



HAL
open science

A Security Method for Cloud Storage Using Data Classification

Oussama Arki, Abdelhafid Zitouni, Mahieddine Djoudi

► **To cite this version:**

Oussama Arki, Abdelhafid Zitouni, Mahieddine Djoudi. A Security Method for Cloud Storage Using Data Classification. *International Journal of Grid and High Performance Computing*, 2023, 15 (1), pp.1-17. 10.4018/IJGHPC.329602 . hal-04722702

HAL Id: hal-04722702

<https://hal.science/hal-04722702v1>

Submitted on 5 Oct 2024


HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.


Copyright

A Security Method for Cloud Storage Using Data Classification


Oussama Arki, University of Abdelhamid Mehri – Constantine II, Algeria*

 <https://orcid.org/0000-0002-5943-1919>

Abdelhafid Zitouni, University of Abdelhamid Mehri – Constantine II, Algeria

 <https://orcid.org/0000-0003-2498-4967>

Mahieddine Djoudi, University of Poitiers, France

 <https://orcid.org/0000-0002-2998-5574>

ABSTRACT

Cloud computing is an information technology model that provides computing and storage resources as a service. Data storage security remains the main challenge in adapting this new model. The common solution to secure data in the cloud is data encryption. However, handling all the data with the same security policy does not appear to be good practice, because they do not have the same sensibility for the data owner. The present research proposes a new method to improve the security of data in cloud storage. It combines two domains represented by machine learning and multi criteria decision making, in order to provide a new classification method, that classifies data before being introduced into a suitable encryption system according to their category. A Cloudsim simulation has been used to demonstrate the effectiveness of the proposed method. The results of the simulation exhibit that our method is more efficient and accurate and takes less processing time, while ensuring data confidentiality and integrity.

KEYWORDS

Classification, Cloud Storage, Confidentiality, Data Security, Encryption, Integrity

1. INTRODUCTION

Cloud computing is a new model for providing diverse services of software and hardware. This paradigm refers to a model for enabling on-demand network access to a shared pool of configurable computing resources, that can be rapidly provisioned and released with minimal service provider interaction (Mell, 2011). It helps organizations and individuals deploy IT resources at a reduced total cost. However, the new approaches introduced by the clouds, related to computation outsourcing, distributed resources and multi-tenancy concept, increase the security and privacy concerns and challenges (Singh, 2016). Besides all the services delivered by the cloud provider, cloud also presents storage as a service (Zardari, 2014). Cloud storage is a model of networked online storage where data

DOI: 10.4018/IJGHP.329602

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

is stored on multiple virtual servers (Pravin, 2013). It allows users to store their data remotely and then access to them at any time from any place (Spoorthy, 2014). Cloud storage services are used to store data in ways that are considered cost saving and easy to use (Wu,2010).

In cloud storage, data is stored on remote servers that are not physically known by the consumer. Thus, users fear uploading their private and confidential files to cloud storage due to security concerns (Ogigau-Neamtii, 2012). The usual solution to secure data is encryption, which makes cloud users satisfied when using the cloud to store their data. However, handling all data with the same security policy, without taking into account the level of data sensitivity does not appear as a good strategy.

Motivated by the above facts, we have proposed a solution to undertake the problem of cloud storage security. In cloud storage, there is public data that does not need any security measures, and there is sensitive data that needs applying security mechanisms to keep them safe. In that context, data classification appears as the solution to this problem. The classification of data into classes, with different security requirements for each class is the best way to avoid under security and over security situation. The existing cloud storage systems use the same key size to encrypt all data without taking into consideration its confidentiality level. Treating the low and high confidential data in the same way and at the same security level will add unnecessary overhead and increase the processing time. In our proposal, we have combined the K-NN (K-Nearest Neighbors) machine learning method and the goal programming decision-making method, to provide an efficient method for data classification. This method allows data classification according to the data owner's security needs. Then, we introduce the user data to the suitable security mechanisms for each class. The use of our solution in cloud storage systems makes the data security process more flexible, besides; it increases the cloud storage system performance and decreases the needed resources, which are used to store the data.

The rest of this paper is organized as follows: Section two is about related work. In Section three, we present our proposed cloud storage security method that is simulated and discussed in section four. Finally, section five concludes this work.

2. RELATED WORK

In the literature, many works have used classification to maintain data security in the cloud, the proposed solutions could be categorized into two classes: the manual classification (defined by the user) and the intelligent or automatic classification (using an algorithm).

From a manual perspective, several contributions have been proposed. In (Yahya, 2015), they proposed a framework focusing on a varied level of security. They considered several security classifications levels: protected, sensitive and top secret. Initially, a protected level involves data protection with a password. A sensitive level may need extra protection such as having multi-factor authentication. Finally, a top-secret level may need to be fully encrypted even by the CSPs system administrators. In their work, they do not conduct any results, and they based only in manual classification.

In (Tawalbeh, 2015), they have proposed a confidentiality-based cloud storage framework. In their work, they have proposed a secure cloud storage model that encrypts data according to its confidentiality degree through three levels: basic, confidential, and highly confidential. The proposed solution is based on the idea of manual classification. For each level, they have used a different mechanism of security, which increases the performance of their framework compared with the existing cloud storage systems.

In (Patel, 2016), they have proposed a secure model for cloud computing by using data classification methodology. The data that resides in the cloud is secure in two ways, firstly by encryption and secondly with authentication scheme. In the proposed framework they have classified the data according to their sensitivity into different levels: basic level, sensitive level, and highly confidential level. Different authentication techniques and security techniques are used at different

levels; besides they have used the concept of file splitting. The main reason for using file splitting is to increase the execution time.

From an automatic perspective, many solutions have been proposed. In (Zardari, 2014), they used the K-NN machine learning technique in the cloud to solve the data confidentiality problem. They classified data into two classes (non-sensitive and sensitive data). To separate sensitive and non-sensitive data, the KNN classifier is used in a designed simulation environment. After classification, the sensitive data is further transferred to the RSA encryption algorithm for data encryption to protect sensitive data from unauthorized users. However, the public data is directly allocated a VM without encryption.

In (Shaikha, 2015), they used data classification for achieving data security in the cloud. In this work, the business value of data is identified based on their usage and access control restrictions. They have identified a set of parameters for data classification. For simulating the classification of data, they based on personal dataset, and used the subjective criteria to classify them and accordingly, security provisions for the storage and communications can be incorporated. Based on that, classification provisions can be given for storage and communication encryption, integrity, and access control mechanisms.

In (Kaur, 2016), they have proposed a data classification model for achieving data confidentiality and integrity in the cloud. They proposed a secure data classification model using a supervised machine learning technique. The classification depends on the attributes of the data. In this work, data is classified according to its sensitivity level, then only the data that is required to be secured is encrypted. The proposed work also ensures the privacy and integrity of data using a hashing approach.

In (Ennajjar, 2017), they have proposed a model for securing data in cloud by classification. It suggests a classification model to categorize data before being introduced into a suitable encryption system according to the category. In technical terms, the approach is based on a multi-criteria classification scheme. It uses different criteria proposed by the data owner or the cloud provider to comprehend their needs in terms of security and cost.

In (Tamanna, 2017), they have proposed a secure cloud model using classification and cryptography. The focus of the research was to characterize the data considering the security prerequisites that divide the data into basic, confidential, and highly confidential data using improved machine learning algorithm. In this work, data is classified according to its sensitivity level, then they encrypted only, the data that is required to be secure using a cryptographic technique in the cloud. The proposed system has been simulated using CloudSim simulator. The results depict that the proposed technique is more relevant than storing the data without deciding the security needs of the data.

In (Kiran, 2017), their work aims to achieve data confidentiality, data access control management and provides an automatic classification of data in the cloud. In this paper, the functionality and performance of existing K-NN technique is improved with modified ensemble learning technique. The data is automatically classified by the machine on the basis of data security parameters. The proposed system proves to be more accurate and economical. And also saves the user time for encrypting and decrypting different classes of data (basic, confidential and highly confidential).

Table 1 shows a comparison of our proposal, and the existing related work. The comparison is based on two main criteria: the used technique, the data owner implication during the classification process.

3. A SECURITY METHOD FOR CLOUD STORAGE USING DATA CLASSIFICATION

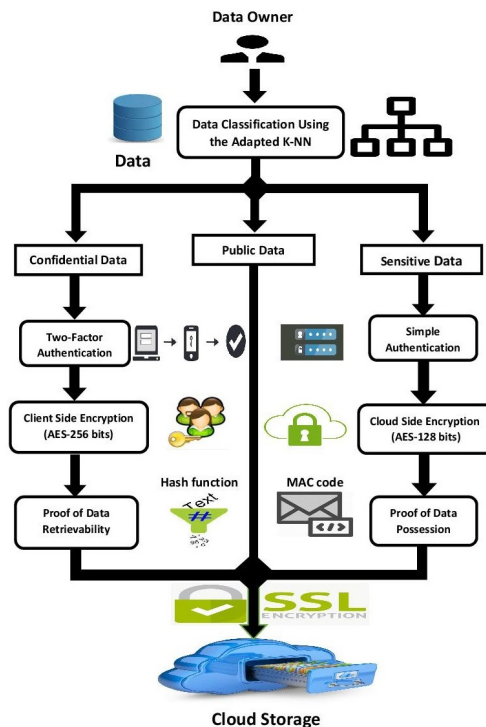
3.1 Overall Architecture

The objective of data classification is to establish the required level of security for data, by providing a sufficient level of security (Sawle, 2016) and identify sensitive data from non-sensitive data (Sood, 2012). Figure 1 shows the proposed method.

Table 1. A comparison table of related work

Work	Technique	Owner Implication	Remarque
(Yahya, 2015)	manual classification	The data owner is implicated	The use of manual classification
(Tawalbeh, 2015)	manual classification	The data owner is implicated	The use of manual classification
(Patel, 2016)	manual classification	The data owner is implicated	The use of manual classification
(Zardari, 2014)	Machine learning (KNN)	No implication	The common problem of KNN method is not treated, witch impact the accuracy of the model
(Shaikha, 2015)	Automatic	No implication	The use of subjective criteria
(Kaur, 2016)	Machine Learning (Naïve Bayes and decision stump)	No implication	The use of decision stump could overfit their model, it could be avoided using Adaboost model.
(Ennajjar, 2017)	Multi-criteria classification scheme	The data owner is implicated	It is just a prototype of the proposed method
(Tamanna, 2017)	Machine learning: Naïve Bayes and KNN	No implication	The common problem of KNN method is not treated,
(Kiran, 2017)	Machine learning: Ensemble learning and KNN	No implication	The common problem of KNN method is not treated,
Our proposal	The use of both machine learning and multi-criteria decision methods	The data owner is implicated during the process of classification	The combination of KNN and goal programming multi-criteria decision method to overcome the common problem of KNN

Figure 1. The data classification method



3.2 The Proposed Method

In our situation, we have defined three classes of data (public, sensitive, and confidential), so we need supervised machine learning to separate sensitive from non-sensitive data. For that, we proposed a new method of data classification, using both the K-NN supervised machine learning method and the goal programming multi-criteria decision-making method. The classification method consists of three steps:

Step 1: Data security requirements specification

Data can be sensitive or non-sensitive, with different security requirements. These security requirements depend on the sensitivity levels of data (Zardari,2014). In this step, the data owners specify their security needs. Those security parameters are sensitivity, access control and integrity of data. For each need defined by the user, a security mechanism is applied as a security measure. Table 2 shows the correspondence between the security requirements and its security mechanisms.

Step 2: Data Classification

It is better to understand the exact security requirements of data before transferring it to the cloud, which is only possible through machine learning techniques (Zardari, 2014). In this step, we classified data using our proposed method into three classes, where each class has a specific sensitivity level.

- **Confidential data:** Those data are highly sensitive and important for the data owner, such as business data, medical data, and financial transactions. No one should be able to access this data, except their owners and the entities authorized by them. In this class, even the cloud provider is not allowed to access this kind of data, so the access is strictly limited and well defined.
- **Sensitive data:** These data are sensitive and important for the owner of data, such as personal data (photo, video, file), nobody should be able to access them, except their owners and the entities authorized by them. Thus, access is limited only for authorized entities.
- **Public data:** This data is not sensitive or important for the data owner, such as shared data for everyone (photo, video, file), so access is allowed for everyone.

In machine learning, the K-NN method is used for classification purpose. To adapt the K-NN method to our context, we have used the goal programming method. Which is used in a multi-criteria decision situation based on the local aggregation of criteria. The idea is to establish a goal level of achievement for each criterion. It requires the decision maker to set goals for each objective that he wishes to obtain. A preferred solution is then defined as the one which minimizes the deviations from

Table 2. The correspondence between the security needs and the security mechanisms

Security needs	The corresponding security mechanisms
The data is public	No Encryption + Transport Layer Security
Data is private	Authentication + Encryption + Integrity check
The cloud provider can access to the data	Cloud side encryption with AES-128 bits
The cloud provider cannot access to the data	Client-side encryption with AES-256 bits
data must be intact in the cloud storage	Proof of Data Possession (MAC Code)
data have to be retrievable from the cloud	Proof of Retrievability(Hash Function)

the set goals (Izadikhah, 2014). In our situation, we have three classes, so for each class, we have defined a set of security mechanisms to respect. Based on the user’s security needs that are defined in the previous step. First, we have evaluated the user’s data compared to the security requirements. Then, we calculated the distances between the evaluation of the user’s data and the predefined evaluations of the three classes, using the formula of the K-NN method. For the distance calculation, we need numbers, so for each security mechanism applied, there is a digital value to represent it, if we apply this mechanism, its digital representation takes this value. Otherwise, the mechanism is represented by the value 0. Table 3 presents the correspondence between the used security mechanisms and their digital values. Table 4 presents the existing classes and the matches between the security mechanisms and their digital values for each class.

The calculation formula:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

- D: Distance between the user data evaluation and the evaluation for each class.
- x: The digital value of the security mechanism that is applied to the user’s data.
- y: The digital value of the mechanism that is applied on the predefined class.

This formula is applied for each class. In the end, we get three different distances and the user’s data is classified into the class of the little distance.

Table 3. The correspondence between the security mechanisms and their digital values

Mechanism of Security	Digital Correspondence
Cloud-side encryption with AES 128 bits	25
Client-side encryption with AES 256 bits	50
Proof of Data Possession (MAC code)	25
Proof of Retrievability(Hash Function)	50
Single authentication	25
Two-factor authentication	50
Transport Layer Security (SSL)	25

Table 4. The correspondence between each class of data and the digital values of their security mechanisms

Mechanism of Security	Public data	Sensitive data	Confidential data
Cloud side encryption with AES-128 bits	0	25	0
Client-side encryption with AES-256 bits	0	0	50
Proof of Data Possession	0	25	0
Pouf of Retrievability	0	0	50
Transport Layer Security	25	25	25
Simple authentication	0	25	0
Two factor authentications	0	0	50

Step 3: Application of the security mechanisms

To address the cloud security issues in cloud storage, we have applied a set of security mechanisms to protect the data that resides in the cloud. Thus, we chose to apply all the security measures that can be used to secure an information system, especially for a cloud-based system, like the encryption of data, authentication, and data integrity check techniques. Those techniques are explained as follows:

- **Authentication:** Authentication remains a fundamental safeguard against illegitimate access to the device or any other sensitive application, whether offline or online (Boyd, 2013) (Mohsin, 2017) (Pathan, 2016):
 - **Single authentication:** As his name shows, that single factor means only one factor is responsible to validate the user. It means “that the user knows”. A single layer of security is provided in this scheme. The most recognized type of single factor is the password. In the username and the password is the factor which makes a single layer of security (Patel, 2016).
 - **Two-factor authentication:** As his name suggests, two factors mean here two factors are responsible for the protection of the data. In addition to the single factor, the user has another factor such as two-step verification. Also, one-time password is also a second factor (Patel, 2016).
- **Encryption of data:** Data encryption is one of the confidentiality techniques. Confidentiality can be achieved through proper encryption technique: symmetric and asymmetric algorithms (Yadav, 2013):
 - **Client-side encryption:** In this model, the data owner becomes responsible for ensuring the security. he has the control over the encryption keys. He also controls who can have access permission to his data (Gaur, 2016).
 - **Cloud-side encryption:** In this model, the security measures are taken up by the cloud-service providers. Here, along with providing cloud services, the cloud service providers adopt certain security measures to safeguard the data that is stored in the cloud (Gaur, 2016).
- **Data Integrity check mechanism:** Data integrity is one of the most critical elements in any information system. Generally, data integrity means protecting data from unauthorized deletion, modification, or fabrication (Sun, 2014):
 - **Proof of data possession:** This scheme checks that an outsourced storage site retains a file. The data owner pre-processes the file, generating a piece of metadata that is stored locally, transmitting the file to the server. The server stores the file and responds to challenges issued by the client. (Shuang, 2014).
 - **Proof of retrievability:** In this scheme, first the file is divided into blocks and then encoded with error correcting codes. Then check blocks called sentinels are embedded for each block. The verifier challenges the prover by specifying the positions of sentinels. The prover returns the respective sentinels (Desai, 2014).

These security mechanisms are applied, depending on the level of data sensitivity. So, for each data class, a set of security mechanisms is applied. Table 5 shows the applied security mechanism for each class.

4. SIMULATION RESULT AND DISCUSSION: CLOUDSIM PROJECT

To simulate our proposed method, we have based it on the use of CloudSim project. The primary objective of this project is to provide simulation framework that enables simulation, and experimentation of Cloud computing infrastructures and application services (Goyal, 2012).

Table 5. The applied security mechanisms for each class

Data levels	Authentication	Encryption	Integrity check	Https(SSL)
Public	No needs	No needs	No needs	SSL handshake
Sensitive	Single Authentication	Cloud-side encryption with AES-128 bits	Proof of data possession	SSL handshake
Confidential	Two-factor authentication	Client-side encryption with AES-256 bits	Proof of retrievability	SSL handshake

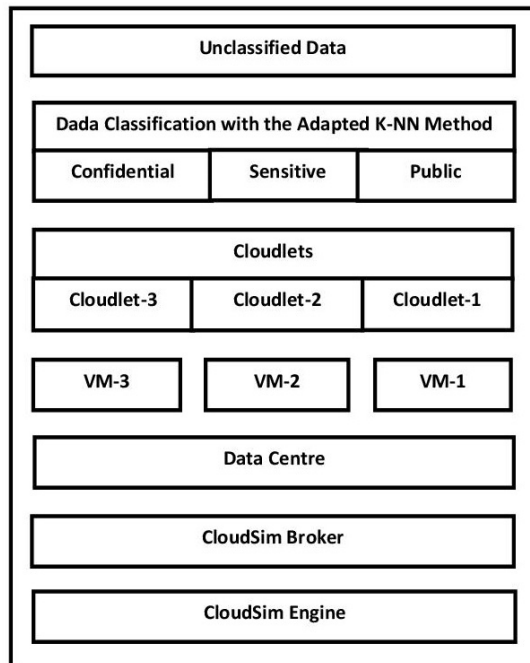
CloudSim supports both system and behavior modelling of cloud components such as data centers, virtual machines (Gupta, 2016).

In our proposed method, we have three classes of data, so we have decided to store each class in a dedicated virtual machine (VM). The storage process in CloudSim is presented with cloudlet. In CloudSim, the cloudlet represents a unit of the user-submitted task in the cloud that is executed in a virtual machine (Sahoo, 2017). Each cloudlet has specific parameters that distinguish each cloudlet, like the length, input file size and output file size. The length of cloudlet represents the needed computation power used by the virtual machine to perform this cloudlet. Figure 2 shows our CloudSim simulation architecture.

4.1 Properties and Description of CloudSim

Before starting the simulation, we have to set the properties of our simulation. In this simulation, we have created one data center with one host, which contains three VMs, each VM can execute one or more cloudlets. Besides, we have to define the allocation policy of the VMs in the data center, the VMs scheduler and the cloudlets scheduler. For that, we prefer to use the default allocation

Figure 2. CloudSim simulation architecture



policy of CloudSim (simple policy). It is the easy way, which means that we assign the VM to the host, which has the needed resource to host it. In our case, all virtual machines are allocated to the same host; because it has enough resources to host all of them. For the scheduler, we prefer to use a space-shared scheduler in the level of the host. Which means that only one VM can be executed by each processing element (Pe), as we have three processing elements in our host, so each one of them is reserved for a dedicated virtual machine. For the scheduler of the cloudlets (tasks), we prefer to use the same scheduler (space-shared), which means that each virtual machine has to finish the current cloudlet first and then move to the next cloudlets. This scheduler is selected in order to know the execution time for each cloudlet separately. Table 6 represents the properties of the data center, table 7 represents the properties of the host, table 8 represents the properties of the virtual machines and table 9 represents the parameters of the cloudlets.

4.2 Performance Discussion

For each class of data, a different cloudlet is executed, we have three cloudlets with different length. The length is selected according to the data class.

The public data is stored in the cloud with the execution of a lower cloudlet because it is stored directly in the cloud. However, for sensitive and confidential data, we apply a set of security mechanisms, which makes the treatment to store this data in the cloud more complex than public data. Therefore, the executed cloudlets to handle these types of data is more than public data.

Table 6. The properties of the data center

Data Center ID	Number of Hosts	Number of VM per Host	Architecture of Data	Operating System	Virtual m/c Manager
2	1	3	X86	Linux	XEN

Table 7. The properties of the host

Host ID	RAM	Storage	Band Width	Number of Cores	MIPS per Core (Pe)
0	16	150	15000	3	10000

Table 8. The properties of the virtual machines

VM ID	MIPS	RAM	Storage	Band Width	Number of VCPU
0	1000	4	50	5000	1
1	1000	4	50	5000	1
2	1000	4	50	5000	1

Table 9. The parameters of the cloudlets

Cloudlet ID	Length	Data class	VM ID	Data Center ID
0	10000	Public	0	2
1	20000	Sensible	1	2
2	30000	Confidential	2	2

We have considered 4 scenarios; the treatment of data as public, like in Table 10, as sensitive data, like in Table 11, as confidential, like in Table 12 and with our proposed method, like in Table 13.

As we can notice in Figure 3, our classification method provides a better result in start time, compared to the other scenarios. Figure 4 shows that our method provides almost the best execution time and Figure 5 shows that our method provides the best finish time than the other scenarios.

4.3 Encryption Discussion

In this step, we have evaluated our classification method from encryption perspective. For that, we have implemented our proposed encryption technique (after classification) and other existing methods, using java language. In this case, the scenarios of experiments are the encryption and decryption of different size files, using our proposed method of encryption and the existing techniques, which are, used in cloud storage systems such AES-128 bits and AES-256 bits algorithms. The encryption and decryption time is considered as the evaluation metrics. We performed experiments on Intel (r)

Table 10. Public data storage scenario

File size(mb)	Data class	Cloudlet ID	VM ID	Start time	Finish time	Execution (ms)
5	public	0	0	0.1	10.1	10
10	public	0	0	10.1	20.1	10
15	public	0	0	20.1	30.1	10
20	public	0	0	30.1	40.1	10
25	public	0	0	40.1	50.1	10
50	public	0	0	50.1	60.1	10

Table 11. Sensitive data storage scenario

File size(mb)	Data class	Cloudlet ID	VM ID	Start time	Finish time	Execution (ms)
5	sensitive	1	1	0.1	20.1	20
10	sensitive	1	1	20.1	40.1	20
15	sensitive	1	1	40.1	60.1	20
20	sensitive	1	1	60.1	80.1	20
25	sensitive	1	1	80.1	100.1	20
50	sensitive	1	1	100.1	120.1	20

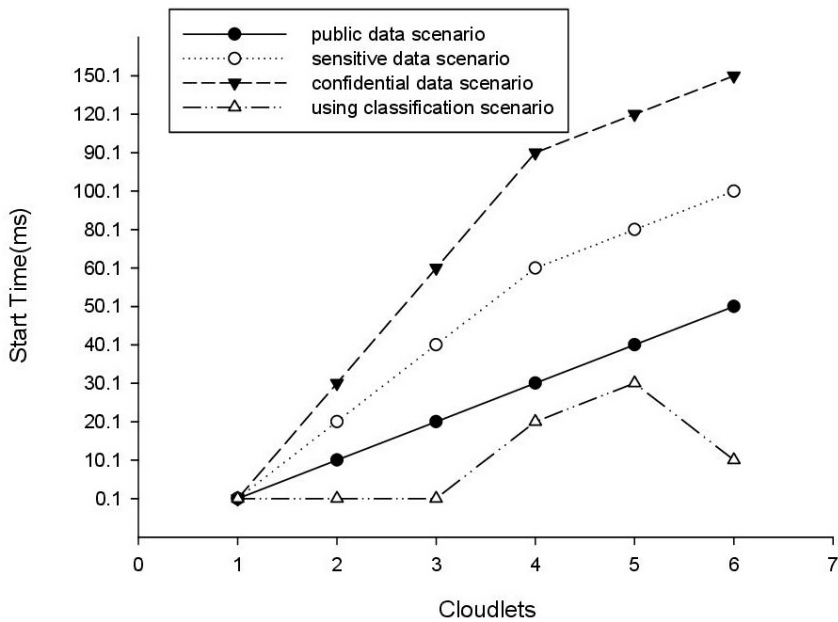
Table 12. Confidential data storage scenario

File size(mb)	Data class	Cloudlet ID	VM ID	Start time	Finish time	Execution (ms)
5	confidential	2	2	0.1	30.1	30
10	confidential	2	2	30.1	60.1	30
15	confidential	2	2	60.1	90.1	30
20	confidential	2	2	90.1	120.1	30
25	confidential	2	2	120.1	150.1	30
50	confidential	2	2	150.1	180.1	30

Table 13. Data storage scenario using classification

File size(mb)	Data class	Cloudlet ID	VM ID	Start time	Finish time	Execution (ms)
5	public	0	0	0.1	10.1	10
10	sensitive	1	1	0.1	20.1	20
15	confidential	2	2	0.1	30.1	30
20	sensitive	1	1	20.1	40.1	20
25	confidential	2	2	30.1	60.1	30
50	public	0	0	10.1	20.1	10

Figure 3. The start time results of cloudlets



core (tm) i5 3230 m CPU 2.6 GHz processor with 4 GB of ram on windows 8.1 operating system. We carried out experiments on 5 Mb, 10 Mb, 15 Mb, 20 Mb, 25 Mb and 50 Mb size text files. In this paper, we used samples, which are already classified using our technique. Table 14 shows the encryption /decryption results.

As we can notice in Figure 6, our classification method provides the best result in encryption time. Figure 7 shows that our method provides almost the best decryption time.

4.4 Classification Discussion

The problem with K-NN is that the number of K influences the result of classification, In K-NN algorithm it is difficult to select an appropriate value for K. This is a big disadvantage of the K-NN algorithm. To undertake this problem; we have adopted the principle of the multi-criteria decision method, called goal programming. The principle of this method is to define first a target like an ideal solution to reach and then looking for the nearest alternatives for this solution. For that, we have defined three targets, because we have three classes. In order to define the suitable target, we have calculated the distance between the new data and the predefined class of data, according to

Figure 4. The execution time results of cloudlets

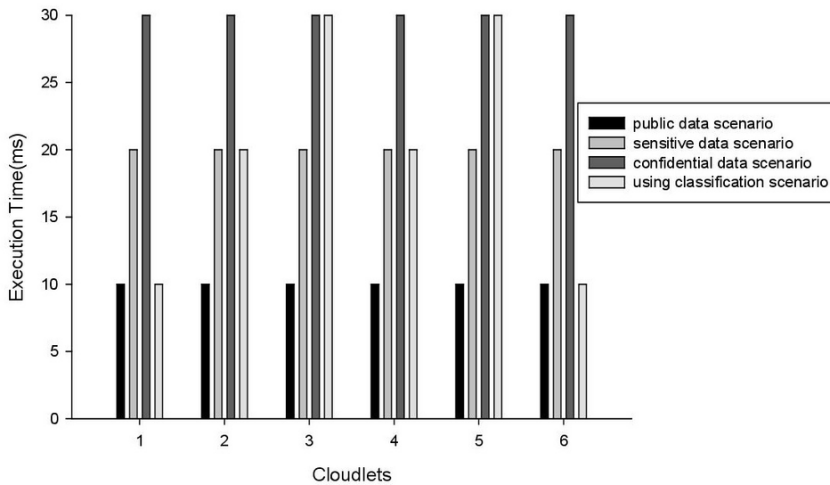
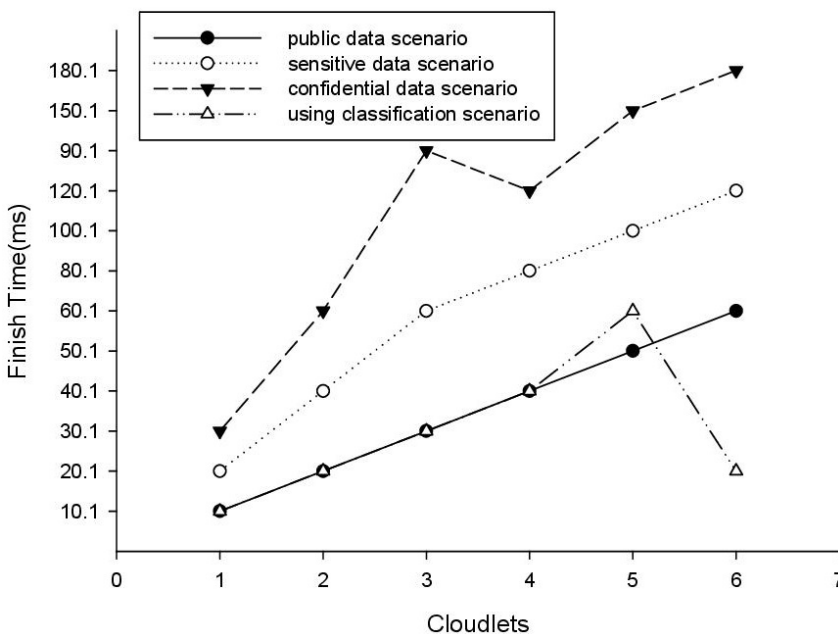


Figure 5. The finish time results of cloudlets

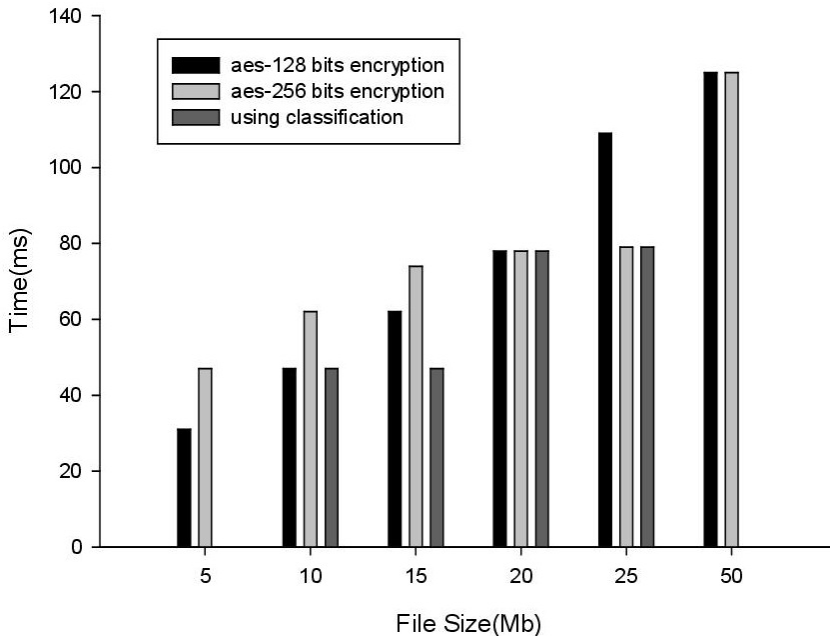


the security requirement defined by the user. Using the principle of goal programming method, we have surpassed the problem of accuracy known in K-NN classification situation. Table 15 shows the probability of the user's security needs definition and its corresponding mechanisms in our system. Table 16 shows the evaluation of those mechanisms for each scenario. Table 17 shows the evaluation of each class according to the security mechanism and Table 18 shows the result of the distance calculation, which shows that in each scenario; we have only one selected class for data. The combination of the multi-criteria decision maker method with the K-NN method improves the accuracy of our proposed method.

Table 14. The encryption /decryption time results

File Size (mb)	Data Sensibility	AES 128 bits		AES 256 bits		Using Classification	
		Encryption (ms)	Decryption (ms)	Encryption (ms)	Decryption (ms)	Encryption (ms)	Decryption (ms)
5	Public	31	62	47	78	0	0
10	Sensible	47	78	62	94	47	78
15	Confidential	62	94	74	109	47	109
20	Sensible	78	109	78	125	78	109
25	Confidential	109	125	79	140	79	140
50	Public	125	172	125	188	0	0

Figure 6. AES-128 vs AES-256 vs. Classification encryption results



5. CONCLUSION

The main concern of this paper was to propose an efficient cloud storage method, that provides confidentiality and integrity through data classification and applying security techniques. The focus of the research paper was to categorize the data according to the security requirements, using an improved K-NN method. In our method, the K-NN classifier is improved using goal programming decision-making method and it is used to classify data based on the security requirements of the data. The best way to improve the confidentiality and the computational capacity of the VMs and servers is through data classification. In this way, it is easy to know which data needs what security and which data does not need any security. Without data classification, the user may over-secure or under-secure his data. The proposed method has been implemented in a designed simulation environment using CloudSim. The simulation results show that our method achieves better processing time, while assuring data

Figure 7. AES-128 vs. AES-256 vs. Classification decryption results

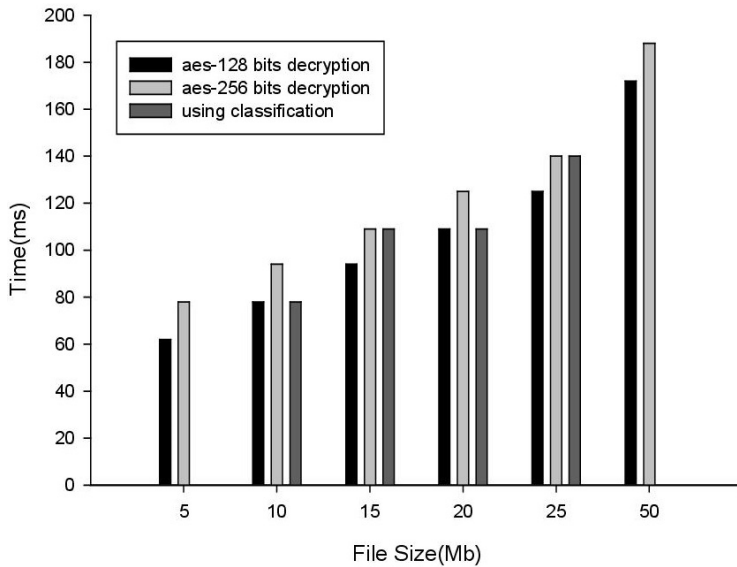


Table 15. The probability of the user's security needs definition

Probability	Sensibility	Access control	Integrity check
Case 00	Public	Not specified	Not specified
Case 01	Private	Cloud allowed	Data must be intact
Case 02	Private	Cloud allowed	Data have to be retrievable
Case 03	Private	Cloud not allowed	Data must be intact
Case 04	Private	Cloud not allowed	Data have to be retrievable

Table 16. Security mechanisms correspondence for each case

Scenario	SSL (https)	Encryption	Integrity check	Authentication
Case 00	25	0	0	0
Case 01	25	25	25	25
Case 02	25	25	50	50
Case 03	25	50	25	25
Case 04	25	50	50	50

Table 17. The evaluation of each class according to the security mechanism

Data classes	SSL(https)	Encryption	Integrity check	authentication
Public	25	0	0	0
Sensitive	25	25	25	25
Confidential	25	50	50	50

Table 18. The result of the distance calculation between predefined classes and new data

Scenario	Sensitive	Sensitive	Confidential	Selected Class
Case 00	$\sqrt{(0^2+0^2+0^2+0^2)}$	$\sqrt{(0^2+25^2+25^2+25^2)}$	$\sqrt{(0^2+50^2+50^2+50^2)}$	Public
Case 01	$\sqrt{(0^2+25^2+25^2+25^2)}$	$\sqrt{(0^2+0^2+0^2+0^2)}$	$\sqrt{(0^2+25^2+25^2+25^2)}$	Sensitive
Case 02	$\sqrt{(0^2+25^2+50^2+50^2)}$	$\sqrt{(0^2+0^2+25^2+25^2)}$	$\sqrt{(0^2+25^2+0^2+0^2)}$	Confidential
Case 03	$\sqrt{(0^2+50^2+25^2+25^2)}$	$\sqrt{(0^2+25^2+0^2+0^2)}$	$\sqrt{(0^2+0^2+25^2+25^2)}$	Sensitive
Case 04	$\sqrt{(0^2+50^2+50^2+50^2)}$	$\sqrt{(0^2+25^2+25^2+25^2)}$	$\sqrt{(0^2+0^2+0^2+0^2)}$	Confidential

security. The results depict that the proposed method is more relevant than storing the data without deciding the security needs of the data. In addition, the results show that the improved K-NN with decision-making method works better than the K-NN classification technique in terms of accuracy.

In future, more security requirements can be taken in order to improve the classification decision using the machine-learning. Furthermore, to enhance the security at the authentication and authorization level, attribute-based access control could be used in order to avoid unauthorized access to the cloud environ.

REFERENCES

- Ashish, S. & Kakali, C. (2016). Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications*.
- Boyd, C., & Mathuria, A. (2013). *Protocols for Authentication and Key Establishment*. Springer.
- Charmee, V. (2014, December). Survey on Data Integrity Checking Techniques in Cloud Data Storage. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(12).
- Ennajjar, I. (2017). *Youness Tabii, Abdelhamid Benkaddour; Securing Data in Cloud Computing by Classification*. Association for Computing Machinery-ACM.
- Gaur, T., & sharma, D. (2016). Divya sharma; A Secure and Efficient Client-Side Encryption Scheme in Cloud Computing; I.J. *Wireless and Microwave Technologies*, 1(1), 23–33. doi:10.5815/ijwmt.2016.01.03
- Goyal, T., & Singh, A. (2012). Cloudsim: simulator for cloud computing infrastructure and modelling. *International conference on modelling, optimisation and computing –ICMIC*.
- Izadikhah, M., Roostae, R., & Hossein, F. (2014). Zadeh Lotfi; Using goal programming method to solve DEA problems with value judgments. *Yugoslav Journal of Operations Research*, 24(2), 267–282. doi:10.2298/YJOR121221015I
- Kaur, K., & Zandu, V. (2016, September). A Secure Data Classification Model for achieving Data Confidentiality and Integrity in Cloud Environment [IJCSE]. *International Journal on Computer Science and Engineering*, 8(9).
- Kiran, Sandeep Sharma, Enhance data security in cloud computing using machine learning and hybrid cryptography techniques; *International Journal of Advanced Research in Computer Science*; Volume 8, No. 9, November-December 2017.
- Lo'ai, T., Darwazeh, N. S., Al-Qassas, R. S., & Al Dosari, F. (2015). A Secure Cloud Computing Model based on Data Classification; *First International Workshop on Mobile Cloud Computing Systems, Management, and Security-MCSMS*. IEEE.
- Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg. doi:10.6028/NIST.SP.800-145
- Mohsin, J., Han, L., Hammoudeh, M., & Hegarty, R. (2017). Two Factor vs. Multi-factor, an Authentication Battle in Mobile Cloud Computing Environments. In *Proceedings of the International Conference on Future Networks and Distributed Systems*. ACM.
- Zardari, M. (2014). Low Tang Jung; Nordin Zakaria; K-NN Classifier for Data Confidentiality in Cloud Computing. IEEE.
- Ogigau-Neamtii F. (2012). Cloud Computing Security Issues. *Journal of Defense Resources Management*, 3, 141-148.
- Patel, R. & Dehariya, S. (2016). Secure Model for Cloud Computing by using Data Classification Methodology. *International Journal of Innovative Research in Computer and Communication Engineering* 4(12).
- Pathan, A. S. K. (2016). *Security of Self-Organizing Networks: MANET, WSN, WMN, VANET*. CRC Press. doi:10.1201/EBK1439819197
- Pravin, O., & Balbudhe, P. (2013). Cloud Storage Reference Model for Cloud Computing. *International Journal of IT, Engineering and Applied Sciences Research (IJEASR)*, 2(3).
- Sandeep, K. (2012). Sood; A combined approach to ensure data security in cloud computing. *Journal of Network and Computer Applications*, 35(6), 1831–1838. doi:10.1016/j.jnca.2012.07.007
- Sawle, P. & Baraskar, T. (2016). Survey on Data Classification and Data Encryption Techniques Used in Cloud Computing. *International Journal of Computer Applications*, 135.
- Shaikha, R. & Sasikumar, M. (2015). Data Classification for achieving Security in cloud computing. *International Conference on Advanced Computing Technologies and Applications –ICACTA*.

- Shuang, T., Zhang, J., & Chen, Z. (2014). Provable Data Possession in Cloud Computing. *Applied Mechanics and Materials*, 513-517, 1406–1413. doi:10.4028/www.scientific.net/AMM.513-517.1406
- Spoorthy, V., Mamatha, M. & Santhosh Kumar, B. (2014). A Survey on Data Storage and Security in Cloud Computing. *International Journal of Computer Science and Mobile Computing*, 3(6).
- Sun, Y., Zhang, J., Xiong Y., & Zhu, G. (2014). Data security and privacy in cloud computing. *International Journal of Distributed Sensor Networks* 10, 1–9.
- Tamanna, R. (2017). Secure Cloud Model using Classification and Cryptography. *International Journal of Computer Applications*, 159.
- Wu, J., Ping, L., Ge, X., Wang, Y., & Fu, J. (2010). Cloud Storage as the Infrastructure of Cloud Computing. *International Conference on Intelligent Computing and Cognitive Informatics (ICICCI)*, (pp. 22-23). doi:10.1109/ICICCI.2010.119
- Yadav, P. (2013). Security issues in cloud computing solution of DDOS and introducing two-tier CAPTCHA. *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, 3, 25–40.
- Yahya, F. (2015). Robert J Walters, Gary B Wills; Protecting Data in Personal Cloud Storage with Security Classifications. *Science and Information Conference*, London, UK.
- Zardari, M., Jung, L., & Zakaria, N. (2014). Data Classification Based on Confidentiality in Virtual Cloud Environment. *Research Journal of Applied Sciences, Engineering and Technology*.

Oussama Arki is a Ph.D. student at the Faculty of New Technologies of Information and Communication, University Constantine II-Abdelhamid Mehri, Algeria. He received his master's degree in Information Systems and Web Technologies (2016) from University Constantine II-Abdelhamid Mehri. His research interests include Multi Agent Systems, Cloud Computing and Information Security.

Abdelhafid Zitouni is a professor of computer science in University Constantine II-Abdelhamid Mehri, Algeria. He received his PhD in computer science in 2008 from the University Mentouri of Constantine, Algeria. Dr. Abdelhafid Zitouni has published many articles in International Journals and Conferences. He supervises many Master and PhD students and peer-reviewed conference and journal papers. His research interests include Software Engineering, Cloud Computing, Multi Agent System, Software Design, and Software Reuse.

Mahieddine Djoudi currently works at Computer Science Department and Techne Labs, University of Poitiers (France). Mahieddine does research in Learning Analytics, Elearning, and Information Literacy. Their current project is 'Digital Practices Awareness'.