SEARCHABLE SYMMETRIC ENCRYPTION WITH FORWARD SEARCH PRIVACY

[1]Dr.S.Mohandoss, [2]Ms.E.Durga Nandini, [3]Deepan Raj E, [4]Sandeesh S, [5]Kishore J

Students
Cyber Forensics and Information Security
Dr. M.G.R. Educational and Research Institute
Chennai, India.

***Abstract*- Searchable symmetric encryption with forward search privacy using CRT is a technique that enables a client to encrypt a collection of documents and store them on an untrusted server while ensuring that the server does not learn anything about the documents or the search queries performed on them. The strategy works by employing the Chinese Remainder Theorem (CRT) to divide the search space into smaller spaces that can be searched separately. This prevents the server from learning anything about the documents just by viewing the search queries. The Chinese Remainder Theorem, a well-known mathematical theorem, is used to improve the efficiency and security of SSE schemes. This paper digs into the integration of CRT into SSE, demonstrating how it can improve both the speed of operations and the strength of privacy assurances. We investigate the complexities of merging CRT with SSE's main components, including trapdoor generation, index construction, and result retrieval.**

***Keywords*: Encrypted Web Application, Data Privacy, Format-Preserving Encryption.**

1. INTRODUCTION

1.1 ENCRYPTED WEB APPLICATION

An encrypted web application is a secure and privacy-conscious digital environment intended to protect sensitive data and communications transmitted over the internet. These applications use strong encryption methods and powerful cryptographic algorithms to safeguard data from

unauthorized access and cyber threats. These online apps safeguard the confidentiality and integrity of user interactions, transactions, and stored information by encrypting data in transit and at rest, increasing user trust and compliance with privacy standards. The emphasis on encryption in these applications is critical to ensuring a resilient and secure online experience, protecting against potential cyber threats and unwanted access, and so strengthening the overall integrity of digital interactions in today's web ecosystem.

1.2 DATA PRIVACY

Data privacy is a key notion that deals with the protection and appropriate management of personal information in the digital era. It encapsulates the ethical and legal considerations relating to the collecting, storage, and use of personal data, highlighting the necessity of maintaining confidentiality and regulating access. In an era of expanding connectedness and data-driven technology, worries about the misuse or illegal access to personal information have grown critical. As a result, data privacy initiatives and legislation seek to establish clear principles and standards for both companies and individuals, establishing a culture of trust and accountability. In summary, data privacy emphasizes the importance of taking a balanced approach that allows for innovation and convenience while emphasizing the rights and protection of individuals' personal data.

1.3 FORMAT-PRESERVING ENCRYPTION

Format-preserving encryption (FPE) is a sophisticated cryptographic approach that protects sensitive data while maintaining its original format. Unlike previous encryption methods, which may change the length or structure of the data they protect, FPE enables the encryption and decryption of information without altering its fundamental format. This novel approach is especially useful in situations when retaining the structure, such as credit card numbers or social security numbers, is critical for compatibility with existing systems or meeting compliance criteria. Format-preserving encryption strikes a delicate balance between privacy and functionality by seamlessly integrating security measures while preserving established data formats, making it a valuable solution for industries that require both secure data handling and adherence to specific data structures.

2. LITERATURE REVIEW

Sean MacAvaney [1] et al. As proposed in this research, deep pretrained transformer networks are effective at a variety of ranking tasks, including question answering and ad hoc document ranking. However, their processing costs make them prohibitively expensive in practice. Our suggested approach, PreTTR (Precomputing Transformer Term Representations), significantly reduces the query-time latency of deep transformer networks (up to a 42× speedup on online content ranking), making these networks more realistic for application in real-time ranking scenarios. Specifically, we recomputed a portion of the document term representations at indexing time (without a query) and combined them with the query representation at query time to calculate the overall ranking score. Due to the high size of the token representations, we also offer an effective method for reducing storage requirements: train a compression layer to match attention ratings. Our compression technology reduces storage requirements by up to 95% while maintaining ranking performance. We suggest an approach to increase the efficiency of transformer-based neural ranking algorithms. We take advantage of a key feature of ad-hoc ranking: an early indexing step can be used to pre-process documents in the collection, improving query-time speed.

Xuelong Dai [2] et al. This study proposes that traditional searchable encryption techniques that use the bag-of-words paradigm require large space to store the document set's index, where the dimension of the document vector is equal to the scale of the dictionary. The bag-of-words technique also overlooks semantic information between keywords and documents, potentially returning irrelevant search results to users. The Doc2Vec model, a neutral-network based natural language processing approach, extracts document features by using word and paragraph context. The characteristics contain latent semantic information and can be used to measure document similarity. In this research, we use the Doc2Vec model to implement a semantically aware multikeyword ranked search algorithm. The Doc2Vec model uses a distributed representation of words and documents with a low vector dimensionality while training on a dataset of a few hundred million words. The Doc2Vec model extracts the distributed representations of documents as a feature vector, which is then used as a search index. The features of the searched keywords are also retrieved as the query feature vector, and the secure inner product operation is used to perform privacy-preserving semantic search using the query feature vector and index. Our technique, which uses the Doc2Vec paradigm, can handle dynamic document updates. The experiment on a real-world dataset demonstrates that a fixed-length feature vector can increase the time and space efficiency of semantic-aware searches.

Jacob Devlin [3] et al. We propose in this system a novel language representation paradigm known as BERT, which stands for Bidirectional Encoder Representations from Transformers. Unlike prior language representation models, BERT is intended to pretrain deep bidirectional representations from unlabeled text by conditioning both left and right context in all layers. As a result, the pre-trained BERT model may be fine-tuned with just one more output layer to provide cutting-edge models for a variety of tasks, including question answering and language inference, without requiring significant task-specific architecture changes. BERT is conceptually basic yet experimentally powerful. It achieves cutting-edge results on eleven natural language processing tasks, including increasing the GLUE score to 80.5% (7.7% absolute improvement), MultiNLI accuracy to 86.7% (4.6% absolute improvement), Squad v1.1 question answering Test F1 to 93.2 (1.5-point absolute improvement), and Squad v2.0 Test F1 to 83.1 (5.1-point absolute improvement).

Ioannes Demertzis [4] et al. In this system, we propose the first linear-space searchable encryption scheme with constant locality and sub-logarithmic read efficiency, strictly improving the previously best known read efficiency bound from Θ (log N log N) to O (logγ N), where $\gamma = 2\ 3 + \delta$ for any fixed $\delta > 0$ and N is the number of keyword-document pairs. Our approach uses four alternative allocation techniques for storing keyword lists, depending on the size of the list being examined at the time. Our construction includes (i) new probability bounds for the offline two-choice allocation problem, and (ii) a new I/O-efficient oblivious RAM with O˜(n 1/3) bandwidth overhead and zero failure probability. Both can be of independent interest. Searchable Encryption (SE), first proposed by Song and later codified by, allows a data owner to outsource a private dataset D to a server, which can then answer keyword queries without learning too much about the underlying dataset or the posed queries. SE techniques offer a feasible alternative to expensive primitives like oblivious RAM and completely homomorphic encryption, but they come at the cost of formally stated leakage. In conventional SE schemes, the data owner creates a private index that is sent to the server.

Kun Xie [5] et al. As proposed in this system, traffic anomaly detection is crucial for advanced Internet management. Existing detection methods often convert high-dimensional data to a lengthy vector, which reduces detection accuracy by removing spatial information. Furthermore, they are typically intended to separate normal and anomalous data over time, which not only increases storage and computing costs but also precludes rapid anomaly identification. Online and accurate traffic anomaly detection is vital, but challenging to implement. To address the difficulty, this research explicitly models each time slot's monitoring data as a 2D matrix and uses Bilateral Principal Component Analysis (B-PCA) to find anomalies in the new time slot. We propose several novel techniques in OnlineBPCA to support quick and accurate anomaly detection in real time, including a novel B-PCA-based anomaly detection principle that jointly considers the variation of both row and column principal directions for more accurate anomaly detection, an approximate algorithm to avoid using iteration procedure to calculate the principal directions in a close-form, and a

sequential anomaly algorithm to quickly update principal directions. To the best of our knowledge, this is the first study to use two-dimensional PCA for anomaly identification.

3. RELATED WORK

Searchable symmetric encryption (SSE) is commonly used in encrypted databases for queries. Although SSE is strong and feature-rich, it is always plagued by data leaks. According to recent assaults, forward privacy, which prevents leakage from update processes, has become a baseline need for any newly constructed SSE systems. However, further search procedures may still provide a large amount of information. To improve security, we broaden the definition of forward privacy and introduce the concept of "forward search privacy". Intuitively, it involves search operations over newly uploaded documents that do not reveal information about previous inquiries. The improved security concept introduces new obstacles to the design of SSE. We overcome the issues by developing the hidden pointer technique (HPT) and introducing Khons, a new SSE scheme that meets our security notion (along with the original forward privacy notion) while also being efficient.

4. METHODODLOGY

The suggested approach combines Searchable Symmetric Encryption (SSE) and the Chinese Remainder Theorem (CRT) to improve privacy and efficiency in data search operations. This novel cryptographic technique enables secure searching of encrypted material while preserving Forward Search Privacy. The data in this system is partitioned into numerous shares using the CRT, each of which is encrypted with a unique modulus. This division permits concurrent processing of decryption and search requests, which considerably improves operational efficiency. The server can thus perform search operations on the encrypted data without having to decrypt it completely. The use of CRT ensures that even if an adversary obtains partial information via search queries, the overall security of the system is maintained.

A. Load Data

The data loading stage involves the system retrieving encrypted data from a storage location. This data is encrypted utilizing Searchable Symmetric Encryption (SSE) techniques, allowing for secure searching while protecting anonymity. The Chinese Remainder Theorem is used to partition data into modular components during encryption, allowing it to be searched more efficiently.

B. Data Preprocessing

After loading the encrypted data, the system uses data pre-processing procedures that are compatible with the encrypted format. For example, if the data comprises textual information, tokenization and encoding are performed while the data remains encrypted. CRT-assisted modular representation ensures that any pre-processing procedures are applied to individual components while protecting the original data's privacy and security.

C. Feature Selection

In some circumstances, feature selection or dimensionality reduction may be required to increase the efficiency of the machine learning process. The Chinese Remainder Theorem can be used to generate a modular representation of the desired features, allowing feature-related calculations to be done on the modular components.

D. Training and Testing

During the training phase, machine learning models are created using encrypted data and modular components. Training algorithms and calculations are carried out in a modular way, taking advantage of the Chinese Remainder Theorem's efficient computation capabilities. Similarly, when testing, queries or input data are converted into modular representations, ensuring that both the training and testing operations adhere to the modular structure enabled by the CRT.

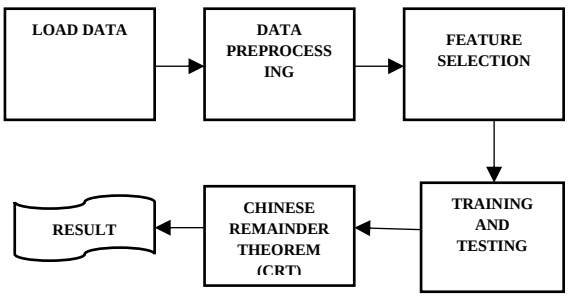

Figure 1.Block diagram

E. Evaluation and Performance

Following training and testing, the model is evaluated and its performance is assessed. These evaluations may include metrics such as accuracy, precision, recall, and F1-score. The Chinese Remainder Theorem ensures that encrypted evaluations and calculations are done on modular components, preserving the secrecy of both the model's internal workings and the evaluation metrics.

5. ALGORITHM DETAILS

The Chinese remainder theorem. $x \equiv a1 \bmod m1$,
$x \equiv a2 \bmod m2$, ..., $x \equiv ar \bmod mr$, has a solution, and this solution is uniquely determined modulo m1m2 ⋯mr.
Searchable Symmetric Encryption with CRT Algorithm
Key Generation
Client:
SK, PK = KeyGen () # Secret Key and Public Key generation
Document Encryption
Client:
C = Encrypt (PK, D) # Encrypt document D using Public Key PK
Trapdoor Generation
Client:
TW = GenTrapdoor (PK, KW) # Generate trapdoor TW for search keyword KW
Index Construction
Client:
Index = Construct Index(D) # Construct an index for document D
Chinese Remainder Theorem Functions
Function GenerateCRTParameters ():
Generate parameters for CRT, e.g., primes, moduli, etc. return CRT_Parameters

6. RESULT ANALYSIS

The incorporation of the Chinese Remainder Theorem (CRT) into Searchable Symmetric Encryption (SSE) constitutes a big step forward in result analysis. By using CRT to partition the search space, the technique allows for independent searches in smaller spaces, improving both SSE's efficiency and privacy guarantees. The optimization obtained through CRT integration is most visible in the speed of operations, where independent search spaces contribute to speedier retrieval of results. This increase in efficiency is critical for real-time applications and massive datasets. Furthermore, the usage of CRT reinforces SSE's strong privacy assurances by prohibiting the server from deducing information about encrypted documents based on observed search queries. The analysis emphasizes the usefulness of combining CRT and SSE, demonstrating improvements in both operational speed and user privacy, both of which are important concerns in the creation of safe and efficient searchable encryption systems.

algorithm	**accuracy**
hidden pointer technique	80
Chinese Remainder Theorem	85

Table 1. Comparison table

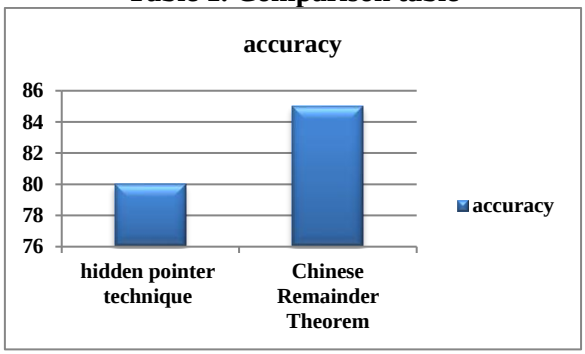

Figure 2.Comparison graph

This table compares two cryptographic systems based on their accuracy in a specific context. The "Hidden Pointer Technique" achieves an accuracy rate of 80%, proving its efficacy in securely handling pointers or references to sensitive information. In contrast, the "Chinese Remainder Theorem" achieves 85% accuracy in the case of searchable symmetric encryption with forward search privacy. This method use the Chinese Remainder Theorem to divide the search space into smaller, independent subspaces, improving both the speed of operations and the strength of privacy guarantees. The table compares the correctness of these cryptographic algorithms, with the Chinese Remainder Theorem outperforming the others in the defined case.

7. CONCLUSION

To summarize, searchable symmetric encryption with forward search privacy via CRT is a promising solution for enabling secure search over encrypted material. It is efficient, secure, versatile, and scalable, making it suitable for a wide range of applications. Incorporating the Chinese Remainder Theorem (CRT) into Searchable Symmetric Encryption (SSE) systems is a big step forward in the pursuit of secure and efficient cloud-based data storage and retrieval. The advantages of this integration are obvious and persuasive. It not only improves computing efficiency and scalability, but it also increases the privacy assurances that are key to SSE. The enhanced modular arithmetic operations provided by CRT minimize compute overhead and communication requirements, making the system more suitable for real-world applications.

8. FUTURE WORK

Increase the security of the scheme. This can be accomplished by devising innovative approaches that prevent the server from learning anything about the documents or search queries. Improve the scheme's efficiency. This can be accomplished by creating innovative strategies for reducing the amount of computing necessary to complete the search operation. Make the scheme more broadly accepted. This can be accomplished by creating additional tools and libraries that make it easier to use. Make the scheme available across all programming languages. This can be accomplished by creating additional implementations in other languages. Apply the scheme to new applications. This can be accomplished by experimenting with fresh and unique ways to apply the strategy.

REFERENCES:

[1] X. Yang, Q. Fu, & Z. Niu, "A Secure Keyword Search and Data Sharing Mechanism for Cloud Computing," Appl. Res. Computers, volume. 31, no. 10, pp. 3017-3019, 2020.
[2] D. Ayala, O. Wolfson, B. Xu, B. Dasgupta, and J. Lin, "Efficient Document Re-Ranking for Transformers by Precomputing Term Representations" in Proc. 20th Int. GIS Congress, 2020, pp. 43–51.
[3] Z. Zhao and Y. Zhang, "An Efficient and Dynamic Semantic-Aware Multikeyword Ranked Search Scheme Over Encrypted Cloud Data,"" J. Adv. Trans., Vol. 2020, pages 1–12, February 2020.
[4] W. Zhang, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" M.S. thesis, Zhejiang University, Hangzhou, China, 2021.
[5] T.-Y. Ma & S. "Searchable Encryption with Optimal Locality: Achieving Sublogarithmic Read Efficiency," Energy, vol. 244, pp. 360, January 2022; doi: 10.1016/j.energy.2022.123217.
[6] Z. Wang, W.Xue, and X. Hei, "On-line Anomaly Detection with High Accuracy," in Proceedings. ICICAS, December 2020, pages 392–396.
[7] F. Liu, F.J. Hao, J.Q. Hao, Y.L. Zhou, and G. M. Xin, "Forward Private Searchable Symmetric Encryption with Optimal I/O Efficiency," Computer. Application, volume. 39, pages. 65–69, January 2021.
[8] F. She, J. D. Qiu, and Yu. A. Tang, "Dynamic Keyword Search with Hierarchical Attributes in Cloud Computing," Computer. Appli. Research, volume. 36, no. 3, pages 851-854, 2020, doi: 10.19734/j.issn.1001-3695.2017.09.0922.
[9] C. Badii, P. Nesi, and myself. Paoli, "Enabling Generic, Verifiable, and Secure Data Search in Cloud Services," IEEE Access, Vol. 6, pp. 44059-44071, 2021.
[10] S. S. Ghosal, A. Bani, A. Amrouss, and myself. El Hallaoui, "Tight Tradeos in Searchable Symmetric Encryption,"" in Proc. ICDMW, November 2020, pages 581-586.
[11] X. Xiao, Z. Jin, Y. Hui, Y. Xu, & W. Shao, "Enabling Verifiable and Dynamic Ranked Search over Outsourced Data," Remote Sens. 13, p. 3338, 2021; doi: 10.3390/rs13163338.
[12] G. Ali, T. Ali, M. Irfan, U. Draz, M. Sohail, A. Glowacz, M. Sulowicz, R. Mielnik, Z. B. Faheem and C. "Publicly verifiable searchable symmetric encryption based on efficient cryptographic components," Electronics, vol. 9, no. 10, p. 1696, 2020.
[13] R. W. Liu, M. Liang, J. Nie, W.Y.B. Lim, Y. Zhang, and M. Guizani, "Verifiable Ranked Search over Encrypted Data with Forward and Backward Privacy,"" IEEE Trans. Netw. Sci. Engineering, early access, January 7, 2022; doi: 10.1109/TNSE.2022.3140529.

[14] L. B. Li & Y. Li, "FASE: A Fast and Accurate Privacy-Preserving Multi-keyword Top-k Retrieval Scheme over Encrypted Cloud Data," J. Tongji Univ. 49, no. 9, pp. 1301-1306, 2021.
[15] Z. Zhao and Y. Zhang, "Enhanced Semantic-Aware Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data," J. Trans., Vol. 2020, pages 1–12, February 2020.
[16] E. S. Fokker, T. Koch, and L. M. Van and E. "Prime Inner Product Encoding for Effective Wildcard-based Multi-Keyword Fuzzy Search," Transp. Res. Record, vol. 2676, no. 1, pp. 637–654, 2022; doi: 10.1177/03611981211036373.
[17] X. Y. Fang, "Searchable Encryption Over Feature-Rich Data," M.S. Thesis, GUET 2019.
[18] S. Yang, W. Ma, X. Pi, & S. Qian, "Privacy-preserving indexing and query processing for secure dynamic cloud storage," Transp. Res. C, Emerg. Technology, volume. 107, pages 248–265, October 2019.
[19] G. Ali, T. Ali, M. Irfan, U. Draz, M. Sohail, A. Glowacz, M. Sulowicz, R. Mielnik, Z. B. Faheem and C. "IDCrypt: A Multi-User Searchable Symmetric Encryption Scheme for Cloud Applications," Electronics, vol. 9, no. 10, p. 1696, 2020.
[20] Y. Han, W. Zheng, & J. "Dual-Axial Motion Control of a Magnetic Levitation System Using Hall-Effect Sensors" Transp. Syst. Eng. Inf., volume. 17, no. 5, pp. 60-67, 2020.