Invalid cloud providers' identification using the support vector machine

Seyedeh Zeynab Mohammadi
Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
and
Nima Jafari Navimipour*
Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
*jafari@iaut.ac.ir

Cloud computing is a relatively new technology by the creation of which, companies and organizations transmit their services to cloud domains in order to do their tasks. With the companies and service provider organizations, profiteers seem not to neglect these areas and with different approaches try to do threats on users information and data security. Hence, it is necessary to adopt an approach relying on cloud providers to distinguish valid from the invalid cloud provider. Recognizing valid and invalid cloud providers can be issued as a classifying subject. In this paper, a support vector machine (SVM) algorithm is used in order to classify cloud providers. The features of SVM are a good generalization, ability to classify input pattern, optimal general pattern, and learning capability. The proposed method converts data to learning vector each of which has a corresponding output value and the ability to find the optimal amount in the non-linear and linear atmosphere. In this research, a data set of 1018 samples was used to classify cloud providers each of which has 10 features of cloud providers. To evaluate the performance of the proposed approach, 80 percent of the data set is randomly considered as a training set and 20 percentages as a test set. The results demonstrated that the proposed approach is efficient as compared to the meta-heuristic algorithms.

Keywords: Cloud Computing, Cloud Provider, Support Vector Machine, Data Clustering.

1. INTRODUCTION

Nowadays, cloud computing has gained vast attention due to its technological advance, cost reduction, and availability [Vakili & Navimipour, 2017]. Cloud computing is the latest emerging trend in distributed computing, where shared resources are provided to end-users in an on-demand fashion that brings many advantages, including data ubiquity, the flexibility of access, high availability of resources, and flexibility [Milani & Navimipour, 2016; Navimipour & Milani, 2015]. Cloud technology connects a network of virtualized computers that is dynamically provisioned as computing resources, based on negotiated agreements between service providers and user [Fouladi & Navimipour, 2017; Keshanchi, Souri, & Navimipour, 2017]. Today cloud computing is arguably one of the most significant advances in information technology (IT) services. It delivers information technology resources in diverse forms of service, and the explosion of cloud services on the Internet brings new challenges in cloud service discovery and selection [Navimipour, Rahmani, Navin, & Hosseinzadeh, 2014; Souri & Navimipour, 2014]. To address these challenges, a range of studies has been carried out to develop advanced techniques that will assist service users to choose appropriate services [Keshanchi et al., 2017; Sun, Dong, Hussain, Hussain, & Chang, 2014]. Several cloud service providers (CSPs) have offered services that have produced various transformative changes in computing activities and presented numerous promising technological and economic opportunities [Keshanchi & Navimipour, 2016; Pichan, Lazarescu, & Soh, 2015]. Cloud computing has emerged as an important paradigm in IT space by enabling cost effective, on-demand provisioning of elastic computing resources [Aznoli & Navimipour, 2017; Chowhan, Shirwaikar, & Kumar, 2016]. Security and privacy are the most critical issues that need to be addressed in designing a computing environment that is reliable and trustworthy [Islam, Manivannan, & Zeadally, 2016; Keshanchi et al., 2017]. However, many cloud customers remain reluctant to move their IT needs to the cloud, mainly due to their concerns about cloud security and the threat of the unknown [Pichan et al., 2015]. Today, numerous cloud services are provided by leading enterprise companies such as Amazon, Microsoft, and Google in the form of customized, reliable and cost-effective web applications. These services attract many individuals and organizations from different disciplines such as health, business, and education [Almishal & Youssef, 2014].

Cloud Computing is clearly one of the today's most enticing technologies thanks to its cost-efficiency and flexibility [Sheikholeslami & Navimipour, 2017]. This technology eliminates the idea of setting up an excessively expensive computing infrastructure for the IT-based solutions and services that are needed in any business. It provides a flexible IT architecture that is accessible via the Internet for lightweight portable devices. In this way, the existing new software is enhanced in both capacity and capability [Medhioub, Hamdi & Kim, 2015]. To increase the profit, a semi-trusted cloud service provider may outsource the files of its client to some low expense cloud service providers, which may violate the wishes of cloud users and impair their legitimate rights and interests [Alamir, Jafari Navimipour, Ramage, Ramage & Ramage, 2016; Jiang et al, 2015].

Currently, cloud computing has considerably solved the limitation problem of low storage capacity and data installation, triggering, backup and process. However, some of the challenges still exist, such as privacy, the coherence of sensitive information and security [Matin Chiregi & Nima Jafari Navimipour, 2016; Matin Chiregi & Nima Jafari Navimipour, 2016]. More precisely, security is a challenge on the way to install and apply cloud computing in large scale; therefore, user satisfaction from security and service quality of cloud providers has an important role in the development of this technology. Obviously, aligned with cloud providers, there are some jobber companies that look for gaining customers' information and data by anyway and for any target. In such an atmosphere by those kinds of security threats, it is required to recognize the invalid cloud providers from the secure ones. As declared in [Rao & Selvamani, 2015], cloud security and privacies are the most important issues in the cloud environments. This research tries to present an appropriate approach to recognize the validity of cloud providers.

In this paper, to classify the cloud providers from 1018 samples of data set which has 10 features. Eight features from the 10 features, introduced previously by other researchers and 2 remaining features are introduced in this paper. Also, in this paper, distinguished valid cloud providers from invalid ones is considered as a classifying process. This process depends on feature observation and the paper offers 10 public and private features. The numeral amounts of these features are sent to SVM classifier in the form of a vector to recognize the validity of cloud provider according to machine learning process on the basis of supposed research data. The result of this method is compared with the results relating to other classifiers to show the prominence of this approach.

The paper is structured as follows: after discussing the problem and subject review, related works are introduced. Then, the presented approach is mentioned including 10 features of cloud providers and SVM classifier. Finally, we conclude the paper with an evaluation of the work and talking about research results.

2. RELATED WORK

In Gholami and Arani [2015], a trust model is presented to choose the best source based on quality criteria of service such as cost, response time, bandwidth, processor speed, and so on. The proposed model has better performance compared with some other trust models such as first in the first out (FIFO), quality of service trust model and other similar models. Taking into account the quality of service measures, the proposed model selects the most reliable source in the cloud environment by considering the speed of things. In forward steps, selecting the best cloud service using rating mechanism via the analytic hierarchy process and the development of trust model based on cost-efficient algorithm are studied. Also, the model has a better performance compared to the trust model of the first input first output (FIFO), trust model of quality of service and some similar models.

Filali and Yagoubi [2015] have presented a more accurate selection solution as a result of computing the trust and the performance of the selected service only based on trust values. The proposed solution fulfills the highest amount of transactions. Furthermore, they have implemented and tested the solution with a performance selection based only on the QOS of the provided service, and a random selection of the services. The proposed model achieves better performance in execution time. However, with the overload of the services demand, the execution time increases significantly; whereas, the proposed solution tends to have a more stable execution time. Hence, the solution can concurrently support numerous services demands with maintaining accurate selection and user satisfaction. In addition, the improved solution is efficient because of the convergence speed and stability compared to other propositions. Furthermore, the feasibility of our approach has been validated by various simulations and comparisons with approved strategies as Eigen trust or subjective logic. The results of simulations show that the service selection model can decide an appropriate service for the user from various cloud services. Currently, the overall system for this model is being developed. This system integrates additional requirements for Cloud Security as authentication, SLA technology, etc. Finally, it would be interesting to examine how the presented trust model can be integrated to treat the security requirements of a real Cloud environment.

Aghdam and Navimipour [2016] have proposed a new framework to select the opinion leaders in online communities. The framework uses the trust relationship between the users and evaluates the total trust value (TTV) of primary opinion leaders between other users to select the highest of them. According to the obtained results, the proposed framework in comparison with the top in-degree method, top out-degree method, top centrality method and hybrid IO-degree method provides better results in the social network marketing (SNM) campaigning. In this paper, a new framework for identifying opinion leaders based on trust relationship among users is proposed. Identified opinion leaders can be used on many issues, such as politics, economy, education, society and etc. In the proposed method, after filtering the values of data sets on three parts, consisting of removing self-trust statements, duplicating comments, and trolling comments, our method identifies those users as opinion leaders which have a top total trust value by using two parts: trust evaluating and opinion leaders selection. The results demonstrated that the returned percent of real opinion leaders (73% of real opinion leaders) and confiding users (29% of all users) using the proposed method is more than that of those users selected by using the top in-degree method (20%of all users), top out-degree method (19% of all users), top centrality method (18% of all users), and hybrid IO degree (22% of all users) methods.

Also, Matin Chiregi and Nima Jafari Navimipour [2016] have evaluated the trust by considering the influence of opinion leaders on other entities and removing the troll entities effect on the cloud environment. Trust value is evaluated by using five parameters; availability, reliability, data integrity, identity, and capability. Also, they proposed a method for opinion leaders and troll entity identification using three topological metrics, including input-degree, output-degree and reputation measures. The method is being evaluated in various situations where it shows the results of accuracy by removing the effect of troll entities and the advice of opinion leaders. Unlike other approaches, in the proposed method, negative correct opinions, negative incorrect opinions, positive correct opinions, and positive incorrect opinions are also considered. The experiments on datasets show that the performances of the approach are better than the other approaches. Furthermore, They obtained that the percentage of opinion leaders and troll entities has a direct relation with reputation. If They increase the threshold of reputation value, low percentage of opinion leaders are selected, but it can identify more troll entities. Furthermore, Navimipour, Nima Jafari [2015] has proposed a new and applicable method for trustworthy human resource Discovery (THRD) in the expert cloud by introducing a resource discovery and trust evaluating method. The proposed method is implemented using ASP.Net and SQL in the expert cloud. Also, the structures and compositions of THRD are verified by describing the behavioral models and state diagrams. A Kripke structure with marked states is used to provide the formal relationship between the expanded model and the original state diagram structures. Moreover, the expected properties of the structures and compositions of THRD by means of temporal logic languages are defined. The results showed that the proposed method can discover trustworthy resources efficiently and is sound, complete, reachable, fair, dead lock-free and consistent. The obtained results also demonstrated the efficiency of the proposed method to discover trustworthy resources.

Shahi and Ghimire [2014] have built a masquerade detection system in cloud computing based on the proposed SLA. The efficiency of back propagation algorithm with SVM is also compared; it is found out that SVM has a better detection rate with a higher false alarm rate compared to the back propagation algorithm. The main concern of this work was to assess the risk of cloud computing and to implement a new methodology in detecting masqueraders involving SLA agreement between cloud computing users and CSPs. This work also proposed a new SLA by mutual consent of CSP and cloud costumers, which will be helpful in detecting masquerades' insider attack.

Finally, Mahendiran, Saravanan, Subramanian and Sairam [2012] have created a database called "K-Mean" and two tables to store the data sets of iris and Blood Transfusion Service Center Data Set in cloud SQL using MySQL. As one of the properties of K-Mean, it assumes the number of clusters, K is known in advance. For the first experiment, They assume that k = 3. Since the value of K is 3, the number of output clusters is three. The contents of each cluster are displayed under the name cluster1, cluster2, and cluster3. The second data set They used was blood transfusion service center data set. It consists of 5 attributes and 748 instances. They have taken first 200 data for analysis. As one of the properties of K-Means is that it assumes the number of clusters, K is known in advance. Here They assume that k = 2. Since the value of K is 2, the number of output clusters was two. This paper focuses on the implementation of K-Means algorithm in the cloud environment and the experimental results show that it works well in the cloud. K-Means algorithm is a more efficient algorithm for mining large databases and cloud computing provides a solution for storing large database with less cost.

3. PROPOSED METHOD

Nowadays, despite removing some limitations from IT world by cloud technology, some challenges came into existence along with it. One of these challenges is confidence on the cloud provider and recognizing the validity of cloud providers. Each cloud provider includes a set of features that give useful information about quality and quantity of services. Recognizing the validity of cloud provider can be considered a classifying problem as the classifying knowledge depends on feature observing. In this research, 10 private and public features are offered to each cloud provider that according to their amounts, cloud providers can be divided into two separate parts: valid and invalid. Undoubtedly the introduced features are not perfect and the researchers can add other ones to the described features' set and get new results.

In order to classify the providers into valid and invalid, the first step is extracting the numeral amounts of each feature of the table to the provider. Then the numeral amounts are sent to the supporting vector machine in the form of a vector that has been trained by a supposed data set. After that, the amount of 0 or 1 is allocated to the input vector. The 0 result shows the invalidity of the cloud provider and 1 shows the validity. The diagram of the suggested approach is shown in figure 1. The proposed method is based on data collection. If valid data is used them, results will be extended to the real environment. Whatever it is more basic information. Its results will be more accurate. At this point, the dataset updates the user. However, in the next step by adding crawler and automatic extraction features, process updates can be done by computer. In fact, always upgradeable software updates the data collection for the proposed method there. For each cloud, the provider has ten characteristics. Users can refer to cloud providers to extract these features. After extracting features, the proposed system introduces them to the user. The system was introduced based on values and learning process to check the validity of the action.

In this part of the paper, after introducing cloud provider features and their calculation method, it is continued by explaining how to classify the cloud providers into two groups of valid and invalid by SVM. The measured reliability of cloud provider uses technical specification and public review about them. For this purpose, the authors considered the first step properties, and for each cloud, a provider is extracted. Then, the obtained information is converted into numerical values. The numerical values in the form of a vector which was trained previously will be sent until a value of zero or one will be assigned to the input vector. Zero result indicates invalid cloud provider and 1 indicates that it will be valid.

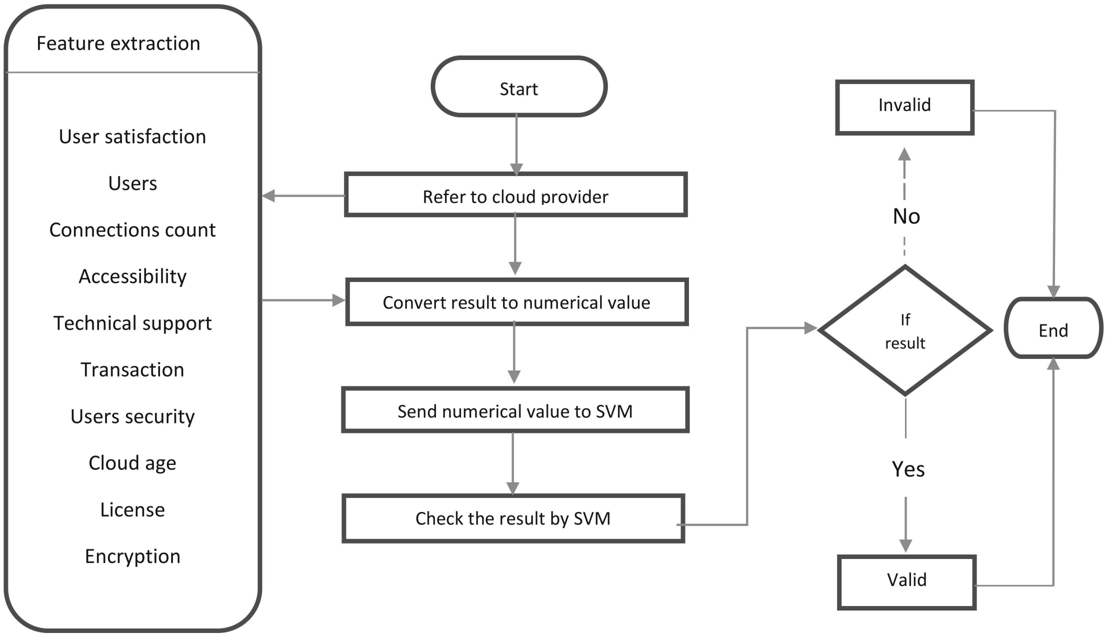

Figure 1. The flowchart of the proposed method

3.1 Features of cloud providers

Ten basic and technical features for each sample (cloud provider) are intended to classify the samples based on extracted features. The method looks for distinguishing valid cloud providers from invalid ones. This requires the use of clustering techniques. Each classification algorithm is based on extracted features for statistical sample works. In order to classify cloud providers to handle the validity, required features are provided for each participant. For this purpose, according to previous studies, ten general and technical features are considered for each cloud provider as follows in the table I.

Table I: Ten general and technical features

Method	Encryption	License	Age	Users	Transaction	Customer Satisfaction	Connections	Security	Technical Support	Accessibility
[Jain and et al, 2014]								*		*
[Medhioub and et al, 2011]								*	*	*
[Medhioub and et al, 2015]						*		*	*	*
[Ahmad and Khan, 2015]								*	*	*
[Miaoand et al, 2015]	*	*		*			*			
[Sun and et al, 2014]				*	*	*		*	*	*
[Wen and et al, 2012]				*			*		*	*
Proposed Method	*	*	*	*	*	*	*	*	*	*

Customer Satisfaction: This parameter indicates the average users future visit within a 30-day period to receive services. It is calculated according to the following equation.

$$ST = \frac{\sum_{i=1}^{i=30} C_i}{AD} \tag{1}$$

Users: This parameter represents the users of the system for a day. The considered period in this study is 30 days. This value is calculated as follows.

$$UCAD = \frac{\sum_{i=1}^{i=3} UCC_i}{AD} \tag{2}$$

Where UCAD is a user caller average rating to the system and UCC_i is user connection count for one day and AD is a count of days in this study, 30 days.

Connection: This parameter represents a rate of established contacts with Cloud users for a period of one day. The considered period is 30 days. This value is calculated as follows.

$$CAD = \frac{\sum_{i=1}^{i=3} CC_i}{AC} \tag{3}$$

Where CAD is an average number of calls of the cloud provider, CC_i is connection count in one day and AC is all connections.

Accessibility: Accessibility of web content to people with disabilities requires semantic information about widgets, structures, and behaviors, in order to allow assistive technologies to make appropriate transformations. This specification provides an ontology of roles, states, and properties that set out an abstract model for accessible interfaces and can be used to improve the accessibility and interoperability of Web Content and Applications. This information can be mapped to access frameworks that use this information to provide alternative access solutions. Similarly, this information can be used to change the rendering of content dynamically using different style sheet properties (Table II).

Table II: The features scoring of cloud provider access

Number	Counteract	Rate
1	Fully maintaining the confidentiality of information.	1
2	Allowing the information to be read and removed if cases are suspected.	0.75
3	Reading the permission to edit and delete the suspicious part.	0.5
4	Enabling the users to analyze and reproduce information.	0.25
5	Allowing unfettered access to all users data and information.	0

Technical support: The purpose of this feature is to provide the ability of technical support for cloud users and their data. To this end, the researchers considered 5 features for each cloud provider. Thus, for each cloud to fit the abilities of the mind a score between zero and 1 is assigned. Features considered for this purpose are as follows (Table III), this calculation is made according to relation (4).

Table III: Technical support capabilities for cloud providers

Number	Capabilities
1	Claude services Claude update
2	Backup user data
3	Member information security against hackers
4	Providing management information for user data
5	Password recovery and data recovery

$$TS = \frac{CFC}{N} \tag{4}$$

Transactions: The purpose of this feature cloud provider is average successful transactions per day.

$$T = \frac{\sum_{i=1}^{i=30} STD}{\sum_{i=1}^{i=30} TD} \tag{5}$$

Where T represents rated the transactions managed cloud provider, STD number of transactions managed cloud provider and cloud provider TD number of daily transactions.

Security: This feature shows the security of services of any cloud providers for users. It is calculated by 3 features of accessibility, transactions, and technical support (Table IV).

$$S = \frac{Ts + T + A}{3} \tag{6}$$

Table IV: Affecting parameters on the security feature

Number	Parameters	Range
1	TS: the technical support services provided by the cloud provider	$0 \leqslant TS \leqslant 1$
2	T: the rate of successful transaction	$0 \leqslant T \leqslant 1$
3	A: Access level to information	$0 \leqslant A \leqslant 1$

Cloud Age: This parameter represents years of providing cloud services. This value is calculated according to the researchers' suggestion. If the provision of services in each of the periods is considered by researchers to be exposed score between 0 and1 is assigned (Table V).

Table V: Rate of each range in the age feature

Number	Cloud Age	Rate
1	One year or less	0
2	Two years or less of it	0.25
3	Three years and less of it	0.5
4	Four years or less of it	0.75
5	More than four years	1

License: This feature represents licenses quality of cloud providers. For each cloud provider, 9 existing certifications are considered which are provided in Table VI.

Table VI: Rate of each range in the age feature

Number	License Name	Number	License Name
1	Google or less	6	Force
2	Microsoft it	7	Serve path
3	Amazon	8	Unisys
4	Rack space	9	Verizon
5	IBM		

$$LS = \frac{CLC}{N} \tag{7}$$

Where N is all count of License and LS is rating certificate provider and CLC is the number of certificates for each provider.

Encryption: This parameter represents a quality of cloud provider encryption protocol. For this purpose, 4 cloud providers encryption protocol was used in the intended method. According to the security level of each of the protocols, Mathias for quality encryption cloud provider goes as follows (Table VII).

Table VII: Rate of encryption protocols

Number	Encryption Method	Rate
1	GPV	0.25
2	Lattice-based cryptography	0.5
3	All haywire encrypt	0.75
4	Threshold secret sharing scheme	1

3.2 Classification

To classify the different features of cloud providers, we apply a method that analyzes data and recognizes patterns known as SVMs (Support Vector Machines), invented by VAPNIK in 1995. The SVM classifier takes a set of input data and predicts one of two possible classes. The SVM classifier is a supervised learning scheme in which we must divide our data into a training set and a testing set. The training set contains the class labels along with the known feature values of each and every training entity. Based on the training set, a model file that is a consolidated repository of the trained data and the rules that are used to classify the test data will be generated. In the testing phase, the data are compared with the trained SVM model, and a decision is made regarding whether a page is legitimate. The SVM classifier requires the list of features to be defined by the domain descriptor, which must be represented as a vector of real numbers. Thus, if there are any attribute values that correspond to discrete categories, they should be converted to equivalent numerical data. Additionally, the attribute values should be scaled before applying SVM to avoid attributes with larger numeric ranges dominating the classification. The input used in this method is a 10-dimensional feature vector inducing the cloud providers feature generator module, representing cloud providers' 10 structural and technical characteristics. For SVM implementation, we Program and support vector machine classification in MATLAB.

3.2.1 **Training process**. The training file is generated by using cloud providers in the training dataset as input, and the feature vector for every cloud provider is generated and stored in the training file like table 8. Each line in the training file contains the feature values (patterns) along with the corresponding class label, which appears in the first column. The training file we have generated consists of 1018 patterns and associated class labels. Table VIII shows an example for training file. Each feature represents a node in the training set; the feature values (Feature value) must be arranged in order, corresponding to the specific training nodes, and these values specify which class belongs to which corresponding training entity.

Table VIII: Example for training file

Class_label	Feature1_value	Feature2_value	...
1	0.7	0.25	...
0	0.99	0.5	
1	0.75	0.26	
1	0.33	0.48	
0	0.12	0.31	
...			

Generating the model file: The model file is a consolidated repository of the training data and the rules that are used to classify the test data. Along with training data, the type of SVM and kernel type used for classification are also given as input to generate the model file. Here, we use linear support vector machine classification.

Generating the testing file: This file is generated by giving the test dataset to the cloud provider, which will analyze a given cloud provider as the proposed models and produce the feature values. The test file is also of the same format as the input training file. However, the only difference between the training and testing file is that the latter will not have a class label.

The classification: Classification in SVM is an example of supervised learning; it deduces rules from labeled training data. The feature vector of the test cloud provider is given to the classifier as input, and the class is predicted using the rules developed during the training period. This stage will produce an output that contains the class label of the corresponding pattern and also classifies the new test patterns.

4. EXPERIMENT RESULT

In this section, we primarily used two metrics to evaluate the performance of the system: the true positive rate (TPR) and false positive rate (FPR). In addition, we used standard measures such as the precision and F-measure. The F-measure combines precision and recall to measure a tests accuracy. The summary of criteria for classifying the valid and invalid cloud providers are illustrated in figure 2. Due to the limitations of real-world implementation, the proposed method is simulated in C# IDE beta version. C# is an object-oriented programming language and .NET is a high level of the family. It was created by Microsoft Corporation and later also was ISO and ECMA standards. C# programming language with high levels of power has managed to attract the attention of many programmers. It simply times, modern, general-purpose, object-oriented was being made. Figure 2 introduces the user interface of the proposed system.

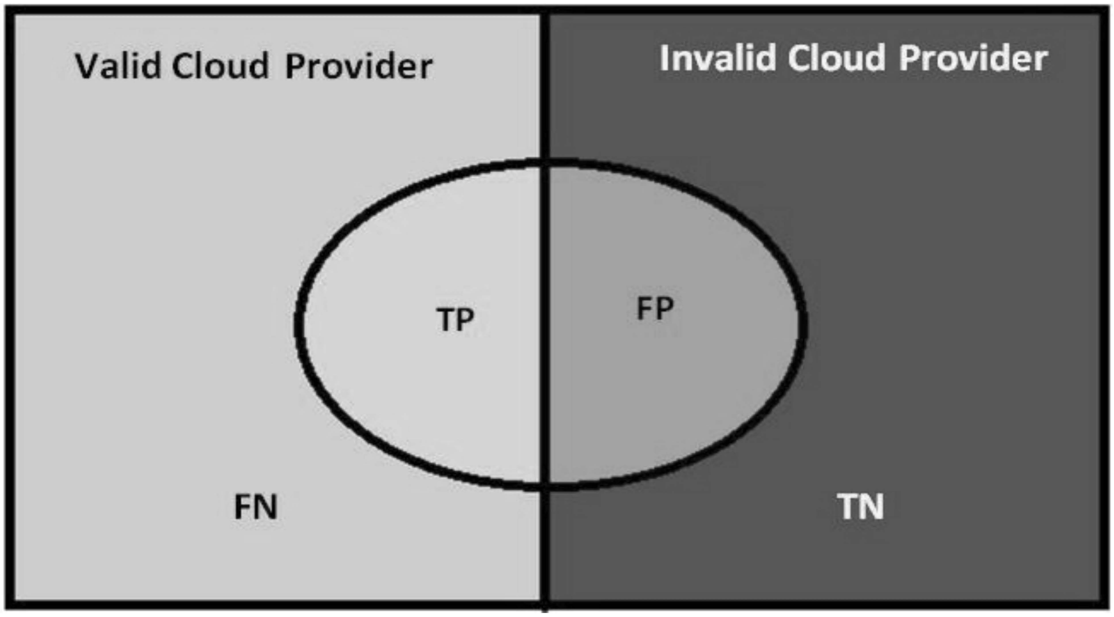

Figure 2. Evaluation criteria for classifying the valid and invalid cloud providers

The precision Rate is calculated by dividing the number of cloud providers identified as Valid by the sum of the number of valid cloud providers and the number of legitimate cloud providers wrongly classified as valid cloud providers. The precision is computed using the following equation:

$$Precision = \frac{tp}{tp + fp} \tag{8}$$

The recall rate is calculated by dividing the number of cloud providers identified as Valid by the sum of the number of valid cloud providers and missed valid cloud providers. The precision is computed using the following equation:

$$Recall = \frac{tp}{tp + fn} \tag{9}$$

The F-measure is a harmonic mean of the precision and recall Rate, as shown in the following equation:

$$F = 2 * \frac{Precision.Recall}{Precision + Recall} \tag{10}$$

These metrics assess the overall performance of the system and its weaknesses, which help us to tune the system Table IX.

Table IX: The assessment matrix

condition	True	False
Predicted condition positive	True positive	False positive
Predicted condition negative	True negative	False negative

Furthermore, a dataset of 1018 samples each of which consists of 10 features are used as illustrated in Table 1. In order to evaluate the performance of algorithms implemented on data collection, all samples contained in the terms of the relationships mentioned in section 3-1 of zero

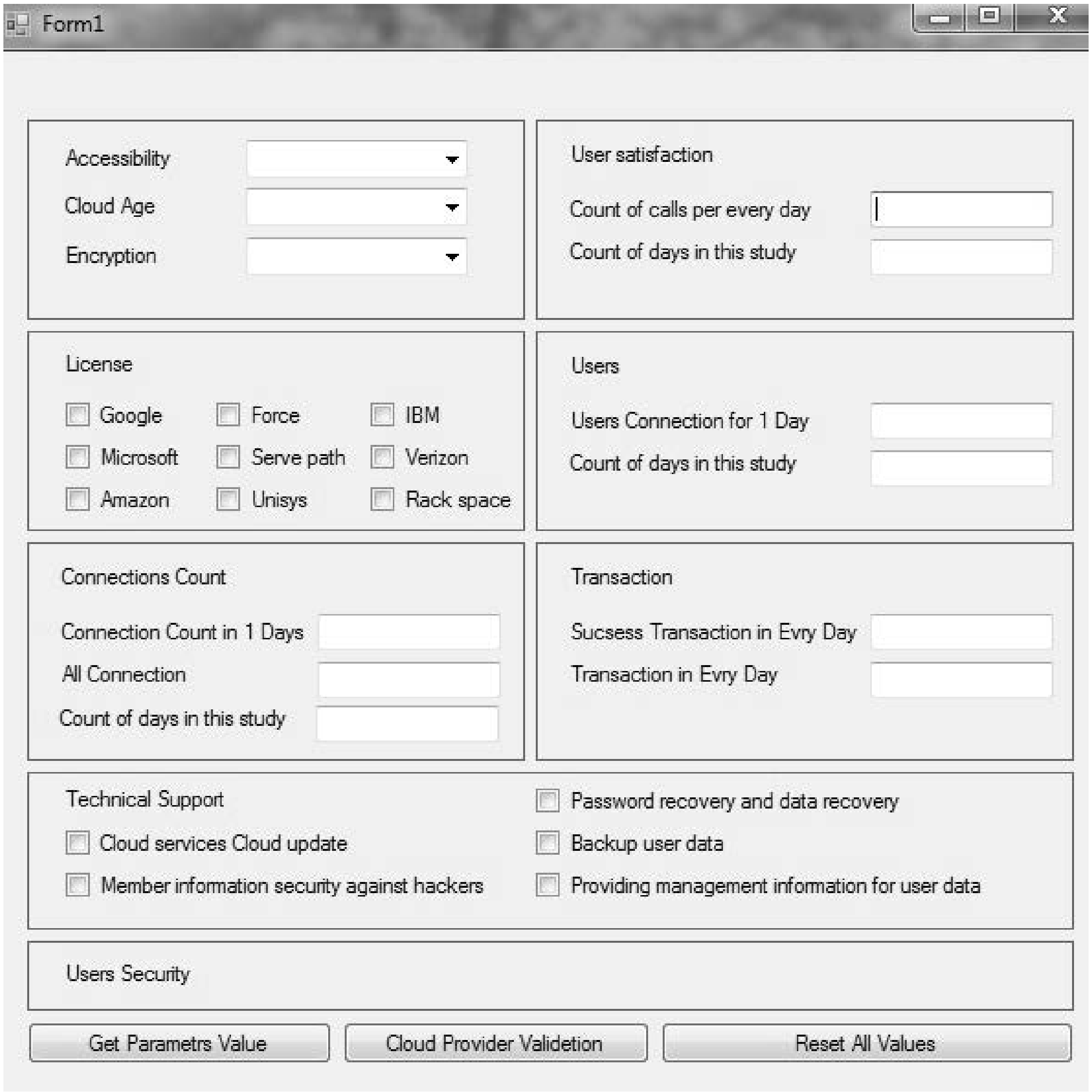
Form1
Accessibility
Cloud Age
Encryption
User satisfaction
Count of calls per every day
Count of days in this study
License
□ Google
□ Force
□ IBM
□ Microsoft
□ Serve path
□ Verizon
□ Amazon
□ Unisys
□ Rack space
Users
Users Connection for 1 Day
Count of days in this study
Connections Count
Connection Count in 1 Days
All Connection
Count of days in this study
Transaction
Sucsess Transaction in Evry Day
Transaction in Evry Day
Technical Support
□ Cloud services Cloud update
□ Member information security against hackers
□ Password recovery and data recovery
□ Backup user data
□ Providing management information for user data
Users Security
Get Parametrs Value
Cloud Provider Validetion
Reset All Values

Figure 3. The user interfaces in .NET framework

and one are labeled with two labels. Samples labeled zero indicate invalid cloud providers, and cloud providers are valid samples labeled 1. The proposed method for assessing the results of the first 80% of the dataset in learning stage was sent to SVM learning. Later on, the remaining 20% were examined as the test data by the proposed algorithm. The results showed that the F-measure approach taken was 0.97(Table X)

Table X: The obtained results of the proposed method

Method	Recall	Precision	F
Support Vector Machine	1	0.95	0.97

In addition, to evaluate the performance of the proposed method in clustering cloud providers, similar methods and algorithms were used for clustering the data of this paper's dataset so that we can claim whether the results obtained were of better performance. Two of these methods were the artificial neural network, and the next two were meta-heuristic algorithms.

The results show that the precision of data clustering of this paper's dataset using the SVM was 0.95, and the precision of the 'back propagation neural network' method with 0.93 had the next ranking (Figure 4).

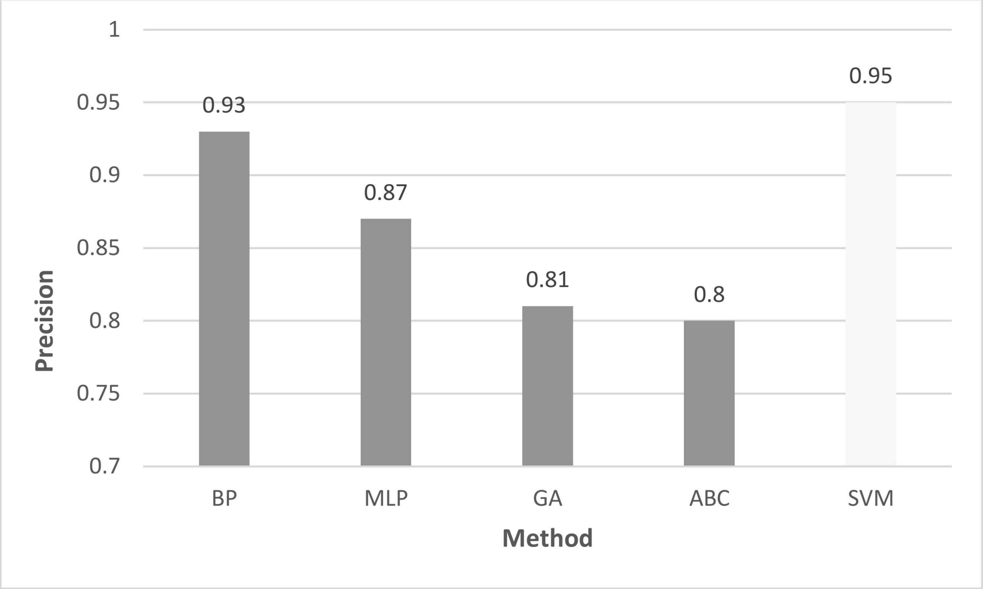

Figure 4. Precision accuracy of the methods

Also based on the results of this study it was shown that the recall of data clustering of this paper's dataset using the SVM back propagation neural network' methods was 1. The recall for the multilayer perceptron neural network' method with 0.95 had the next ranking (Figure 5)

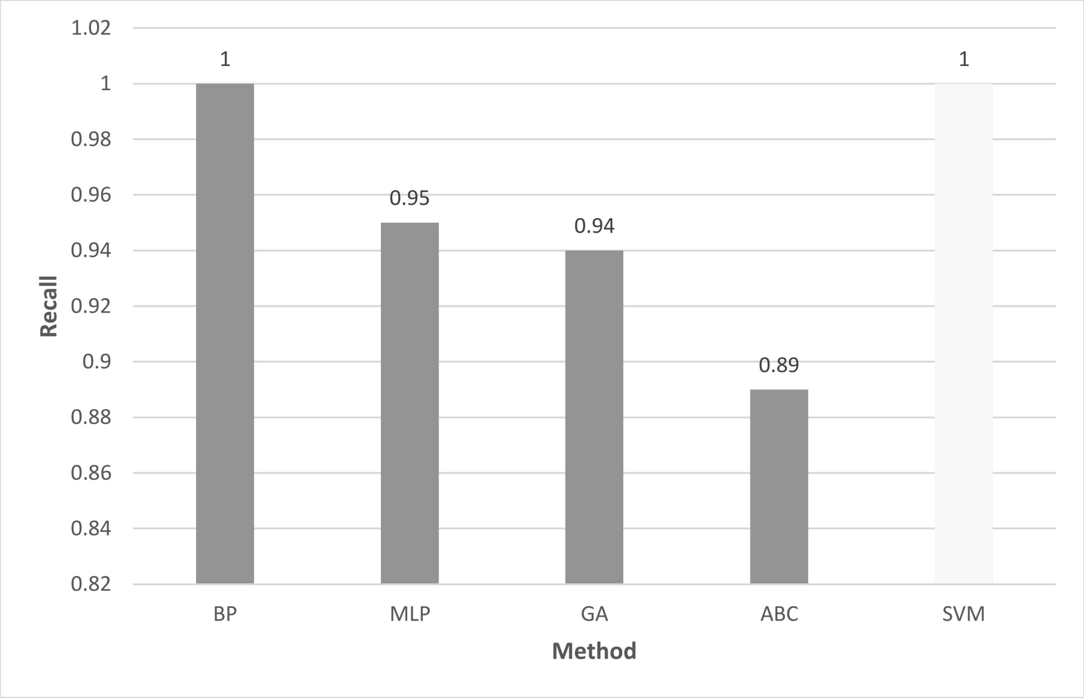

Figure 5. Recall accuracy of the methods

According to the findings of this research, the F-measure of data clustering of this paper's dataset using the SVM was 0.97. The F-measure for the 'Back Propagation neural network' method with 0.96 had the next ranking (Figure 6).

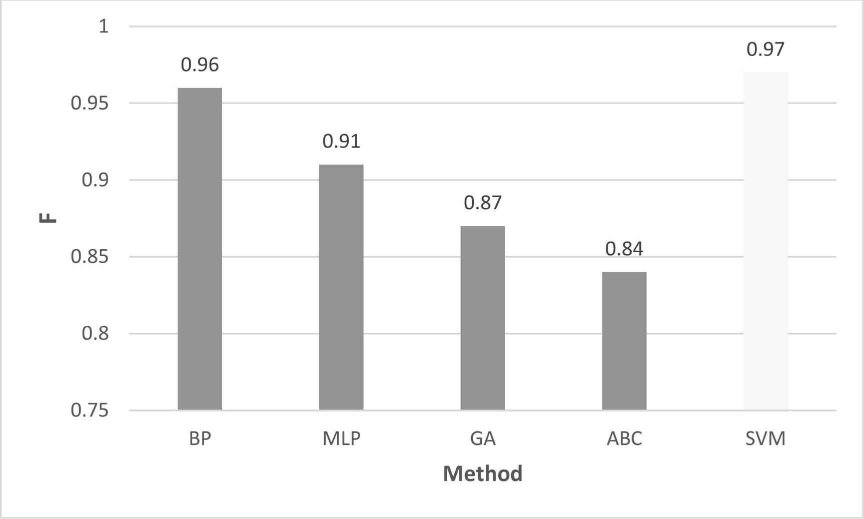

Figure 6. F-measure of the methods

The overall performance of the proposed method compared with other methods shows that SVM algorithm performance of data clustering of this paper's dataset is desirable. This is shown in table XI and Figure 7.

Table XI: The obtained results for precision, recall, and F

Methods	Method	Precision	Recall	F
Neural network	BP	0.93	1	0.96
	MLP	0.87	0.95	0.91
Metaheuristic	GA	0.81	0.94	0.87
	ABC	0.80	0.89	0.84
Proposed method	SVM	0.95	1	0.97

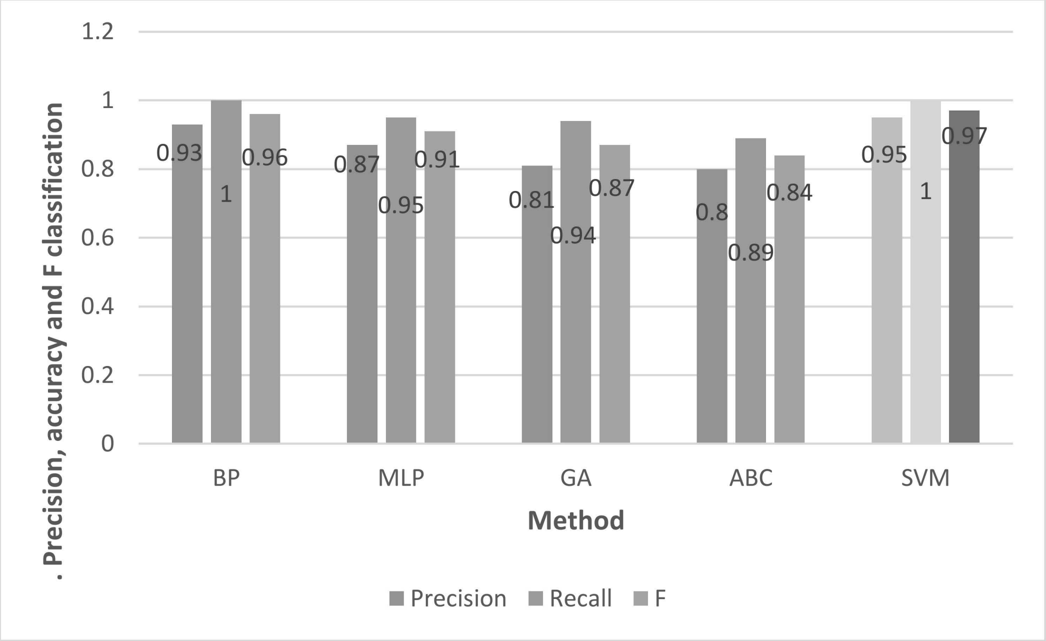

Figure 7. Precision, accuracy and F criteria of the methods

5. CONCLUSION

In this proposed method, determining cloud providers validity was considered as a classification/clustering issue. F-measure was used to examine the performance of the proposed approach. The results revealed that F-measure of the proposed method with 0.97 was desirable. Given that the subject is in the area of clustering, we examined the data in the article's dataset with 4 other algorithms used for clustering. Two of them were related to artificial neural networks. And the other two cases were related to the clustering of meta-heuristic. The results showed SVM compared to other methods functions desirably in identifying valid and invalid cloud providers. Despite these good results, SVM has some inherent limitations. These limitations are:

—SVM is based on complex calculations and it is a time-consuming process.

—Due to computational complexity, SVM consumes a lot of memory.

References

Aghdam, Samad Mohammad, Navimipour, Nima Jafari 2016. Opinion leaders selection in the social networks based on trust relationships propagation. *Karbala International Journal of Modern Science Vol.2,* No.2 .pp.88–97.

Ahmad, Mohammad Oqai, Khan, Rafiqul Zaman 2015. The Cloud Computing: A Systematic Review. *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE) Vol.3,* No.5 .

Almishal, Abdulelah, Youssef, Ahmed E 2014. Cloud Service Providers: A Comparative Study. *International Journal of Computer Applications & Information Technology Vol.5,* .

Aznoli, Fariba, Navimipour, Nima Jafari 2017. Cloud services recommendation: Reviewing the recent advances and suggesting the future research directions. *Journal of Network and Computer Applications Vol.77,* .pp. 73–86.

Chiregi, Matin Navimipour, Nima Jafari 2016. A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll entities. *Computers in Human Behavior Vol.60,* .pp. 280–292.

Chowhan, Seema Sunil, Shirwaikar, Shailaja Kumar, Ajay 2016. Predictive Modeling of Service Level Agreement Parameters for Cloud Services. *International Journal of Next-Generation Computing Vol.7,* No.2 .pp.115–129.

Filali, Fatima Zohra, Yagoubi, Belabbes 2015. Global trust: A trust model for cloud service selection. *International Journal of Computer Network and Information Security Vol.7,* No.5 .

Gholami, Atoosa, Arani, Mostafa Ghobaei 2015. A trust model based on quality of service in cloud computing environment. *International Journal of Database Theory and Application Vol.8,* No.5 .PP.161–170 .

Islam, Tariqul, Manivannan, D Zeadally, Sherali 2016. A classification and characterization of security threats in cloud computing. *Int. J. Next-Gener. Comput Vol.7,* No.1 .

Jain, Siddharth, Kumar, Rakesh Anamika, Sunil Kumar Jangir 2014. A comparative study for cloud computing platform on open source software. *ABHIYANTRIKI: An International Journal of Engineering & Technology (AIJET) Vol.1,* No.2 .pp28–35 .

Jiang, Tao et al 2015. Towards secure and reliable cloud storage against data re-outsourcing. *Future Generation Computer Systems Vol.52,*pp86–94 .

Keshanchi, Bahman Navimipour, Nima Jafari 2016. Priority-Based Task Scheduling in the Cloud Systems Using a Memetic Algorithm. *Journal of Circuits, Systems and Computers Vol.25,* No. 10 .

Keshanchi, Bahman Souri, Alireza Navimipour, Nima Jafari 2017. An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: Formal verification, simulation, and statistical testing. *Journal of Systems and Software Vol.124,* pp.1–21 .

Mahendiran, A et al 2012. Implementation of K-means clustering in cloud computing environment. *Research Journal of Applied Sciences, Engineering and Technology Vol.4,* No. 10 pp.1391–1394 .

Mahjoub, Meriam et al 2011. A comparative study of the current cloud computing technologies and offers. *Network Cloud Computing and Applications (NCCA), 2011 First International Symposium on* pp.131–134 .

Mahjoub, Meriam et al 2015. Requirements Capture and Comparative Analysis of Cloud Security Techniques. *International Journal of Grid and Distributed Computing Vol.8,* No. 2 pp.285–308 .

Miao, Meixia Jiang, Tao You, Ilsun 2015. Payment-based incentive mechanism for secure cloud deduplication. *International Journal of Information Management Vol.35,* No. 3 pp.379–386 .

Milani, Alireza Sadeghi Navimipour, Nima Jafari 2016. Load balancing mechanisms and techniques in the cloud environments: systematic literature review and future trends. *Journal of Network and Computer Applications Vol.71,* pp.86–98 .

Navimipour, Nima Jafari 2015. Task scheduling in the cloud environments based on an artificial bee colony algorithm. *Proceedings of the Paper Presented at the International Conference on Image Processing, Production and Computer Science, Istanbul (Turkey)* .

Navimipour, Nima Jafari 2015. A formal approach for the specification and verification of a trustworthy human resource discovery mechanism in the expert cloud. *Expert Systems with Applications Vol.42,* No 15 pp.6112–6131 .

Navimipour, Nima Jafari Milani, Farnaz Sharifi 2015. Task scheduling in the cloud computing based on the cuckoo search algorithm. *International Journal of Modeling and Optimization Vol.5,* No 1 .

Pichan, Ameer Lazarescu, Mihai Soh, Sie Teng 2015. Cloud forensics: Technical challenges, solutions and comparative analysis. *Digital Investigation Vol.13,* pp.38–57 .

Rao, R Velumadhava Selvamani, K 2015. Data security challenges and its solutions in cloud computing. *Procedia Computer Science Vol.48,* pp.204–209 .

Shahi, Tej Bahadur Ghimire, Dadhi Ram 2014. Comparison of BP and SVM on SLA based Masquerader Detection in Cloud. *International Journal of Computer Applications Vol.91,* No 14 .

Sun, Le Dong, Hai Hussain, Farookh Khadeer Hussain, Omar Khadeer Chang, Elizabeth 2014. Cloud service selection: State-of-the-art and future research directions. *Journal of Network and Computer Applications Vol.45,* pp.134–150 .

Navimipour, Nima Jafari Vakili, Asrin 2017. Comprehensive and systematic review of the service composition mechanisms in the cloud environments. *Journal of Network and Computer Applications* .

Wen, Xiaolong Gu, Genqiang Li, Qingchun Gao, Yun Zhang, Xuejie 2012. Comparison of open-source cloud management platforms: OpenStack and OpenNebula. *Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on* pp.2457–2461 .

Seyyedeh Zenab Mohammadi received her B.S. in computer engineering, software engineering, from Jaber Ebn Haiyan Institute of Higher Education, Rasht, Iran, in 2011, the M.S. in computer engineering, software engineering, from Islamic Azad University, Tabriz Branch, Tabriz, Iran, in 2016. Her research interests include Cloud Computing and Data Mining.

Nima Jafari Navimipour received his B.S. in computer engineering, software engineering, from Tabriz Branch, Islamic Azad University, Tabriz, Iran, in 2007; the M.S. in computer engineering, computer architecture, from Tabriz Branch, Islamic Azad University, Tabriz, Iran, in 2009; the Ph.D. in computer engineering, computer architecture, from Science and Research Branch, Islamic Azad University, Tehran, Iran in 2014. He is an assistance professor in the Department of Computer Engineering at Tabriz Branch, Islamic Azad University, Tabriz, Iran. He has published more than 100 papers in various journals and conference proceedings. His research interests include Cloud Computing, Social Networks, Fault-Tolerance Software, Computational Intelligence, Evolutionary Computing, and Network on Chip.