB Endow. 11, 11 (July 2018), 1729–1741. DOI:https://doi.org/10.14778/3236187.3236218

[11] Tao Feng and Weiyou He. 2018. Research on Privacy Preserving of Searchable Encryption. In Proceedings of the 2018 2nd High Performance Computing and Cluster Technologies Conference (HPCCT 2018). Association for Computing Machinery, New York, NY, USA, 58–68. DOI:https://doi.org/10.1145/3234664.3234665

[12] Lei Xu, Xingliang Yuan, Ron Steinfeld, Cong Wang, and Chungen Xu. 2019. Multi-Writer Searchable Encryption: An LWE-based Realization and Implementation. In Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security (Asia CCS '19). Association for Computing Machinery, New York, NY, USA, 122–133. DOI:https://doi.org/10.1145/3321705.3329814

[13] Florian Hahn and Florian Kerschbaum. 2014. Searchable Encryption with Secure and Efficient

Updates. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14). Association for Computing Machinery, New York, NY, USA, 310–320. DOI:https://doi.org/10.1145/2660267.2660297.

[14] Cédric Van Rompay, Refik Molva, and Melek Önen. 2018. Secure and Scalable Multi-User Searchable Encryption. In Proceedings of the 6th International Workshop on Security in Cloud Computing (SCC '18). Association for Computing Machinery, New York, NY, USA, 15–25. DOI:https://doi.org/10.1145/3201595.3201597.

[15] Shi-Feng Sun, Xingliang Yuan, Joseph K. Liu, Ron Steinfeld, Amin Sakzad, Viet Vo, and Surya Nepal. 2018. Practical Backward-Secure Searchable Encryption from Symmetric Puncturable Encryption. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18). Association for Computing Machinery, New York, NY, USA, 763–780. DOI:https://doi.org/10.1145/3243734.3243782.

[16] Shangqi Lai, Sikhar Patranabis, Amin Sakzad, Joseph K. Liu, Debdeep Mukhopadhyay, Ron Steinfeld, Shi-Feng Sun, Dongxi Liu, and Cong Zuo. 2018. Result Pattern Hiding Searchable Encryption for Conjunctive Queries. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18). Association for Computing Machinery, New York, NY, USA, 745–762. DOI:https://doi.org/10.1145/3243734.3243753.

[17] Peter Baumann. 2011. Accelerating computationally intensive queries on massive earth science data: (system demonstration). In Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases (AD '11). Association for Computing Machinery, New York, NY, USA, 31–35. DOI:https://doi.org/10.1145/1966895.1966899.

[18] Paul Brown and Michael Stonebraker. 1995. BigSur: A System For the Management of Earth Science Data. In Proceedings of the 21th International Conference on Very Large Data Bases (VLDB '95). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 720–728.

[19] Alejandro Aguilar-Sierra. 2009. Visualization laboratory for Earth Sciences: a multidisciplinary visual learning environment. In SIGGRAPH '09: Posters (SIGGRAPH '09). Association for Computing Machinery, New York, NY, USA, Article 96, 1. DOI:https://doi.org/10.1145/1599301.1599397

[20] Dorian Gorgan. 2014. Spatial data processing on high performance computation architectures. In Proceedings of the 7th Euro American Conference on Telematics and Information Systems (EATIS '14). Association for Computing Machinery, New York, NY, USA, Article 1, 1–4. DOI:https://doi.org/10.1145/2590651.2590653

[21] Sriram Krishnan, Christopher Crosby, Viswanath Nandigam, Minh Phan, Charles Cowart, Chaitanya Baru, and Ramon Arrowsmith. 2011. OpenTopography: a services oriented architecture for community access to LIDAR topography. In Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications (COM.Geo '11). Association for Computing Machinery, New York, NY, USA, Article 7, 1–8. DOI:https://doi.org/10.1145/1999320.1999327.

[22] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, pages 79–88. ACM, 2006.

[23] D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In 2000 IEEE Symposium on Security and Privacy, pages 44–55. IEEE, 2000.