Application of Intelligent Data Mining Approach in Securing the Cloud Computing

Hanna M. Said[1]
[1]Faculty of computer and information science,
Ain Shams University
Cairo, Egypt

Bader A. Alyoubi[3]
[3]Management Information Systems (MIS)
College of Business, University of
Jeddah, Jeddah, Saudi Arabia

Ibrahim El Emary[2]
[2]Information Science Department
King Abdulaziz University,
Jeddah, Saudi Arabia

Adel A. Alyoubi[3]
[3]Management Information Systems (MIS)
College of Business, University of
Jeddah, Jeddah, Saudi Arabia

***Abstract*—Cloud computing is a modern term refers to a model for emerging computing, where it is possible to use machines in large data centers for delivering services in a scalable manner, so corporations has become in need for large scale inexpensive computing. Recently, several governments have begun to utilize cloud computing architectures, applications and platforms for meeting the needs of their constituents and delivering services. Security occupies the first rank of obstacles that face cloud computing for governmental agencies and businesses. Cloud computing is surrounded by many risks that may have major effects on services and information supported via this technology. Also, Cloud Computing is one of the promising technology in which the scientific community has recently encountered. Cloud computing is related to other research areas such as distributed and grid computing, Service-Oriented Architecture, and virtualization, as cloud computing inherited their limitations and advancements. It is possible to exploit new opportunities for security. This paper aim is to discuss and analyze how achieve mitigation for cloud computing security risks as a basic step towards obtaining secure and safe environment for cloud computing. The results showed that, Using a simple decision tree model Chaid algorithm security rating for classifying approach is a robust technique that enables the decision-maker to measure the extent of cloud securing, and the provided services. It was proved throughout this paper that policies, standards, and controls are critical in management process to safeguard and protect the systems as well as data. The management process should analyze and understand cloud computing risks for protecting systems and data from security exploits**

Keywords—Cloud computing; Cloud security issue; Data mining; Naive Bayes; multilayer percepton; Support vector machine; decision tree (C4.5); and Partial Tree (PART)

I. Introduction

Security is a basic requirement for cloud computing [1]. This view is shared by many distinct groups, such as business decision makers [4], academia researchers [2, 3], and government organizations [5, 6]. The several similarities in these opinions illustrate a high concern on crucial legal and security obstacles for cloud computing including availability of service, confidentiality in data, provider lock-in and status fate sharing [7]. The security is considered one of the major issues which reduces the growth of cloud computing and complications with data privacy and data protection continue to plague the market. The advent of an advanced model should not negotiate with the required functionalities and capabilities present in the current model. Cloud computing embraces cyber-infrastructure and builds on grid computing, utility computing, virtualization, distributed computing, networking, web and software services [11].

The illusion of infinite computing resources available on demand, the elimination of up-front commitments by cloud users, and the ability to pay for use of computing resources on a short-term basis are needed [3]. Cloud computing has been defined in many and diverse ways according to the point of view that deals with the cloud computing. One of these definition is adapted by [52] where cloud computing has been defined as a pool of highly scalable, abstracted, and managed infrastructure that is capable of hosting end-customer applications and billed by consumption. Another definition has been suggested by NIST [6], where cloud computing has been defined as a model for enabling on-demand network access to a shared pool of computing configurable resources in convenient way (e.g. servers networks, applications, storage, and services) that can be rapidly released and provisioned with minimal service provider interaction or management effort.

According to [42], a cloud is a pool of virtualized resources across the Internet that follows a pay per-use model and can be dynamically reconfigured to satisfy user requests via on-the-fly. Also, Cloud computing is known as a method for increasing the capabilities or adding capacity in dynamic manner without investing in new training new personnel, infrastructure or licensing new software, accordingly it works towards extending the existing capabilities of Information Technology (IT) [8]. Recently, cloud computing has been converted from being a concept of promising business to one of the fast growing segments of the IT industry. It is possible to say that the more the spread of cloud computing, the more the concern about the security and safety of the cloud computing environment. The security problem can be considered as a barrier against the deployment of cloud computing in business environment [9].

In cloud computing, security is considered a critical and important aspect, where security in cloud computing has many problems and issues related to it. Both the cloud service consumer and cloud service provider should be sure that the cloud is sufficiently safe from external threats in order to make the customer avoid any problem such as data theft or loss of data [10]. The penetration through a malicious user into the cloud can be occurred through impersonating a legitimate user, thus infecting many customers.

The users of cloud service should understand the risks of data breaches in the new environment of cloud computing; as the architecture of cloud forms a threat to the existing technologies security when deploying such technologies in a cloud environment [11].

Data mining can be considered one of the most important for discovering the knowledge from large data. Different algorithms and techniques are available in data mining. For finding the mine rule, classification technique can be used in large data. In general decision tree technique can be utilized as efficient technique for classification; because it owns simple hierarchical structure for the decision making and user understanding. This paper evaluates and investigates the decision tree as data mining techniques and as an intrusion detection mechanism.

This paper focus on a survey about the risks and threats that faces cloud computing followed by deep analysis of cloud security major issues such as: trust, encryption, multi-tenancy and compliance and finally utilizing the intelligent data mining and attack classification methodology in analyzing the cloud security.

In the last few years, it was shown that Cloud Computing depends on gathering few new and many old concepts in several fields of research such as Service-Oriented Architectures (SOA), grid, and distributed computing as well as virtualization. This was a result of its high probability for substantiating advances in other technologies while presenting supreme advantage over the current under-utilized resources that are deployed at data centers [12].

It is possible to say that cloud computing can be understood as a new computing paradigm that give users the temporal ability to use computing infrastructure over the network, that is provided as a service by the cloud-provider at one or more than one of abstraction levels. Thus, several models of business are rapidly evolved to utilize this technology by introducing programming platforms, software applications, computing infrastructure, data-storage, and hardware as services. Their inter-relations have been ambiguous. The feasibility of enabling their inter-operability has been debatable while they refer to the core cloud computing services, taking into consideration that each cloud computing service has a unique interface and uses a different protocol of access [13-15].

This paper is organized as follows: Section 2 shows the Literature Survey. Section 3 describes the Cloud environment layers, Service Models and Deployment Models. In section 4, we cover the gaps and security issues in service models, Denial-of-service attack classification, and Decision tree: C4.5 as well as Performance evaluation. Finally, section 5 provides conclusions.

II. LITERATURE REVIEW

Nowadays, Small and Medium Business (SMB) organizations have been understood that simply by getting into the cloud computing, they can obtain fast access to best business applications or increasingly reinforce the resources of their infrastructure, all at negligible cost. Gartner [53] defined cloud computing as ‘‘a style of computing where massively scalable information technology- enabled capabilities are delivered ‘as a service’ to external clients using Internet technologies’’. [42] Mentioned that nowadays cloud providers obtain benefits from opportunity in the marketplace. The providers should ensure that they obtain the right security aspects; therefore they will bear the responsibility if things are wrong.

The cloud offers many benefits such as pay-for- use, fast deployment, scalability, lower costs, rapid elasticity, rapid provisioning, ubiquitous network access, greater resiliency, low-cost disaster recovery and data storage solutions, hypervisor protection against network attacks, on-demand security controls, real time detection of system tampering, and rapid re-constitution of services. The unique attribute of the cloud computing, poses several new challenges from security point of view [44]. The posed challenges include virtualization vulnerabilities, accessibility vulnerabilities, web application vulnerabilities such as SQL (Structured Query Language) injection, cross-site scripting, privacy and control issues arising from third parties having physical control of data, physical access issues, issues related to credential management, identity and issues related to data verification, tampering, integrity, data loss and theft, issues related to authentication of the respondent device or devices and IP spooling.

III. CLOUD ENVIRONMENT LAYERS

Cloud computing attracts many managers and organizations. There are many similar terminologies that are usually utilized for describing cloud computing, these terms such as: distributed, grid, cluster, virtualization, on-demand, utility, and software-as-a-service. In other words, cloud computing refers to end-users connecting with applications running on sets of shared servers, often hosted and virtualized, instead of a traditional dedicated server.

For over thirty years client-server computing has provided applications that were assigned to specific hardware, often residing in on-premise data centers. On-demand cloud computing enables its end-users through allowing them using their selection of Internet-connected device, at any time [29].

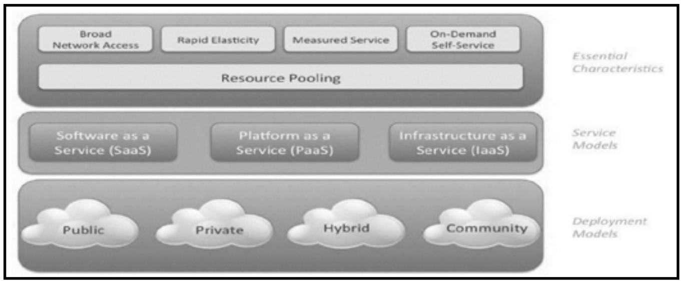


Fig. 1. Layers of cloud environment [19]

From Figure 1, it is shown that the lower layer of the cloud computing layers represents the different models of deployment for the cloud as follows: community, private, public and hybrid cloud models of deployment. The second layer above the deployment layer represents the different models of delivery that are used within a certain model of deployment. These delivery models are the IaaS (Infrastructure as a Service) delivery, PaaS (Platform as a Service) and SaaS (Software as a Service) models. These delivery models represent the core of the cloud and they show certain features like multi-tenancy, on-demand self-service, ubiquitous network, measured service and rapid elasticity that are illustrated in the upper layer. These basic elements of the cloud computing require security that depends and varies with respect to the used deployment model, the method of delivery and the character it shows. Some of the basic security challenges can be summarized in data transmission security, data storage security, security related to third-party resources and application security [19].

A. Service Models

It is possible to categorize cloud services into three categories namely: Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS).

- **Software-as-a-Service (SaaS),** usually called on-demand software, is a software deployment and a model of subscription-pricing that gives an enterprise application as a managed service by a software vendor. The SaaS provider bears all responsibilities related to the implementation and maintenance of the system from the customer; this in turn, is useful when adding new hardware as it minimize the addition cost and complexity. One example of SaaS is the Salesforce.com CRM application. According to the Forrester study, ’’The State of Enterprise Software: 2009,’’ security concerns are the most commonly cited reason why enterprises are not interested in SaaS. Consequently, addressing enterprise security concerns has emerged as the biggest challenge for the adoption of SaaS applications in the cloud [45].

- **Infrastructure-as-a-Service (IaaS),** usually called on-demand infrastructure, it gives computing infrastructure as a utility service. IaaS users buy or rent software, servers, network equipment, data-center space etc. IaaS usually gives networking with immense possibility for extensibility, scale and raw storage. One example of (IaaS) is the Amazon web services [45].

- **Platform-as-a-Service (PaaS)** can be defined as a rich ecosystem that is used for database development along with other applications, programmer communities, and application development, as a solution stack or service. One of the major advantages of PaaS is that it empowers developers of an application to build their own applications on top of the platform. PaaS usually overcome the gaps in functional holes within a SaaS solution. An example of (PaaS) would be Google Apps (Cloud Security Alliance, 2011) [4, 13, and 14].

All models of cloud service (IaaS, SaaS and PaaS) should be strongly well-defined service level agreements (SLAs) that are able to protect the cloud user. According to the CSA, “security, governance, service levels, compliance, and liability expectations of the service and provider should be stipulated from contracting point of view, enforced and managed” (Cloud Security Alliance, 2011).

B. Deployment Models

The deployment models can be categorized into four categories namely Public cloud, Private cloud, Hybrid cloud and Community cloud [6]. These categories will be described in details as follows:-

- **Public Cloud,** this model is owned by an organization for selling the cloud services and the design of infrastructure is made in order to be available for industries, organizations and businesses.

- **Private Cloud**, this model is managed by the organization itself or by a third party. Private cloud may be either off or on premises. The major characteristic of this model is that the infrastructure of the cloud is private, in addition to its availability to a single organization.

- **Hybrid Cloud, this model** is similar to the private cloud as it is managed by third party or by organization itself and may exist off or on premises. But the cloud infrastructure may combine two or more clouds (public, private or community).

- **Community Cloud, this model is similar to the previously mentioned private and hybrid cloud** as the organization or third party are allowed to manage it and also exists off or on premises. But in community cloud, multiple organizations with common interests, requirements, or considerations share the infrastructure.

The security of the cloud needs testing, it is important for organizations that want to ensure the optimal product before distributing it. The results are used in finding out security weakness points and to patch them before the occurrence of penetration. However organizations' lack of time and resources, computer related crime is usually on the rise. Consequently penetration investigators (testers) have to reduce the amount of resources. This motivates testers to widely adopt automatic tools, as it is demonstrated by the continuous release of platforms finalized to automate this process, discovering gaps in compliance, verifying secure configurations, finding holes now before somebody else does, Report problems to management and testing new technology.

IV. Gaps and Security Issues in Service Models

Although cloud computing has huge promising future but unfortunately it had not been adopted in enthusiasm and pace manner by the customers. This may refer to the reality of the existing gaps. The National Institute of Standards and Technology (NIST) [9] confirmed that security, portability and interoperability are the major barriers to wide adoption of cloud computing. Armbrust et al. [3] identified 10 major obstacles to cloud computing as follows: data lock-in, data confidentiality, availability of service, and audit ability, performance unpredictability, bugs in large distributed systems, data transfer bottlenecks, scalable storage, reputation fate sharing, scaling quickly, and software licensing. Ness [10] determined three major obstacles to cloud computing given by: first, cloud can break static networks; second, cloud is based on the new security approaches; and third, the criticalness of network automation.

Leavitt [11] mentioned six barriers as follows: latency and reliability; control; performance; vendor lock-in and standards; related bandwidth costs; security and privacy; and transparency. There may be many methods for defining gaps, and many parties are also embedded other than customers and cloud providers and. But, practically, what real situation at the end is that it refers to the customer whether he/she or his/her company is desire to join the cloud. The reputation of a company and the type of services expectations one is going to receive from a certain provider are the basic elements in selecting a cloud provider. According to [3, 7–11], it is possible to define cloud computing gaps as follows: The factors that slow down joining cloud computing from the current system are defined as gaps of cloud computing.

Fig 2 shows the gaps between expectations and perceived services by cloud customers' based on our understanding [2–5, 7–14]. Also a gap between customers' expectations and deliverable services has been witnessed. In our opinion, many of the potential clients are aware of this gap and consequently, they are waiting on the sidelines. Convincing these customers (clients) that the cloud will meet their expectation will encourage them to join the cloud computing. [2, 7, 8].

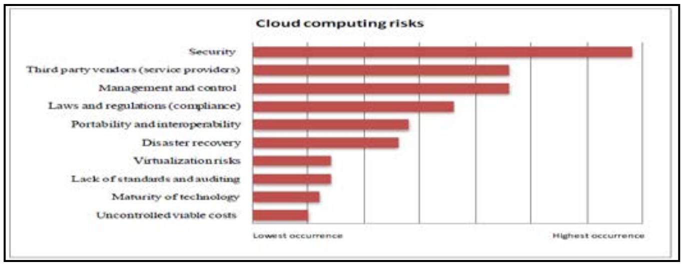


Fig. 2. Cloud Computing Risks [55]

According to the recent survey by Cloud Security Alliance (CSA) & IEEE, we can conclude that guaranteeing the security of corporate data in the ''cloud'' is difficult. There are different levels of security required by the different service models in cloud environment. **IaaS** represent the base of all cloud service, upon which the **Paas** is built and thus **SaaS**, in turn, is built upon the **PaaS,** this is shown in Figure 2. Tradeoffs should be taken into consideration for each model in terms of complexity and integrated features versus the security and extensibility. This means that the cloud service provider should take all aspects in account and should not concentrate only on security only at the lower part of the architecture of security as this may lead to make consumer more liable for managing and implementing the capabilities of securities [41, 42].

Each service has its own security issues [43]. The SaaS model provides the customers with important benefits, such as efficiency, reduced costs and improved operational. But according to the Forrester study, ''The State of Enterprise Software: 2009,'' security concerns are the most commonly cited reason why enterprises are not interested in SaaS. Thus, enterprise security appears to the strongest challenge for adapting SaaS applications [45]. Regarding the security of IaaS, the only basic security measures introduced by IaaS are (perimeter firewall, load balancing, etc.) but these measures are not enough as applications, that move into the cloud, need higher levels of security that are supplied by the hosts [43-45]. Despite the various advantage of the PaaS layer, it has key disadvantage represented in that, these advantages itself can be used by the hackers to expose the PaaS cloud infrastructure to malware control, command and going beyond IaaS applications [46- 48].

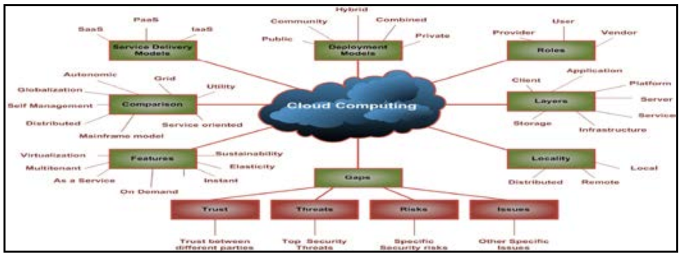


Fig. 3. Understanding cloud computing [8]

A. *Denial-of-service attack classification*

It is very important to understand and determine the most probable method used by attacker for attacking the application or a network [49, 50]. Determining the location of weakness points in network or application defenses is important as this will help in knowing how an attacker could use these weaknesses [38- 40], this is shown in Figure 4.

Machine learning techniques can be used successfully for classification of any activity depending on predefined classes. Machine learning techniques are available from the computational intelligence community Figure 4. From the available list of algorithms in machine learning, we have selected Naive Bayes [32], multilayer percepton [33], support vector machine [34], decision tree (C4.5) [35] and Partial Tree (PART) [36] for classifying our data. Naïve Bayes is a probability-based technique, multilayer perceptron and support vector machine are function estimation based techniques, and decision tree and PART are rule-based machine learning techniques. All these techniques have been implemented in Weka [37], which is a Java-based popular machine learning tool. Weka uses C4.5 [37] algorithm for decision tree implementation.

1) Tree Augmented Naive Bayes. This algorithm can be mainly used for the classification processes. It efficiently creates a simple Bayesian network model. The model is an improvement over the naïve Bayes model as it allows for each predictor to depend on another predictor in addition to the target variable. The main advantages of this algorithm are the accuracy of its classification and efficient performance compared with general Bayesian network models. The major disadvantage of this algorithm can be summarized in that although its simplicity; it generates more restrictions on the uncovered dependency structure among its nodes [32].

2) Multilayer perceptrons. The neurons are arranged in layers in such networks. Typically, one layer is assigned as input layer for the neurons, on the other hand, one or more layers are assigned for internal processing units that represent the hidden layers, and another one layer is assigned for neurons output that represents the output layer. There are interconnection between the different layers, e.g., in a network with an input layer, a single hidden layer, and an output layer, each neuron in the input layer should be connected to all neurons in the hidden layer, and each neuron in the hidden layer is connected to each one in the output layer. The strength of influence one neuron has on another can be determined through giving weights for the connections between neurons. The prediction generation can be obtained through information flows from the input layer through the processing layer(s) to the output layer. Adjusting the weights of connection during training leads to cope predictions to target values for specific records, the network "learns" to generate better and better predictions [33].

3) The Support Vector Machine (SVM) is a technique used for supervised learning that reproduces input-output functions for mapping from a group of training labeled data. The mapping function can be either a regression function or a classification function. For classification, nonlinear kernel functions are usually exploited to transform input data to a high-dimensional feature space in which the input data become more separable compared to the original input space [34].

4) Decision tree (C4.5); the voting for boosted C4.5 classifiers' algorithm is as follows, For each record, each composite classifier (decision tree or rule set) assigns a confidence and a prediction. The sum of confidence figures for each output value is computed, and the value with the greatest confidence sum is selected as the final prediction [35].

5) Partial Tree (PART): This algorithm provides only a partial specification of the data. A model of executable data should always contain the binary type for each field so, that the output and input data can be marshaled in correct manner. The sufficiently input model that is specified to allow a peer to compute plan of execution is called executable for that peer [36].

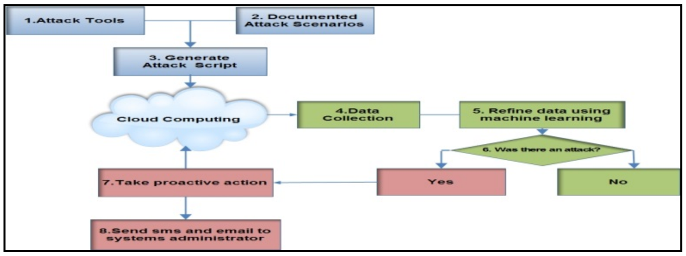
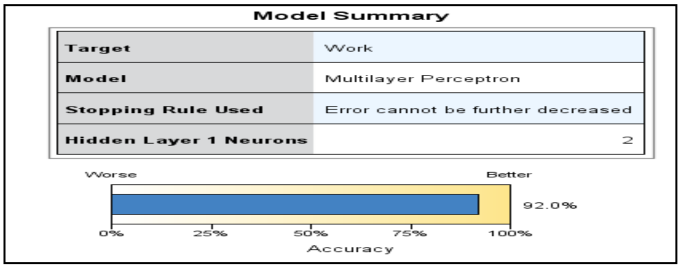

1.Attack Tools
2. Documented Attack Scenarios
3. Generate Attack Script
Cloud Computing
4.Data Collection
5. Refine data using machine learning
6. Was there an attack?
Yes
No
7.Take proactive action
8.Send sms and email to systems administrator

Fig. 4. Attack detection and proactive resolution in single cloud environment Using machine learning (modified by the authors) [8]

All the above algorithms have been implemented in Weka [37], which is a Java-based popular machine learning tool. C4.5 [37] algorithm has been used in Weka for implementation of decision tree [51]. At the beginning, some experimental tests have been carried out in order to determine the best-suited technique for classifying the attack. However, we suggest using C4.5 algorithm "this algorithm needs further explanation" in the cloud system, because it is a comparatively established algorithm and is computationally cheaper than PART. The next selection for our task "what do you mean by this" is multilayer perceptron.

TABLE I. CLASSIFICATION ACCURACY OF DENIAL-OF-SERVICE ATTACK

	Naïve Bayes	Multilayer perceptron	Support vector machine	part	Decision tree
Classification accuracy (%)	75	92.0	92.45	93	94
No .of unclassified instances	0	0	0	0	0
Model building time (s)	0.02	4.55	0.67	0.11	0.06
Model testing time (s)	0.01	0.01	0.01	0.01	0.00


Model Summary

Target	Work
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	2

Fig. 5. Classification accuracy

At the beginning, we carried out some experimental tests to identify the best-suited technique for attack classification. The details of performances are available in Table 1. We primarily consider classification accuracy, number of unclassified instances and computational complexity. The classification accuracy calculated the percentage of activities that were classified correctly by the machine learning techniques. The number of unclassified instances basically measured the technique's limitations, which means it failed to classify any attack as shown in Figure 5. We are also aware of the computational efficiency of the techniques and how well they learn because we are dealing with comparatively large data sets. Therefore, we observe the model building and testing time, which are listed in Table 2.

On the basis of the classification accuracy, number of unclassified instances and computational complexity, we found decision tree C4.5 could be a preferred choice for DoS attack classification in the cloud computing area. The classification accuracy and number of unclassified instances essentially summaries the average performances of the techniques for our attack classification task. So we tried to observe the details of performance about the attack classification scenario. As a result, we employed confusion matrix [38] analysis to see the details of the techniques' performance measures.

B. C4.5 Decision Tree Algorithm

Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. [35, 40] has developed C4.5 algorithm. A large tree can be constructed by C4.5 taking into account all attribute values and finalizes the decision rule by pruning [38]. This algorithm uses a heuristic methodology for pruning, depending on the statistical significance of splits [39]. The process of tree construction essentially calculates information gain and the entropy to finalize the decision tree. Depending on this gain information, the C4.5 can determine the occurrence or non-occurrence of an attack. The expected information or entropy depends on the set partitioning process into subsets by the equation [1]:-

$$E(S) = -\sum_{j=1}^{n} f_{s}(j)\log_2 f_{s}(j) \qquad (1)$$

Where:-

- E(S) is the subset information entropy (S);
- n is the number of different values of the attribute in S (entropy is computed for one selected attribute);
- $f_S(j)$ is the frequency (proportion) of the value j in the subset S; and
- Log_2 is the binary logarithm. Entropy of (0) defines a perfectly classified subset, whereas (1) indicates a completely random composition. Entropy is used for determining the next node to be split in the algorithm. This means that raising the entropy, leads to increase in the potential to improve the classification.

The encoding information that would be gained by branching on A is given by the following:-

- G(S, A) is the gain of the subset S after a split over the A attribute;
- E(S) is the information entropy of the subset S;
- M is the number of different values of the attribute A in S;
- f_S (Ai) is the items frequency that possess Ai as a value for A in S;
- Ai is the i^{th} possible value of A; and
- SAi is a subset of S that contains all items, where the value of A is Ai.

Gain quantifies the entropy improvement through splitting over an attribute: higher is better. For to constructing the final decision tree, the algorithm computes the information gain of each attribute [40].

$$E = \frac{1}{N}\sum_{i=0}^{N} E_i \qquad (2)$$

We build a model based on data mining for evaluating the security state of cloud computing through simulating an attack from a malicious source. This process involves identification and utilization of vulnerabilities in real world scenario which may occur in the cloud due to improper configuration, known or unknown weaknesses in software systems, or hardware, operational weaknesses or loopholes in deployed safeguards.

We will use how strategy of inferring and analyzing the data, searching for them in the cloud by one of the technology tools (data mining) this paper shows the vision of the insurance. and the general arrangement for extracting the required data, through the cloud , enabling fighting terrorism to limit the harms in advance by making the relief arrangements from the view of comprehensive security and through the analysis of the results for the data survey.

This process of assigning predictions to individual records is known as scoring. By scoring the same records used to estimate the model, we can evaluate how accurately it performs on the training data—the data for which we know the outcome. This example uses a decision tree model, which classifies records (and predicts a response) using a series of decision rules.

C. Testing and verification of security and Integrity using a simple decision tree model

Using a simple decision tree model Chaid algorithm security rating for classifying the data including the fields of entry or the variables, Decision tree is the structure of the tree on the shape of tree branches that represents sets of decisions. These decisions generate rules for classification of the set of the data. It includes limited forms for the branches of the branches, which includes the decision of classification, or decline, it includes the space of the automatic discovery of the mistakes.

TABLE II. FIELD NAME DESCRIPTION

Data security		*N*	Marginal Percentage
sec	Y	*13*	52.0%
	No	*12*	48.0%
variable	X1	*6*	24.0%
	X2	*11*	44.0%
	X3	*8*	32.0%
Valid		*25*	100.0%
Missing		*0*	
total		*25*	
subpopulation		*25(a)*	
a. The dependent variable has only one value observed in 25(100.0%)			

Table 2. shows the Coding Input data (0, 1) and the independent variables [$(x_{1}, x_{2}, x_{3}, x_{4}, x_{5})$ & Y]

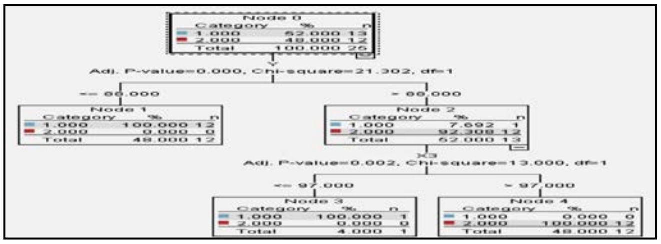


Fig. 6. Analyses of Decision Tree Model “by the authors"

Results and discussion: Figure 6. shows the upper part of the first node of the tree in C4.5, it summarizes all the records in the set of the data. We can find the rate of secured data is 48% that corresponds to 12 scores only in the cases of the set for the data sample representing the secured data (that has protection). While the rate of unsecured data is 52% that represents 13 scores for the data exposed to risk (unsecured), it needs to improve the performance for protection and security, it is exactly the first part of the analysis, so let us see if each tree can give us any evidence to what are the factors that may be responsible for the attack. Figure. 5 shows that the first division is according to the level of the input data. So, it will be possible to assign or to determine the scores on terms that the income level in allow class to (node 2) it is not surprising to see that this classification contains the highest rate of the unsecured data, it is a clear indicator for the data of this class, to contain high risks and needs a solution thus the rate of 52% for the data of this class represents a risk actually, if not supposedly, consequently, the prediction model practically cannot respond but that the model must be good and allow us to expect and to respond more likely for each score based on the available data by the same way . According to the analysis of node (2) shown in Figure 5, we can find that the vast majority (92.308%) appears unsecured that represents a risk and needs to set a new mechanism for security. So, the standards of security can be improved in this set of data to reduce the risk, accordingly we learned that each score is an indicator for this model. We will determine the points of weakness in the cloud through assigning particular node. The new predictions assignment (either good or bad), depending on the prevailing response for this node, this process is known for assigning the predictions of the individual scores as it is the aim, by recording the same scores that are used for assessing the model. According to the percentage, we can assess the extent of accuracy for the training data. This model is used for the decisions tree that classifies the scores. It's expected the response by using a group of rules for taking the decision.

TABLE III. CASE PROCESSING SUMMARY

Field name	Description,
Input variable	$(x_{1}, x_{2}, x_{3}, x_{4}, x_{5})$ & Y
Security rating	Security rating : 0 = attack 1 = security
Data risk	Number of test range of security 1 = < 88.00 , 0 > 88.00

Table 3. Shows the cloud needs to be improved and to enhance its sufficiency and taking the necessary arrangements to raise the efficiency of the security. As the data is exposed for the occurrence of violations at the rate of 48.0% is no secure.

V. CONCLUSION AND FUTURE WORKS

Although cloud computing is a new emerging technology that introduces a number of benefits to the users, but unfortunately it faces lot of security challenges. In this paper data security challenges and solutions are provided for these challenges in order to overcome the risk included in cloud computing. In this paper, a review on cloud computing with the main focus on gaps that hinders cloud adoption has been undertaken, and at the same time, a review about threat remediation challenges has been mentioned.

The paper presented the performance of machine learning techniques used in attack identification in a cloud computing environment. A statistical ranking approach has been used for the final selection of a learning technique for the task. C4.5 technique's performance has been evaluated through different performance evaluation matrices that included the rigorous testing of 10-fold cross-validation, true positive rate, false positive rate, precision, recall, F-measure and the area of receiver operating characteristic. In another phase, we also counted computational complexity for our final selection.

Our experimental results showed that, using a simple decision tree model Chaid algorithm security rating for classifying approach is a robust technique that enables the decision-maker to measure the extent of cloud securing, show that the cloud needs to be improved and to enhance its sufficiency and taking the necessary arrangements to raise the efficiency of the security indicated the fact that C4.5 gives a better performance and the level of performance has acceptable standard. It is found that rule-depending technique (C4.5) is efficient technique for solve the problem of security. However, on the basis of the computational performance, we suggest C4.5 as the better technique for real-time attack protection in a cloud environment.

The paper presented some recommendation regarding the customers and vendors; however, to overcome the customer concerns about application and data security, vendors should deal with these issues head-on. In the future, concrete standards for cloud computing security should be improved. In the future, we will continue and follow up the study in this field through using search in data to be an active way in decision making. It is expected that there will be several challenges related to operation and development of cloud computing system. The use of data mining techniques in cloud computing will be an effective tool that will help in securing the data.

REFERENCE

[1] Schubert L, Jeffery K, et al. The future for cloud computing: opportunities for European cloud computing beyond 2010. Expert Group report, public version 2010; 1. http://cordis. europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf

[2] Khorshed MT, et al. Trust issues that create threats for cyber attacks in cloud computing. In Proceedings of IEEE ICPADS, December 7–9, 2011, Tainan, Taiwan, 2011.

[3] Armbrust M, et al. Above the clouds: a Berkeley view of cloud computing. EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009- 28, 2009.

[4] Brunette G, Mogull R. Security Guidance for critical areas of focus in Cloud Computing V2. 1. CSA (CloudSecurity Alliance), USA. Disponible en: https://cloud securityalliance.org/csaguide.pdf, vol. 1, 2009.

[5] Catteddu D, Hogben G. Benefits, risks and recommendations for information security. European Network and Information Security Agency (ENISA), 2009.

[6] Mell P, Grance T. The NIST definition of cloud computing. National Institute of Standards and Technology 2009; 53(6): 50. http://csrc.nist.gov/publications/nistpubs/ 800-145/SP800-145.pdf

[7] Khorshed MT, et al.Monitoring insiders activities in cloud computing using rule based learning. In Proceedings of IEEE TrustCom-11, Nov. 16–18, Changsha, China, 2011.

[8] Khorshed MT, et al. A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Generation Computer Systems 2012.

[9] NIST. (2011, 21 May 2011). NIST Cloud Computing Program. Available: http://www.nist.gov/itl/cloud/

[10] Ness G. (2009, 22 May 2011). 3 Major Barriers to Cloud Computing. Available: http://www.infra20. com/post.cfm/3-major-barriers-to-cloud-computing

[11] Vouk, M.A., "Engineering of Telecommunications Software", High-Speed ... Journal of Computing and Information Technology, Vol 16 (4), 2008, pp 235-246.

[12] Brodkin J. Gartner: seven cloud-computing security risks, 2008. Available: http://www.infoworld.com/d/ security-central/gartner-seven-cloud-computing-security risks- 853 (Retrieved: 6th August 2012).

[13] Archer J, Boehm A. Security guidance for critical areas of focus in cloud computing. Cloud Security Alliance 2009. Available: https://cloudsecurityalliance.org/ guidance/csaguide.v1.0.pdf

[14] Archer J, et al. (2010, 7 May 2011). Top Threats to Cloud Computing,Version 1.0. Available:http://www.cloud securityalliance.org/topthreats/csathreats.v1.0.pdf

[15] Monfared AT. Monitoring intrusions and security breaches in highly distributed cloud environments. 2010.

[16] Grosse E, et al. Cloud computing roundtable. Security & Privacy, IEEE 2010; 8: 17–23.

[17] Wrenn G. (2010, 25 May 2011). Unisys Secure Cloud Addressing the Top Threats of Cloud Computing. Available:http://www.unisys.com/unisys/common/download.jsp?d_id=1120000970002010125&backurl=/unisys/ri/wp/detail.jsp&id=1120000970002010125

[18] Grobauer B, et al. Understanding cloud-computing vulnerabilities. IEEE Security and Privacy 2010; 50–57. DOI: 10.1109/MSP.2010.115

[19] Thomas Sommer, et al. The Conundrum of Security in Modern Cloud Computing. (2012) Communications of the IIMA: Vol. 12: Iss. 4, Article 2. Available at: http://scholarworks.lib.csusb.edu/ciima/vol12/iss4/2

[20] Yildiz M, et al. A Layered Security Approach for Cloud Computing Infrastructure. Publisher IEEE, 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks, Kaohsiung, 2009; 763–767. DOI: 10.1109/I-SPAN.2009.157

[21] Dahbur K, et al. A survey of risks, threats and vulnerabilities in cloud computing. In Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications, ISWSA '11, ACM, New York, USA, 2011; 12.

[22] Wang C, et al. Ensuring Data Storage Security in Cloud Computing. Publisher IEEE, 17th International Workshop on Quality of Service, 2009. IWQoS, Charleston, SC, 2009; 1–9.

[23] Yan L, et al. Strengthen cloud computing security with federal identity management using hierarchical identity-based cryptography. Cloud Computing 2009; 5931: 167–177. DOI: 10.1007/978-3-642-10665-1_15

[24] Ristenpart T, et al. Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. In Proceedings of the 16th ACM conference on Computer and communications security Chicago, Illinois, USA, November 09–13, 2009; 199–212.

[25] Chonka A, et al. Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks. Journal of Network and Computer Applications 2010. DOI: 10.1016/j.jnca.2010.06.004

[26] Danchev D. (2011, 31 May 2011). Dancho Danchev's Blog—Mind Streams of Information Security Knowledge. Available: http://ddanchev.blogspot.com/

[27] Grossman J. (2011, 19 June 2011). Jeremiah Grossman. Available: http://jeremiahgrossman.blogspot.com/

[28] Company H.-P. D. HP ProLiant DL380 G4 server - specifications, 2012. Available: http://h18000.www1. hp.com/products/servers/proliantdl380/specifications-g4. html (Retrieved: 6th August 2012).

[29] Knorr, E., & Gruman, G. (2009). What cloud computing really means. Retrieved from http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031

[30] Corporation M. Windows 7, 2012. Available: http://windows. microsoft.com/en-au/windows7/products/home (Retrieved: 6th August 2012).

[31] McDowell M. (2009, 21 June, 2011). Understanding Denial-of-Service Attacks. Available: http://www.uscert. gov/cas/tips/ST04-015.html

[32] John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995; 338–345.

[33] Lopez R, Onate E. A variational formulation for the multilayer perceptron. Artificial Neural Networks–ICANN 2006 2006; 4131: 159–168. DOI: 10.1007/11840817_17

[34] Platt JC. Fast training of support vector machines using sequential minimal optimization. 1999; 185–208.

[35] Quinlan JR. C4. 5: Programs for Machine Learning. Morgan Kaufmann: San Mateo, CA, 1993.

[36] Frank E, Witten IH. Generating accurate rule sets without global optimization. In FifteenthInternational Conference on Machine Learning, 1998; 144–151.

[37] Witten IH, et al. Data Mining: Practical Machine Learning Tools and Techniques (3rd edn). Morgan Kaufmann: San Francisco, 2011.

[38] Kohavi R, Provost F. Glossary of terms. Machine Learning 1998; 30: 271–274.

[39] Ali ABMS, Wasimi SA. Data Mining: Methods and Techniques. Thomson. 2007.

[40] Quinlan JR. Induction of decision trees. Machine Learning 1986; 1: 81–106.

[41] Ali ABMS, Smith KA. On learning algorithm selection for classification. Journal on Applied Soft Computing, Elsevier 2006; 6: 119–138.

[42] Shafiullah G, et al. Prospects of renewable energy—a feasibility study in the Australian context. Renewable Energy, ELSEVIER 2012; 39(1): 183–197.

[43] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner. A break in the clouds: towards a cloud definition, in: ACM SIGCOMM Computer Communication Review, 2008.p.50-55.

[44] M.B. Mollah, K.R. Islam, and S.S. Islam. Next generation of computing through cloud computing technology, in: 2012 25th IEEE Canadian Conference on Electrical Computer Engineering (CCECE), May 2012.p.1-6.

[45] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable and fine-grained data access control in cloud computing, in: IN-FOCOM, 2010 Proceedings IEEE, 2010.p.1-9.

[46] Kresimir Popovic and Zeljko Hocenski. Cloud computing security issues and challenges, in: MIPRO, 2010 Proceedings of the 33rd International Convention, 2010.p.344-349.

[47] Akhil Bhel, Emerging Security Challenges in Cloud Computing. Information and Communication Technologies, in: 2011 World Congress on, Mumbai, 2011.p.217-222.

[48] Farzad Sabahi. Cloud Computing Security Threats and Responses, in: IEEE 3rd International Conference on Communication software and Networks(ICCSN), May 2011.p.245-249.

[49] Eman M.Mohamed, Hatem S Abdelkader, Sherif EI Etriby. Enhanced Data Security Model for Cloud Computing, in:8th International Conference on Informatics and Systems(INFOS), Cairo, May 2012.p.12-17.

[50] Wentao Liu. Research on Cloud Computing Security Problem and Strategy, in: 2nd International Conference on Consumer Electronics. Communications and Networks (CECNet), April 2012.p.1216-1219.

[51] Eystein Mathisen. Security Challenges and Solutions in Cloud Computing, in: International Conference on Digital Ecosystems and Technologies (IEEE DEST 2011), 2011.p.208-212.

[52] R. Velumadhava Raoa, K. Selvamanib , Data Security Challenges and Its Solutions in Cloud Computing , 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Conference Organized by Interscience Institute of Management and Technology, Bhubaneswar, Odisha, India ,Peer-review under responsibility of scientific committee of International Conference on Computer, Communication and Convergence (ICCC 2015)

[53] J. Staten, “Is Cloud Computing Ready For The Enterprise,” Forrester ... 17th International workshop on Quality of Service, pp.1–9, July 13–15, 2009.

[54] Jay Heiser and Mark Nicolett, " An Engineering Process to Address Security Challenges in Cloud Computing" ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference, Stanford University, May 27-31, 2014

[55] Seccombe A, et al. Security guidance for critical areas of focus in cloud computing. Cloud Security Alliance, 2009.

[56] Paula Kotzé , et al. Secure cloud computing: Benefits, risks and controls, ieeexplore.ieee.org/iel5/6017604/.../06027519.pdf