International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

September 2015 Issue 24 Volume 7 Number 3 Pages 47-52

SOFTWARE ENGINEERING AND APPLIED CRYPTOGRAPHY IN CLOUD COMPUTING AND BIG DATA

M.I. Mihailescu [1] **S.L. Nita** [2]

1. Information Technology and Communications Department, University of South-East Europe LUMINA, Bucharest, Romania, mihmariusiulian@gmail.com
2. Computer Science Department, University of Bucharest, Bucharest, Romania, stefanialoredanani@gmail.com

Abstract- In this article we have decided to cover the most important trends and challenges from software engineering and cryptography fields by presenting some important results achieved already (as state-of-the-art). We will propose some new research directions, which are absolutely necessary to be covered and focused on if we want to move to cloud computing environment and using big data. Due to these research directions, we have figure out that if we don't have a multi-disciplinary research activity based on these two directions, the actual technologies, such as *cloud computing* and *big data,* will be in deep bucket and they will become useless. The mentioned technologies, represents the peak of the spear, in which we can actually say that we have a research work to do.

Keywords: Cryptography, Chaos-Based Cryptography, Fully Homomorphic Encryption, Searchable Encryption, Cloud Computing, Big Data.

I. INTRODUCTION

Cloud computing represents the delivery process of computing services over the Internet. Cloud services give the possibility to individuals and businesses to use software and hardware which are managed by third parties at locations that can be accessed remotely. Examples of cloud services online file storage, social networking sites, webmail, and online business applications [10].

The cloud computing give the possibility to access information and computer resources from any location where a network connection is available. Cloud computing provides a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications [10]. The following definition of cloud computing has been developed by the U.S. National Institute of Standards and Technology (NIST):

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models [11].

We are facing to one of the most problem from nowadays and which will have a strong impact on the future of encrypted search. Encrypted search is a very sensitive subject from many aspects and it becomes an important problem in security and cryptography. From my point of view, this aspect is happening because of a combination of three things:
- Searching represent the main way of accessing the data.
- Outsourcing is winning more and more faith to third parties.
- The trust in these third parties is limited, and the reasons are obvious.

Now, searching over encrypted data is of interest for many other sub-fields of computer science, such as cryptography, privacy, databases) but also we can find applications in governments and industry.

A-priori searching on encrypted data could sound impossible and even contradictory; there are several ways of doing it. Some methods are practical, some of them are more secure, and some of them are more functional and/or flexible.

In 2001, for the first time the problem of searching on encrypted data was proposed and taken into consideration by Song, Wagner, and Perrig [1]. The authors describe a set of schemes for the mentioned problem and provide proofs of security for the proposed crypto systems. The schemes proposed by them are focused on:
- The scheme is provably secure [];
- Query isolation for searches is provided;
- Controlled searching for the untrusted server is provided;
- Hidden queries.

Beside the advantages provided above, their schemes are very efficient and practical. The algorithms developed by them are simple to understand and to implement, and also are very fast.

A. Characteristics

The characteristics of cloud computing are based on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. On-demand self-service is represented by the customers (usually organizations) that can request and administrate their own resources used in computing. Broad network access give the possibility for the services to be offered over the Internet or private networks. Pooled resources means that customers are able to draw from a pool of computing resources, usually in data centers place in other location (remotely). Services are scaled larger or smaller; and use of a service which is measured and customers are billed according with the plans.

B. Service Models [10]

The cloud computing service models are Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). In a Software-as-a-Service model, a pre-made application, along with any required software, operating system, hardware, and network are provided. In PaaS, an operating system, hardware, and network are provided, and the customer installs or develops its own software and applications. The IaaS model provides only the hardware and network; the customer installs or develops its own operating systems, software and applications.

C. Deployment of Cloud Services [10]

Cloud services are typically made available via a private cloud, community cloud, public cloud or hybrid cloud. Generally speaking, services provided by *a public cloud* are offered over the Internet and are owned and operated by a cloud provider. Some examples include services aimed at the general public, such as online photo storage services, e-mail services, or social networking sites. However, services for enterprises can also be offered in a public cloud.

In a private cloud, the cloud infrastructure is operated solely for a specific organization, and is managed by the organization or a third party.

In a community cloud, the service is shared by several organizations and made available only to those groups. The infrastructure may be owned and operated by the organizations or by a cloud service provider.

II. CLOUD COMPUTING AND ITS IMPORTANCE

As we can observe, the needs of the market are very important when we try to fix some patterns regarding the necessities and to find the right patterns and solutions for enterprises and not only.

The market around Google, AWS, and Microsoft will become more and more convergence. The enterprises that wish to move in public clouds, they have to deal with these three companies. In the foreseeable future, the mentioned companies are placing their technology bets on cloud computing.

Anybody, besides of Google, AWS, and Microsoft, will have to obtain a small piece from the market and hold it as long as possible. The private cloud only, will face a fall if they will not emerge with public clouds. In this way, after the emerging process is done, a new hybrid or multi-cloud(s) is borne. In the next few years, the companies that will focus on private clouds will fall substantially.

Another important aspect is that the new providers which offer public access on cloud will have to find very fast a niche. The market is highly dominated by some players who are willing to spend billions on research and development, and marketing. Startups are focusing on new and emerging cloud technology. The same pattern will be observed around data storage services, analytics services, Internet of Things (IoT), and other new technologies. Regarding the Internet of Things (IoT) [8], we have covered some important aspects which are important to take into consideration when we are facing with this concept. The paper is case study for biometrics technologies and eLearning applications.

In Figure 1 we can observe that 58% is characterized by a top-line growth and the collaboration among employees, and there is also 56% for the supply chain, three areas on which the enterprises expect cloud computing to have a huge impact in the following three years. Also we can see the evolution in three years (light-blue line), and the top-line growth will have the most powerful evolution, around 58%, and also an important role will be played by collaboration with employees, around 58%.

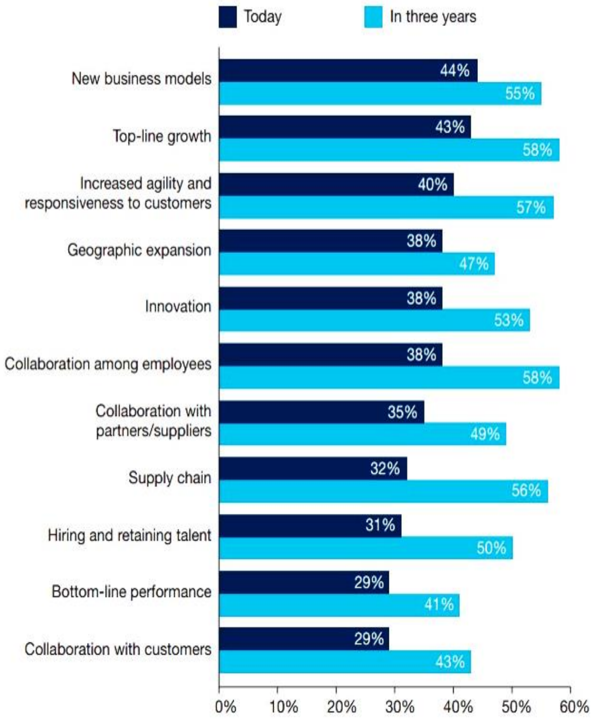

Figure 1. Evolution of the impact of cloud on business, from today and in three years [12]

A. Software Developing in Cloud Computing

Agile Development is the best way (method) to use when we want to develop software for cloud computing. Cloud computing and virtualization technologies create an easy way for agile development teams to combine multiple environments based on development, test and production. These environments are combined with other cloud services, a very important aspect which need to be taken into consideration.

In the followings, we present six important ways in which cloud computing and virtualization technologies enhance agile software development.

1. Cloud computing gives an unlimited number of testing and staging servers.
2. The Agile development becomes a truly parallel activity.
3. Innovation and experimentation are encouraged.
4. The Delivery and continuous integration is enhances.
5. We have more development platforms and external services become more and more available.
6. Code branching and merging becomes very easy to use.

B. Security in Cloud Computing

This topic is a very important one. When we wish to have a secure cloud, security becomes a sensible discussion which needs to give the right focusing.

The favorite topic in this area is *homomorphic encryption* or *fully homomorphic encryption* (FHE). FHE is a form of encryption which give the possibility for computations to be carried out on the ciphertext, in this way by generating an encrypted result which, when is decrypted, matches the result of operations performed on the plaintext [1].

Because of the design, the homomorphic schemes are malleable [2], a property of some specific cryptographic algorithms. We say that an algorithm is malleable is it is possible for an adversary to transform the cipher text into another cipher text which decrypts to a related plaintext. We suppose, given an encryption of a specific plaintext *message,* it is obvious to generate another cipher text which decrypts to *f* (message), for a known function *f*, without necessarily knowing or learning *message.*

Some examples of malleable cryptosystems need to be listed because represents the foundation and the main entry point of the homomorphic encryption schemes and research that can be used in cloud computing in order to assure the security of the data.

In a *stream cipher,* the ciphertext is generated by using $\oplus$ (exclusive or) on the plaintext and a pseudorandom stream which is based on a secret key *k,* as $E(message) = message \oplus S(k)$. The adversary is able to construct the encryption of $message \oplus t$ for any *t,* as $E(message) \oplus t = message \oplus t \oplus S(k) = E(message \oplus t)$.

In *RSA cryptosystem,* the plaintext *message* is encrypted as $E(message) = message^e \bmod n$, where (e,n) represent the public key. Starting from ciphertext, the adversary is able to construct the encryption of $message.t$ for any *t*, as $E(message).t^e \bmod n = (mt)^e \bmod n = E(message\ .t)$.

In *ElGamal cryptosystem,* the plaintext *message* is encrypted using $E(message) = \left(g^b, message.A^b\right)$, where (g,A) represent the public key. Starting from this ciphertext (c_1, c_2), the adversary is able to compute $(c_1, t.c_2)$ which represent a valid encryption of $t.message$, for any *t.*

In 2005, we are facing with a new concept, called *non-malleable,* introduced by Mare Fisclin [3], an ability characteristic to the system to keep their non-malleable property while giving some extra power to the adversary in order to choose a new public key, a key that is represented by a function based on the original public key. So, in this case, the adversary shouldn't be able to come up with a cipher text whose underlying plaintext is in connection with the original message which also take into consideration the public keys.

Starting from the algorithms mentioned above, now we are ready to go further with the evolution of the security mechanisms used in cloud computing.

First modern technique that is used in cloud computing is *crypto cloud computing* [5] which is used in order to assure the security of the customer's data which relies on the security of service from the cloud computing providers, however, the current structure of cloud computing service that are provided by independent operators.

Crypto Cloud Computing (CCC) represents one of the newest secure cloud computing architecture. Using CCC we have protection of information security at the system level, and the users are allowed to access shared services in a convenient way and also accurately. The architecture is able to:

1. Protects individual's connection with the rest of world.
2. The personal privacy is protected without any delay of the exchanging information.

The method, on which CCC is based, is known as *Quantum Direct Key* (QDC) [4]. QDC represent a set of asymmetric offline mechanism, everything regarding the entities will go public and the private key pair is attached to their ID. Each entity will have his own private key and the entity will have also a public key generator in order to generate any public key.

In the followings, we will present how DeTron Inc. describes the algorithm behind the system.

QDK functions are used by generating a pair of public and private key from two public and secret seed matrices M_p (public) and M_s (secret). The *Key Management Center* (KMC) is the single owner of M_s and it is always offline. The private key is generated using the KMC based on the use's ID:

$$Key_{cs} = g_s(ID, M_s) \tag{1}$$

The public is generated any user which use a known ID:

$$Key_{CP} = g_p(ID, M_p) \tag{2}$$

C. Homomorphic Encrption

The second modern technique is represented by homomorphic encryption [6]. Here are facing with two categories of homomorphic cryptosystems, such as *partially homomorphic cryptosystems* (PHC) and *fully homomorphic encryption* (FHE).

In PHC, we have some strong examples which are absolutely necessary to know and to understand. They represent the main core of frameworks, such as cloud computing architecture and QDK. In the following examples, the *E* (message) denote the encryption function of the *message*.

C.1. First Example – Unpadded RSA

In case that the RSA public key is modulus *message* and exponent *e*, then the encryption of the *message* is given by $E(message) = message^e \bmod m$. The homomorphic property is then $E(message_1).E(message_2) = message_1^e message_2^e \bmod m =$ $= (message_1 message_2)^e \bmod m = E(message_1\ .message_2)$.

C.2. Second Example – ElGamal

When we deal with ElGamal cryptosystem, in a cyclic group G of order *q* with generator *g*, if the public key is represented as (G, q, g, h), where $h = g^x$, and x represent the secret key, then the encryption of the *message* is $E(message) = (g^r, message.h^r)$, for some random $r \in \{0, \ldots, q-1\}$. The homomorphic property becomes $E(message_1).E(message_2) = (g^{r1}, message_1.h^{r1}).$ $.(g^{r2}, m_2.h^{r2}) = (g^{r1+r2}, (message_1.message_2)h^{r1+r2}) =$ $= E(message_1.message_2)$

C.3. Third Example – Goldwasser-Micali

In Goldwasser-Micali cryptosystem, if the public key is represented by the modulus *m* and a quadratic non-residue *x*, then the encryption of a bit is $E(bit) = x^{bit} r^2 \, mod \, m$, where $r \in \{0, \ldots, m-1\}$. The homomorphic property is then $E(bit_1).E(bit_2) = x^{b1} r_1^2 x^{b2} r_2^2 = x^{b1+b2}(r_1 r_2)^2 = E(bit_1.bit_2)$, where **.** denotes addition modulo 2.

In FHE the computation are done on some operations over the ciphertexts, such as additions, multiplications, quadratic functions, etc). If a cryptosystem which supports arbitrary computation on the cipher texts is known as fully homomorphic encryption (FHE) and is very powerful.

One of the most important examples in this direction is represented by the Gentry's cryptosystem, proposed by Craig Gentry, which is using lattice-based cryptography. The author described the first plausible construction for a fully homomorphic encryption scheme. The scheme proposed by Gentry, supports both addition and multiplication operations on cipher texts, from which is possible to develop and build circuits for performing arbitrary computation.

Another example is represented by the cryptosystem over the integers, proposed in 2010 by Marten van Dijk, Craig Gentry, Shai Halevi and Vinod Vaikuntanathan. The scheme uses many of the tools of Gentry's scheme, but does not require ideal lattices. This scheme uses integers.

III. BIG DATA

Big data is a compound term for data sets so large and complex where the classic applications are not the right ones that are able to process the data. New challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy.

Big data can be described by the following characteristics (we will not go into deep for each characteristic, because is not the purpose of the article):

- Volume
- Variety
- Velocity
- Variability
- Veracity
- Complexity

In [7] is presented a survey for securing data analytics in the cloud.

One of the most important challenge on which will stop in this work is represented by security, information privacy, by focusing on *searchable symmetric encryption* (SSE).

A SSE scheme is correct if the protocol used for search returns the (current) correct results for the keyword that is begin searched $(DB(w))$, except the negligible probability. In order to simplify the formalism we will ignore the case when a client will attempt to add a file with an existing identifier or delete/edit with an identifier which is not present in *DB*. The protocols that already exist can handle these situations in a clean manner.

Let's see an example of algorithm used in this direction. Definition: Suppose we have $\Pi = (Setup, Search, Update)$, which represent a dynamic SSE scheme and $\mathcal{L}$ represent a leakage functions. Let's start from two algorithms A and S, we define some games $Real_A^{\Pi}(\lambda)$, where λ represent the security parameter, and $Ideal_{A,S}^{\Pi}(\lambda)$ as following:

A. $\boldsymbol{Real_A^{\Pi}}(\lambda)$

$A(1^\lambda)$ choose DB. The game will run $(K, EDB) \rightarrow Setup(DB)$ and will give EDB to A. So, A will repeat requests to engage in the protocols for Search and/or Update. This is happening when A picks a client input as entry. In order to respond, the game will runs the Search or Update protocols with the input from the client (K, in) and the server input EDB and will receive a transcript to A (the server we supposed to be deterministic). So, A will return a bit that the game uses as its own output.

B. $\boldsymbol{Ideal_{A,S}^{\Pi}}(\lambda)$

$A(1^{\lambda})$ will choose DB. The game will run $EDB \leftarrow S(L(DB))$ and will give EDB to *A*. In this case the *A* will repeat the requests to engage in the Search or Update protocols. This will take place when *A* picks a as client input in. In order to respond, the game will give as output of $L(in)$ to S. This is taken place when S will outputs a simulated transcript which is passed to *A*. So, *A* will return a bit which is used by the fame as his own output.

In order to demonstrate that Π is L-secure against adaptive attacks for any adversaries for *A*, there will be an algorithm, S, that exist in such way that

$$\Pr\left[Real_{A}^{\Pi}(\lambda)=1\right]-\Pr\left[Ideal_{A,S}^{\Pi}(\lambda)=1\right]\leq neg(\lambda) \quad (3)$$

In [9] we find more details about the mentioned algorithm. In [13] the authors discuss about a set of algorithms such as Zernike and support vector machine (SVM) in order to detect the moving targets. This example represents a very interesting case study for data which can be stored and made them available to subscribers through a big data environment. Another interesting aspect presented in the mentioned article refers to reconstructing binary image, a nice example of using Zernike moments in order to find the most suitable parameters when they are compared with the original image. This idea can be used in combination with [16] in order to create a strong authentication mechanism. The authentication mechanism is based on face recognition and an analysis is strongly recommended before the authentication process is executed.

IV. CONCLUSIONS

By presenting the main trends and challenges in cloud computing and big data, we can say that we fulfill our goal to cover the most important technologies and algorithms from software engineering and cryptography that can be applied.

Our incursion in cloud computing and big data we will not stop here. Our research will be conducted in order to optimize and to find the best ways to develop software and cryptographic algorithms for the mentioned environments.

Cloud computing offers benefits for organizations and individuals. There are also privacy and security concerns. If you are considering a cloud service, you should think about how your personal information, and that of your customers, can best be protected. Carefully review the terms of service or contracts, and challenge the provider to meet your needs [10].

The presented article addresses some important aspects which are necessary to be take into consideration when we will have software applications which will work for technical and physical problems in electrical engineering. Such examples we have seen in [13] and [14].

ACKNOWLEDGEMENTS

We would like to show our gratitude to Assoc. Prof. Adrian Beteringhe and Prof. Ion Sima for their support and courage that they give it to us for writing this paper. In the end we would like to thank Dr. Cristian Kevorchian for his advices that he gives us in these directions, for opening our minds and encouraging us to explore the new challenges and research directions in areas of cloud computing and big data.

REFERENCES

[1] "Homomorphic Encryption", https://en.wikipedia.org/wiki/Homomorphic_encryption, 17 August 2015.

[2] "Malleability (Cryptography)", https://en.wikipedia.org/wiki/Malleability_%28cryptography%29, 10 February 2015.

[3] "Completely Non-malleable Schemes, Marc Fischlin", Automata, Languages and Programming, Lecture Notes in Computer Science, Vol. 3580, pp. 779-790, 2005.

[4] F. Ladinois, DeTron, "Introduces its QDK Cryptosystem to Enable True Trusted Identity for the Cloud Era", Posted on September 26, 2012, https://en.wikipedia.org/wiki/Crypto_cloud_computing.

[5] "Crypto Cloud Computing", https://en.wikipedia.org/wiki/Crypto_cloud_computing, 27 December 2014.

[6] V. Vaikuntanathan, "Computing Blindfolded: New Developments in Fully Homomorphic Encryption", http://www.cs.toronto.edu/~vinodv/FHE-focs-survey.pdf.

[7] S. Yakoubov, V. Gadepally, N. Schear, E. Shen, A. Yerukhimovich, "A Survey of Cryptographic Approaches to Securing Big-Data Analytics in the Cloud", http://www.ieee-hpec.org/2014/CD/index_htm_files/FinalPapers/28.pdf.

[8] P.V. Corneliu, M.M. Iulian, "Internet of Things and its Role in Biometrics Technologies and Elearning Applications", 13th International Conference on Engineering of Modern Electric Systems (EMES), pp. 1-4, ISBN 978-1-4799-7649-2, INSPEC Accession Number: 15295927, 11-12 June 2015.

[9] D. Cash, et al., "Dynamic Searchable Encryption in Very-Large Databases: Data Structures and Implementation", http://www.internetsociety.org/sites/default/files/07_4_1.pdf.

[10] V.C. Pau, M.I. Mihailescu, "Introduction to Cloud Computing", https://www.priv.gc.ca/resource/fsfi/02_05_d_51_cc_e.pdf

[11] "NIST Cloud Definition", Version 15, http://csrc.nist.gov/groups/SNS/cloud-computing/.

[12] "55% of Enterprises Predict Cloud Computing Will Enable New Business Models in Three Years", http://www.forbes.com/sites/louiscolumbus/2015/06/08/55-of-enterprises-predict-cloud-computing-will-enable-new-business-models-in-three-years/.

[13] E. Yakhti Fard, A. Amiri, "Finding Specific Targets Based on Fuzzy Logic Using Zernike Moments and Support Vector Machine", International Journal on Technical and Physical Problems of Engineering (IJTPE), ISSN: 2077-3528, Issue 23, Vol. 7, No. 2, pp. 60-64, June 2015.

[14] S.H.R. Alemohammad, M. Saniei, E. Mashhour, "Reconfiguration of a Distribution Network in a Restructured Power Industry for Minimizing the Cost of Energy Loss" International Journal on Technical and Physical Problems of Engineering (IJTPE), ISSN: 2077-3528, Issue 23, Vol. 7, No. 2, pp. 90-95, June 2015.

[15] M.I. Mihailescu, "New Enrollment Scheme for Biometric Template Using Hash Chaos-Based Cryptography", Elsevier - Procedia Engineering, ISSN: 1877-7058, Vol. 69, pp. 1459-1468, 2014.

[16] M.I. Mihailescu, "Research on Biometric Synthetic Faces", Indian Journal of Research (PIJR), ISSN: 2250-1991, Vol. 2, Issues 9, pp. 38-40, September 2013.

BIOGRAPHIES

Marius Iulian Mihailescu was born in Bucharest, Romania, 1985. He received two B.Sc., first B.Sc. degree from University of Titu Maiorescu (Bucharest, Romania), second B.Sc. from University of Southern Denmark (Odense, Denmark) and the M.Sc. degrees from University of Bucharest (Bucharest, Romania) and Military Technical Academy (Bucharest, Romania). He has Ph.D. degree from University of Bucharest (Bucharest, Romania), all in Information Communication Technology, Computer Science, Computer Engineering, and Information Security, in 2008, 2009, 2010, 2011 and 2014, respectively. Currently, he is Lecturer of Information Technology and Communications at University of South-East Europe LUMINA (Bucharest, Romania).

Stefania Loredana Nita was born in Bucharest, Romania, on January 10, 1991. He received the B.Sc. in Mathematics from University of Bucharest (Bucharest, Romania) in 2013 and M.S.E. degrees in Software Engineering in 2016 (expected). Currently, he is an Assistant Lecturer (associate) at University of Bucharest, Romania. Her research interests are in the application of cryptography, programming languages and databases, cloud computing and big data, applied mathematics, software development management.