ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL
Volume 6 (Issue 1) (2022) Pages 66-80

Exploiting the Pallier Cryptosystem for Secure Distributed Spatial Search Using a Distributed Ring Algorithm

Jimmy Katambo, Mayumbo Nyirenda and David Zulu

Department of Computer Science, University of Zambia, Lusaka, Zambia.
{jimmy.katambo, mayumbo.nyirenda, david.zulu}@cs.unza.zm

Abstract **- The problem of lack of anonymity and confidentiality can be experienced by those who collect statistical data online as well as those who provide the data. One end may be secure, for example, the one providing data, and yet the other end, for example, the one collecting data, may not be secure. In another scenario, both the data provider and collector may seek anonymity. Preventing the decryption of data provided while providing aggregated results is the best solution for such scenarios. To achieve this, this paper proposes a protocol that puts into application Homomorphic Encryption and a Distributed Ring algorithm, to ensure data anonymity of both parties involved in a spatial search that is a data provider and a searcher. Firstly, we identify a Homomorphic Encryption technique that can work best for a spatial search by reviewing literature on Homomorphic Encryption techniques. Among the Homomorphic Encryption techniques reviewed were Rivest, Shamir and Adleman (RSA), El Gamal cryptosystem, Goldwasser-Micali cryptosystem, Benaloh cryptosystem, Paillier cryptosystem and Fully Homomorphic Encryption (FHE). After a comprehensive study, Paillier Homomorphic Encryption technique was identified as the best approach to be employed in securing a spatial search. Secondly, we propose a protocol for distributed spatial searching using Paillier cryptosystem and distributed ring algorithm principles. Finally, a proof of concept prototype using the proposed approach was implemented. From initial experiments conducted using the proposed approach, it is evident that the bigger cost comes from the communication over the network and less from the encryption algorithm and protocol itself. A 39.7% overhead when compared to the usefulness of the approach, is outweighed making the solution highly practical and useful.**

Keywords: Homomorphic Encryption, Paillier Cryptosystem, Confidentiality, Spatial search, Ring Algorithm.

I. INTRODUCTION

Spatial searching gives us an opportunity to search for data based on location preferences. The result is limited to the physical area of the location. Some common examples of spatial searches are finding all highways within a distance of a city, finding the cities above a specified population nearest to a highway, and mapping applications to search libraries, museums, archive collections, etc. However, our research will emphasize more on the following key examples. In secure geocoding, a geocoder can be made available to the cancer surveillance community where participants are able to provide their physical locations in the form addresses to the geocoder which then transforms these into a pair of latitude/longitude coordinates which are used to come up with an aggregate data structure that is appropriate for calculations and analysis. Another good example of the applicability of our research is sharing of private patient disease data across several facilities to support overall reports which highlight information such as case counts and so on. Such aggregations like during the Covid-19 pandemic can be used to find the total number of cases based on geographic/spatial boundaries.

Statistical information such as election results, the number of HIV AIDS patients in certain geographical areas, data for use in cancer clustering and surveillance, etc. can be accessed by researchers online. The reason for accessing it are numerous including sharing and combining information, broadcasting it, identifying patterns for statistical analysis, logistical support and planning, market surveys and business initiatives, and further scientific research to mention but a few. Approaches like that proposed by [1] [2] have made access to spatial searches a highly feasible approach. However, this access can be problematic to both a data provider and a searcher and can result into compromised privacy and sometimes exposure of confidential information. Research in Mathematical Cryptography has opened many doors for researchers to study Homomorphic Encryption techniques to identify possible solutions so that anonymity is fully guaranteed on both ends of the equation, that is, the anonymity of the person doing the searching and also the anonymity of the data provider [3]. Homomorphic Encryption offers capability to do highly complex mathematical computations on data in its encrypted form without the need of first decrypting the data. [4]. Homomorphic Encryption allows users to make changes to the data without decrypting it [5]. The key principle in Homomorphic Encryption is we can achieve addition and subtraction of the data using Additive Homomorphism and we can also get multiplication and division of the data using Multiplicative Homomorphism. Paillier Cryptosystem was identified as the best for supporting a spatial search because it is computationally cheaper to be used in practice. Paillier's scheme is the most efficient among currently known additively homomorphic schemes, that is, it requires simple operations in the encryption, decryption, and addition procedures and hence

achieves high performance. Another observation is that calculations on the encrypted data can be performed without necessarily reconstructing the original message and without having access to the private key. This article, therefore, proposes a distributed protocol that implements Paillier Homomorphic Encryption, to hide the identity of the searcher and data provider participating in a spatial search. The work is presented in the rest of the paper as follows: Section II gives an overview of the literature reviewed; Section III discusses materials and methods used; Section IV presents the results of the research, Section V presents a discussion based on the results and finally, Section VI presents a conclusion.

II. Literature Review

A. Homomorphic Encryption Techniques

1) Rivest, Shamir and Adleman (RSA): Rivest, Shamir and Adleman proposed RSA in 1977 [6]. It is a widely used public key cryptosystem. The underlying principle of RSA's security is the factorization of big numbers of modulus. The suggested length of the value of the modulus is 2048 bits. Usage of 640 bits for the key is considered not to be secure. RSA uses a modulus value which is 1024 bits in size [7] and adopts the public and private key pair concept. The public key is used for encryption whereas the private key is used for decryption. In this paper we adopt the notation used in [6] and symbolize the public key as *ku = {e, n}* and the private key as *kr = {d, p, q}*. A description of variables and operators in the stated equations is illustrated in full in the following steps below:

Step 1: p and q are two relatively prime and large random numbers.

Step 2: A positive integer n is defined as a product of p and q.

Step 3: Eulers value of

$$\varphi(n) = (p-1)(q-1). \quad (1)$$

Step 4: Choose e such that $1 < e < M < n$ and

$$C = Me \bmod n. \quad (2)$$

Note that M denotes the plaintext, C is the encrypted message or cyphertext while e and n are public keys.

Step 5: In RSA e and n are public keys and d and (p, q) are private keys so the plaintext M is encrypted by:$1 < M < n$ and $C = Me \bmod n$.

Step 6: The cipher text C is decrypted by

$$M = Cd \bmod n \quad (3)$$

Similarly, M denotes the plaintext while C is the encrypted message or cyphertext while **mod** refers to the modulo function which returns the remainder or signed remainder of a division after one number is divided by another (called the modulus of the operation).

Homomorphic encryption schemes have been around for a while and RSA is one the earliest multiplicative ones [8]. It supports applications that require multiplication of encrypted private data [8]. There are several methods that potentially increase the speed of the RSA encryption process, but they also make the decryption process more computationally complex [9]. Consequently, like other encryption methods, RSA is mathematically unstable. RSA has several constraints and numerous effective attacks can be formulated to defeat this algorithm [7], [10] with its biggest obstacle being the factorization. The whole algorithm is broken once the process of factorization is done. When this occurs, RSA can no longer ensure the security of the secrets it protects [7].

2) El Gamal Cryptosystem: El-Gamal is a public key encryption cryptosystem proposed by Taher El-Gamal in 1984. It portrays a multiplicative homomorphic encryption property [12] built on the Diffi-Hellman problem's hardness [11].

3) Goldwasser-Micali Scheme: According to R. Shruthi, P. Sumana and A. K. Koundinya, Shafi Goldwasser and Silvio Micali created the Goldwasser-Micali cryptosystem, an asymmetric key encryption technique, in 1982 [13]. It is the first probabilistic public-key encryption system to be provably secure under accepted cryptographic premises [13]. However, due to the possibility of ciphertexts being hundreds of times larger than the original plaintext, it is not an efficient cryptosystem.

The fact that the RSA encryption method leaks one bit of plaintext per ciphertext in its original form is one of its flaws [14], [15]. The first cryptosystem to indubitably resolve this issue was the GM cryptosystem. Goldwasser and Micali introduced it along with a strict concept of security known as semantic security and evidence that the GM cryptosystem is semantically safe against plaintext attacks.

R. Shruthi, P. Sumana, and A. K. Koundinya's study demonstrates that while Goldwasser-Micali requires a little bit more time for encryption than RSA, it is also more resistant to assaults. Second, they demonstrate that the Goldwasser-Micali had more variable encryption times, reaching a maximum of 27.5 milli seconds for plaintext of 6 bytes and a minimum of 3.6 milli seconds for plaintext of 20 bytes. In contrast, the decryption time of RSA remained almost constant between plaintext sizes 2 to 26 bytes. Thirdly, they discovered that the number of blocks of ciphertext created for growing plain text sizes increased linearly for RSA and Goldwasser Micali. However, the increase is more noticeable in the case of Goldwasser Micali, which can be explained by the fact that every piece of plain text in Goldwasser Micali is given a random value [13].

In their paper's conclusion, they show how Goldwasser and Micali create a bit encryption function based on the hardness of the quadratic residuosity problem. Although the method has many advantageous characteristics, it has one significant flaw: for a given security parameter N, each bit's probabilistic encryption requires N random bits and involves several operations on N bit integers. The message expansion by a factor

of lg n bits of the Goldwasser-Micali system is a significant drawback [13]. In a probabilistic encryption technique, some message extension is inevitable because several ciphertexts correspond to each plaintext [13]. The homomorphic property of the Goldwasser-Micali public-key encryption scheme can treat plaintext bit after bit, and the security is based on its semantic security, namely the quadratic residuosity assumption [16].

4) Benaloh Cryptosystem: Benaloh proposed an extension of the Goldwasser-Micali (GM) Cryptosystem by improving it to encrypt the message as a block instead of bit by bit [17]. Benaloh's proposal was based on the higher residuosity problem. A higher residuosity problem (x^n) is the generalization of quadratic residuosity problems (x^2) that is used for the GM cryptosystem. Any multiplication operation on encrypted data corresponds to an addition on plaintext, according to the homomorphic property of Benaloh. The Benaloh cryptosystem is additively homomorphic because it is possible to calculate the encryption of the addition of messages directly from encrypted messages E(m1) and E(m2) [17].

5) Additive Homomorphic Encryption (Paillier Cryptosystem): Another innovative probabilistic encryption approach based on the composite residuosity problem was developed by Paillier in 1999 [17]. The quadratic and higher residuosity issues utilized in the GM and Benaloh cryptosystems are quite similar to the composite residuosity problem. It questions whether there exists an integer x such that $x^n \equiv$ a $(mod\ n^2)$ for a given integer a. Other authors have referred to it as the Decisional Composite Residuosity Assumption (DCRA) [11]. This enables Paillier cryptosystem to have numerous applications such as threshold schemes and e-voting systems.

KeyGen Algorithm: For large primes p and q such that gcd (pq, (p − 1) (q − 1)) = 1, compute n = pq and λ = lcm (p − 1, q − 1). Then, select a random integer g ∈ $Z^*_{n^2}$ by checking whether gcd (L($g^\lambda\ mod\ n^2$),n)=1, where the function L is defined as L(u) = (u−1)/n for every u from the subgroup $Z^*_{n^2}$ which is a multiplicative subgroup of integers modulo n^2 instead of n like in the Benaloh cryptosystem [17].

Pick two large primes p and q and let n=pq. Let λ denote the Carmichael function, that is $\lambda(n) = lcm\ (p-1, q-1)$. Pick random g∈$Z^*_{n^2}$ such that $L\left(g^\lambda\ mod\ n^2\right)\ is\ invertible\ modulo\ n$ ($where\ L(u) = \frac{u-1}{n}$). n and g are public; p and q (or λ) are private. For plaintext x and resulting ciphertext y, select a random r ∈ Z^*_n. Then,

$$e_k(x, r) = g^m . r^n\ mod\ n^2, \quad (4)$$

$$d_k(y) = \frac{L(c^\lambda\ mod\ n^2)}{L(g^\lambda\ mod\ n^2)}.\text{mod } n \quad (5)$$

Finally, the public key is (n, g) and the secret key is (p, q) pair. The Homomorphism: Suppose x1 and x2 are plaintexts. Then,

$$e_k(x1,r1)\ e_k(x2, r2) = g^{x1}.r1^n g^{x2}.r2^n\ mod\ n^2$$
$$= g^{x1+x2}.(r1r2)^n\ mod\ n^2$$
$$= e_k(x1+x2, r1r2) \quad (6)$$

Note the following representations; c stands for the ciphertext, p and q are large prime numbers, KP is a public key, SK or λ is a private key, m is the plaintext message, L is the auxiliary function used in the decryption method to obtain m (plaintext message), E shows the encryption function, D is the Decryption function. Z_N denotes the set of nonnegative integers less than n. Z^*_N denotes the set of integers that are relatively prime to n. g is a random number where it has an ordered multiple of n. n is the product of two large primes p and q. gcd is the greatest common divisor which is the same as lcm or lowest common multiple. **mod** denotes the modulo function which returns the remainder or signed remainder of a division after one number is divided by another (called the modulus of the operation).

To perform addition and multiplication on encrypted data stored in the cloud provider, the client must have two different key generators (one for RSA and one for Paillier) [12].

6) Fully Homomorphic Encryption Schemes: According to A. Acar et al., an encryption scheme is said to be a Fully Homomorphic Encryption (FHE) scheme if the encrypted data may be subjected to an infinite number of evaluation operations and the output remains within the ciphertext space [17]. A long-standing unresolved challenge, acquiring an FHE scheme, was finally addressed by Gentry in his seminal PhD thesis about 30 years after the privacy homomorphism concept was initially introduced. The plan put forth by Gentry includes both an FHE method and a broad framework for obtaining one. As a result, several scholars have tried to build a safe and useful FHE system on top of Gentry's work. Although Gentry's ideal lattice-based FHE system is highly promising, it also has a number of drawbacks, such as its computational cost in terms of application in real life and some of its sophisticated mathematical principles that make it complex and difficult to implement [17].

According to V. Biksham and D. Vasuma, another limitation is that FHE does not cater to multiple users [18]. The practical applications which involve the running of enormously large and complex algorithmic computations homomorphically have a massive computational overhead, which makes the intermediate complex functional computations impractical [18]. In order to address the aforementioned bottlenecks, numerous new schemes and optimization have been developed in the wake of his work. The hard challenges of lattices are the main foundation for the security of new methods to obtain a new FHE scheme [17].

B. Ring Algorithms in Distributed Systems

B. K. Saraswat et al defines a Distributed System as a gathering of networked computers, which seems like one large computer where there is no shared memory in the distributed system so they communicate and coordinate their actions with the help of message passing [19]. This can be demonstrated by using an example of a Ring Election Algorithm. The Ring Election Algorithm is based on the ring topology with the

processes which are logically ordered and each process knows its successor either clockwise or anticlockwise direction [20].

Distributed systems, according to M. Al-Refai et al., are made up of numerous independent computers that work together to complete tasks that are distributed among them [21]. The main idea is to spread out jobs among several computers to speed up calculation when solving problems. When interacting over different network topologies, a leader process is necessary in distributed systems to organize and direct the communications and activities of a set of processes. Any process in a purely distributed system needs to communicate with every other process in order to perform a certain action. This communication eventually becomes complex. The fundamental concept behind lowering communication complexity is to select one of the currently active processes to serve as a centralized process, which in turn controls all processes' system-wide communications [21].

Le Lann provided an algorithm to choose a leader in a ring network, which makes the assumption that the processes are logically ordered and arranged in a ring and that each process has a unidirectional communication link with the next process in the ring [21]. A process will initiate an election message with its ID and transmit it to the next process in the ring if it notices that the leader has stopped functioning. Each process acknowledges receipt of the message and adds its ID to the message's list, designating itself as a contender for election as the message's leader. The message then returns to the initiator process that sent it, which selects the member of the list with the highest ID as the new leader while simultaneously sending out a new message to announce the new leader and the members of the new ring [21].

According to S. Basu, we must also make some assumptions about the communications network as we think about distributed systems [22]. This is crucial since the only means of communication between nodes is through the exchange of messages. The dispersed communications network's reliability should be taken into account in terms of the following factors.

1. That in a limited amount of time, messages are appropriately sent to their destination without being lost or modified.
2. Though the timing of arrival may vary, messages can only arrive at their destination in a limited amount of time.
3. Nodes are aware of one another's physical configurations and communication routes.

In his paper, S. Basu presents an algorithm for achieving mutual exclusion in a Distributed System [22]. The proposed algorithm is an improvement of the already existing Token Ring Algorithm, used to handle mutual exclusion in a Distributed system. His proposed algorithm does not permit the circulation of the token along the ring when there is no need (i.e. when no process wants to enter its Critical Section (CS)). There is easy detection when there is a loss of a token in the ring and that token can easily be regenerated in this algorithm. Not only is there an easy management of the process crash and recovery of a crashed process using this algorithm but also there is no chance of the creation of duplicate tokens in the ring.

According to A. Dadlani, in Distributed Systems, nodes communicate with each other using shared memory or via message passing [23]. Coordination is the key requirement for nodes to execute any distributed task effectively. There exists no central controlling node in a purely Distributed System that determines decisions. As a result, to make an appropriate decision, every node must communicate with the rest of the nodes in the network [23].

In her study, S. Naseera points out that nodes in distributed systems are connected at various geographical locations [24]. The data and resources are shared among these nodes in order to efficiently utilize resources. She highlights in her essay the requirement of choosing a leader to oversee this resource-sharing procedure by resolving disputes among the nodes. She then suggests a brand-new distributed ring technique for a distributed system's coordinator election as a result. The algorithm is an extension of the standard ring method.

The algorithm's performance was evaluated on a simulated distributed network and compared to the traditional algorithm. The proposed algorithm performs better than the traditional ring algorithm and is fault-tolerant. The distributed election process used by the proposed approach to choose the new coordinator takes less time overall than the traditional ring algorithm, according to their results [24].

A Ring Algorithm, according to H. Shaheen, offers the easiest technique to set up mutual exclusion across N processes without requiring an additional process to set them up in a logical ring [25]. The P(i+1)/mod N communication channel connects each process Pi to the process after it in the ring. A unique token is transmitted in message form in a clockwise direction between processes. A process immediately forwards the token to its neighbour if it doesn't require to enter the Critical Section (CS) when it receives the token. A process requires the token, waits until it receives it, but retains it. To exit the Critical Section, the process sends the token to its neighbour.

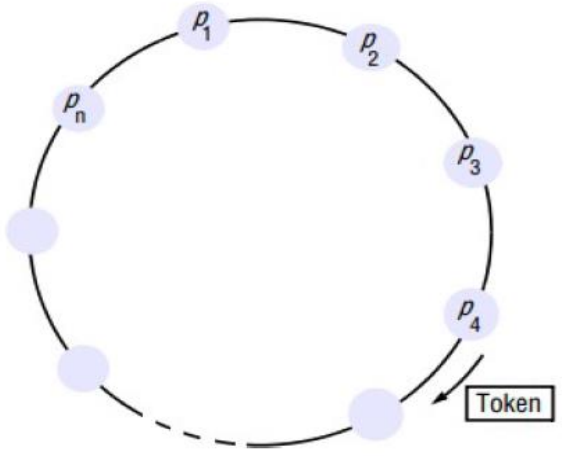

Figure 1: Ring-based Algorithm [25]

In Google's distributed search system, each computer is involved in indexing crawls and reviews a portion of the Web, taking URLs and sending that information, in a format that is compressed, back to a centralized server [25]. This data is then coordinated in a database by the central server along with data from additional indexing systems. Google's Domain Name Server (DNS) software routes a user's search query to the closest or busiest cluster of computers based on many parameters when the user types a query into the search field.

The Web server software distributes the query to hundreds or thousands of machines at the recipient cluster so they may all search at once. A database index is scanned by hundreds of computers to locate all required records. The document server gathers the titles and summaries, the index server aggregates the results, and the page builder creates the search result pages [25].

C. *Related Works*

Data guardians continue to be concerned about access to geographically referenced health data because of privacy violations or the accidental disclosure of sensitive information, according to G. M. Jacquez et al. [26].

In many countries, census data has been a staple of spatial analysis, and continues to be viewed as vital [27]. The data is not only accurate typically and detailed with spatial granularity highly but it is also easily accessible and presented with highly standardised documentation. Consumer data that results from the customers' and service providers' interactions are becoming ubiquitous. These data are inviting in as far as research is concerned because they are not only collected frequently but are also released quickly. They also deal with a wide variety of attitudes, lifestyles and behavioural characteristics; and they are often dynamically replenished and longitudinal. Demonstrated also is the fact that consumer data can importantly contribute to understanding problems in Transport Geography and in dealing with applied problems ranging from migration, infrastructure investment and retail service provision to commuting and individual mobility. However, to effectively exploit these data more, there is a need to construct bridges to permit greater freedom in data transfer from the commercial to the academic sector. There is also the necessity to develop frameworks for privacy and ethics in the way personal data is used secondarily [27].

In order to realise the potential of consumer data, careful ethical controls and well-designed security protocols are needed. New methods from data science may be worthy, but opportunities to restore and reinvigorate classical techniques in the light of new data should not be neglected [27].

A. Tondwalker and P. V. Jan identify the problem that is connected with data collection based on localization [28]. Battlefield surveillance or enemy tracking, rescue operations as well as monitoring of military facilities demand security and reliability of the location information. Similarly, spatial searching involves searching for data limited to a physical location or Geographical area. In their paper, A. Tondwalker and P. V. Jan propose to deal with this problem of lack of security to sensor networks and reliability of the location information by using steganography.

In their research, J. Ajayakumar and K. Ghazinour use readily accessible open-source tools for reverse geocoding and mapping to illustrate the spatial privacy risks caused by posting Location Based Services (LBS) check-ins on Twitter [29]. Along with highlighting the flaws, they also examine Twitter's and the LBS Swarm app's privacy regulations and offer recommendations for enhancing spatial privacy based on those analyses. They conclude by offering recommendations for how to enhance spatial privacy using policies and algorithmic tools.

M. Kiedrowicz categorizes confidentiality as one of the security attributes that affect spatial data [30]. He writes that creation of spatial information is made possible by the spatial data collected through remote sensing, photogrammetry, geodesy, cartography, services performing observations and measurements, monitoring, inventory taking, and statistical surveys, which are then used for expertly executed spatial analyses. Facts, events, things, phenomena, processes, and commercial operations are all covered by this knowledge. The relevant databases store the characteristics (sets of features and traits) of these things. Practically speaking, the spatial database—or geodatabase—mentioned here is a physical representation of actual objects. It facilitates the storing of spatial data (geometric, descriptive, raster, etc.) in databases or repositories of the IT system in Spatial Information Systems (SIP). He provides a summary of the procedures for guaranteeing the security of databases in SIP or Geographic Information Systems (GIS) as part of his article's conclusion [30].

According to R. M. M. Pradeep and N. T. S. Wijesekera, spatial data portrays information about the physical location, characteristics and shape of geometric objects [31]. These objects can be line and polygon entities representing nations, roads, lakes, etc., or point features indicating geographic places. Additionally, topology and coordinates are the primary data types used to store associations between geographic entities. A computer system called a Geographic Information System (GIS) can be used to collect, store, query, analyze, and present geographical data. The spatial data that goes along with a GIS is separate from the GIS program.

In their paper, R. M. M. Pradeep and N. T. S. Wijesekera also point out that the security issue is one of the major challenges in spatial decision support systems, especially when land ownership is part of the associated modifications, as R. M. M. Pradeep and N. T. S. Wijesekera point out in their study. Therefore, any purposeful or accidental modifications to spatial data would have an impact on ownership, which may be beneficial or detrimental. Therefore, the objective of their study is to create a data security mechanism that enables users to verify the legitimacy of GIS systems using desktop software [31].

In their paper, K. Chaturvedi et al focus on securing spatial data infrastructures for distributed smart city applications and services [32]. They describe Smart Cities as being complex distributed systems that may involve multiple stakeholders,

applications, sensors, and Internet of Things (IoT) devices. Spatial data infrastructures for Smart Cities can play a crucial role in ensuring interoperability between systems and platforms to handle the connecting and use of such heterogeneous data. The Smart District Data Infrastructure (SDDI) idea integrates various sensors, IoT devices, simulation tools, and 3D city models under a single operating framework. It is based on the open and global standards of the Open Geospatial Consortium (OGC). If these distributed systems are not protected, they could pose a serious risk by giving sensitive information to unauthorized or untrusted parties [32].

As location-based applications are flourishing, R. Guo et al observe that privacy preservation is a critical issue as far as spatial data is concerned [33]. A cloud server's outsourced data should be protected, especially when it is queried, to safeguard the privacy of people's geolocation information. Although the issue of secure range query on outsourced encrypted data has been thoroughly researched, the present solutions lack efficiency and scalability.

By creating a flexible multi-level search structure for spatial data that can quickly retrieve points inside a given geometric range while protecting private data and range queries from a cloud server, R. Guo et al. present a novel geometrically searchable encryption system, MixGeo, in their article. Their method greatly decreases the relationship between the search time and the amount of the recovered points in order to provide quick retrieval, effective updates, and steady performance, as opposed to directly performing compute-then-compare operations over a dataset.

According to D. Chen et al, Location Based Service (LBS) are becoming more popular on smartphones [34]. One of the common LBS is a range search. A range search is used to retrieve all Points of Interest (POIs) within a user-specified boundary. However, when employing LBS like range search, users also put their location privacy at risk. Chen et al aim to address the interesting but difficult topic of how a user can utilize such a service without disclosing their location. The majority of current methods obscure a person's location into a cloaked area, making it impossible for LBS to determine the precise location of the asking user. However, this would lead to incorrect findings that included some POIs that were outside of the acceptable range. To achieve this, D. Chen et al. provide a novel technique to give range search users location privacy. The user can encrypt her location using their method, which uses homomorphic encryption, and the LBS server can calculate distances using the ciphertext. The POIs within the defined range are precisely what LBS returns as results in this manner, and LBS gains no knowledge of the user's actual location [34].

According to A. Talha et al. the rise in spatial data has prompted businesses to upload their data to outside service providers [35]. With the use of cloud computing, database owners can outsource their operations, saving money on storage and processing power. The key difficulty is keeping query responses accurate and timely for authenticated users while protecting data secrecy regarding untrusted parties. To solve this problem, A. Talha et al. suggest using a dual transformation technique on the spatial database while the service provider runs queries and provides consumers with the results. Their method begins by mapping each spatial point in the multidimensional space to a one-dimensional space using the space-filling Hilbert curve. This method of space transformation preserves spatial proximity while being simple to compute. Next, the clustered data is subjected to the order-preserving encryption algorithm. A secret key is used to decrypt the query response when the user makes spatial range requests to the service provider on the encrypted Hilbert index. This makes data protection possible and consequently the cost of query communication between the user and service provider is decreased [35].

Furthermore, B. R. Pushpa highlights that secrecy guards against unwanted access to private data [36]. It guarantees that the required level of secrecy is upheld. One way of ensuring that data being transmitted is kept confidential is the application of cryptography. Cryptography is a technology that converts plain text into an unreadable format called cipher text (encrypt), then converts the ciphertext back to plain text (decrypt).

M. Modak and R. Shaikh observe that while data mining is an important and useful emerging trend, the possibility of it being distributed among various parties raises the issue of privacy [37]. According to X. Liao and C. Shu, nowadays, trust management is a new security problem that cannot be solved by traditional techniques such as data backup, recovery backup, and firewalls but by employing certain data hiding methods [38].

Data anonymization plays a major role in privacy preservation in non-interactive data sharing and releasing process. Data anonymization refers to hiding the identity of sensitive data so that the privacy of an individual is effectively preserved even though certain aggregate information can be still exposed to data users for diverse analysis and mining tasks [39].

III. Materials and Methods

A. *Selecting an HE Technique*

A systematic literature review, or SLR, was used to identify a HE technique that enables a spatial search in order to meet objective 1's main goal. An SLR provides a thorough summary of the literature. We adhered to the well-known Kitchenham rule [40, 41]. The key stages for carrying out an SLR are shown in Figure 2. We go into great depth about each stage in this section.

1) Stage 1: Setting up the Review: The main aim of this stage was to assess studies that focus on Homomorphic Encryption techniques so that the best technique is selected to meet the initial objective of finding the best technique to use which can support a spatial search among the various HE techniques studied. According to F. Fakhfakh et al, setting up an SLR requires developing a review protocol in terms of the review objective, protocol search questions, search strategy and inclusion and exclusion criteria [42].

Protocol Search Questions: We examined the reviewed publications using the following Protocol Search Questions (PSQs):

PSQ1. Which and how many HE techniques are we going to base our study on?

PSQ2. What kind of properties do these HE techniques possess?

PSQ3. Which among those properties can be analysed in comparison with the other HE techniques we have chosen to study?

PSQ4. What factors influence a spatial search?

PSQ5. Which technique among the techniques we have studied would be the best suited in conducting a spatial search to ensure the anonymity of data and the data provider on either one side of the equation, that is, the client side or the server side or on both sides?

PSQ6. What are some of the advantages of this technique that stand out among the rest of the techniques?

Search Strategy: The incorrect identification of the papers under examination could result in erroneous and inconsistent conclusions. Therefore, understanding the key terms is essential to choosing the best resources. The terms "Homomorphic Encryption Techniques," "Types of Homomorphic Encryption," "Factors that Influence a Spatial Search," "Properties of Paillier Encryption Technique," "Properties of Benaloh Encryption," "Properties of RSA," "Advantages of FHE," "Homomorphic Encryption Algorithms," etc. were used in our thorough search.

Inclusion and Exclusion Criteria: The search results were filtered to select only the papers published in the English language from 2013 to 2019. The results were further filtered so select only papers published in journals and conferences. Our search base was relevant scientific sources such as Science Direct, ResearchGate, Springer, Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE).

2) Stage 2: Performing the Review: 200 papers, in total, were found and added to the INITIAL LIST. We then eliminated duplicate publications or papers unrelated to the research questions in order to choose the pertinent studies. The remaining articles were then graded using some quality standards, including writing style, clarity of results, viability of solution, and clarity of study purpose. The 56 papers that met our exclusion criteria were kept in the FINAL LIST. The information found, obtained, and discussed in this publication were taken from the articles on the FINAL LIST.

3) Stage 3: Reporting the Review Findings: In this stage, we used the five protocol search questions that we have already presented to assess each solution under the subheading "Protocol Search Questions." Based on the solution collected, the first objective of our research was achieved by us identifying the best HE technique that can support a spatial search. Some of the comparisons we came up with can be seen later on in Table I and Table II in section IV which follows this chapter.

B. Developing a Protocol

To achieve objective 2 which was to do with developing a protocol for distributed spatial searching based on the identified Homomorphic Encryption technique, a protocol using the Paillier cryptosystem and distributed algorithm based on the Ring Algorithm (distributed ring algorithm principles) was designed. The protocol developed used the Paillier Encryption Algorithm which was imported as a free library online. The Paillier Encryption was used on both ends of the platform, that is, on the client side and also on the server side. On the client end, the researcher's attribute value which is a random value was encrypted before it was broadcast to the next machine which represents a data custodian. Encrypted also was the attribute value for the data custodian. The one thing in common was that both ends were provided with the public key but their only difference was that the server side where the data custodian lies didn't have a private key while the client side where the researcher or searcher sends requests from both the public and private keys.

C. Developing a proof-of-Concept Prototype

To achieve objective 3 which was to do with developing a proof-of-concept prototype using the proposed protocol, a prototype implemented as a distributed application was written in Java using the proposed protocol. Note that among the characteristics of a Distributed System is one that has to do with it presenting a single system image in which the entire network is visualised as a computer. This is true because a Distributed System deals with two aspects namely; the hardware and the software. While the machines connected in a distributed system are autonomous in terms of hardware, users of the software have the sense that they are interacting with a single system. Therefore, we were able to work with two computers and yet established several participating nodes in the network. The other key principle in our discussion is that a computing element, which we will generally refer to as a node, can be either a hardware device or a software process [15]. A second element is that users (be they people or applications) believe they are dealing with a single system. So, it was several nodes in terms of software processes that were generated by the use of two computers that formed a ring and not necessarily the two computers. This discussion helps to deal with the issue of distinguishing whether this was based on a peer-to-peer topology or a Distributed System. Yet in reality, it was based on the latter. While two computers cannot form a Ring in a peer-to-peer topology, two computers can form a Ring in a Distributed System based on the software processes the two computers can generate which eventually appear as a single system.

Three experiments were conducted in Java using two computers that were acting as two ends, that is, the client end and the server end. The first experiment involved doing the spatial search without the proposed protocol. In this one, the Java program was executed without having to send it to the next node or computer.

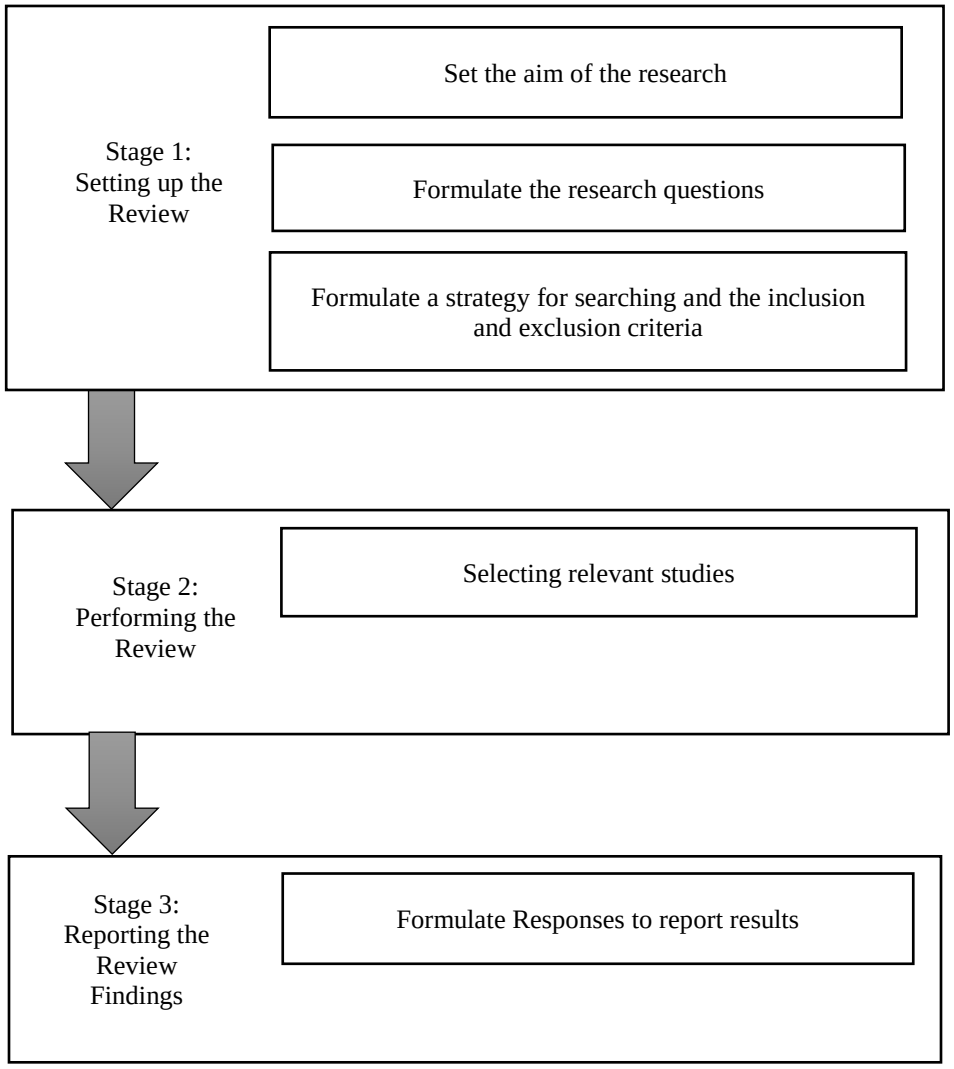

Figure 2: Protocol for Systematic literature review process [42]

The second one involved conducting the spatial search using the protocol with Paillier Encryption. In this one, using a Java program, a request was sent from one node to another where the random value for the client end is encrypted and the attribute value for the server side is also encrypted so that the server side sends back an encrypted value to the client side to demonstrate what happens in a Ring algorithm. Recall that one of the key principles in our discussion under Distributed Systems is that a computing element, which we will generally refer to as a node, can be either a hardware device or a software process. As a result, it was several nodes in terms of software processes that were generated by the use of two computers that formed a ring and not necessarily the two computers. The application during the implementation was coded in such a way that it could generate an n number of nodes to participate in one Distributed System. Note that to discover the next node to send the encrypted value to, each node participating in the distributed computation checks its list of participating nodes to find the next active node.

The third option had to do with running it with a protocol without Paillier encryption at all. In this one, we ensured that both ends didn't use the Paillier Encryption Algorithm at all. The researcher sent a value as a request that was not encrypted and vice versa the recipient could neither encrypt its value nor send back an encrypted sum.

IV. RESULTS

A. *Cryptosystem Selection*

After going through the literature on various HE techniques, an analysis of the factors that influence a spatial search was done through a literature review. Here are some of the factors that were identified.

The most popular class of such services is k-nearest neighbor (kNN) queries where users search for geographical points of interest, for example, restaurants and hospitals, and the corresponding directions and travel times to these locations. Accordingly, numerous algorithms have been developed to efficiently compute the distance and route between objects in large road networks [43].

In the query processing of spatial-keyword search, indexing techniques for both text and geographic data are used [44].

In a GeoSN, a variety of spatial items are marked on the map and tagged with user-generated tags, such as amenities. Users of GeoSN can share information about their location and activity as well as search for fascinating geographical amenities. More importantly, users with similar interests can collaborate to plan social activities like going out to eat and shop or going on a bike ride. Coordinating such plans requires pinpointing a collection of spatial objects of amenities that may best meet the needs of the users [44].

Another factor that influences a spatial search is the scalability of the dataset with regards to size. J. Zhong, X. Meng, X. Zhou, and D. Liu [44] carried out a series of tests to assess the scalability of three algorithms by altering the number of spatial objects. Their main goal was to simulate real geosocial networking in which the number of spatial objects and tags continuously increases.

A summary of this discussion is presented in Table I.

TABLE I. FACTORS THAT INFLUENCE A SPATIAL SEARCH

S/N	FACTORS THAT INFLUENCE A SPATIAL SEARCH
1.	Point Data (Longitude/Latitude)
2.	Geocoding-When there are no coordinates, the search can still be done by the ggmap package
3.	Search Algorithms -such as the kNN (k-nearest neighbor)
4.	Time Dependent-how fast and how slow after running the query
5.	Query Length-Handling queries of various length
6.	Scalability

The Paillier cryptosystem was selected for distributed spatial search for the following reasons:

a) It has a smaller expansion rate and lower cost of encryption and decryption than FHE and better security than RSA, El Gamal, Goldwasser-Micali and Benaloh. The desirable lower bound of this expansion rate is preferably four in order to keep the system safe and secure. The expansion factor has been decreased in improved designs to boost efficiency [4]. For instance, the Paillier cryptosystem enabled the encryption of several bits during a single calculation with a better expansion rate of two [4], which allowed for efficient decoding.

b) It is computationally cheaper and can be used in practice.

c) Paillier's scheme is the most efficient among currently known additively homomorphic schemes, that is. it requires simple operations in the encryption, decryption and addition procedures and so achieves high performance.

d) Even without having access to the private key, calculations can be made on the encrypted data to decipher the original message.

e) Analysis of encrypted data can be done with little to no danger of disclosing private information [26].

f) It possesses a self-blinding property which property allows mapping a plaintext into possibly many different ciphertexts and the same plaintexts cannot be recognized from their ciphertexts [45].

g) Since Paillier encryption is probabilistic, the encrypted files on the different peers are not linkable to each other for anyone not knowing the private decryption key [46].

h) Paillier Encryption is probabilistic in nature. Therefore, many encryptions of the same message will result in various ciphertexts, making it difficult for even a knowledgeable intruder to compare encrypted messages in order to determine the original message that was encrypted [26].

The discussion presented is high level. Details can be found in the literature cited. A summary of this discussion is presented in Table II.

Here is a brief account on the meaning of the subheading, Security Assumption, Computational Cost, Probabilistic Nature, Expansion Factor and Encryption/Decryption Costs as identified in Table II below. The word "security assumption" is used in Table II to describe how the most viable homomorphic encryption algorithms are secure. The security of these schemes is based on Ring-Learning With Errors (RLWE). Encryption techniques that use RLWE work under the premise that if the encryption method can be efficiently overcome, so can the RLWE problem. High-dimensional lattices are the subject of the challenging mathematics in RLWE. For example, Paillier Encryption is secure under the assumption of the Decisional Composite Residuosity Assumption (DCRA) [4]. DRCA states that for given integers N ∈ Z and x ∈ $Z_{n^2}^*$, it is "hard" to decide whether there exists y ∈ $Z_{n^2}^*$ such that x ≡ y^n mod N. Here,

the notation Z_n^* denotes the set of integers modulo N for all N ∈ N [47]. We are certain that these systems are at least as safe as any standardized encryption scheme because of a lengthy line of peer-reviewed research that demonstrate the hardness of the RLWE problem.

Computation cost is the total time taken for the algorithm to run which can be measured in milliseconds, seconds, minutes, or even hours. X. Wang, T. Luo and J. Li write that computational cost is the amount of time an algorithm takes to complete all additions and multiplications [48]. Computational cost associated with the features can be measured in computing power, memory occupation and Central Processing Unit (CPU) time. Commonly, a more computationally expensive feature usually provides more discrimination power in classifying ambiguous cases than a cheap feature.

An encryption procedure is a set of steps of converting a message from a readable form into an unreadable form. In other words, it is a process used to convert plaintext into ciphertext taking into account that without a secret key, no unauthorized users can access the original message [49]. Therefore, encryption cost is the time taken for the algorithm to complete the encryption procedure. Similarly, a decryption procedure is a process of taking encrypted or encoded text and then, converting it back into text that you can read and understand. Basically, decryption is the inverse process of encryption. Therefore, decryption cost is the time taken for the HE algorithms in our context to complete the decryption procedure. However, it must be noted that the difference between computational cost and encryption/decryption cost is that computational cost involves the whole length of time the algorithm runs from step A to step Z while encryption/decryption involves the total time the algorithm takes to finish the encryption or decryption process alone respectively, either of which may be a subset of the whole computational process (computational cost).

The probabilistic encryption nature is a property of encryption in which a given plaintext and encryption key will result into a different ciphertext for each encryption because the algorithm uses pseudorandom number generators, for example, Paillier Encryption is probabilistic. The majority of famous cryptosystems are deterministic. Deterministic systems essentially mean that for a given plaintext and fixed encryption key, the same ciphertext will always be produced. However, there can be some security issues as a result. The RSA system is a nice illustration of this idea.

J. Sen defined the expansion factor as the bit-to-bit ratio between the length of the ciphertext and the corresponding plaintext [14]. The value of this parameter plays a crucial role in deciding how a probabilistic encryption system balances security and efficiency. An effective probabilistic encryption system with a value of expansion less than two has been proposed in Paillier's scheme.

In addition to this, Table II shows that the security standard of Paillier is relatively higher than that of RSA, El Gamal, Goldwasser-Micali and Benaloh because it has higher encryption and decryption costs than these encryption algorithms [50]. However, its security standard is relatively lower than FHE since its encryption and decryption costs are lower than FHE. The higher the encryption and decryption cost the better the security since it is hard and takes more time to break the algorithm. Therefore, computational costs, encryption and decryption costs are thus key to the selection of a good algorithm to use. FHE may have better encryption and decryption costs in terms of security strength but it has higher computational costs due to complex mathematical computations created by its ability to perform both multiplications and additions. Hence, Paillier Encryption was preferred in the choice of the best algorithm to use in the designing of our protocol for a spatial search for it has a higher security standard than RSA, El Gamal, Goldwasser-Micali and Benaloh and also a better computational cost than FHE.

TABLE II. SECURITY ASSUMPTION OF THE HE SCHEMES AND COMPARISONS ON PROPERTIES.

			PROPERTIES OF THE SCHEMES				
HE SCHEME	**HOMOMORPHIC NATURE**	**SECURITY ASSUMPTION**	***COMPUTATIONAL COST***	***PROBABILISTIC NATURE***	***EXPANSION FACTOR***	***ENCRYPTION COSTS***	***DECRYPTION COSTS***
RSA	Multiplicative	Integer factorization problem [49].	High [51]	Deterministic	3	Low	Low
El Gamal	Multiplicative	Diffi-Hellman problem [11]	Low	√	3	Low [50]	Low [50]
Goldwasser-Micali	Additive	Quadratic residuosity problem [16]	Low	√	3	Low	High (heavier) [14]
Benaloh	Additive	Higher residuosity problem [17]	Low	√	3	Low	Complex
FHE	Additive & Multiplicative	Sparse Subset Sum (SSSP) assumption [14]	Higher [18]	√	Complex or >3	Complex/Higher	More complex (Higher)
Paillier	Additive	Decisional Composite Residuosity Assumption (DCRA) [11]	Low	√	2 [14]	High [50]	High

B. Sample Application

To achieve objective 2 which was to do with developing a protocol for distributed spatial searching based on the identified Homomorphic Encryption technique, we had to illustrate how our platform works by a demonstration of our platform using a simple example of computing the mean. The example uses four data custodians namely P_1, P_2, P_3 and P_4, but the approach applies in general when there are more than two P_i.

1. The client end has a searcher of data or a researcher.
2. The researcher or person doing the spatial search issues an online request for particular data by sending an encrypted random number which can be symbolic for a particular search request. He also sends a public key to all the locations identified, i.e. P_1, P_2, P_3 and P_4 when you consider the four distributed points as per the four locations. Hence the value of n = 4.
3. The request is then sent as an encrypted random value to the first point P_1.
4. P_1 receives the encrypted value of the request, i.e. E(R).
5. P_1 encrypts its attribute value and sends $E(R) \otimes E(P_1)$ to P_2.
6. P_2 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2)$ to P_3.
7. P_3 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2) \otimes E(P_3)$ to P_4.
8. P_4 encrypts its attribute value and sends $E(R) \otimes E(P_1) \otimes E(P_2) \otimes E(P_3) \otimes E(P_4s)$ to R.
9. The final stage is the subtraction of the random number R from the number obtained after decrypting the received encrypted number to retrieve the total.

Finally, by dividing by n the researcher retrieves the mean value of the selected attribute. This whole process is illustrated in Fig 3.

To discover the next node to send the encrypted value to, each node participating in the distributed computation checks its list of participating nodes to find the next active node. Extensive detail about this mechanism can be found in literature on ring algorithms in distributed systems.

C. Proof of Concept Prototype

Furthermore, a prototype implemented as a distributed application was written in Java using the proposed protocol. We demonstrated by using the BigInteger values in Java how large mathematical computations can be carried out to transfer encrypted values, through a distributed system, that is sent as requests without decrypting them since we ensured that both ends didn't have the secret key. The fact that the Paillier Cryptosystem uses modulo arithmetic (mathematical calculations involving functions that return the remainder) and BigInteger values in Java make it complicated and impossible for any third party or intruder to decrypt and obtain the mean an of attribute whose value needs to be kept confidential.

Table III. shows the analysis of performance when we consider three scenarios as we did the research, that is, the first one is doing the spatial search without the proposed protocol, the second one is doing the spatial search using the protocol with Paillier Encryption and the third option of running it with a protocol without Paillier encryption at all. Table III. shows the average run time in milliseconds calculated after executing each experiment for ten (10) different runs. From the results, it can be concluded that the encryption and protocol result in an increased running time. It can, however, be argued that they collectively introduce an overhead of 39.7% in processing time. The bigger expense comes from the network communication overhead.

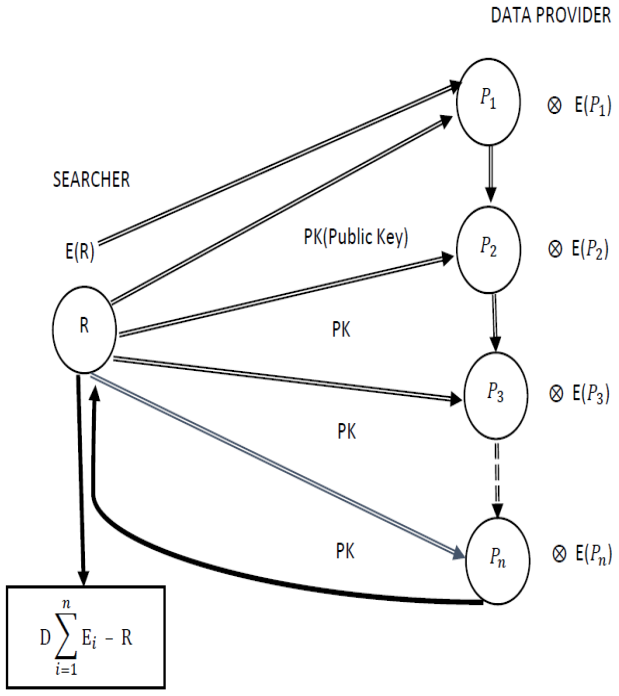

Figure 3: Proposed protocol uses a distributed ring algorithm

conducting a spatial search with protocol without Paillier Encryption was done to calculate the running time for the protocol without the encryption algorithm applied to it.

TABLE III: ANALYSIS OF PERFORMANCE DURING A SPATIAL SEARCH.

DESCRIPTION	AVERAGE RUN (EXECUTION TIME)-JAVA PROGRAM	ANONYMITY/SECURITY
i. Spatial search without protocol but without Paillier Encryption	702 milliseconds	NIL
ii. Spatial search with protocol and Paillier Encryption	981 milliseconds	Guaranteed
iii. Spatial search with protocol but without Paillier Encryption	767 milliseconds	NIL

D. Validity Test

In addition to these results in general, a validity test was done by doing experiments to calculate an average using anonymous data from a varying number of participating nodes. Experiments carried out involved running the proposed protocol in Java using n number of processes where n > 1. Data passed between nodes was encrypted in nature. It was only the result that was decrypted. More than 10 experiments were done. Experiment No. 1 had 2 participating nodes, experiment no. 2 had 3 participating nodes and experiment No. 3 had 4 nodes, experiment No. 4 had 5 nodes, etc. As explained in the methodology, it was several nodes in terms of software processes that were generated by the use of two computers that formed a ring and not necessarily the two computers. For each experiment, it was run ten times to obtain the average run time in milliseconds for the time taken to run the Java application as shown in Table IV, Table V, Table VI, etc. below. Here are the tables for Experiment No1., Experiment No.2, Experiment No. 3., respectively;

TABLE IV: RESULTS FOR EXPERIMENT NO.1

S/N	EXPT 1: INPUT VALUES	RESULTS FOR 10 RUNS (RUN-TIME IN MILLISECONDS)	NUMBER OF PARTICIPATING NODES
1	10	681	2
2	20	679	
3		674	
4		776	
5		663	
6		740	
7		717	
8		696	
9		697	
10		700	
AVERAGE	**15**	**702**	

TABLE V: RESULTS FOR EXPERIMENT NO.2

S/N	EXPT 2: INPUT VALUES	RESULTS FOR 10 RUNS (RUN-TIME IN MILLISECONDS)	NUMBER OF PARTICIPATING NODES
1	10	794	3
2	20	802	
3	30	799	
4		879	
5		820	
6		819	
7		805	
8		791	
9		810	
10		804	
AVERAGE	**20**	**812**	

TABLE VI: RESULTS FOR EXPERIMENT NO.3

S/N	EXPT 3: INPUT VALUES	RESULTS FOR 10 RUNS (RUN-TIME IN MILLISECONDS)	NUMBER OF PARTICIPATING NODES
1	10	906	4
2	20	826	
3	30	888	
4	40	863	
5		978	
6		962	
7		876	
8		886	
9		911	
10		980	
AVERAGE	**25**	**908**	

The results of the above tables and many more tables were grouped to come up with the results to prove the validity of the proposed protocol. These results were then plotted in a graph as shown in Figure 4 below.

The validity test results proved that the proposed protocol is scalable, i.e. it can be used, for example, for 1,000, 10,000, 100,000 nodes, etc. The results of the validity test are displayed graphically in Figure 4 below. Figure 4 below shows results for ten different experiments that were conducted for n nodes where $2 \leq n \leq 11$.

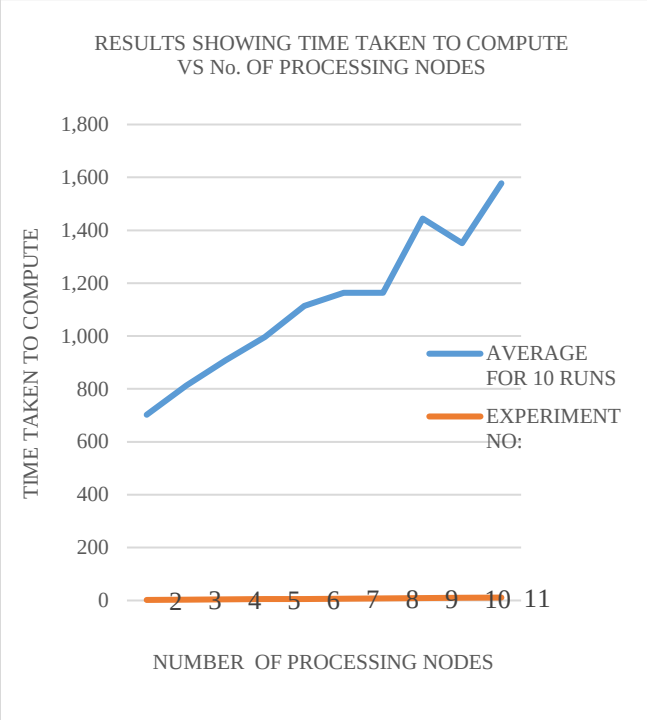

Figure 4: Showing results compiled when experiments are conducted on n number of machines where n > 1

V. Discussion

Digitalisation is attracting a lot of attention and is used in a lot of fields such as agriculture [53], finance and so on. Many scholars have considered different aspects of security in information systems [54],[54]. In this paper we however focus on data in transit. After analysing the encryption schemes using existing literature on Homomorphic Encryption, Paillier Encryption Scheme became the focus for the three experiments. In the first experiment, the spatial search was conducted without using the protocol and encryption. The average execution time after running the program ten times is shown in Table III. In the second experiment, the spatial search was conducted by using the protocol with Paillier Encryption Scheme. After running the program ten times, the average execution time is shown in Table III. In the third experiment, the spatial search was conducted using the protocol but without Paillier Encryption Scheme. The average result after running the program ten times is displayed in Table III. Furthermore, a validity test was done by doing experiments to calculate an average using anonymous data from a varying number of participating nodes. During the validity test, more than ten experiments were conducted. Each experiment done had n > 1 number of participating nodes with each node being a process that represents one system image. In addition, each experiment was run ten times so that we could calculate the average run time in milliseconds as displayed in Table IV, Table V, Table VI, etc. It was these processes that formed a ring. This test proved that the proposed protocol is scalable, i.e. it can be used for any number of nodes that are in communication with each other through the network where n > 1 and n represents the number of nodes.

Note that experiments were not conducted with other schemes here because the focus was on using the Paillier Encryption scheme after selecting Paillier when the other encryption schemes were analysed during the literature review of Homomorphic Encryption putting into consideration desirable qualities of a spatial search. It was during the literature review that the comparisons in Table II were made and Paillier was chosen as the best scheme to be used in our research.

The key principle in Homomorphic Encryption is we can achieve addition and subtraction of the data using Additive Homomorphism and we can also get multiplication and division of the data using Multiplicative Homomorphism. Paillier Cryptosystem was identified as the best for supporting a spatial search because it is computationally cheaper to be used in practice. Paillier's scheme is the most efficient among currently known additively homomorphic schemes, i.e. it requires simple operations in the encryption, decryption and addition procedures and hence achieves high performance. Another observation is that calculations on the encrypted data can be performed without necessarily reconstructing the original message and without having access to the private key.

VI. Conclusion

The risk of lack of anonymity and confidentiality is what a client or a data provider may experience. A key limitation is that both the user and the data provider focus on either getting information or providing data without being careful about their anonymity respectively. Therefore, protecting both the searcher and the data provider side is of greater importance for spatial searches.

Homomorphic Encryption, particularly Paillier Homomorphic Encryption, supports a spatial search by hiding identity of searcher and data provider participating in a spatial search while allowing analyses to be conducted on encrypted data in the encrypted space. Statistical data such as data to be aggregated in secure geocoding, data aggregation in cancer registries, and election results were key examples in our research where analysis of data can be used to find total sums, averages, differences, and so on. In secure geocoding, data aggregation such as the computation of the total number of cases based of geographical boundaries can be done based on the physical addresses provided. Other complex calculations like finding an average number of cases per geographic boundary can also be performed. Analyses of large data sets such as election results can be achieved while ensuring the anonymity of the one casting a vote and protecting the numbers of the votes cast. This can be realised when such a scheme as a Paillier cryptosystem is used to encrypt both the running ends of the searcher and the data provider. The Paillier cryptosystem is cheap and computationally capable and viable compared to other cryptosystems. Even though the proposed solution introduces

a 39% overhead this is outweighed by the benefits of the proposed approach.

FUTURE WORK

A proof-of-concept prototype was developed that can be implemented in most of the programming languages that can make use of Paillier Homomorphic Encryption. There is a need to complete the work and build a full-fledged system that can be used by data providers and online researchers as they conduct spatial searches, and search for or exchange data online.

There is also a need to find an alternative for the operation limitation of the Paillier cryptosystem. Paillier only allows for one computational operation i.e. either addition or multiplication and not both addition and multiplication. It cannot perform both [52].

ACKNOWLEDGMENT

Many thanks to the University of Zambia, Department of Computer Science for providing us with an environment to carry out this study through their technical support and the necessary advice and expertise.

REFERENCES

[1] M. Nyirenda, H. Arimura and K. Ito, "Relaxing the data access bottleneck of geographic big-data analytics applications using distributed quad trees," 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, 2016, pp. 189-195.

[2] M. Nyirenda and D. Zulu, "Speeding up construction of distributed quadtrees for big-data analytics applications using dilated integers and hashmaps," 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, 2017, pp. 1-6.

[3] N. Hamlin, "(2017) Number in Mathematical Cryptography," Open Journal of Discrete Mathematics, 7, 13-31, pp. 1-19, Jan. 2017. [Online]. Available. Accessed on: Nov 12, 2019. http://dx.doi.org/10.4236/ojdm.2017.71003

[4] M. Alkharji, H. Liu, M. A. Hammoshi, "A Comprehensive Study of Fully Homomorphic Encryption Schemes," International Journal of Advancements in Computing Technology (IJACT) Volume10, Number1, pp.1-24, Mar. 2018.

[5] J. Ashok, K. N. Dheeraj, C. Subhedar and R. Tiwari, "Homomorphic Encryption over Databases," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8, June 2019.

[6] T. Zhao, Q. Ran, L. Yuan, Y. Chi and J. Ma, "Key Distribution and Changing Key Cryptosystem Based on Phase Retrieval Algorithm and RSA Public-Key Algorithm," Mathematical Problems in Engineering, pp.1-13, Jun 2015. [Online]. Available. Accessed on: Oct 8, 2019.

[7] N. F. H. Al Saffar, "Steganography Algorithm Based RSA Cryptosystem," Journal of Engineering and Applied Sciences, pp. 1-5, April 2019. [Online]. Available. Accessed on: Jan 20, 2021. DOI: 10.36478/jeasci.2019.2240.2243

[8] K. Mallaiah and S. Ramachandram, "Applicability of Homomorphic Encryption and CryptDB in Social and Business Applications: Securing Data Stored on the Third Party Servers while Processing through Applications," International Journal of Computer Applications (0975 – 8887) Volume 100– No.1, p1-15, August 2014.

[9] X. Ye, C. Liu and D. Ga, "Weakness of RSA Cryptosystem Characteristic," AIP Conference Proceedings 2040, 130005 (2018), pp. 1-9, Dec 2018. [Online]. Available. Accessed on: Jan 20, 2021. doi: 10.1063/1.5079187.

[10] G.A.V.R. Rao, P.V. Lakshmi and N. Ravi Shankar, "RSA Public Key Cryptosystem using Modular Multiplication," International Journal of Computer Applications (0975 – 8887) Volume 80 – No5, p1-5, October 2013.

[11] M. Alkharji and H. Liu, "Homomorphic Encryption Algorithms and Schemes for Secure Computations in the Cloud," ICSCT 2016 - International Conference on Secure Computation and Technology, Virginia International University, Fairfax, VA, 2016. Available. Accessed on: Oct 11, 2019.

[12] M. Tebaa and S. EL Hajii, "Secure Cloud Computing through Homomorphic Encryption," International Journal of Advancements in Computing Technology(IJACT) Volume5, Number16, December 2013.

[13] R. Shruthi, P. Sumana and A. K. Koundinya, "Performance Analysis of Goldwasser-Micali Cryptosystem," International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2013.

[14] J. Sen, "Theory and Practice of Cryptography and Network Security Protocols and Technologies," Homomorphic Encryption-Theory and Application, Chapter1, pp. 2-34, July 2013. [Online]. Available. Feb 23, 2021. DOI: 10.5772/56687

[15] M. Steen and A. S. Tanenbaum, "A Brief Introduction to Distributed Systems," Computing 98. pp. 1-43, June 2016. [Online]. Available. Feb 18, 2021. DOI 10.1007/s00607-016-0508-7

[16] J. Bringer, H. Chabanne, M. Izabache'ne and D. Pointcheval, "An Application of the Goldwasser-Micali Cryptosystem to Biometric Authentication," Jun 2014. [Online]. Available. Accessed on: Oct 8, 2019.

[17] A. Acar, H. Aksu, A. S. Uluagac and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," pp.1-35, Oct. 2017. [Online]. Available. Accessed on: Oct 9, 2019.

[18] V. Biksham and D. Vasuma, "Homomorphic Encryption Techniques for securing Data in Cloud Computing: A Survey," International Journal of Computer Applications (0975 - 8887) Volume 160 - No.6, February 2017.

[19] B. K.Saraswat, R. Suryavanshi, and D.S.Yadav, "A Comparative Study of Checkpointing Algorithms For Distributed Systems," International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 1595-1603, pp. 1-11, Sep. 2019

[20] P. B. Soundarabai, J. Thriveni, K. R. Venugopal and L. M. Patnaik, "An Improved Leader Election Algorithm for Distributed Systems," International Journal of Next-Generation Networks (IJNGN) Vol.5, No.1, March 2013, pp. 1-9.

[21] M. Al-Refai, Y. Alraba'nah, M. Alauthman, A. Almomani, M. Al-Kasassbeh and M. Alweshah, "A Novel Leader Election Algorithm for Honeycomb Mesh Networks," Journal of Theoretical and Applied Information Technology 31st July 2019. Vol.97. No 14, pp. 1-14.

[22] S. Basu, "Token Ring Algorithm to Achieve Mutual Exclusion in Distributed System – A Centralized Approach," IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814, pp. 1-6.

[23] A. Dadlani, A. Khonsari, M. Effatparvar, N. Yazdani and M. Effatparvar, "Improved Algorithms for Leader Election in Distributed Systems," pp. 1-6, 2010. DOI: 10.1109/ICCET.2010.5485357 · Source: IEEE Xplore.

[24] S. Naseera, "A Distributed Ring Algorithm for Coordinator Election in Distributed Systems," ICTACT Journal On Communication Technology, September 2016, Volume: 07, Issue: 03, ISSN: 2229-6948(ONLINE) DOI: 10.21917/ijct.2016.0197.

[25] H. Shaheen., Distributed Systems. Hyderabad: BONFRING Intellectual Integrity, 2019.

[26] G. M. Jacquez et al., "Geospatial Cryptography: Enabling researches to access private, spatially referenced human subjects data for cancer control and prevention, pp. 1-26, Jul. 2017. [Online]. Available. Accessed on: Oct 8, 2019. www.https://ncbi.nlm.nih.gov/pmc/articles/pmc5659297/

[27] M. Birkin, "Spatial data analytics of mobility with consumer data," Journal of Transport Geography 76 (2019) 245–253, pp. 1-9, Apr. 2019.

[28] A. Tondwalker and P. V. Jan, "Secure localisation of wireless devices with application to sensor networks using steganography," International

Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA, pp. 1-7.
[29] J. Ajayakumar and K. Ghazinour, "I am at home: Spatial Privacy Concerns with Social Media Check-ins," The 4th International Symposium on Emerging Information, Communication and Networks (EICN-2017), Procedia Computer Science 113 (2017) 551–558.
[30] M. Kiedrowicz, "Methodology of Ensuring the Security of GIS Spatial Data," 26th Geographic Information Systems Conference and Exhibition "GIS ODYSSEY 2019" Conference proceedings, pp. 1-13, Nov. 2019.
[31] R. M. M. Pradeep and N. T. S. Wijesekera, "Development of Security Stamp for Desktop Spatial Data Modification in Unrestricted Access Platform," pp. 1-8, Sept. 2015.
[32] X. Ma, "Spatial Data," pp. 1-7, Jan. 2017. [Online]. Available: DOI: 10.1007/978-3-319-32001-4_192-1.
[33] R. Guo, B. Qin, Y. Wu, R. Liu, H. Chen and C. Li, "MixGeo: Efficient Secure Range Queries on Encrypted Dense Spatial Data in the Cloud," pp. 1-11, Jun. 2019.
[34] D. Chen, P. Zhang, C. Hu, H. Wang, S. Wu and N. Xing, "PAPERS: Private and Precise Range Search for Location Based Services," IEEE ICC 2015 - Communication and Information Systems Security Symposium, pp. 1-6, 2015.
[35] A. Talha, I. Kamel and Z. A. Aghbari, "Enhancing Confidentiality and Privacy of Outsourced Spatial Data," pp. 1-8, Nov. 2015.
[36] B. R. Pushpa, "Enhancing Data Security by Adapting Network Security and Cryptographic Paradigms," (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 1319-1321, pp. 1-4, Aug. 2019.
[37] M. Modak and R. Shaikh, "Privacy Preserving Distributed Association Rule Hiding Using Concept Hierarchy," 7th International Conference on Communication, Computing and Virtualization 2016- Procedia Computer Science 79 (2016) 993 – 1000, Mar 2016. [Online]. Available. Accessed on: Aug 13, 2019.
[38] X. Liao and C. Shu, "Reversible data hiding in encrypted images based on absolute mean difference of multiple neighboring pixels," J. Vis. Commun. Image R. 28 (2015) 21–27. [Online]. Available. Accessed on: Aug 13, 2019.
[39] S. Vennila and Priyadarshini. J, "Scalable Privacy Preservation in Big Data a Survey," Procedia Computer Science 50 (2015) 369 – 373, Apr 2015. [Online]. Available. Accessed on: Aug 16, 2019.
[40] B. Kitchenham, "Procedures for performing Systematic Reviews," Keele, UK, Keele University 2004; 33 (2004): 1-26.
[41] K. Abawi, "Systematic Review-From Research to Practice: Training in Sexual and Reproductive Health Research 2015," pp. 1-12. [Online]. Available: https://www.gfmer.ch/SRH-Course-2015/research-methodology/pdf/Systematic-review-Abawi-2015.pdf
[42] F. Fakhfakh, M. Tounsi, M. Mosbah and A. H. Kacem, "Formal Verification Approaches for Distributed Algorithms: A Systematic Literature Review," International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3-5 September 2018, Belgrade, Serbia. Procedia Computer Science 126 (2018) 1551–1560.
[43] U. Demiryurek, F. Banaei-Kashani and C. Shahabi (2010), "Efficient K-Nearest Neighbor Search in Time-Dependent Spatial Networks," International Conference on Database and Expert Systems Applications DEXA 2010: Database and Expert Systems Applications, pp 432-449. https://doi.org/10.1007/978-3-642-15364-8_36
[44] J. Zhong, X. Meng, X. Zhou and D. Liu, "Co-spatial Searcher: Efficient Tag-Based Collaborative Spatial Search on Geo-Social Network," DASFAA 2012, Part I, LNCS 7238, pp. 560–575, 2012. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.9314&rep=rep1&type=pdf
[45] I. San, N. At, I. Yakut, H. Polat, "Efficient Paillier Cryptoprocessor for privacy-preserving data mining," Feb. 2016, www.https://doi.org/10.1002/sec.1442
[46] M. Nassar, A. Erradi, Q. M. Malluhi, "Paillier's Encryption: Implementation and Cloud Applications," pp. 1-6, Oct 2015. [Online]. Available. Accessed on: Feb 9, 2021. DOI: 10.1109/ARCSE.2015.7338149
[47] F. Farokhi, I. Shames and K. H. Johansson, "Private and Secure Coordination of Match-Making for Heavy-Duty Vehicle Platooning," IFAC PapersOnLine 50-1 (2017) 7345–7350.
[48] X. Wang, T. Luo and J. Li, "A More Efficient Fully Homomorphic Encryption Scheme Based on GSW and DM Schemes," Security and Communication Networks, pp. 1-15, Dec 2018. [Online]. Available. Feb 23, 2021. https://doi.org/10.1155/2018/8706940
[49] G. A. AL-Rummana, G. N. Shende, "Homomorphic Encryption for Big Data Security: A Survey," International Journal of Computer Sciences and Engineering, pp. 1-10, Oct 2018. [Online]. Available. Feb 23, 2021. DOI: 10.26438/ijcse/v6i10.503511
[50] M. Zhao and Y. Geng, "Homomorphic Encryption Technology for Cloud Computing," 8th International Congress of Information and Communication Technology (ICICT-2019)- Procedia Computer Science 154 (2019) 73–83. [Online]. Available. Accessed on: Aug 16, 2019.
[51] H. Yu and Y. Kim, "New RSA Encryption Mechanism Using One-Time Encryption Keys and Unpredictable Bio-Signal for Wireless Communication Devices," Electronics 2020, pp. 1-10, Feb 2020. [Online]. Available. Feb 23, 2021. doi:10.3390/electronics9020246
[52] T. Oladunni and S. Sharma, "Homomorphic Encryption and Data Security in the Cloud," Proceedings of 28th International Conference on Software Engineering and Data Engineering EPiC Series in Computing Volume 64, 2019, pp. 129–138, Oct. 2019.
[53]Simukanga, A., Phiri, J., Nyirenda, M., & Kalumbilo-Kabemba, M. (2018). E-Governance Systems: A Case Study of the Development of a Small-Scale Farmer Database. *Zambia ICT Journal*, *2*(1), 7–15. https://doi.org/10.33260/zictjournal.v2i1.41
[54]K. Kamusweke, M. Nyirenda and M. Kabemba, "Data mining for fraud detection in large scale financial transactions", EasyChair, no. 1729, Oct. 2019, [online] Available: https://mail.easychair.org/publications/preprint_download/G5sK.
[55]Musonda, C., 2019. Security, Privacy and Integrity in Internet Of Things - A Review. ICTSZ Int. Conf. ICTs Lusaka, Zambia (12th -13th December 2018 146–152.