A Comprehensive Study of Traditional and Deep-learning Schemes for Privacy and Data Security in the Cloud
Mohammed fawzi Sheet¹ Melad Jader saeed²

mohammed.20csp63@student.uomosul.edu.iq¹, Meladjader@uomosul.edu.iq

Abstract

After the availability of Internet infrastructure all over the world, and connectivity is no longer an obstacle over the Internet, cloud computing has emerged as a practical and ideal solution. A huge revolution has taken place in the field of cloud computing, where it is now an industry. However, it faces great difficulties in ensuring data confidentiality and privacy. People hesitate to use it due to the risk of innumerable attacks and security breaches. This article has covered several directions relayed to cloud computing ideas. This research would focus on traditional and deep-learning based schemes to secure user’s data in the cloud. This study concluded some points about the capabilities of the traditional and deep learning-based scheme. The comparison showed that both of them increased the levels of security and privacy of the cloud. The study conclude that the Deep learning-based method had been implanted to secure clouds’ data in combination with other technique performed better than others.

Keywords:
Cloud computing, deep learning, Biometric, Privacy in cloud computing

1. INTRODUCTION (HEADING 1)

Cloud computing is defined as a technology that relies on transforming programs from products to a websites’ services. It is characterized by solving problems of program maintenance and development for variety of businesses. It turned information technology programs from products to services by transferring computer processing, storage, and data to the so-called cloud, which is a server device accessed via the Internet. It is characterized by eliminating challenges of program maintenance and development for enterprises. As a result, the beneficiaries’ efforts are solely focused on using these services. Thus, cloud computing is defined as a technique that relies on shifting the computer's processing and storage space to the so-called cloud, which is a server device that can be accessed via the Internet. From items to services, information is available. [¹]

Cloud computing have many types included in terms of prevalence, the National Institute of Standards and Technology (NIST) has identified four cloud computing models:
A- Public cloud computing
B- Private cloud computing
C- Community cloud computing
D- Hybrid cloud [²]
We can classify the types of services provided by cloud computing as:
A- "Infrastructure as a Service (IaaS)"
B- "Platform as a Service (PaaS)"
C- "Software as a Service (SaaS)"
D- Cloud Computing "Data as a Service (DaaS)"
E- "HARDWARE AS A SERVICE (Haas)" [³]
The cloud computing has the following important issues:

- Cyber security and security risks: Despite the benefit of multiple data entry points in cloud computing, it faces a significant challenge of security threats pertaining to this data. So, it is preferable to store data of each client in a private server under your own. So, you do not have to worry about data leakage if it is stored on other servers. You may also secure your cloud using "multi-factor authentication" to safeguard it from any attack or security breach.

- Data migration is difficult: The stage of data migration from conventional to cloud computing is one of the most challenging and time-consuming issues for users. It necessitates a change in management approach.

- Ability to comply to operational guidelines: As this challenge is currently regarded as one of the most significant, companies are concerned about the future growth of their business and their suitability with cloud computing. As well as the existence of gaps in operational processes and traditional information technology infrastructure. Thus, the nature of the companies’ work can be hampered if they do not have the infrastructure.

### 2. The Related works

Several surveys and comparative studies had been presented to evaluate performance of the recently proposed methods. Those studies relied on the security issues that might be implemented for unauthorized access or data deletion. In this section related works would be discussed to reshape the current study.

(Basu et. Al., 2018), the authors had presented a review of the security schemes that enhanced the security issues of the cloud. The listed comparisons among them according to data confidentiality and virtualization confidentiality of the proposed security of the cloud [1] (Kumar, 2019) had presented a review of the recently proposed solutions that might prevent the unauthorized access to users’ data in the cloud. In fact, the reviewed papers focused on the traditional encryption strategies with no pure comparative among them [8]. (Al-Sit, et al., 2019) had presented a review study to evaluate the security issues of several cloud computing systems, the focused on the encryption system that implemented in those cloud systems. Several attacks had been discussed that may prevent unauthorized person to access the users’ data. All of these systems do not presented deep learning methods [9]. (Xu, et al. 2019) had presented a review study that dealt with some of cloud computing security issues in deep learning-based clouds. They are attacks, countermeasures and their opportunities [10]. (Balani & Varol, 2020) had presented a survey for the challenges and threats that faced both users and providers of clouds. They listed some of the most important security issues in cloud computing. No comparison among the proposed security schemes at all [11]. (Abdulateef et al., 2020) presented a survey in which machine learning algorithm had been implemented to enhance cloud security. The authors present advantages and disadvantages of each listed paper but no comparison had been presented [12]. (Tariq, et al., 2020) had proposed some important countermeasures to evaluate the risk and challenges of the deep learning-based models that may be vulnerable for several attacks in cloud computing [13]. (Tahirkhel et al., 2021) had presented a survey that dealt with the security of clouds according to several issues. The study listed nine different clouds and presented the comparative criteria among them. Only one announce about Deep learning algorithm to prevent malware activities [14]. (Mohammed & Taha, 2021) presented a comprehensive study about homomorphic encryption to enhance the security and privacy of the cloud computing. No deep learning had been listed among the study papers [15]. (Andi, et al., 2021) had presented a review of the implemented methods and systems that are implemented in cloud computing security. They had listed the common clouds and the security methods, they are (blockchain, Deep Learning and Cryptography. Limited information had been reviewed for the deep learning papers with limited criteria [16].

### 3. The method:

This paper focused on securing the cloud using deep learning schemes. It is one of the very successful methods, especially since the methods of attack are always evolving and changing. It is needed to neural networks that are constantly evolving with every penetration attempt and every attempt to steal data that is stored in the cloud. The researcher also did not ignore the security of the cloud in different ways such as encryption, where the reader has the possibility to compare between the methods and take advantage of the methods that are more successful.

#### 3.1 Works related to deep learning field

In [17], the authors proposed a new deep learning model that based on CNN and RNN to detect intruders in cloud computing security. Their model was trained and tested by NSL-KDD dataset. Using deep learning model allowed to detect unapproved traffic and prevent access to the cloud. The result showed 99.86% accuracy for 5-classes.

In [18], the researchers presented a method to prevent risk in the cloud to quantify level of risk based on fine-grained model in co-residency. A large-scale dataset had
been built and implemented. Data had been collected via Microsoft Azure Platform. The method proposed some features to describe the patterns of normal subscribers (tenants) behavior. The Tenants are clustered into multiple categories by using DBSCAN algorithm and MinPts algorithm. Then, Deep Neural Network (DNN) is implemented to detect high-risk group according to quantification of risk component and control normal behavior pattern among the Tenants. The results showed the robustness of the proposed model specially with new data. Its accuracy was between (94.98%) and F-score were (0.99, 0.773 and 0.609) for Normal, Periodically Active and Extremely Active respectively.

In [19], the authors implemented the Distributed Deep Learning DDL to enhance the security of the fog-cloud computing with preserving privacy. The proposed scheme based on a trusted authority that provides parameters and public keys. Then, users’ authentication would be performed and local gradient uploaded the out put of CNN would be distributed on users. It is similar to other works and it is promising scheme.

In [20], a modified Distributed Deep Learning (DDL) had been implemented in fog-cloud computing environments. The authors proposed Secure and Privacy-Preserving deep learning (SPDDL) fog-cloud computing. The SPDDL had been implemented Trusted Authority (TA) to provide private and public keys for the encryption algorithm. The results showed that the proposed method granted the privacy of users’ information and perform a secured authentication by adding more complexity to the key generation in comparing with other DDL algorithms.

In [21], the authors had improved the authentication process in the cloud computing by implementing Iris as biometric authentication. The proposal authentication is based on neural network model, CNN for features’ extraction and MLP for iris feature matching. The authentication was performed in the cloud. The results sowed that the proposed models performed good according to the time.

In [22], Authors presented a new framework to secure health-care systems that include huge private information about users. MSCryptoNet is proposed to secure the information via implementing the Convolutional Neural Network (CNN) in Key generation. The CNN improved the encryption system of the framework that aims to implement classifier in deferent encryption scheme and minimize the computational and communication cost. The results showed that the proposed frame work had high performance and CNN may be used in another encryption scheme.

In [23], the authors presented a Long Short-Term Memory (LSTM) based system to monitor the network traffic. This would allow to notice and detect the abnormal network traffic. In fact, it is a promising idea that enhance the security of the cloud via extracting features from the network traffic.

In [24], the authors proposed four deep learning models to detect the DDoS attack on the cloud. CNN, LSTM MLP and CCN+LSTM are implemented and compared with machine learning algorithms. The proposed models had less level of performance than the machine learning algorithms except the CNN+LSTM that had high performance level. This model may enhance the cloud security against DDoS attack.

In [25], the authors proposed Intrusion Detection Systems (IDS) in cloud computing security to detect some types of attacks. In which, the IDS can get information about other suspected attack from other IDSs then made a decision using the aggregation algorithm. This matter increased execution time. Therefore, Denoising Autoencoder (DA) is used to build block to construct a deep neural network. The results showed that the system had 95% of accuracy.

In [26], the authors presented a detection method for anomaly traffic in the cloud that based on combination of two deep learning models: CNN and Gray Wolf Optimizer GWO. Those models detect abnormal traffic in a cloud datacenter. The result showed that the proposed method had improve the accuracy of anomaly detection that reach to 8% in compared with other metods.

In [27], the authors proposed a Privacy-Preserving Deep Learning Model (PDLM) to add multiple keys to the encryption system in the cloud. The initial optimization of data had been done by Stochastic Gradient Descent (SGD) to normalize the users’ data. The PDLM determine extract the perfect Tylor Series that implemented to calculate the accuracy and weight. Both private and public key for each user would be produced based on fit series that discovered by PDLM. The results showed that PDLM is effective and efficient.

In [28], the authors proposed a deep learning model that implemented to evaluate the security parameters of a multi-party cloud system. This model encrypted the databases of the users locally then preform the CNN to extract the security features of the hole databases. In fact, it is mater of evaluation not an enhancement.

In [29], the authors presented a scheme based on Privacy-Preserving Machine Learning to generate multiple keys for each user of the cloud. Firstly, users’ data had been optimized by SGD. the optimized data used in keys.
generation by Privacy-Preserving Machine Learning. Users’ data had been encrypted by two encryption systems. they are “Multi-Key Fully Homomorphic Encryption (MK-FHE)” and “the double decryption mechanism with Fully Homomorphic Encryption” (FHE). The results showed that both proposed schemes are efficient but the second scheme needed interaction among the users.

3.2. Work relate to cloud Security field:

Many researchers had developed and presented their proposed schemes that based on developed and recently proposed encryption algorithms. These algorithms did not implement neither Deep learning nor machine learning methods. The following are some of these studies. In [10], the author proposed a complex encryption system based on both symmetric and asymmetric encryption systems. A Blowfish block cipher as a symmetric encryption had been used in combination with ECC as asymmetric encryption. The blowfish encrypt the data then, ECC would encrypt the key of Blowfish. This proposal is suppose to increase the security of the user data.

While In[11], the author proposed an effective approach based on homomorphic encryption that implemented in banking application. The approach relies on encoding the bank's data then, it may be transmitted securely to the cloud. This approach may be used with confidential data. In [12], the authors presented a scheme based on two concepts. Firstly, the user verification would be performed by depending fingerprint to authenticate user. Secondly, encryption process which is based on One-Time-Password (as secret key) that had been exchanged between the cloud server and user by which AES block cipher used this secret key to encrypt/decrypt the user’s data. The results showed that implementing biometric is a good solution for cloud security.

For the security of cloud computing, in this research, a hybrid verification technique is made which is based on biometric and encryption systems. In order to achieve strong and secure technique, this work uses fingerprint as biometric technique and advanced encryption standard as a trustworthy encryption system. The primary goal of this paper is to avert information access from cloud information stockpiling focuses by unapproved clients. This new data security system can provide efficient authentication of cloud computing.

Data is a collection of information that is combined into one and has a very important meaning. In the study, the object to be secured is the password data, the encryption method Advanced Encryption Standard (AES) with a key length of 256 bits, before the data is encrypted with method of AES, the first password will be encrypted using MD5, and the second one will be encrypted again using the AES256 method. Based on trial data conducted through two sites about the complexity of passwords, it can be concluded that the original data (before encryption) no 1 to 5 increased to 32 bytes after being encrypted by the MD5 method, and its size increased again to 88 bytes after being encrypted by the AES256 method. Data can be obtained by value original data AES256 is 9, 96 times larger than its original size, would be but the value of the complexity of her also increased in line with the increase in the number of characters or the byte size of the data password above, thereby increasing the level of difficulty by the party that will hack the login data in the cloud.

In [13], the authors proposed a combination of encryption system to encrypt passwords during authentication. Their proposal based on encrypting password by MD5 encryption then re-encrypted by AES. The store password of the cloud is encrypted using AES too. This scheme can protect password of the user and added more complexity.

In [14], The authors proposed a backend application to provide immediate solution to prevent unauthorized access to the cloud. Footprint Notifications, Logging Encryption and activities of the authenticated users had been used to authenticate the user access. It is similar to behavior detection but it used events of the cloud to authenticate user.

In[15], the authors proposed an enhancement for cloud security. Two steps had been performed. Firstly, the users’ data would be encrypted by AES with it private key, then the AES key would be encrypted by Elliptic Curve Cryptography (ECC). Secondly, the encrypted key would be hide in a user image by HLSB algorithm of steganography. The implementation of the second encryption secured the AES key which is embedded in an image. This prevents eavesdroppers to get the private key.

In [16], the authors enhanced scheme that implemented Fingerprint as biometric to the identification phase of the scheme. Basically, the proposed scheme is called B³EIS-RG. It consists of two phases. Firstly, a key generation had been performed by DIV to determine the multi keywords among the huge data of the users in a cloud. The key generation depended on the extracted keywords. Then, Paillier cryptosystem had been implemented to encrypt users’ data. In the identification phase, the fingerprint features had been extracted and converted to vectors. These vectors had been encrypted during the Identifying process. The results showed that depending on vectorized biometric features reduced the required space to save biometric images and granted authenticated access to the user’s data.

In [17], the authors proposed a security scheme to protect the medical records in a public server that are connected to several medical institutes. The idea of their proposal based on additional cloud security scheme that allows access for the authenticated institutes. The user data had been encrypted by Shamir’s Secret Share (SSS). The proposed scheme allows whole institute to share medical data via Virtual private Network VPN. The results showed that the security scheme preform better than other schemes.

In [18], the authors enhanced the security of cloud computing by allow Third Party Auditor (TPA) which can
check the integrity of users’ data. Data check would be performed in the cloud on behalf of the users. In this scheme, when a user tried to check the integrity of his stored data, he would send a request to the check from TPA. The TPA requested the signature of the user data from the cloud server. The server built a signature for the user data. A hash code is implemented to match both of TPA and server signatures to perform data check. The results showed that the TPA checked the user’s data without knowing any details about it. In [40], the authors presented a complicated secured scheme that consisted of three phases. Firstly, the enrolment phase in which the user’s data and fingerprint had been encrypted by RSA key of the server then sent to a third-party server. The third part server decrypted data by RSA. The key and fingerprint template had been encrypted by AES and stored in the cloud. In Authentication Phase, user provided his user-name, password and fingerprint; they all would be encrypted by RSA then send to the third-party server. The user authentication data decrypted from the cloud by AES to be compared with the provided data. If they matched, the scheme granted access. If not, access would be denied. In the Data protection phase, the key and fingerprint would be encrypted by 3DES then the user’s data would be encrypted by Blowfish. After saucerful authentication, the third party would decrypt the key and the fingerprint template by 3DES then the key would be decrypted by Blowfish. The results showed that the complexity of proposed design granted high security for the cloud users. In [29], the authors proposed a security model based on AES encryption system to encrypt the transmitted data. The proposed model had been implemented in the Heroku, as a cloud platform of (PaaS) type. The author concluded that the AES encryption needed more time to perform encryption and decryption processes. In [31], the authors presented fully data encryption in the cloud. The idea based on encryption of the data before user sends it to the cloud. The implemented encryption system was AES as a robust encryption. The proposed scheme provided cloud security and user privacy. It also reduced the noise of encryption in the cloud.

Both of the deep learning based and the traditional schemes had been illustrated in table1 and 2. The common countermeasures had been discussed for each presented work in each category. The deep learning-based scheme had some countermeasures that can be listed as a promising idea for further enhancement of security of the cloud computing. They consist of strategy, Deep learning techniques, other requirements, goals and achievements as shown in table 1.

Table 1. The Deep learning-based papers countermeasure.

<table>
<thead>
<tr>
<th>No.</th>
<th>strategy</th>
<th>Deep learning techniques</th>
<th>other requirements</th>
<th>goals</th>
<th>Achievements</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Anomaly traffic</td>
<td>CNN, RNN</td>
<td>-</td>
<td>Detect abnormal network traffic by CNN and RNN to extract traffic features and prevent the anomaly traffic</td>
<td>The scheme had 99.86% of accuracy</td>
</tr>
<tr>
<td>19</td>
<td>Abnormal behavior</td>
<td>DNN</td>
<td>DBSCAN MinPts</td>
<td>Prevent abnormal activities</td>
<td>The scheme had 94 - 98% of accuracy</td>
</tr>
<tr>
<td>20</td>
<td>Preservation privacy</td>
<td>DLL, CNN</td>
<td>Parameter and public key providing</td>
<td>Provides trusted Authority</td>
<td>Secure fog cloud</td>
</tr>
<tr>
<td>21</td>
<td>Preservation privacy</td>
<td>SPDDL</td>
<td>Public key providing</td>
<td>Provides trusted Authority</td>
<td>Secure fog cloud</td>
</tr>
<tr>
<td>22</td>
<td>Biometric Authentication</td>
<td>CNN, MLP</td>
<td>Iris biometric</td>
<td>More secured authentication</td>
<td>preformed good according to the time</td>
</tr>
<tr>
<td>23</td>
<td>Preservation privacy</td>
<td>CNN classifier</td>
<td>Key generation</td>
<td>Maintain privacy of health care cloud</td>
<td>preformed good with high privacy</td>
</tr>
<tr>
<td>24</td>
<td>Anomaly traffic</td>
<td>LSTM</td>
<td>network traffic’s feature</td>
<td>Detect abnormal network traffic by LSTM</td>
<td>preformed good</td>
</tr>
<tr>
<td>25</td>
<td>Anomaly traffic</td>
<td>CNN, LSTM, MLP, CCN+LSTM</td>
<td>Machine learning methods network traffic’s feature</td>
<td>enhance the cloud security against DDoS attack</td>
<td>The machine learning performed better than deep learning but the combination of CCN+LSTM is performed better than all.</td>
</tr>
<tr>
<td>26</td>
<td>Anomaly traffic</td>
<td>Denoising Autoencoder</td>
<td>aggregation</td>
<td>Build Intrusion Detection Systems for the cloud computing</td>
<td>The scheme has 95% of accuracy</td>
</tr>
<tr>
<td>27</td>
<td>Anomaly traffic</td>
<td>CNN, Gray Wolf</td>
<td>-</td>
<td>detect abnormal traffic in a cloud datacenter</td>
<td>More accuracy reaches to 8%</td>
</tr>
</tbody>
</table>
The illustrated papers above are mostly dealt with preserving privacy of users’ data and anomaly traffic in the cloud computing. Just one paper dealt with biometric authentication and another dealt with user’s behavior. Most of the papers implement CNN with other deep learning algorithms. The combination had performed better than CNN based scheme alone. Deferent encryption system had been used to encrypt user’s data. The listed achievements may enhance the privacy and security and trusted authentication process.

The traditional schemes are illustrated in table 2. Another two countermeasures had been used. The main method and the auxiliary methods instead of the Deep learning method and other requirements.

<table>
<thead>
<tr>
<th>No.</th>
<th>strategy</th>
<th>main method</th>
<th>auxiliary methods</th>
<th>goals</th>
<th>Achievements</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Preserving privacy</td>
<td>PDLM</td>
<td>Tylor Series Private and public key</td>
<td>add multiple keys to the encryption system in the cloud</td>
<td>PDLM is effective and efficient</td>
</tr>
<tr>
<td>29</td>
<td>Preserving privacy</td>
<td>CNN</td>
<td>the security features of the hole databases</td>
<td>Enhance privacy of the cloud based on whole data</td>
<td>Evaluate the performance for the features</td>
</tr>
<tr>
<td>30</td>
<td>Preserving privacy</td>
<td>CNN SGD</td>
<td>MK-FHE DFHE</td>
<td>Secured Users’ data</td>
<td>More interaction of users is needed</td>
</tr>
</tbody>
</table>

Table 2. The Traditional paper based countermeasure.
The authors proposed encryption scheme in most of the papers, while some other proposed biometric authentication. Signature authentication and abnormal behavior had been proposed one paper for each. Double encryption schemes had a good implementation as well as whole data encryption but they require more time. Biometric authentication is a good solution based on extracted string. Both schemes are implemented in deferent fields and businesses because of their high performance.

5. CONCLUSION
Cloud computing has emerged as a practical and optimal solution following the availability of Internet infrastructure in various parts of the world, and the issue of communication does not constitute an impediment to touching the cloud, specially, in light of the massive increase in smartphone issuance, which always carry with it the characteristics of Internet connection and the ability to deal with a wide range of data and files over the network, particularly multimedia, and one of the most significant challenges confronting the cloud computing system are security and privacy, where there is concern about confidentiality violations. Both of traditional and deep learning-based schemes had the capability to increase the levels of data security and privacy. It is obvious that biometric based schemes had met the deep-learning schemes in their capabilities. Behavior strategies can offer more security to the cloud to prevent unauthorized access. While the privacy issue had great interest in the cloud’s data privacy. Deep learning-based method had been implanted to secure clouds’ data in combination with other technique performed better than others.

Reference
[19] Jin Han, Wanyu Zang, Meng Yu and Ravi Sandhu "Quantify Co-Residency Risks in the Cloud through Deep Learning" 1545-5971 (c) IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission 2020.
[23] Jin Han, Wanyu Zang, Meng Yu and Ravi Sandhu "Quantify Co-Residency Risks in the Cloud through Deep Learning" 1545-5971 (c) IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission 2020.


