International Journal of Developments in Big Data and Analytics
Volume 1 No. 1, 2014, pp. 19—40

Big Data Analytics and the Epitome of fully Homomorphic Encryption Scheme for Cloud Computing Security

VICTOR O. WAZIRI, JOHN K. ALHASSAN
OLALERE MORUFU & IDRIS ISMAILA
Federal University of Technology, Minna, Niger State, Nigeria

ABSTRACT This paper studies the issues of Big Data Analytics (BDAs) or the Internet of Things (IOT) and its implementation in the Cloud Computing Environment with a full thrust consideration to the Fully Homomorphic Encryption Scheme (FHE). The FHE is a computational algorithm that allows the computations on encrypted data; and yet, when the outputs of the computations of the encrypted data are decrypted, they still conform to the operations of the original plaintext. The Bootstrapping of the Somewhat Homomorphic Encryption Scheme (SWHE) of the Craig Gentry's PhD thesis was studied; an algorithm that continuously reduces the inherently noisy ciphertext parameters of the Ideal Circuit at each computational iteration stage, and thus, gives more room for the less noisy ciphertext computation in the cloud based the application of addition and multiplication operations. The FHE implementation on Big Data Analytcs Security on the cloud computing was discovered to have boosted confidentiality, Integrity and Availability of the Big Data resources in the cloud computing environment. Most especially, the issue of privacy was considered achieved. This achievement of Bootstrapping algorithm for the FHE is now applied to compute ciphertexts with the private-key embedded in the ciphertext and upload in the cloud. Fully Homomorphic Encryption Scheme is able to decrypt and write new ciphertexts in the Cloud without the private-key being compromised. The paper concluded that the advent of Bootstrapping in FHE has accelerated the implementation BDAs as most security issues previously in existence based on the traditional FHE has been fully ameliorated. With this mitigation for effectual security based on the new advancement, the implementation of the Big Data Analytics has been adequately improved upon. The paper also examines the application of the FHE that will bootstrap cloud-based businesses and e-governance in the developing economies. The paper concludes that Big Business Associations such as the google.com and Amazon.com, amongst other enterprises; are benefitting on this BDAs drift. FHE can also be implemented in counting-voting and many more.

Keywords: Fully Homomorphism, Bootstrap, Data Analytics, Confidentiality, Integrity, Privacy

Introduction

Data are collected and analyzed to find patterns and associations (correlations) that may not be that initially may not be found or observed. These observed patterns and correlations can help in decision making to well-organized business enterprises, and or Government institutions or agencies. This process is called the Big Data Analytics. However, as we go on in the paper, real definitions would surface that give further insights for the nebulous issues of Big Data Analytics are in divergence in conceptions and applications. Let us think in big data; what is big data? There is no generic construe on the definition of Big Data Analytics; big we can start from Big Date inducement. Big Data be adduced from the following experts' documentaries as researched by Irwin, et el (2014).

i) From Wikipedia, the free encyclopedia, "A collection of datasets so *large* and *complex* that it becomes *difficult to process* using on-hand database management tools or traditional data processing applications."

ii) Teradata-Wiki, "Big Data *exceeds* the reach of commonly used hardware environments and software tools to *capture, message,* and *protect* it with a tolerable elapsed time for its populations."

iii) The Mickinsy Global Institute (2011) defined "Big Data as datasets whose size is beyond the ability of typical database software tools to *capture, store, message and analyse."*

From these propounded barrage of definitions, we see that Big Data refers to datasets that grow so *large* and *Complex* that it becomes so immensely difficult to *capture, store, message, share and visualize* within the current ambience and computational aggregations of modern utilities of architectural technologies.

The Features of Big Data Analytics

In furthering the building of the topological structure of the paper, the following definitions this significant phrase, "Big Data Analytics by Irwin, et al. (2013) may be considered expedience. The authors opined that Big Data Analytics is a process of *inspecting, cleaning, transforming, and modeling* big data with the central target of *discovering useful information, suggesting conclusions and supporting decision making*. They went a step forward by corroborating that it is a data mining with the following supporting objectives in analytics:

a) That include both data analysis (mining) and communication to guide decision making; and

b) It does not so much impart much concern with individual analyses or analysis steps, but with the *entire methodology* to achieve unified aim.

Big Data is characterized with seven features; but we discuss four here in line with postulates of Irwin, et al. (2014). These include; volume, velocity, variety and veracity.

Big Data is a term used to describe the collection of large and complex data sets that are difficult to process using on-hand database management tools or traditional data processing applications to meet of today's analytical aspirations. Big Data spans across seven dimensions which include volume, variety, volume, value, veracity, volatility and complexity:

Volume: The volume of data here is very huge and is generated from a lot of different devices. The size of the data is usually in terabytes and petabytes. All this data also needs to be encrypted for privacy protection. The volume stipulates that by the year a large dataset would have aggregated datasets (Big Data) in volumetric of about 40 teratabytes and in the next after five years (2020), the total datasets will exceed 5.2 terabyte per person

Velocity: This describes the real time attribute found in some of the data sets for example streaming data. The result that misses the appropriate time is usually of little value. The velocity of data movement between storing and analyzing will be to short. Data sharing per individuals would amount from 2.5 GB (billion) per day in 20015 and 500TB (terabyte) per day in 2015. However, Mike (2013) gives a more insightful estimates previously unknown questions, defining new insights into velocity of Big Data analytics and reducing the time between when an event happens somewhere in the world and responds or reacts to that events. In his own declaration, he opines that Big Data analytics is a rapidly emerging universe of newer technologies that has drastically reduced data processing cycle time, that makes it possible to explore and experiment with data in ways that would not have been practical or even possible a few years ago.

Despite the availability of new tools and systems for handling massive amounts of data at incredible velocities, however, the real-time Big Data analytics gives more reliant promises of advanced data analytics that lies roundly beyond the realm of pure technology currently in use. Time and velocity in Big data analytics is not just about a process for storing petabytes or exabytes of data in a data warehouse as harangued by Michael, et al. (2013).

Variety: Big data consists of a variety of different types of data i.e. structured, unstructured and semi-structured data. The data may be in the form of blogs, videos, pictures, audio files, and location information etc.

Value: This refers to the complex, advanced, predictive, business analysis and insights associated with the large data sets.

Veracity: This deals with uncertain or imprecise data. It refers to the noise, biases and abnormality in data. This is where we find out if the data that is being stored and mined is meaningful to the problem being analyzed. Variety demands that different datasets is now available commonly than the previous years. These variety in abundance signifies the needs for more accommodative storage spaces and faster analyzing the data. Veracity connotes that the authenticity of datasets and analysis from various information technologies should be standardized to pave room for legitimacy and uniformity of products worthiness

Volatility: Big Data volatility refers to how long the data is going to be valid and how long it should be stored.

Complexity: A complex dynamic relationship often exists in Big data. The change of one data might result in the change of more than one set of data triggering a rippling effect

So far, we have talked a little about the phrase or term "Big Data" in this section. Big Data, was however, idealized and traceable to John Mashey (1997). This was further supported by Weiss and Indurkhya (1998) who advocated the big data concept as a data mining process. They suggested practical difficulties in its applications. Today, Big Data Analytics is still being conceived in this context of data mining. Mohammed et al (2012) look at Big Data in the Context Data Mining. By Mohammed et al., (ibid), Big Data analytics is a process that discovers patterns and interesting trends from large collections of data. Therefore, one can see through their window minds and contrive the term big data analytics as data mining that transcends into the extensive outlook of a very large datasets that are measured in terabytes and petabytes; and have faster analytical tools if their Big Data Analytics perceptive definition should weigh squarely within the armpit definition of Michael, et al. (ibid).

Motivation

The increase of technological advancement and the advent of vast utility of the Internet in businesses and governance, many issues in secrecy and privacy have come to the frontage in cloud computing environment. Thus in this paper, the paramount aim is to protect the organization databases from unnecessary disclosure of the data information. The Big Data is normally stored and accessed in the cloud computing environment. Conceived security model strives in the building bridge of secrecy and privacy as propounded by Gentry (2009) encryption scheme "Fully Homomorphic Encryption (FHE) scheme; and Kantarcioglu and Clinton (2004) that provided an efficient solution for encrypted codes in the cloud environment using the "Somewhat Homomorphic Encyption (SHE) without decryption private key. Other public encryption scheme like the El Gamal (1985) and Rivest, et al, (RSA) SHE shall be presented first. The SHE is partial because it operates under one operation in the addition or multiplication operator. Gentry (ibid) made a breakthrough masterpiece that provided the first secure FHE scheme. FHE allows an individual without access to the decryption private key, to apply any Boolean circuit to the ciphertext and decrypting the ciphertext which could yield a correct plaintext.

In our train of thoughts of the modern advancement in business environment, the golden mine of knowledge has expanded new business enterprises and so are new techniques of gathering or generating business intelligence. Hence the use of Big Data Analytics for Information and Technology (IT) frontier has equally generated massive amounts of aggregated Big Data algorithms that have been developed to meet different inspirational needs of Big Data organizations and individuals. Algorithms are provided for allowing for the discovering of correlations (or inter-relations) between different data items in the universal database. For two or more associations, in the Big Data Analytics formations, and like all human interactive associations, the rules binding these associations through policies or protocols will need modification to accommodate privacy of each participant in the business enterprises. For want of space, the researchers could refer to Andras (2014), Cliffer (2011), Williams Pryor (2011) Michael and Company (2011), Mineli at al. (2013) and Randal, et al. (2010). These businesses implementation abound and are growing by the day in the Internet of Things (IOT) or Big Data Analytics) As we already know, Cloud Computing is now the drift that dominates the modern association businesses under the Big Data umbrella. In a nutshell, the robustness of privacy in the any associations is to prevent misuse Clinton, et al., (2002) and Lindel (2000).

At the moment, many business associations are leveraging cloud services due to their specified simplicity and cost effectiveness. There is, however, concerns over government inspection of all data due in security and terrorisms which they perceive that the service providers may breach by conniving assorted shades of insurgents. This is seen as act of enervating the right of the cloud providers to go fully cloud computing. Despite this enervation, and because of large accumulations of Big Data for analysis and with limited computational resources, most enterprises proffer to storing their data on the cloud. Doing so, nevertheless, has its adverse cost because of insufficient access controls are driving conversations about information security controls in the cloud. For instance, enterprise and software as a Service (SaaS) providers have particularly high interest in using cryptographic techniques for protecting data in the cloud.

Data Security in the Cloud

The risks embodiments in the cloud computing environment come in with a lot of unobservable risks to any sensitive data being stored. The most outstanding and conspicuously observable risks are conceived from the need to trust data protection to a third party cloud provider. Different servers in some environment could be in control or administered by different untrusted parties, and could have some vulnerability that are given to attacks by their cloud tenants, or malicious insiders or some external adversaries. It is in good face deduction that when data owners (clients) give up control of their data to the Cloud computing infrastructure, they should require the guarantees that their data remains protected. As provided and outlined in a Service Level Agreement (SLA); these SLA guarantees are legal though promises that may not be enforced when adversely exploited. Cryptography allows data owners to protect their sensitive data proactively instead of relying primarily on legal agreements through SLA that are difficult to monitor or enforce by appropriate law agencies. Most developing nations lack laws to enforce cybercrimes; and cyber security laws are not applicable to all nations. The laws attracting certain crimes in some country may be lacking in another country.

In view of the aforesaid outlook on SLA, it is necessary we overview some fundamental description of cryptography in the basic level of CIA (Confidentiality, Integrity and Availability). These three basic traditional cryptography import are succinctly articulated by Sophia, et al. (2014):

a) *Confidentiality:* Confidentiality is a profound process of disguising plaintext contents or all sight sensitive data (Computational input, output, and intermediate state) into gibberish scripts that remain secret from any potential adversarial or untrusted entities

b) *Integrity:* This process enables the authentication of the sensitive data after transmission by the service providers to the user or recipient**.** It ensures that upon verification, the data is no tampered with by any unauthorized person or modification of the data can be verified through Hash function technology. Verification should ensure that the outputs of the computation on the sensitive data correct (that is consistent with the input data).

c) *Availability:* In this context, Data owners (e.g. the output recipients described above) are assured access to their data legitimate data and compute resources as at when promptly demanded

The issue of availability is always straight away addressed in today's cloud environments through non-cryptographic means. Therefore, we are left with the issues of looking into confidentiality and integrity of the encrypted data in the cloud computing environment. The application of confidentiality and integrity depends on the cloud deployment, the service providers and the trust that is reposed between entities. They review categorically the above endorsed following the conceptual scenarios as in Sophia, et al., (2014).

i) *Untrusted Cloud:* In this encounter, the data owners do not have absolute trust on the cloud servers to maintain the confidentiality or integrity of data or computations outsourced to the cloud. As a result of this prevailing outcome on untrusted situation, the client-side protections are necessary to ensure that confidentiality and integrated are enshrined in the aftermath of an adversarial cloud. This practical setting scenario would commonly be associated correspondently to the public cloud development

ii) *Trusted Cloud:* This enticed second scenario is observable in government use cases, In such used cases, the cloud is deployed in an air-gapped environment and it is completely isolated from the outside networks and adversaries. Clients can put their data in the cloud with the assurance that it will remain confidential against outside adversaries. In spite of seemingly watertight secrecy of privacy, in some isolated environment, some of the nodes may become corrupted due to malware or insiders on the side of the service providers. These evidences observable instances, cannot insinuated any private data. This model may be scenario may corresponds to the private cloud deployment attack.

iii) *Semi-entrusted Cloud:* This scenario is infiltrated as corresponding to the hybrid deployment that mixes the public and private cloud deployment models. Hybridized because it is idealized as partially trusted, that it has a mixture of trusted and untrusted. Put in another permutated way, some parts of the cloud may be under the control of an adversary at any given time and space. However, a sufficient fraction of the resources will remain adversarial-free.

The rest of the paper is as follow: Section presents the general concepts of the literatures, Section 3 treats the concept of homomorphism and examines some encryption scheme in Somewhat Homomorphic Cryptostems, Section 4 reviews the Fully Homorphic Cryptosytem with relevant computational Craig Gentry in the Fully Homomorphic scheme and the theoretical idea of bootstrapping. Section 5 give some remarks on the application of FHE implementation in Big Data Analytics

Related Works

In this section, we are reviewing some cardinal working trends that are being used in modern cryptography and that have been executed to take into the account issue of privacy for Big Data Analytics in the cloud. The terms *Big Data Analytics*, as we observed is mostly concerned with two main targets-Data Analysis and security of the data on transmission over the open and unsecure Internet provided by some cloud provider or just some internet provider. Cryptography, a useful computational instrument in the hands spies and modern enterprise, is as old in application for security as human history. The modern introduction of digital computation through the applications of various cryptographic algorithms, has made information exchange fast, securely and simple. Nevertheless, maintaining confidentiality is a challenge in the systems. Most crypto -algorithms rely on mathematical problems that are considered very difficult to solve especially in applying the algorithmic functionalities of heuristics models and number theories, and or abstract algebra as number theory is its child in the parlace of graph theory. The first widely used encryption algorithm was Digital Encryption Standard (DES). It made use of 56 bit cipher keys and hence the possible number of keys was 2^{56}. Advanced Encryption Scheme (AES) is a symmetric key encryption algorithm with 2^{128} or 2^{194} and 2^{256} bits that are the variant of the AES known also as RIJNDAL which represents the evolution of DES .It was defined in 2000; it consisted of Cipher keys of 2^{128} bits that are adopted and the possibility of brute force decryption is made practically impossible, to this day, this variant has not been cryptanalyzed. However, the problem of key distribution management is a herculean task for Symmetric distribution between two communicating systems that need a tight secret medium to exchange the unique private key that is used for developing encryptexts and decrypting the ciphertext respectively. This, in most cases, is impossible as the monogamous Internet for transmitting encrypted data is deficient, and therefore, insure for such private keys sharing. The development of the RSA, Elgamal cryptosystems help to ameliorate the inherent deficiencies found in DES and AES in key management.

It is a known knowledge that encryption is a well-known techniques for preserving the privacy of sensitive information. Rivest, et al (1978)stated that main problem with encryption techniques is that an information system working with encrypted data could not at most store or retrieve such data for the user, any more complicated operations would seem to require that such data stored be decrypted before being operated on. They concluded this limitation follows from the choice of encryption scheme used. They however opined that there exist encryption functions which permit encrypted data to be operated on without preliminary decryption of the operands in the stored information system, for many sets of interesting operations. To sum it all, they those seamless sets of special encryption scheme, "Privacy Homomrphism".

The basic RSA and Elgamal cryptosystems need to be modified to enhance security applications in computing encrypted codes saved and controlled by the Cloud Computing Providers. The observed weakness of these encryption schemes is that they are Somewhat Hormomorphic Encryption (SWHE); meaning that such operators use one operator at a time for its data evaluation. Another observed problem of the RSA and Elgamal cryptosystems; and their variants are that they are slow and power consuming. Such algorithms cannot meet the modern space definition of Big Data Analytics as described having the properties for faster velocity and a vast volume analytics speed. Although there are many modified brands of modern cryptography, they still have the slow effects as observed in the operations of RSA and Elgamal algorithms as they cannot meet certain cryptographic demands; thus the advent of Fully Homomorphic Cryptosystem, which is the central concern of our works, was conceived and developed by Gentry (2009).

This phrase, "Big Data" as earlier considered in the introductory section, has an initial traceable record that was formally used in the modern context by John Mashey (1997) in his various talks circa 1997-1998, then, he was an electronic Engineer at the Silicon Graphics. This was then followed by Weiss and Indurkhya (1998) who referred to Big Data in a data-mining context and suggest practical difficulties for its application.

Laney (2001) defined what would be the now ubiquitous three V's of Big Data: Volume, Velocity and Veracity. This is the first definition of the characteristics of Big Data and Gartner's definition of Big Data mentions these V's exclusively. From here, the Big Data phenomenon begins to pick up speed with more and more organizations adopting, co-opting and expanding on Laney's concepts, and over the years, many attempts to append or expand this definition to include characteristics such as Variability, Value, Veracity, and Volatility, among others can be found in Francis (2012).

As earlier noted, the current world trend in scientific and commercial ventures is to go to the Cloud. Cloud computing now leverages a lot of computing difficulty including scientific data analysis, scientific computations and information storing platform. Despite its seamless simple construct by definitions, the safe keeping platform, there are problems that inform security issues such as confidentiality, Integrity and privacy of the owners' data. Although encrypted data could be stored on the platform of the Cloud computing framework Rivest (1978), this does not guarantee its safety, for encrypted data can be brute-forced by either Cloud Providers or by new modern technology that involve by the day. Another conspicuous observation is that analysis of encrypted data on the cloud cannot be carried out without decrypting the ciphertext; and therefore, expose the content of the plaintext before real computational processes could be carried out, Rivest, et al. (ibid). In the process of decrypting the ciphertext, the private key could be accessed by the inner circle or an external adversarial; thus compromised the cryptosystem. To forestall this compromise syndrome., Rivest, Addleman and Dertouzos (ibid) introduced the notion of a *privacy homomorphism;* a notion that was expanded by the introduction of ring concept by Gentry (2009) and called fully homomorphic Encryption (FHE).

The basic RSA is a multiplicatively homomorphic encryption scheme, that is, given RSA public key (N, e) and ciphertexts $\{\psi_i \leftarrow \pi_i^e \bmod N\}$, one can efficiently compute $\prod_i \psi_i = (\prod_i \pi_i)^e \bmod N$, a ciphertext that encrypts the product of the original plaintexts, Dentry (ibid).

Basic Terminologies

In this section, we describe some basic terminologies used in the paper.

Encryption and Decryption

Encryption is the conversion of data into a form, called a ciphertext that cannot be easily understood by unauthorized people and decryption is the process of converting encrypted data back into its original form, so that the authorized recipient can understand it by converting it using a private secret key. According to Kerckoffs' principle, Rijanarayanna and Idiebold (2012), Fontaine and Galand (2009), Larkin, et al., (2012), security must rely upon the secrecy of the scheme, but not on the obfuscation of the code. A cryptography scheme is assumed to be publically known whereas the secret piece of information such as key is responsible for the secrecy of the scheme. According to key management, encryption schemes are of two types: Symmetric and Asymmetric encryption schemes. Common key that is used to encrypt and decrypt the message is called as Symmetric Encryption. Symmetric-key systems are faster, but their main drawback is that two parties wishing to communicate have to exchange the key in a secure way. In addition, scalability is problem as the number of users increase in the network. Due to its secret nature, symmetric-key cryptography is sometimes referred as secret-key cryptography.

Public Encryption

An encryption system in which the sender and receiver of a message share asset of two different keys is referred to as public key cryptosystem. An important element of the public key system is that the public and private keys are related in such a way that only the public key can be used to encrypt messages and only its corresponding private key can be used to decrypt them. Moreover, it is virtually impossible to deduce the private key even if the public key is known. Public key cryptography was invented in 1976 by Diffie and Hellman (1976) and the scheme was called Diffie-Hellman encryption as the progenitors called it "New Direction in Cryptography". Security of this type of scheme is based on hard problems in mathematics, which are difficult to solve in probabilistic polynomial time. However, the downside is that they are slower than the symmetric schemes due to non-trivial mathematical computations. That is why this encryption scheme is used only for encryption of small data or keys while symmetric scheme can be used for larger ones or both are hybridized for efficiency.

The Concept of Homomorphism Encryption

In this section we look into three brands of Homomorphic Encryption Schemes. We present this scheme as in the variant of Silverberg (2013). Fully Homomorphic Encryption (FHE) Scheme can be used to query a search engine on the Service Provider's platform without revealing what is being searched for; the search engine does the computation on the encryption of information blindly that it does not know. Gentry (2009) in his PhD thesis proposes the first fully homomorphic encryption scheme, solving a central open problem in cryptography. Such a scheme allows one to compute arbitrary functions over encrypted data without the decryption key. For instance, given encryptions $E(m_1), E(m_2), E(m_3), ..., E(m_t)$ of $m_1, m_2, m_3, ..., m_t$ one can efficiently compute a compact ciphertext that encrypts $f(m_1, m_2, ..., m_t)$ for any efficiently computable function f . This problem was posed by Rivest et al. in 1978 referred above. Alice (2013) makes clearer work on Dentry(2009) in this way:

First and in precise format, the FHE has two properties as it is doubly homomorphic, meaning that it has two operators which are addition and multiplications. FHE searches the information system engine, without revealing what is being searched. During this process of search, the information system computes the encrypted data surreptitiously and output the result C_i . In a simplified exposition, the information will compute C_i , which it sends to the owner to decrypt to yield the plaintext.

$$Decrypt(C_i) = m_i$$

(3.1)

$Decrypt(C_i) = m_i$ where the $m_i's$ and $c_i's$ are the elements of some ring.

Also by the Gentry (2009), FHE is doubly homomorphic, meaning that it can carry out two operations in addition and multiplication at the same time. A ring in group theory can operate on two operands using this technique. Such a group is said to be an Abelian. Abelian group in abstract algebra is a commutative group. An Abelian Group (Wikipedia, the free Encyclopedia) is a commutative group in which the operation to the two operands or elements does not depend on their orders of the operands.

The operation of (3.1) can easily be extended to more encrypted ciphertexts on the information system or the Cloud Computing environment. Therefore, in a FHE, suppose we want the output of two ciphertexts $C_1 \, and \; C_2$ these could be computed on the Cloud and output the sum or product of the ciphertexts which the owner would decrypt to yield the following results:

$$Decrypt(C_i + C_2) = m_1 + m_2 \quad (3.2)$$

$$Decrypt(C_i * C_2) = m_1 * m_2 \quad (3.3)$$

In FHE, whenever a function f is computed with many finitely additions and addition, this can generically be expressed as follow:

$$Decrypt(f(c_1, c_2, \ldots, c_n)) = f(m_1, m_2, \ldots, m_n) \quad (3.4)$$

Equation (3.4) is the as result computing the multiple ciphertexts $c_i's \; \forall \; i = 1,2, \ldots, n)$

and then decrypted the result output $f(c_1, c_2, \ldots, c_n)$ to yield the function plaintexts $f(m_1, m_2, \ldots, m_n)$. We note in passing that output of the function of the plaintext should yield a single value which is the complete plaintext. As noted, the output of a real function which is a value.

With the phenomenal idea of FHE, there are gaining immense momentum in recent years due to the Big Data Analytics facilities on the cloud. However, but for the problems of confidentiality, Integrity and privacy in the cloud computing platform; hence many a business enterprises are reluctant to go clouding through Big Data Analytics. Notwithstanding, larger enterprises are already in the cloud which is their economy more than. The FHE is exemplified in Pardo and Luis (2013) and Silverberg (2013). This simplified computational process could be given as follow:

Suppose that (p_k / s_k) is a public/private key pair and $c_1, c_2, \ldots, c_t$ are ciphertexts such that

$$c_i = Enc(p_k . s_k) \;\; \forall \; i = 1,2,\ldots,t \quad (3.5)$$

and $m_1, m_2, m_2, \ldots, m_t$ to be able to compute a certain function f of the ciphertext $f(c_1, c_2, \ldots, c_t)$ such that

$$Dec(s_k, f(c_1, c_2, ..., c_t)) = f'(m_1, m_2, ..., m_t)$$

(3.4)

Taking the left hand equation, we have to compute the function of the ciphertext using the secret key s_k thereafter, we decrypt the result to obtain the right hand side of (3.4) f' where is a known function acting on t-tuples of plaintexts. This gives a general concept of homomorphic encryption scheme, with the simplest example as derived in Pardo and Luis (2013). The simplest example of such schemes could be the basic RSA scheme because of its multiplicative property and the Elgamal which also has a multiplicative property. As introduced earlier in this section, for homomorphic encryption, the plaintext and the ciphertext spaces have a group operation and hence define a group homomorphism property. In these cases, the roles of f are played by the group operations as observed in Luis and Padro (2013) and Silverberg (Ibid) respectively.

Based on the above last statements adduced in the last immediate paragraph, one can multiply two ciphertexts (the group operation in the ciphertext space) to obtain a new ciphertext that is a valid encryption of the operation defined; upon decrypting this cipher, then the plaintext space function is known).

In this section, we present an overview of the quadratic residue and modular square roots that are having increase interest in modern cryptographic security. This would enable us construe the concept of HFE as established in Paillier encryption Scheme. It would also lay the bedrock for Goldwasser and Micani (1996) homomorphic cryptosystem.

This subsection describes the encryption in the form of probability exposure as propounded by Goldwasser and Micali (1996).They were the first authors that saw into the invention of the probabilistic public-key cryptosystems based-on deterministic algorithmic trapdoor function. Trapdoor function is easy to compute but hard to invert unless some information known as trapdoor-key is known. Thus, while everybody can use the function called to encrypt messages, only the legal receiver knows the trapdoor, which serves as a decryption key. For instance, the trapdoor function used for RSA is exponentiation to the power of the public key in Z_n, where $n = pq$ is the product of some two large composite primes. That is, the encryption of plaintext $m \in Z_n^*$ is $c = m^e \bmod n$. The prime factors of n can be considered as the trapdoor Fuchsbaur (2006); and Pardo and Luis (2013). To go further in comprehending this algorithm in detail, let us peruse over Quadratic residuosity problem (QR problem, for short).

Theory of Homomorphism

In this subsection, we present the theory of homomorphism and its mathematical representations from the following page.

The Term Homomorphism

The term "Homomorphism" as we apply in this paper is acquired from the field of group theory in abstract Algeria. Therefore, the first impact is to semantically consider its definition:

Definition I (Homomorphism):

Consider $(G,*)$ and (H,o) as two groups. A mapping $\Phi : G \to H$ is homomorphism if by definition

$$\Phi(x * y) = \Phi(x) \circ \Phi(y) \ \forall \ x, y \in G$$

Definition II (Ring Homomorphism)
Let R and S be rings with addition and multiplication operators. The map $\Phi : R \to S$ is a homomorphism if the following are satisfied:

a) Φ is a group homomorphism on additive group $(R,+)$ *and* $(S,+)$

b) $\Phi(x, y) = \Phi(x)\Phi(y)\, x, y \in R$

These two definitions are useful in the cryptographic space, especially in modern enhancement in Big Data Analytics. Homomorphic encryption function allows for the manipulation of encrypted data in information systems without the seemingly inherent loss of the encrypted data. It's application could be understood more in these applications e-cash, e-voting, private information retrieval, and in cloud computing that is strengthened by Big Data Analytics.

Fully Homomorphic Encryption

A Fully Homomorphic Encryption functions on two operators (addition and multiplication) as previously mentioned. The onset of this problem has brought along an open problem in cryptography, and most especially in the area of Confidentiality and Integrity security. The first ever system to offer computational succor to this area of study was proposed by Gentry (ibid). However, encryption systems that operate on only operation had been in existence long after Rivest, et al. (1978) seminar paper.

Definition III (RSA Homomorphism)

Consider $n = pq$ where p and q are large primes, meaning that $gcm(p, q) = 1$, that are selected randomly. Select a from

$\varphi(n) = (p-1)(q-1) \ni ab = 1 \bmod \varphi(n)$, implying that the parameters (n,b) are public while the three parameters $p, q \ and \ a$ are private. Compute the encryptions:

$$En_k(x) = x^b \bmod n = y$$

where x is the plaintext and y is the ciphertext computed;

$$Dec_k(x) = y^b \bmod n = c$$

The homomorphism of the RSA is computed in simplified steps as follow:

Consider two plaintexts $x_1 \ and \ x_2$. Then their Homomorphism is expressible as follows:

$$\begin{aligned} En_k(x_1)En_k(x_2) &= x_1^b x_2^b \bmod n \\ &= (x_1 x_2)^b \bmod n \\ &= En_k(x_1 * x_2) \end{aligned}$$

Definition IV (El Gamal):

Randomly select a large odd number and a generator $\alpha \in \mathrm{Z}_p^*$, select $a \ and \ \beta$ such that

$$\beta = \alpha^a (\bmod p)$$

Make $p, \ \alpha \ and \ \beta$ public; a the private key. Provide a parameter $r \in \mathrm{Z}_{p-1}^*$ also a secret random number. Then,

$$En_k(x, r) = (\alpha^r \bmod p, x\beta^r \bmod p)$$

The homomorphism of El Gamal Homomorphic Cryptosystem can be computed theoretically from its cryptosystem as follow:

For two arbitrarily plaintexts $x_1 \ and \ x_2$ and compute:

$$\begin{aligned} En_k(x_1, r)En_k(x_2, r) &= (\alpha^{r_1} \bmod p, x_1\beta^{r_1} \bmod p)(\alpha^{r_2} \bmod p, \ x_2\beta^{r_2} \bmod p) \\ &= (\alpha^{r_1}\alpha^{r_2} \bmod p, x_1, \beta^{r_1}\beta^{r_2} \bmod p) \\ &= (\alpha^{r_1+r_2} \bmod p,, \ (x_1 + x_2)\beta^{r_1+r_2} \bmod p \\ &= En_k(x_1 + x_2, r_1 + r_2) \end{aligned}$$

The usefulness of El Gamal additive Homomorphism usefulness to Big Data Analytics: E-cash and e-voting will benefit from the additive homomorphism in the cloud computing environment. To achieve this aim, however, then algorithm will look theoretically in this abstractive formation:

$$En_k(x,r) = (\alpha^{r_1} \bmod p, \alpha^x \beta^r \bmod p)$$

The generic Additive El Gamal Homomorphic Cryptosystem will compute uniquely to:

$$En_k(x_1,r_1)En_k(x_2,r_2) = (\alpha^{r_1}\beta^{r_1} \bmod p, x_1\alpha^{r_1} \bmod p)(\alpha^{r_2} \bmod p,\ \alpha^{x x_2}\ \beta^{r_2} \bmod p)$$

$$= (\alpha^{r_1}\alpha^{r_2} \bmod p, x_1, \alpha^{r_1}\beta^{r_2} \bmod p)$$

$$= (\alpha^{r_1+r_2} \bmod p,,\ \alpha^{(x_1+x_2)}\beta^{r_1+r_2} \bmod p$$

$$= En_k(x_1 + x_2, r_1 + r_2)$$

The modified homomorphism is that $Dec_k = \alpha^x$, introduces the discrete logarithm problem into the Somewhat Homomorphic Encryption into the decryption scheme. For large enough ciphertexts, this will become impractical. Due to the complication of the El Gamal Cryptosystem, let us envision further another alternative, that takes into account the additive property of exponentiation; which can operate with extra decryption time; Paillier (1999) cryptosystem

Definition V (Paillier Homomorphism):

First let us face some preambles. Pascal Paillier presented a seminar paper "Public-key Cryptosystem Based-on Composite Residuosity Classes" that is an asymmetric PKE (Public Key Cryptosystem) that could be applied on probabilistically as a Somewhat Homomorphic Encryption Scheme based on multiplication operation. This polynomial composite residuosity assumption runs as follows:

Select a composite number that is an integer; it is hard to determine if if y exists such that

$$\mathrm{Z} = y^n (\bmod n^2)$$

This works was further extended by Damgard anJurik (2001), a Homomorphism we will not discuss in this paper.

Now the race to Paillier runs in this sequential order:

Select randomly two large primes $p\ and\ q$ and compute $n = pq$. Suppose λ denotes the Carmichael function:

$$\lambda(n) = lcm(p-1, q-1)$$

Select a random number $g \in \mathrm{Z}^*_{n^2}\ \ni L(g^\lambda \bmod n^2)$ is invertible modulo n where $(L(u) = \dfrac{u-1}{n})$ and g are public key; $p\ and\ q\,(or\ \lambda)$ are private entities. For a given plaintext x and resulting y cihertext, select a random $r \in \mathrm{Z}^*_n$. Then we have the ciphertext

$$En_k(p,r) = g^m r^n \bmod n^2 = c$$

And the decrypted computed to obtain the plaintext in this manner:

$$Dec_k(y) = \frac{L(y^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n$$

Fully Homomorphic Cryptosystem (FHE)

The homomorphism systems that we have abstracted so far through description so far have been plainly within the region of Somewhat Homomorphism Encryption scheme that operates with ever the addition or multiplication operation axioms. The FHE is implemented based on the concept of applying the two operators in a given homomorphic encrypted problem homomorphism. The ring Homomophism modulo 2 directly correspond to the binary operation exclusive or (XOR) and the AND operations of a circuit.

The Craig Gentry FHE

Gentry (2009) proposed the first Fully Homomorphic Encryption scheme. The description of the scheme:

This is centered around a function which introduces a certain level of noise into the encryption scheme that generates the following compounded issues:

1. Each operation on the ciphertext results in compounding noise;
2. Resolved this problem by bootstrapping the encryption;
   a. Each re-encryption cuts down the down the noise;
3. The operation of FHE is based on Ideal Lattices: that is Ideal in number theory which
   i) allows for new complex circuit implementation;
   ii) corresponds to the structure of ring homomorphism

The FHE as outlined include Somewhat Homomorphic scheme as part of its operation. The intention is to disallow the algorithm to not to converge locally. Let us idealize concretely bootstrap noise reducing algorithm the SWHE.

1. KEYEN Evaluation: Output a random odd integer p;
2. For bit $m \in \{0,1\}$, let $m' = m \bmod n$; where m' is even if m=0, odd if m-
   1. Select a random q.

$$ENCRYPT_E(m,p) = C = m^1 + pq$$

m' is the noise associated with the plaintext

3. Let $C' = C \bmod p \ \ and \ \ C' \in (-\frac{p}{2}, \frac{p}{2})$; then

$$DECRYPT(p,c) = c' \bmod 2$$

C' is considered to be the noise associated with the ciphertext (that is, the shortest distance to a multiple of p)

The Homomorphism; Multiplication:

Let $m_1,\ m_2 \in \{0,1\}$ then,

$$Enc_{\in}(m_1, p)Enc_{\in}(m_2, p) = (m_1' + pq_1)(m_2' + pq_2)$$

The decryption :

$$Dec(c) = (m_1' + pq_1)(m_1' + pq_2) \bmod p \bmod 2$$
$$= m_1'.m_2' \bmod 2 = m_1.m_2$$

The compounding noise $m_1', m_2')$ stems in the loss of some homomorphic property after a certain number of operations. To disallow the computation to stop it's computational processes due to the inherent noise, Gentry introduced bootstrapping into FHE. This allows the process to take in more inputs and process for a longer period and therefore improves performance. Let us review bootstrapping theoretically without any mathematical pronouncement:

Gentry, in his PhD thesis (2009), introduced the working idea for bootstrapping. He opined that to go bootstrapping is to go from somewhat Homomorphism encryption to Fully Homomorphism. He gave a vivid illustration on to go through bootstrapping. One can include as a part of the public key, include a private. The idea is this, when the ciphertext gets too large or noise during computation in a remote server, the encryptor can then use the somewhat homomorphic encryption scheme to evaluate the decryption function that is applied to the ciphertext, using the encrypted key. This re-encryption process produces a new ciphertext that is compact and less noisy. However, for this sequential method to work, it is necessary to for the Somewhat Homomorphic Scheme to be "circular secure". Circular secure implies that it must be able to re-encrypt itself. This is the main reason that motivated him into introducing the Ideal Lattice into the Fully Homomorphic Scheme.

In a final stepping down, bootstrapping of the Somewhat Homomorphic Cryptosystem allows for the reduction of the noise and therefore makes the ciphertext more compact and less noise which allows no limit of operations. Nonetheless, the combination of the noise production that follows the noise reduction completely makes scheme terribly impractical. Ever since the observation of this *bootstrappability*, more schemes have been introduced to try and decrease this complexity, but all depend largely on the bootstrapping.

Despite all the impracticability, Gentry (2009) has made much breakthrough to Cryptography with his bootstrapping Somewhat Homomorphic Encryption Scheme that strengthen the implementation of the Fully Homomorphic Cryptosystem Scheme. With this onerous contribution and its security attendance, the security of Big Data Analytics or Internet of Things has been boosted on the Cloud computing environment.

Conclusion

This paper has intentionally set out to give edification to the application of the Fully Homomorphic Encryption scheme that could be applied to enhance the security of the Big Data Analytics. Ideally, certain modeling and simulation should have complemented the work, but for time constraints. The paper has, however, provide some details on how some modern cryptography algorithms encrypted data can be still be saved and necessary computational procedures taken without compromising the data and the privacy of the owners of the data. In furtherance of the security of Big Data Analytics, it is important to also recognize and accept some fundamental truth that could enhance further acceptability of the Big Data Analytics on the cloud computing environment. Issue of data security in government establishments is one thing all together that could also benefit from the discourse in this paper. If applied intelligence, Fully Homomorphic cryptosystem can be used for computation of sensitive information securely.

As highlighted in the preceding paragraphs, the year 2009 was an eventful year in which Gentry in his PhD Thesis constructed a Fully Homomorphic Encrypion Scheme (FHE). The scheme provides a framework in which the client evaluates any circuit on plaintext values by computing only the ciphertexts The FHE scheme has since morphed into some parameters evolution, namely: [Gen 09,DGHV10, SV10] and [GH11, BV11a, BV11b, CMENT11, CENT12, CCK^{+}13]. The latter parametric details are best applied on the computation of the ciphertext in the cloud. These adapt the blueprint of Gentry to meet the implementation of the FHE.

As stated earlier, the steps in the construction of FHA are as follow: First, construct a Somewhat Homomorphic Encryption Scheme (SWHE) that could evaluate some "low degree" polynomial homomorphically. The reason for the application of the SWHE procedural pattern is the inherent contents of noisy states in the ciphertexts. The noises hinder further computations on the ciphertexts after a few iterations. This is more slightly observant in homomorphic additions but exponentially pronounced in homomorphic multiplication iterations. The SWHE renews and maintains a low noise level, as a result, allows subsequent homomorphic operations. To ensure a Fully Homomorphic Encryption Scheme, which is the Gentry (2009) main contribution to the developmént of FHE, *Bootstrapping,* emphasizes that SWHE is capable of evaluating its own decryptional procedures (addition and multiplication operations) which can further be transformed into the FHE. Therefore, Bootstrapping is a process that may be conceived as a process that consists in the evaluating the decryption circuit of the SWHE Scheme by using the decryption key bits in the encrypted ciphertexts. This Means that the private-key can be sent out to the cloud embedded in the ciphertext without the adversely affecting the compromise of the ciphertext. In view of this, this would result in different encryption of the same plaintext with some drastic noise reduction. In practice, the scheme parameters are generally specified so that the new cphertexts can handle one additional homomorphic multiplication [GH11, BV11a, BV11b, CMENT11, CENT12, CCK^{+}13]. Gentry and Halevi (2010).

Additionally, developing countries need to go back to their syllabus designs and implementation of abstract algebra and programming codes if they must meet the demands of Big Data Analytics.

The paper would therefore expect Open and trusted security architecture that could be extended for business Intelligence and Big Data Systems. Any organization that wants to run the Big Data Analytics based on Fully Homomorphic Encryption Scheme must give maximum compliance to legislation of that nation. Thus no country should give up in legislating laws that would give civil penalties to any organization breaching the laws. Thus, if FHE is treated appropriately, homomomorphism can be used to treat both internal and external security of a country.

The application of FHE is numerous in the Big Data Analytics and in Governance and Search Engines. For example:

1. Private queries on search engines. The search engine will be able to return encrypted data without ever decrypting the query data in an information system.
2. Cloud Computing environment can store encrypted data that looks seemly useless. Computation on the encrypted data on the cloud can be done and encrypted output delivered that can be decrypted by the owner to obtain the original plaintext.

The scenario of models for the security in the cloud computing environment as outlined in other sections of this paper could be applied to construct models for the cloud computing Security based on the standard of modern cryptography.

Correspondence

Victor Onomza Waziri, PhD
Cyber Security Science Department
School of Information and Communications
Technology, Federal University of Technology,
Minna, Niger State, Nigeria
Email: victor.waziri@futminna.edu.ng.

John K. Alhassan
Department of Computer Science
School of Information and Communications
Technology, Federal University of Technology,
Minna, Niger State, Nigeria
Email: jkalhassan@futminna.edu.ng.

Olalere MORUFU
Cyber Security Science Department
School of Information and Communications
Technology, Federal University of Technology,
Minna, Niger State, Nigeria. Email:
lerejide@futminna.edu.ng.

Idris ISMAILA
Cyber Security Science Department
School of Information and Communications Technology, Federal University of Technology, Minna, Niger
State, Nigeria
Email: ismi_idris@yahoo.co.uk.

References

Balow, M. (2013) Real-Time Big Data Analytics: Emerging Architecture, CA: O'Reilly Inc [Online] http://oreilly.com/catalog/erra.csp?isbn=97k44-9364121.

Bryant, E. R., Hensel, C., Katz, R. H., Gianchandani, E. P. (2010) From Data in Knowledge to action: Enabling the Smart Grid, *Computing Community Consortium*, [Online] http://www.cra.org/ccc/files/docs/init/Energy.pdf.

Burt, K. (Date Not Stated) "The Mathematics of the RSA Public-Key Cryptosystem", RSA Laboratories, pp.1-9. [Online] http://www.mathaware.org/mam/06/Kaliski.pdf.

Chaum, D. (1983) "Blind signatures for untraceable payments," *Advances in Cryptology*, pp. 199-203, Springer-Verlag.

Cliffer J. (2011) Adopting to the New Regulatory Environment; the evolving Role of Technology in Financial Services [Online] http://www.statestreet.com/vision/technology/pdf/TheEvolvingRoleTech.pdf.

Clinton G., Kantarcioghu M., Vaidya J. (2002) Defining Privacy in Data Mining in National Science Foundation Workshop on Next Generation, *Data Mining*, 1 (26), 126–133.

Diffie, W., Hellman, M. E. (1976) New Directions in Cryptography, *IEEE Transactions on Information Theory*, Volume IT-22, No. 6, [Online] http://www-ee.stanford.edu/~hellman/publications/24.pdf.

Francis, X. D. (2012) A personal Perspective on the origins and Development of "Big Data"; the Phenomenon, the term and the discipline, Social Science Research Network. [Online] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2202843.

Fontine, C., and Garland, F. (2009) A Survey of Homomorphic Encryption Scheme, *Journal of information Security*, pp. 41-50

Gentry C. (2009). A fully homomorphic encryption scheme. PhD Thesis. Stanford University, [Online] http://crypto.stanford.edu/craig.

Gentry, C. (2010) "Computing Arbitrary Functions of Encrypted Data." *Communications of the ACM*, Vol. 53 Issue 3, pp. 97-105.

Gentry, C. and Helevi, S. (2010) Implementing Gentry's Fully homomorphic Encryption Scheme, Preliminary Report, [Online] http://researcher.watson.ibm.com/researcher/files/us-shaih/fhe-implementation.pdf.

Irwin, K., Michael, R. and Yang, H. (2013) Online Learning for Big DT Analytics, CA, http://googl.gl/BOSGS.

Lancy, D. (2001) Application Delivery Strategies, *Meta Group* Inc, Stamford, [online] http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Larkin K. G., Fletcher P. A., Hardy S. J. (2012) "Tenacious Tagging of Images via Melin Monomials". [Online] http://arxiv.org/ftp/arxiv/papers/1208/1208.5842.pdf.

Luis, O. and Pardo, L. (2013) – sources from Wikipedia, the free encyclopedia. [Online] http://en.wikipedia.org/wiki/Main_Page.

Lindell, Y., and Pinkas, B., (2000) Preserving Data Mining Advances in Cryptology. Advances in Cryptology—CRYPTO 2000, 36-54, Springer Berlin Heidelberg. [Online] http://scholar.google.com/citations view_op=view_citation&hl=en&user=tpMNnPwAAAAJ &citation_ for_view=tpMNnPwAAAAJ:u-x6o8ySG0sC.

McKinsey Global Institute (2011) Big Data: The Next Frontier for Innovation, Computation, and Productivity, *McKinsey Global Institute* [Online] www.mckinsey.com/~/.../big%20data/mgi_big_data_full_report.ashx.

Mineli, M., Michele, C. and Ambiga D. (2013) Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today, John Wiley & Sons, available: http://eu.wiley.com/WileyCDA/WileyTitle/productCd-111814760X.html.

Laura, L. (2006) "Symmetric Private Information Retrieval via Additive Homomorphic Probabilistic Encryption". RIT Department of Computer Science [Online] http://hdl.handle.net/1850/2792.

Pallier, P. (1999) "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes." Advances in Cryptology — EUROCRYPT '99, *Lecture Notes in Computer Science*, Volume 1592, pp. 223-238. [Online] http://link.springer.com/chapter/10.1007%2F3-540-48910-X_16.

Pryor W. (2011) "Technology with a Purpose. The next Generation Today" in Adopting to the New Regulatory Environment; the evolving Role of Technology in Financial Services [Online] http://www.statestreet.com/vision/technology/pdf/TheEvolvingRoleTech.pdf.

Rijanarayanna, S. and Pushparaghaven, A. (2012) Present Development in Signal Encryption: A Critical Survey, *International Journal of Scientific and Research publications*, Vol. 2, Issue 6, pp. 1-7; available: http://www.ijsrp.org/research_paper_jun2012/rp74.html.

Rivest, R., Adleman, L., and Dertouzos, M. L. (1978) "On data banks and privacy homomorphisms," Academic Press Inc. [Online] http://people.csail.mit.edu/rivest/pubs/RAD78.pdf.

Silverberg, A. (2013) Fully Homomorphic Encryption for Mathematicians, Cryptology ePrint Archive, Report 2013/250. [Online] https://eprint.iacr.org/2013/250.pdf.

Sophia, Y., Vijay, G., Nabil, S., Emily, S., and Arkady, Y. (2014) "A Survey of Cryptography Approaches Approach to Securing Big Data Analytics in the Cloud", MIT Lincoln Laboratory, MA [Online] http://www.ieee-hpec.org/2014/CD/index_htm_files/FinalPapers/28.pdf.

Szakal, A. (2014) "The Emergence of the Third platform", *PEN Newsletter*, The Open group, available: http://blog.opengroup.org/2014/10/17/the-emergence-of-the-third-platform/?

utm_content=l.baynes%40opengroup.org&utm_source=VerticalResponse&utm_medium =Email&utm_term=The%20Emergence%20of%20the%20Third%26nbsp% 3BPlatform&utm_campaign=November%20Newslettercontent.

Weis, S. (2007) “Verifying Elections with Cryptography”. Google Tech Talks: Theory and Practive of Cyptography. December 2007; available on Youtube: https://www.youtube.com/ watch?v=ZDnShu5V99s.

Weiss, S. M. (1998) *Predictive Data Mining; A Practical Guide*, Morgan Kaufmann Publisher, Massachusetts.