PIR - Research Initiation Project (4th year)

Electronic voting II

"Doctolib"

Cléa AUBERY
Maxime DAVID
Loïc THOMAS
Aymeric SHINAWI--CROS
Léo VENDEVILLE

Tutors: Eric ALATA – Vincent MIGLIORE 4IR - SI

***Abstract*—Web services are more and more used to take any kind of appointment, and the medical field is no exception. They might even replace phone calls in the near future. However, this kind of service can present threats to users' privacy. Sensitive data such as health conditions or medical documents should not be accessed by anyone. In the best-case scenario, it should not even be accessible to the service provider. This study aims to find a secure, reliable online service that fully protects data privacy. We further explore the methods and issues to date. We focus on guarantees and requirements, current architectures, and cryptographic algorithms. Our goal is to have an overview before implementing our own medical appointment management app.**

Keywords: E-Health Services, Data Privacy, Encrypted Database, Homomorphic Encryption, Third-party architecture, Cloud, Blockchain, Ring-LWE, Cryptographic algorithm

1. Introduction

The COVID-19 pandemic made the use of medical appointment applications widely popular. However, such applications often require access to sensitive medical data, which should not be needed. The users thus have no trustworthy alternative and must choose between their privacy and convenience. We want to build a similar application where only patients and healthcare professionals can access this data. For instance, no system authority could access the appointment record of any patient. We looked for privacy issues in existing applications and the requirements for data protection and security in Europe. We went through possible approaches to address privacy issues, such as quantum computing on encryption, cryptographic algorithms, or blockchain systems. We want to ensure that the data stays encrypted during operations, i.e., the server must be able to correctly treat encrypted data.

2. Privacy and security issues

The recent development of telecommunications technologies and IoT has induced an incredible boost in the creation of mobile applications. Health condition management and wellbeing apps are a great example, with an average of 250 new health care apps every day for more than 350,000 in total, representing 47% of all apps in 2020 [1]. The reason for such growth is their accessibility and usefulness. Users can manage their health themselves from home and often for free, while ill patients can exchange messages with their doctor directly from the app and receive a diagnosis or advice on their condition.

However, the race to develop and publish mobile health apps should not be done at the expense of security and privacy. In many applications, service providers have full access to the customers' information. This is a privacy issue because private information should be disclosed only to the practitioner. This also poses a security threat if the third-party integrity is compromised. These issues are accentuated when doctors and patients use health apps that have not been thoroughly tested and secured or without reading the security and privacy policies.

2.1. What are the issues?

To operate, health care apps collect Personal Health Information (PHI) such as medical conditions, general practitioners, treatment plans, and contact details. Those data are stored on the cloud and can be used in long-distance transactions with other apps, doctors, or third parties. This is a source of privacy concern since the users do not know if their data is used legitimately or sold to other companies.

As well as being misused, data can also be stolen. Health apps store a lot of data on their servers, exposing them to breaches or hacks by cybercriminals if they are not protected. Information is also often exchanged between practitioners, which increases their risk of being compromised. When breaches occur, users' data is exposed and can be used for malicious purposes.

2.2. How to address these issues?

To address these concerns, different laws have been established, but because of disparities among regions, EU and US laws are different from Asian ones. Those laws are covered in [2] for the EU and USA:

- Cover data: which data should be protected in priority, namely those that can be used to identify a person.
- Information requirements: users should be informed about the identity of the entity using their PHI as well as their motive.
- Consent requirements: users' written consent must be obtained.
- Data retention: PHI should be kept only while necessary then erased when their purpose is reached.
- Security: entities must ensure data are protected against accidental and intentional loss or access by implementing the necessary technical, administrative, physical, and organizational security measures.
- Breach notification obligations: in case of a breach, competent authorities should be notified, as well as concerned users. If the breach is massive, the media should also be informed.
- Data transfers: entities need the user's consent to transfer their data to a third party.

To respect these laws, [2] gives several requirements as well as recommendations that must be followed for better security, especially in a sensitive domain such as health care:

- Access control: Users should be able to allow or forbid access to their information at any moment.
  Using a role-based access policy is recommended.
- Authentication: Authentication must be done with a unique ID and password and encrypted with a sufficient algorithm (RSA).
  Passwords must be complex, and the use of multifactor authentication is recommended.

- Security and confidentiality: Use the Advanced Encryption Standard (AES) to encrypt PHI. The cryptographic key must have at least 128 bits.
  It is recommended to use a 192- or 256-bits key for more security.
- Integrity: Developers must at least use a symmetric key-based authentication code like AES.
  Public key-based digital signatures are preferred. Watermarking should never be used with medical images since it can deteriorate their quality.
- Inform patients: Before the collection of data, the app should present a privacy policy to its users.
  Policy should be concise, easy to understand, and accessible at any time.
- Data transfer: Use Transport Layer Security (TLS) with 128-bit encryption methods or Virtual Private Networks (VPNs).
  It is preferable to use TLS with 256-bit encryption methods and to show an icon in the app notifying the transfer of data.
- Data retention: The retention policy should be included in the privacy policy.
  When the purpose is achieved, the data must be erased and the user notified. Users should be able to check that the data has been deleted.
- Breach notification: The competent authority as well as the affected user must be notified as soon as possible (1–3 days).
  If possible, the entity should compensate the affected user to restore the damage done. In cases of large breaches, the media must be notified to inform them about the problem.

While mobile health apps can offer many benefits and ease the lives of their customers, it is crucial to consider the privacy and security risks linked to the exposure of private data to the internet and take the necessary precautions to protect personal health information.

3. Guarantees and requirements

3.1. Data protection

A. Data protection in Europe

Data usage is protected in Europe with the General Data Protection Regulation (GDPR) since 2018. It applies to all businesses and other legal entities that process personal data. The GDPR requires that businesses take measures to secure the processing of personal data to protect basic human rights. The data should not be used to discriminate against people based on their ethnicity or genre. It focuses on data protection and privacy and requires that all systems that handle personal data be designed with data protection in mind. Since this was previously not a legal requirement, it presents a challenge for businesses that did not consider privacy when designing their system. Less than 20% of systems would be considered “designed with privacy in mind” if analyzed [3]. Despite its importance and the overall improvement in terms of data usage, GDPR also has limitations, the most important one for us being the lack of specific requirements for certain industries (here, medical). We should not only consider GDPR for the treatment of data, but also set rules on our own, creating a system based on privacy.

B. Data security

Cloud-based systems are vulnerable to various risks that must be considered to ensure data security. While some risks are similar to those found in other IT environments, others are unique to the cloud environment due to the nature of data circulation on servers. [4] Data security is something very large, and to ensure it, we must consider security concerns at multiple levels, including the network, host, and application. These levels can be matched to steps of the data life cycle: creation, transfer, storage, and computation. In terms of concrete warranties needed for the data, we find the CIA triad: confidentiality, integrity, and availability, which guides information security policies. We can use this basis to describe the assurance we need in our system. In every step of data treatment, we must ensure confidentiality and integrity: the data must be kept private and must not be altered. We can also add a special guarantee for the stored data: it

must be available when needed. Some of these warranties have existing solutions. For instance, encryption for data confidentiality, with the problem of the key storage, repartition and communication. Some are trickier, like the integrity of the data in long-distance storage.

C. Specific needs for medical data

E-Health, the use of the internet for medical purposes, can be very convenient for health-care professionals and patients. However, it brings new problems and challenges concerning data protection. Everyone could want to use those web services, but not everyone is comfortable with the use of the Internet or aware of the risks of data collection. There is the issue that the service itself should follow some ethics, and we need to be sure that is the case. There exist some standards specific to the medical field on the internet; they are grouped under the Health On the Net Foundation Code of Conduct (HONcode). It sets out eight principles that health and medical websites should follow, including providing accurate and trustworthy information, protecting users' privacy, and clearly indicating the qualifications and identities of authors and contributors. When a website meets the standards, it can display the HONcode seal, which stamps the website as reliable. It is widely known and recognized, which makes it a first step to indicating to users which websites are more secure. However, this stamp isn't sufficient: in a recent study of 30 HON websites, 10 were deemed not secure [5].

3.2. Towards post-quantum

A. Quantum computation vs current cryptographic algorithms

The security of many current cryptographic algorithms is based on some kind of computational complexity. For instance, RSA, AES, ECC. They are considered secure because traditional computers can only break this kind of system in an exponential amount of time [6]. Regardless, everything changes when we consider quantum computers. They will drastically change computing capabilities, allowing for computations that are not feasible on classical computers. In fact, their development will pose a threat to a lot of systems currently in use, as they could easily crack current cryptographic algorithms.

Still, we are not entirely destined for failure. We can't sum up current cryptography by some algorithms like RSA, DSA and ECDSA. We must keep in mind that even though some algorithms will become inefficient, they are not the only ones used. Algorithms that depend on hash-based cryptography, code-based cryptography, or lattice-based cryptography remain unbreakable using quantum computing. Also, the cost of the hardware needed for quantum is not to be neglected. A contemporary cryptographic system can operate on a standard CPU and produce gigabytes of data per second using a $200 processor. On the other hand, quantum cryptography only generates kilobytes of data per second and requires specialized hardware that costs around $50,000. [7]

B. Quantum algorithms

Since the first International Workshop on Post-Quantum Cryptography in 2006 and onwards, some algorithms have been designed specifically for quantum cryptography. We can cite Shor's, Grover's, and Simon's algorithms. Each one of these solves a different problem that is impossible to solve in a reasonable time with classical algorithms. Other algorithms, like McEliece encryption, provide an alternative to the classical encryption theme.

The threat of quantum computers is very distant, but it exists. Still, we are not yet replacing every encryption scheme with quantum-resistant ones. It is very expensive, and we need more time to improve the efficiency and usability of post-quantum cryptography as well as build confidence in it.

4. Currently Explored Architectures

Two main types of architecture are popping up in papers on this subject. On one hand, databases that are using blockchain architecture principles

[8]; on the other hand, architectures that use cloud databases [9].

4.1. Blockchain

The blockchain is a distributed database, where data is stored on several machines rather than on one or a few centralized servers. Its principle is based on “blocks” of encrypted data. Every single block of data is linked to its predecessor with a cryptographic hash. This creates a chain of blocks, or a blockchain. The blockchain has real advantages concerning the storage transparency of the public database [9]. It is also a great way to provide confidentiality, integrity, and authenticity for the data, because the community verifies the validity of the data of each user and each transaction thanks to consensus algorithms.

Concerning healthcare data, it could be interesting for us to use a blockchain-based architecture because of its privacy and transparency benefits. The article about the Decentralized Self-Management of data Access Control (DSMAC) presents a new blockchain-based model specialized for health care data [8]. This architecture is like bitcoin’s and provides full confidentiality for all users. However, there are some big issues with a blockchain database. This kind of architecture would not cost a lot of course, but it would imply a huge amount of latency because of verification time [9]. Every transaction has to be verified by a subcommunity of the network’s nodes. In the paper presenting the *PMRS,* searchers concluded from their analysis that efficient blockchain-based architectures such as Hyperledger or Ethereum have mostly higher latency when they process a lot of users' transactions, while cloud-based architectures remain more constant [9]. It is mostly due to the need for the system to validate the data and the transactions in a blockchain-based architecture, which requires more processing power and time as the amount of data grows.

4.2. Third Party Architectures

A third-party-based database is a database that is not stored by a company itself but by a third-party company specialized in data storage and cloud management. In this way, the company that needs a database will not have to worry about hosting its data itself and will save money and time. Plus, the third-party company will provide constant availability, maintenance, security with adapted firewalls and encryption mechanisms, and scalability in case a company needs to expand its database. Nonetheless, it involves full trust in the third-party provider because the whole set of data, including sensitive ones, is stored outside of the company. Some researchers tried to solve this problem by finding different ways to hide data from the cloud provider so they would not have full access to sensitive data. In our case, the efficiency of a third-party architecture would be a real benefit, but we also strongly want to find a way to protect patients' healthcare data.

A paper already provides a secure way to communicate with point-to-point encryption and warranted access using a key management service as a trusted third party (TTP) service [10]. First, the key exchange between the two parts is based on the Diffie-Hellman algorithm, providing a secure way to set up any encryption protocol with public and private keys. This algorithm will ensure both communicators have the same shared secret while preventing any other malicious party from being able to guess it by sniffing the key exchange protocol. Then, the two parties will agree on a digital signature algorithm and choose two values, p and g, that they will both use to encrypt and decrypt their messages. This article also gives a set of 13 typical requirements on a TTPs concerning the use of this method and improvements that are needed to be fully worth it. Some issues are presented; for instance, the trust between the sending and receiving parties is supposed to be valid. It implies the need for a certification hierarchy. The choice of values (such as p and g) is really controversial too. Finally, authors also explain that TPPs require a lot of financial resources to be held, so this model could be too expensive for little businesses. In our case, a huge infrastructure would of course find the necessary resources. However, this model doesn’t provide something that our structure would need: the treatment of the data by the third-party provider. Here, everything is about

communicating, but we also need to show the available slots of a doctor according to a patient's availability and region. We cannot use only such a model.

Another study proposed a model where the blockchain's advantages are kept while using a third-party provider for cloud management [9]. Their aim was to improve the latency and throughput of e-healthcare records. They mainly use steganography and roles to protect patients' data. This architecture provides full control for users over their health care data so that only certified doctors or relatives can access it. The steganography module could be a good way to protect patients' data in our case, but we still have an issue concerning the data processing: the third-party provider would have access and would process, in this case, patients' E-Healthcare records. What we want is an architecture like that where even the cloud provider would ignore those data.

We didn't find any articles concerning a system exactly like that. Though one article is about processing encrypted data in the case of an untrusted environment, it is an approach that we will discuss in Section 5.2 [11].

We could imagine an architecture where a first third-party provider A containing an encrypted e-healthcare database would be linked to another third-party provider containing all keys to e-healthcare records. In this configuration, the third-party provider A would not have any understanding of its data, but could only send them to their treatment part to B. B would only have requests like "add three to this value", so data would be hidden from B. Then, B would encrypt and send the modified data to A. This architecture could be used to solve our problem.

5. Cryptographic algorithms

5.1 Cryptographic algorithms types

In this document, we focus on cryptography for privacy in cloud services. We want to perform operations on encrypted user data without revealing it. For this purpose, we can identify four main types of cryptographic algorithms. [12] Firstly, homomorphic encryption (HE) is a special kind of encryption scheme that allows third parties to operate on the encrypted data without decrypting it in advance. We also have fully homomorphic encryption (FHE), which allows an unlimited number of operations an unlimited number of times. It can be set up with cloud-based data services, and the first achievable model was made in 2009. However, it is conceptually and practically a non-realisable scheme. Another cryptographic algorithm that we can use is partial homomorphic encryption (PHE). This algorithm is already deployed in some e-voting applications, but it allows only a few operations with an unlimited number of repetitions. So, it can be used only in applications that need addition or multiplication. The last type is the somewhat homomorphic encryption (SWHE) ; it allows some type of operations like addition or multiplication. The size of the ciphertexts grows with each homomorphic operation; thus, the maximum number of allowed homomorphic operations is limited.

5.2 Fully homomorphic encryption and ring LWE problem

Fully homomorphic encryption (FHE) is a powerful cryptographic tool that enables computation on encrypted data. One of the FHE schemes can be simple and secure; this scheme is based on the ring learning with errors (RLWE) assumption [13]. Basically, the RLWE problem is defined like this: given a ring R and a secret vector s with coefficients in R, with several noisy equations $a_i\, s + e_i = b_i\, mod(q)$ where a_i and b_i are random vectors with coefficients in R and e_i an error vector. We want to recover the secret vector through the noise. It is a hard computational problem and if we have a "polynomial ring $R_q = Z_q[X]/f(X)$, and a random polynomial $\omega \in R_q$, it is computationally hard to distinguish the uniform

distribution over $R_q \times R_q$ from ordered pairs of the form $(a_i, a_i\omega + e_i)$ where a_i are uniformly distributed in R_q and e_i are polynomials in R whose coefficients are independently distributed Gaussians" [14]. The original definition of Ring-LWE creates theoretical and implementation difficulties. However, in the case when f(X) = X^n + 1 with *n* being a power of two, it may be simple. Transformation can be used to reduce the complexity [14] The resulting scheme is circularly secure, which means it can be used to encrypt its own secret key. This is achieved by transforming a somewhat homomorphic encryption scheme into a fully homomorphic one using standard techniques [13]. The FHE scheme based on RLWE has many potential applications, including secure cloud computing. However, the implementation of RLWE involves generating a random polynomial that has a noise coefficient, which is a random value added to the product of a secret key and a public parameter. The correct noise distribution is important, as it can affect the security and efficiency of the scheme. To simplify the noise distribution, a Gaussian distribution can be generated over a particular extension ring of *R* [14]. The noise coefficient can also be scaled by a factor of *m* and then taken modulo *R,* but doing so can increase the magnitude of the noise and make decryption more difficult [14]. In summary, the RLWE problem is a powerful tool that can be used to construct FHE schemes and pseudo-random functions. The FHE scheme based on RLWE is simple and secure and can be used in applications such as secure cloud computing. However, implementation issues related to the noise coefficient and the fractional ideal *R* may arise and must be addressed to ensure the security and efficiency of the scheme.

5.3 Using a cloud database while preserving privacy

Database as a Service (DBaaS) enables companies to run their databases on a cloud server, provided by other big companies (Amazon, Oracle, IBM, etc.). We might use this solution for a future implementation as it is both cost-effective and time-saving. However, it also raises obvious privacy concerns. As a storage medium, all data can be encrypted, but we still need an algorithm to run the queries. The database needs to know where to perform operations and what to retrieve without knowing anything about the content itself. In a 2015 paper [15], the authors propose a new algorithm to answer this problem. Former approaches include Query Optimal Bucketization (QOB) and Binary Query Bucketization (BQB). Here, Shastri, Kresman, and Kwan Lee introduce Parallel Binary Query Bucketization (PBQB). The parallelization mainly increases performance, but what is important in our case is the Bucketization. To retrieve encrypted data X, the server will divide the records into several parts and send back the last record of each part. The client will decrypt them and tell the server in which part X should be, based on a binary search. The operation is done until the client retrieves only the wanted data. Figure 1 illustrates the process.

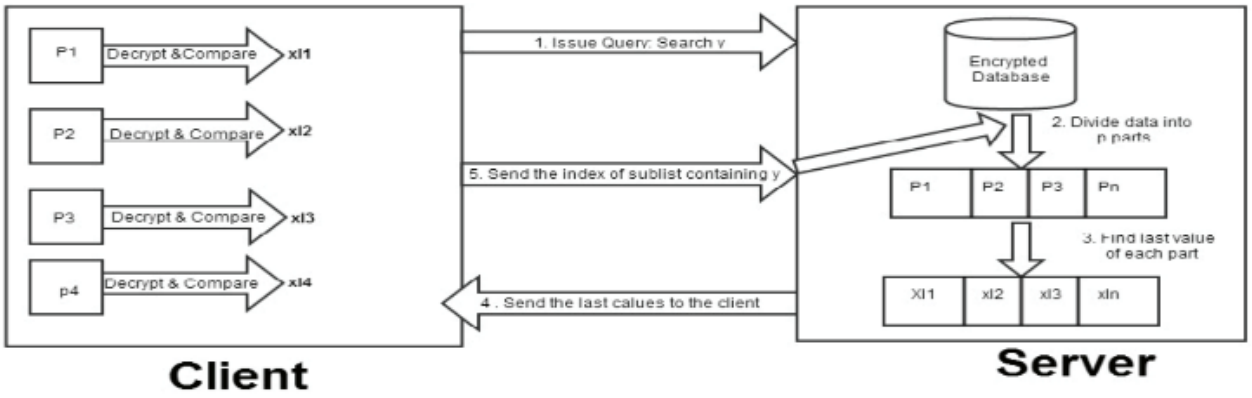

Figure 1. Architecture of Parallel Binary Query Bucketization

Other researchers have also developed a bipartite online evaluation protocol [16] analogous to our topic. In their paper, they address the problem of an online course evaluation system where a student S must give a score to each university course he is enrolled in. Like a medical appointment, one professor P must be able to see the score given by their students, but no one between these two parties can have access to the information. Several constraints must be checked to ensure the user's privacy:

1. S must be enrolled in the course, and thus need to authenticate.
2. To be fair, S can give a score only if they have minimum class attendance.
3. It is impossible to trace who submitted a given score or what other scores come from the same evaluator.
4. S can only evaluate each course once.
5. S cannot perform the evaluation for another student.
6. S cannot evaluate the same source twice.
7. No fake evaluation can be posted by any user or authority.
8. No submitted evaluation can be modified by any user or authority.

The authors answer the problem with three servers between the student and the university. One is used to authenticate the student, another to check the eligibility for evaluating each course, and the last one to submit the scores. The protocol uses primitives such as encryption keys, hash chains, and blind signatures to ensure data privacy. As a result, their protocol enabled several hundred users to submit an evaluation with their anonymity protected against honest but curious adversaries.

6. Conclusion

This literature review exposes the major issues that come with the development of a health care application as an intermediary between patients and doctors. Due to the nature of the application, privacy and security must be thoroughly studied to offer guarantees to their users. This requires different architectures from the other applications. Cryptography also plays an important role in ensuring the safety of its use. Various kinds of algorithms are available, each with their advantages and disadvantages. Considering the development of quantum computers, it is worth mentioning quantum cryptographic algorithms, which are currently unnecessary but might become a necessity in the future.

7. Bibliographical references

[1]. "Digital apps in the health experience", in *Digital Health Trends 2021*, IQVIA Institute, United States, 2021. Available on: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021/. [Visited: 15-fév-2023].

[2]. B. Martínez-Pérez, I. de la Torre-Díez, M. López-Coronado, "Privacy and Security in Mobile Health Apps: A Review and Recommendations", J Med Syst 39, 181 (2015). Available on: https://doi.org/10.1007/s10916-014-0181-3. [Visited: 12-Dec-2022].

[3]. F. Blix, S. A. Elshekeil and S. Laoyookhong, "Data protection by design in systems development: From legal requirements to technical solutions," 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST), Cambridge, UK, 2017, pp. 98-103, doi: 10.23919/ICITST.2017.8356355.

[4]. D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," 2012 International Conference on Computer Science and Electronics Engineering, Hangzhou, China, 2012, pp. 647-651, doi: 10.1109/ICCSEE.2012.193.

[5]. Y. Hong, T. B. Patrick and R. Gillis, "Protection of Patient's Privacy and Data Security in E-Health Services," 2008 International Conference on BioMedical Engineering and Informatics, Sanya, China, 2008, pp. 643-647, doi: 10.1109/BMEI.2008.331.

[6]. R. Bavdekar, E. Jayant Chopde, A. Bhatia, K. Tiwari, S. Joshua Daniel, Atul, "Post Quantum Cryptography: Techniques, Challenges, Standardization, and Directions for Future Research", 2022, arXiv:2202.02826v1 [cs.CR].

[7]. D. J. Bernstein, J. A. Buchmann, E. Dahmen, "Post-Quantum Cryptography", 2008, in First

international workshop PQCrypto 2006, Leuven, The Netherland. Available on: https://www.researchgate.net/publication/267139573_Post-quantum_cryptography_First_international_workshop_PQCrypto_2006_Leuven_The_Netherland_May_23-26_2006_Selected_papers. [Visited: 02-fev-2023].

[8]. H. Saidi, N. Labraoui, A. A. A. Ari, L. A. Maglaras and J. H. M. Emati, "DSMAC: Privacy-Aware Decentralized Self-Management of Data Access Control Based on Blockchain for Health Data" in IEEE Access, vol. 10, pp. 101011-101028, 2022, doi: 10.1109/ACCESS.2022.3207803.

[9]. K. Zala, H. K. Thakkar, R. Jadeja, P. Singh, K. Kotecha, M. Shukla, "PRMS: Design and Development of Patients' E-Healthcare Records Management System for Privacy Preservation in Third Party Cloud Platforms," in IEEE Access, vol. 10, pp. 85777-85791, 2022, doi: 10.1109/ACCESS.2022.3198094.

[10]. N. Jefferies, C. Mitchell, M. Walker, "A Proposed Architecture for Trusted Third Party Services" in Cryptography: Policy and Algorithms, 1995.

[11]. C. Boyens,O. Günther, "Using online services in untrusted environments: a privacy-preserving architecture" in European Conference on Information Systems, ECIS 2003, Naples, Italy. Available on: https://www.researchgate.net/publication/221407800_Using_online_services_in_untrusted_environments_a_privacy-preserving_architecture. [Visited: 30-Jan-2023].

[12]. A. Acar, H. Aksu, A. S. Uluagac, M. Conti, "*A Survey on Homomorphic Encryption Schemes: Theory and Implementation",* Vol 51, No 4, Article 79, 2013. Available on: https://dl.acm.org/doi/10.1145/3214303/. [Visited: 12-Jan-2023].

[13]. L. Ducas, A. Durmus, "Ring-LWE in Polynomial Rings" in Lecture Notes in Computer Science, vol 7293. Springer, Berlin, Heidelberg. Available on: https://doi.org/10.1007/978-3-642-30057-8_3. [Visited: 18-Jan-2022].

[14]. Z. Brakerski, V. Vaikuntanathan, "Fully Homomorphic Encryption from Ring-LWE" in: Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference. Lecture Notes in Computer Science, vol 6841, Springer, Berlin, Heidelberg. Available on: https://doi.org/10.1007/978-3-642-22792-9_29. [Visited: 09-fev-2023].

[15]. S. Shastri, R. Kresman and J. K. Lee, "An Improved Algorithm for Querying Encrypted Data in the Cloud," 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, India, 2015, pp. 653-656, doi: 10.1109/CSNT.2015.250.

[16]. N. Petrakos, S. Monachos, E. Magkos, and P. Kotzanikolaou, "Design and Implementation of an Anonymous and Secure Online Evaluation Protocol," Electronics, vol. 9, no. 9, p. 1415, Sep. 2020, doi: 10.3390/electronics9091415. [Online]. Available on: http://dx.doi.org/10.3390/electronics9091415. [Visited: 25-Dec-2022].