# TPT1201
# Research Methodology in Computer Science

## ASSIGNMENT 2

## Cloud-Based Health Account Management

Prepared by

Group ID K8

| Members | Student ID | Contact Number | Tutorial Section |
|---|---|---|---|
| Ayat Abdulaziz Gaber Al-Khulaqi | 1191202335 | +60 18-293 6294 | TT9L |
| Roha Ali | 1191102484 | +60 11-2615 5759 | TT5L |
| Nur Irdina Binti Hassan | 1191202351 | +60 10-860 4152 | TT4L |
| Mohammad Harez Bin Hafez | 1191202413 | +60 12-389 0207 | TT4L |

**Abstract**

There is an increasing need for healthcare environments where a patient can receive collaborative treatment from several healthcare professionals as smart healthcare services are developed. Due to unauthorized users' illegal access to private health information belonging to a data owner, this system, nevertheless, may give rise to major privacy problems. The E-health system is hoped to be enhanced in order to promote its use in Malaysia's healthcare sector, including those outlined by the National Institute of Standards and Technology (NIST) and the International Organisation for Standardisation (ISO) Cybersecurity Frameworks. The goal is to increase the security and privacy of systems that use cloud storage for personal health records (CB-PHR). Furthermore, the goal is to provide a system encryption technique to maintain data privacy in e-health management. It is also expected that cloud-based health account management will work securely in the cloud using improved MSOPE with blockchain. This contribution will make an impact to the research community to continually study cloud security.

# 1 Introduction

In this introduction, the main focus is on the problems underlying when coming up with a solution to counter these challenges. The main problem found when doing research on this topic is the need to investigate a cloud computing design for a proof-of-concept for healthcare services, implement and evaluate the proof-of-concept. Besides, health-care organisations do not have a methodical tactic to allow them to assess, manage and improve their information and physical cybersecurity posture when using the cloud. Ergo, in every problem, an objective is needed to counter it. The objective is to improve the privacy and security of Cloud-based Personal Health Record systems (CB-PHR). Therewith, the aim is also to propose a system encryption method to sustain data privacy in e-health management. A propound system for the method called Multi source Order Preserving Symmetric Encryption (MOPSE) is proposed in order to keep up with the problems that many practitioners go through. The only downside is that this method, however, does not protect the EHRs order information against cloud servers, allowing adversaries to extract EHR information from the leaked order.

# 2 Motivation of the Research

This section identifies the problem that is worth tackling where this paper strives to fix a concurrent problem of poor usability in the cloud based for e-health management. The main problem discussed in this research proposal is to investigate a cloud computing design for a proof-of-concept for healthcare services. Health-care organisations also do not have a methodical tactic to allow them to assess, manage and improve their information and physical cybersecurity posture when using the cloud. Ergo, it is hoped that the system for E-health to be improved to promote its use in the healthcare industry in Malaysia, standards such as National Institute of Standards and Technology (NIST) Cybersecurity Frameworks and the International Organisation for Standardisation (ISO). The aim is to further resolve the privacy concerns about data indexes and queries, searchable encryption schemes.

# 3 Research Objectives

- To improve the privacy and security of Cloud-based Personal Health Record systems (CB-PHR).
- To propose a system encryption method to sustain data privacy in e-health management.

# 4 Literature Review

| Authors | Data Loss | Data Security Issues | Data Leaks | Client or Provider Privacy | Maintainability |
|---|---|---|---|---|---|
| Akinsanya, Papadaki, and Sun (2020) | X | X | X |  | X |
| Brown and Randall (2020) | X | X | X | X |  |
| Kim, Edemacu, and Jang (2019) | X |  |  |  |  |
| Seol, Kim, Lee, Seo, and Baik (2018) |  | X |  |  | X |
| Xia et al. (2017) |  | X |  |  |  |
| Edemacu, Park, Jang, and Kim (2019) | X | X |  |  | X |
| Zhang et al. (2018) |  |  | X |  | X |
| Yeh, Chiang, Tsai, and Huang (2018) |  |  |  |  | X |
| Yao, Lin, Liu, and Zhang (2018) |  |  | X | X |  |
| Kong, Zhou, Xia, Pan, and Zhu (2019) |  | X | X |  |  |
| Idoga, Toycan, Nadiri, and Çelebi (2018) |  | X |  |  | X |
| Bi, Liu, and Kato (2022) |  | X |  | X |  |
| Zheng et al. (2022) |  | X |  |  |  |
| Li et al. (2022) |  | X |  | X | X |
| Chinnasamy and Deepalakshmi (2018) |  | X |  | X |  |

In this section, discussions about the literature review where data security issues are the most common challenges encountered by the 11 out of 15 authors. As stated by [Chinnasamy and Deepalakshmi (2018)] medical related data need to store examples such as patient data, x-ray, therapy procedure, medical prescription, and many more information concentrated data. Therefore, cloud data should be secured. Many countries' researchers have stated issues with healthcare maintainability, [Idoga et al. (2018)] said that inadequate health institutions in some countries make a cloud-based system almost impossible to maintain. Other similar issues are data leaks, client or provider privacy concerns, and data loss issues were also encountered by authors [Brown and Randall (2020)]. Authors [Edemacu et al. (2019)] encountered data loss and data security issues and authors [Yao et al. (2018)] faced data leaks and client or provider privacy issues. Many methods were proposed by the authors to solve the challenges and problems. [Yao et al. (2018)] authors used a Multi-source Order-Preserving Symmetric Encryption (MOPSE), this method was designed so the data in the cloud could be merged with the encrypted data from multiple sources without knowing the content. This method tackles the security and privacy concerns which is the most common challenge. The outcome is positive where MOPSE enables an outstanding privacy-preserving process. Other authors have used various methods such as [Bi et al. (2022)] are using deep learning-based privacy preservation to overcome the security issues of privacy information. However, this method not as efficient as MOPSE and does not have any clear outcome. Privacy preserving forward algorithms are one of the methods use by the author [Zheng et al. (2022)], with the main objective to construct a monitoring scheme over the cloud. In addition, the authors conducted simulations for the proposed system and the result shows efficiencies of the system. However, the MOPSE system shows the most reliable results. Some authors also used system encryption and authors Seol et al. (2018)] proposed a system that carry out an attribute-based access control which utilises expandable access control markup language. Since the main objective is to concentrate on security and performing partial encryption, this model works effectively. Authors [Edemacu et al. (2019)] also use an attribute-based encryption that contains the required building block in restraining against privacy violation and data protection in the cloud. The authors' objective is to sustain data privacy in the cloud storage. The EHR model works more effectively than attribute-based encryption since the EHR from authors [Seol et al. (2018) ] only redirect the mandatory data to the suppliant.

# 5 Research Methodology

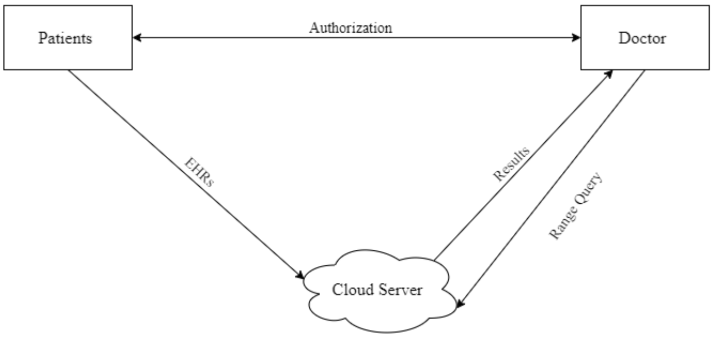

Fig 1.1 Cloud-based eHealth system model

In Fig (1), the cloud server stores the EHRs gathered from multiple patients and offers a range query service to an authorised doctor. Patients are EHR owners who outsource their EHRs to a cloud server and grant some doctors access to their EHRs.The doctor is an authorised data user who submits an EHR range query request and receives the corresponding results. The Multi-Source Order-Preserving Symmetric Encryption (MSOPE) scheme by [Liang et al. (2020)], includes a privacy-preserving range query scheme that allows the doctor to query the encrypted EHRs. The MSOPE scheme has four polynomial-time algorithms: Secret State Generation, Encryption, Decryption and Range Query. However, the proposed MSOPE scheme by [Liang et al. (2020)], does not protect the EHRs order information against cloud servers, allowing adversaries to extract EHR information from the leaked order.

To overcome this problem, a blockchain was proposed from authors [Liu, Wang, Jin, Li, and Li (2019)]. First, blockchain can do autonomy. The blockchain discusses an accordant protocol which validates all nodes in the system to transfer data securely. So, humans cannot intervene. All data allocated in the blockchain is encrypted by the asymmetric encryption scheme, which could put a stop to unauthorised nodes from gaining access to the medical information. When uncertain historical data of a user, the proxy re-encryption technology is used. It allows the stored information in the blockchain to be transmitted in the ciphertext state. The security of the proposed scheme is additionally advanced. One of the main features of the blockchain is Decentralization. Transaction records are done by multiple nodes that are assigned in separate places, and each node records and keeps a complete account. All nodes can separate the transaction and jointly testify for it. The decentralisation features are very suitable for the protection and sharing of medical data.

## 6 Expected Outcomes

The method proposed to further improve and secure high-risk information is by using blockchain in the MSOPE.MSOPE has some security issues that need improvement, and blockchain can securely share data. The expected outcome stated will contribute to the research community and allow further research, experiment with cloud security and use this new design of MSOPE that utilizes blockchain.To conclude, this system is new to the field and more research is needed to discover better tactics. With every research continuity, there will be challenges to be tackled. The proposed blockchain will protect data, making it useful in securing the cloud and solve the security corncerns.Blockchain is expensive to develop, however it is worth it to protect more valuable data.

# Part 2 : Statistical Data Analysis

Two datasets were given : "ppv.csv" which contains 100 vaccination centres' locations and "people.csv" which consists of 10000 vaccines' locations. Based on both datasets, we were tasked with matching each person from the "people.csv" file to their nearest vaccine centre. This matching was done by using the address coordinates(latitude and longitude). In table 2.1 it illustrates our computer specification, ASUS Vivobook has the highest performance and HP Notebook has the lowest performance.

| Computer | RAM | Hard Disk Type | CPU |
|---|---|---|---|
| ASUS Vivobook | 16 GB | KINGSTON SNVS1000GB | AMD Ryzen 9 5900HX withRadeon Graphics 3.30 GHz |
| MSI GF65 Thin 9SEXR | 16 GB | KIOXIA KGB40ZNV512G | Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz |
| ASUS G14 | 16GB | INTEL SSDPEKNW512GB | AMD Ryzen 7 4800HS 2.90 GHz |
| HP Notebook - 14q-cs0001tx | 8GB | 1 TB 5400 rpm SATA | Intel ® Core(™) i5- 7200U 2.50Hz |

Table 2.1 Computer specification

For the program we have created four methods to calculate the distances: Great Circle, Haversine, Haversine Vector and Euclidean. Three different libraries were used: scipy, haversine and geopy. In the program we have three for loops:the first loop runs the program 50 times, which will calculate the execution time for each loop then save the inside time array. To find execution time a time library was import, and to calculate the execution time we subtracted the end time from start time. In the second for loop, it will loop based on the number of indexes in the "people.csv" file. Then we created a distance empty array and coords_1 tuple which consist of the latitude and longitude of each person depending on the loop number.

For the third for loop, it will loop based on the number of indexes in the "ppv.csv" file. After we created coords_2 tuple which contains the latitude and longitude of each vaccination centre. Then it calls the distance method and passes coords_1 and coords_2 tuples lastly saves it at “distance list”. The third for loop will run 100 times before going to the next person. After the third for loop finishes running, the min function it’s used to find the minimum distance in the “distance list”. By passing the minimum distance through “distance.index” we can find the index of the vaccination centre. Lastly the minimum index is saved in the “results_list array”. After the first for loop finishes running, using the times array we created a dataframe then passed the data to a csv file using the “to_csv”.

As seen in Table 2.2, We ran the program using four laptops on jupyter lab using the Great Circle Method. Since Asus Vivo laptop has the highest performance as seen in Table 2.1, it also has the fastest running time around 15 seconds comparing to the other laptops. HP laptop had the lowest performance due to it had the slowest running time around 28 seconds.

| No. | AsusVivo | AsusG14 | MSI | HP |
|---|---|---|---|---|
| 1 | 15.1516415 | 19.4719234999999 | 25.0857417 | 29.397804809 |
| 2 | 14.9314581 | 19.4345244 | 25.0074996 | 29.181429164 |
| ... | ... | ... | ... | ... |
| 49 | 15.4913099 | 19.3207436999999 | 24.8445856999999 | 28.8775885719999 |
| 50 | 15.0620423 | 19.2319279999999 | 25.0975813 | 28.8684226919999 |

Table 2.2 Execution Time Comparison using 4 Computers Using Great Circle

As seen in the (Fig 2.1), Asus Vivo laptop has the fastest time for running using the Great Circle method and HP laptop has the slowest running time. However, Asus G14 laptop has the second fastest running time and then MSI laptop.

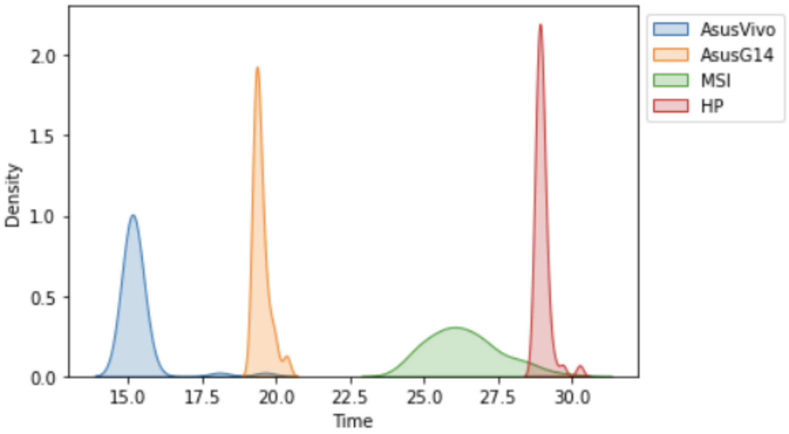

Fig 2.1 Shaded Density Plot for Execution Time Comparison using 4 Computers Using Great Circle

We ran the four different methods using Asus Vivo laptop on jupyter lab. In Table 2.3, shows the first and last two running time values and as seen Haversine has the fastest running time around 9 seconds.

| No. | Great Circle | Haversine | Haversine Vector | Euclidean |
|---|---|---|---|---|
| 1 | 15.1516415 | 9.1994434 | 39.7250152 | 21.1622154 |
| 2 | 14.9314581 | 9.0981446 | 37.2212594 | 18.1441643 |
| ... | ... | ... | ... | ... |
| 49 | 15.4913099 | 9.1751192 | 37.5652991 | 18.5466206 |
| 50 | 15.0620423 | 9.3623552 | 37.6719922 | 18.5205681 |

Table 2.3 Execution Time Comparison for 4 Different Methods

As illustrated in (Fig 2.2), Haversine has the shortest time to run and Haversine Vector takes the longest to run. However, the second best to run is Great Circle then Euclidean.

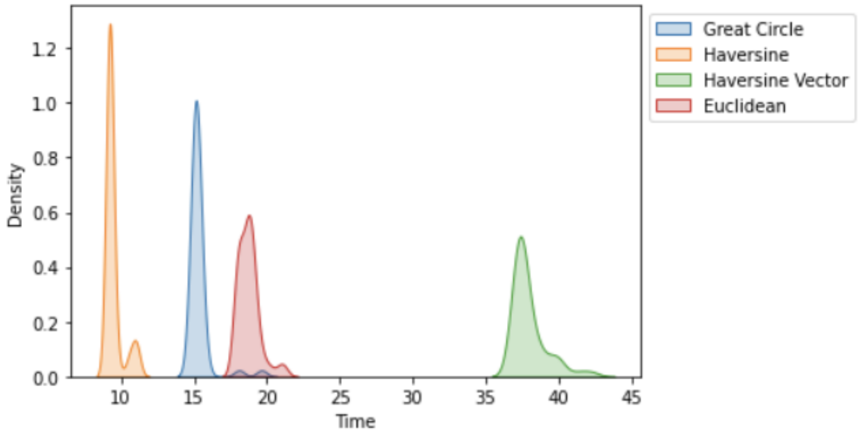

Fig 2.2 Shaded Density Plot for Execution Time Comparison for 4 Different Methods

# References

Akinsanya, O. O., Papadaki, M., & Sun, L. (2020). Towards a maturity model for health-care cloud security (m2hcs). *Inf. Comput. Secur.*, *28*, 321-345.

Bi, H., Liu, J., & Kato, N. (2022). Deep learning-based privacy preservation and data analytics for iot enabled healthcare. *IEEE Transactions on Industrial Informatics*, *18*(7), 4798-4807. doi: 10.1109/TII.2021.3117285

Brown, A. P., & Randall, S. M. (2020, Sep 23). Secure record linkage of large health data sets: Evaluation of a hybrid cloud model. *JMIR Med Inform*, *8*(9), e18920. Retrieved from `http://medinform.jmir.org/2020/9/e18920/` doi: 10.2196/18920

Chinnasamy, P., & Deepalakshmi, P. (2018). Design of secure storage for health-care cloud using hybrid cryptography. In *2018 second international conference on inventive communication and computational technologies (icicct)* (p. 1717-1720). doi: 10.1109/ICICCT.2018.8473107

Edemacu, K., Park, H. K., Jang, B., & Kim, J. W. (2019). Privacy provision in collaborative ehealth with attribute-based encryption: Survey, challenges and future directions. *IEEE Access*, *7*, 89614-89636. doi: 10.1109/ACCESS.2019.2925390

Idoga, P. E., Toycan, M., Nadiri, H., & Çelebi, E. (2018). Factors affecting the successful adoption of e-health cloud based health system from healthcare consumers' perspective. *IEEE Access*, *6*, 71216-71228. doi: 10.1109/ACCESS.2018.2881489

Kim, J. W., Edemacu, K., & Jang, B. (2019). Mppds: Multilevel privacy-preserving data sharing in a collaborative ehealth system. *IEEE Access*, *7*, 109910-109923. doi: 10.1109/ACCESS.2019.2933542

Kong, F., Zhou, Y., Xia, B., Pan, L., & Zhu, L. (2019). A security reputation model for iot health data using s-alexnet and dynamic game theory in cloud computing environment. *IEEE Access*, *7*, 161822-161830. doi: 10.1109/ACCESS.2019.2950731

Li, X., Liu, S., Lu, R., Khan, M. K., Gu, K., & Zhang, X. (2022). An efficient privacy-preserving public auditing protocol for cloud-based medical storage system. *IEEE Journal of Biomedical and Health Informatics*, *26*(5), 2020-2031. doi: 10.1109/JBHI.2022.3140831

Liang, J., Qin, Z., Xiao, S., Zhang, J., Yin, H., & Li, K. (2020). Privacy-preserving range query over multi-source electronic health records in public clouds. *Journal of Parallel and Distributed Computing*, *135*, 127-139. Retrieved from `https://www.sciencedirect.com/science/article/pii/S074373151930053X` doi: https://doi.org/10.1016/j.jpdc.2019.08.011

Liu, X., Wang, Z., Jin, C., Li, F., & Li, G. (2019). A blockchain-based medical data sharing and protection scheme. *IEEE Access*, *7*, 118943-118953. doi: 10.1109/ACCESS.2019.2937685

Seol, K., Kim, Y.-G., Lee, E., Seo, Y.-D., & Baik, D.-K. (2018). Privacy-preserving attribute-based access control model for xml-based electronic health record system. *IEEE Access*, *6*, 9114-9128. doi: 10.1109/ACCESS.2018.2800288

Xia, Q., Sifah, E. B., Asamoah, K. O., Gao, J., Du, X., & Guizani, M. (2017). Medshare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE Access*, *5*, 14757-14767. doi: 10.1109/ACCESS.2017.2730843

Yao, X., Lin, Y., Liu, Q., & Zhang, J. (2018). Privacy-preserving search over encrypted personal health record in multi-source cloud. *IEEE Access*, *6*, 3809-3823. doi: 10.1109/ACCESS.2018.2793304

Yeh, L.-Y., Chiang, P.-Y., Tsai, Y.-L., & Huang, J.-L. (2018). Cloud-based fine-grained health information access control framework for lightweightiot devices with dynamic auditing andattribute revocation. *IEEE Transactions on Cloud Computing*, *6*(2), 532-544. doi: 10.1109/TCC.2015.2485199

Zhang, Y., Xu, C., Li, H., Yang, K., Zhou, J., & Lin, X. (2018). Healthdep: An efficient and secure deduplication scheme for cloud-assisted ehealth systems. *IEEE Transactions on Industrial Informatics*, *14*(9), 4101-4112. doi: 10.1109/TII.2018.2832251

Zheng, Y., Lu, R., Zhang, S., Guan, Y., Shao, J., & Zhu, H. (2022). Toward privacy-preserving healthcare monitoring based on time-series activities over cloud. *IEEE Internet of Things Journal*, *9*(2), 1276-1288. doi: 10.1109/JIOT.2021.3079106

# Reference Links

Welcome to GeoPy's documentation! GeoPy 2.2.0 documentation

Calculating distance between two geo-locations in Python — by Ashutosh Bhardwaj — Towards Data Science

Finding distance between two latitudes and longitudes in Python — by Zolzaya Luvsandorj — Towards Data Science

haversine · PyPI

geopy · PyPI

Finding distances based on Latitude and Longitude (hedges.name)