Research Article

Network Intrusion Detection System (NIDS) in Cloud Environment based on Hidden Naïve Bayes Multiclass Classifier

Hafza A. Mahmood, Soukaena H. Hashem
Department of Computer Science, University of Technology, IRAQ.
*Correspondent Author Email: h_adel_89@yahoo.com

ArticleInfo

Received
02/06/2016

Accepted
05/10/2016

Abstract

Cloud Environment is next generation internet based computing system that supplies customizable services to the end user to work or access to the various cloud applications. In order to provide security and decrease the damage of information system, network and computer system it is important to provide intrusion detection system (IDS. Now Cloud environment are under threads from network intrusions, as one of most prevalent and offensive means Denial of Service (DoS) attacks that cause dangerous impact on cloud computing systems. This paper propose Hidden naïve Bayes (HNB) Classifier to handle DoS attacks which is a data mining (DM) model used to relaxes the conditional independence assumption of Naïve Bayes classifier (NB), proposed system used HNB Classifier supported with discretization and feature selection where select the best feature enhance the performance of the system and reduce consuming time. To evaluate the performance of proposal system, KDD 99 CUP and NSL KDD Datasets has been used. The experimental results show that the HNB classifier improves the performance of NIDS in terms of accuracy and detecting DoS attacks, where the accuracy of detect DoS is 100% in three test KDD cup 99 dataset by used only 12 feature that selected by use gain ratio while in NSL KDD Dataset the accuracy of detect DoS attack is 90 % in three Experimental NSL KDD dataset by select 10 feature only.

Keywords: Cloud Environment, IDS, NSL KDD Dataset, KDD CUP 99 Dataset, Multiclass Classification, Hidden Naïve Bayes (HNB), and Dos.

الخلاصة

ان بيئة الغيمة هي الجيل الجديد الذي يعتمد على نظام الحوسبة والذي يجهز مختلف الخدمات الى المستخدمين للوصول والعمل على تطبيقات الغيمة المتعددة. لكي يتم توفير الامن وتقليل الاضرار التي تحصل لنظام المعلومات و الشبكات ونظام الحاسبة فأن من الضروري توفير نظام كشف التطفل في بيئة الغيمة. لقد اصبحت بيئة الغيمة حاليا تحت تاثيرالمتطفلين على الشبكة ويعد DoS من اكثر الانواع انتشارا وهجومية هو الذي يسبب تاثير خطير على بيئة الغيمة. في هذا البحث تم اقتراح المصنف HNB ليتم اكتشاف DoS والتي تعتبر احدى طرق DMالتي تستخدم للتخلص من افتراضية الاستقلالية المشروطة الموجودة في NB, النظام المقترح يستخدم HNB مع discretization و feature selection حيث يتم اختيار افضل حقول لتحسين اداء النظام وتقليل وقت التنفيذ, لتقييم النظام المقترح تم استخدام KDD Dataset, NSL KDD. حيث اظهرت النتائج ان استخدام HNBيحسن من اداء نظام كشف التطفل من ناحية اكتشاف Dos حيث اصبحت نسبة اكتشافه 100% في ثلاث مجاميع من KDD Cup 99 استخدمت لفحص النتائج بأستخدام اثنى عشر صفة فقط تم اختيارها بأستخدام GR بينما في NSL KDD كانت نسبة اكتشاف DoS 90% في ثلاث مجاميع مختلفة اخرى بأستخدام عشرة صفات فقط.

Introduction

Cloud computing enables the customers to access and use resources that are distributed in the internet to make processing or computations without installing in their own computer and they must to pay just for the service they consumed, it is a modern technology that provide immediately access to resources as per the needs of the users [1]. Cloud environment started in the mid of 2007 and it is developed rapidly to satisfy infusion and diffusion of IT in systems, it's important to provide IDS in cloud environment because of the Distributed model of cloud that makes it susceptible and prone to sophisticated attacks like DoS, ID is process of examining the events happen in a network resources or computer system and analyzing them to determine the presence of intrusion and possible accident that can cause threats to security measures [2]. While the IDS are defined as the hardware or software product that detecting attacks over network, computer systems or against information systems [3]. ID methods can be classified into misuse detection and anomaly detection, in misuse detection that is also called rule-based detection or signature-based the user's activities are compared with known behaviors of attackers, its gathered information, analyzed and compared with huge databases for attack signatures [4]. While in anomaly detection is used to identify abnormal behavior on a network or host, where assume that intrusions are different from legitimate events and therefore can be detected by the systems that identify these differences [5].

Data mining (DM) is used for extracting relevant information from huge database; DM techniques are used to analyze and monitor large network data and classify these data into anomalous and normal data. DM commonly involves four classes of task. Clustering, Classification, Regression and Association rule learning [6]. A classification is process of taking each instance in dataset and determines it to a specific class attack or normal, that means known structure will be used for new instances [7].

In last two decades, there are several studies focused on reducing the independence assumption of NB classifier, one of these studies introduced HNB classifier, this new model depends on build additional layer, this layer represents a hidden parent for every feature as shown in Figure 1. The benefit of using hidden parent (A_{hpi}) is to gather the weighted influences from all other features (A_i), where i j= 1, 2,… n and i is not equal to j, and P(C) is the probability of class. Joint distribution is defined as Equation 1, while the hidden parent defined as Equation 2, and HNB classifier is defined as Equation 3 [8].

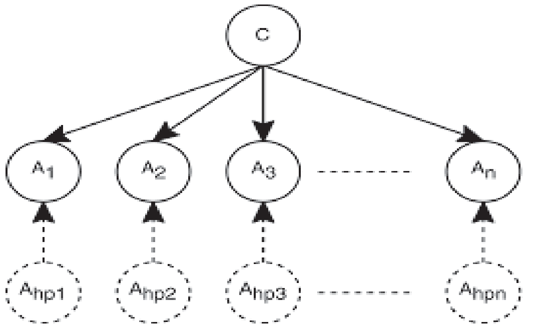

Figure 1: HNB Structure.

$$P(A_{1,\ldots,}A_2|\mathrm{C}) = \mathrm{P(C)} \prod_{i=1}^{n} P(A_i|A_{hpi,}\, C) \quad (1)$$

$$\mathrm{P}(A_i|A_{hpi}, C) = \sum_{j=1, j\neq i}^{n} W_{ij} * P(A_i|A_j, C) \quad (2)$$

$$c(E) = \arg\max_{c\in C} P(c) \prod_{i=1}^{n} P(a_i|a_{hpi}, c) \quad (3)$$

The method to calculate the weights W_{ij}, is by using conditional mutual information (CMI) between every two features A_i and A_j as shown in Equation 4, The CMI is defined as Equation 5 [8]:

$$W_{ij} = \frac{I_p\ (A_i; A_j|C)}{\sum_{j=1, j\neq i}^{n} I_p(A_i; A_j|C)} \quad (4)$$

$$I_p(A_i; A_j|C) = \sum_{a_i, a_j, c} P(a_i, a_j, c)\ log\ \frac{P(a_i, a_j|c)}{P(a_i|c)P(a_j|c)} \quad (5)$$

Feature selection is an essential data processing step prior to applying a learning algorithm. Feature selection is a process of finding a subset of significant features from the original set of features and reduces the number of irrelevant redundant features from dataset to improve the performance of the classification and also decreases storing of memory space, one of the most common methods in feature selection is Information Gain that measures the amount of information in bits about the class prediction. It measures the expected reduction in entropy. Entropy measure is considered as a measure of systems unpredictability which is usually used in information theory measure.

The Expected information (Entropy) of a feature A is defined as Equation 6, where a is a value of feature, and a =1, 2, …, n. The Information needed to classify D after using A for divide D into n partitions is mention in Equation 7. Information gained by branching an attribute A as in Equation 8 [9]:

$$\text{Info D} = \text{H(A)} = -\sum_{a=1}^{n} \text{P(a)} \,\text{Log}_2\, \text{P(a)} \quad (6)$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} * I(D_j) \quad (7)$$

$$\text{Gain (A)= Info D - } Info\ _{A}\text{(D)} \qquad (8)$$

Gain ratio (GR) is an enhancement of the information gain to solve the matter of bias towards features with big set of values that appeared in Information Gain. GR should be small when all data belong to one branch attribute and large when data is equally spread. GR selecting an attribute by takes size and number of values into account.

It's correct IG by taking the substantial information of a split into account (i. e. How much information is needed to determine which branch the instance belongs to?) Where substantial information is the entropy of distribution of instances into branches based on Equation 9. This value generated by splitting the training data set as in Equation 10 where represents the substantial information [10]:

$$Split\ Info_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \log_2 \frac{|D_j|}{|D|} \qquad (9)$$

$$\text{Gain Ratio (A)} = \frac{\text{Gain (A)}}{\text{Split Info (A)}} \qquad (10)$$

Related Work

Mukherjeea S. et al., 2012, discussed the importance of reduce features to build effective and efficient IDS. They checked performance of (Information Gain, Gain Ratio and Correlation-based Feature Selection methods, they propose Feature Vitality Based Reduction Method to identify the importance of reduce feature. They applied NB classifier on NSL KDD dataset for ID. Experimental results showed that select Features enhance performance to design effective and efficient NIDS [11].

Koc L. et al., 2012, introduced HNB model as a solution of ID problem. To decrease the resource requirements and enhance the accuracy, they used NB and structurally extended Naïve Bayes methods augmented with feature selection and discretization. They compared the performance of the NB classifier and leading extended Naïve Bayes approaches with the HNB classifier as an IDS, they uses KDD99 dataset, The results proved that HNB model enhance the accuracy of detecting DOS attacks, where the accuracy of detect Dos is 0.99 [12].

Padmakumari P. et al., 2014 presented IDS in a cloud environment, to detect most occurring attacks in several network environments by applying the Apriori algorithm using k-means clustering and combine it with a frequent attacks generation module.

Experimental results showed that applying a clustering algorithm separately for different attributes enhance the accuracy of detection. The frequent attack detection module increases the reliability and achieve low false alarm rate, they used KDD 99 CUP dataset to evaluate their system [13].

Koc L. et al., 2015, they discussed that the HNB binary classifier model can be applied to ID problem. They used KDD Cup 99 dataset to prove that the HNB binary classification model with CONS feature selection method and EMD discretization enhance performance of system in terms of accuracy and error rate than the traditional NB model, where the accuracy of detect normal and attack events is 0.93 [8].

Datasets and Attacks in Cloud Environment

The KDD Cup 99 dataset is widely used in IDS which consist of 10% of the original dataset that containing 494,020 records each record consist of 41 features and class feature labeled either normal or attack. It has 80.31% attack and 19.69% normal. The NSL KDD data set solve some of the ingrained problems of the KDD CUP 99 dataset, which selected records of the complete KDD data set that contain the same features as KDD cup 99. The class feature contains 21 kinds of attacks within four types: DOS, Probe, R2L attacks and U2R attacks as mention in Table 3 [13].

Table 3: description of attacks in KDD Cup 99 and NSL KDD Datasets.

Attack type	Description	Types
DOS	Denial of services attacks	Pod, Land, smurf, back etc.
Probe	Surveillance and probing	Satan, ipssweep, nmap etc.
R2L	Unauthorized access from remote machine to local machine	Guess_passwd, ftp_write, imap, phf etc.
U2R	Unauthorized access to local superuser privileges by a local unprivilege user	Rootkit, buffer overflow, loadmodule etc.

The benefit of using NSL KDD over the original KDD data set, that it doesn't contain redundant records in the train and test dataset and from every difficulty level set, the number of records that selected is inversely commensurate to the percentage of records in KDD 99 dataset [14].
Since large size of data translates between cloud environments, the intrusions are eager to exploit the vulnerabilities in cloud and by this way they can gain the important data. DoS attacks are the dangers attack among numerous threats in cloud computing, even the Cloud Security Alliance has been indicated as one of the nine major attacks. DoS make the system cannot respond to any requests by overloads the system with requests and that leads to make the resources unavailable to its users [1].

Proposal Network Intrusion Detection System

The proposed system is multiclass NIDS in Cloud environment based on HNB classifier, as we mention above the attackers in cloud environment is different from traditional network, where is usually from DOS attack which is the most dangerous attack that effect the availability of resource, the reason of used NIDS instead of host intrusion detection system HIDS is that the HIDS can be detected by use antivirus, to evaluate the system we used the well-known dataset KDD Cup 99 and NSL KDD Dataset. Figure 2 describes the general structure of the proposed NIDS, for more understanding see Algorithm 1. The proposed NIDS consists of the following steps:

1. Normalization.
2. Discretization.
3. Feature Selection method.
4. Training and testing

Algorithm 1: General structure of the proposed system.

```
Input: training dataset
Output: evaluation for three test dataset
Begin
  1. Normalization process
  For each Attribute in Dataset
        select Maximum value (Max)
        select Minimum value (Min)
        For each value v in Attribute
     Combine the new value by use Equation 11
          $\text{New v} = \frac{\text{Old v} - \text{Min}}{\text{Max} - \text{Min}}$        (11)
        End For
  End For
  2. For each continues feature in dataset
        Discrete the values into specific range
    End for
  3. Feature selection
    Find the size of training dataset D
    Find the Probability of each class
    compute the entropy of five class (c) to find info
      D by use Equation 6
          $\text{Info D} = \text{H(A)} = -\sum_{a=1}^{n} \text{P(a)} \operatorname{Log}_2 \text{P(a)}$        (6)
    For each Feature F in training dataset
     For each value j in Feature F
     compute the frequency of value in all training
          dataset Ft
     compute the frequency of value with each
          class $Fn$
     compute the entropy for each value with five
          class by using Equation 6
          $\text{I}(D_j) = -\sum \frac{Fn}{Ft} \operatorname{Log}_2 \frac{Fn}{Ft}$
     End For
     compute info A by used Equation 7:
          $Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} * I(D_j)$        (7)
     compute gain for each Feature as in Equation
          8
```

```
            Gain (A)= Info D - $Info_A(D)$          (8)
        compute Split Info by use Equation 9
            $\text{Split Info}_A(D) = -\sum_{j=1}^{v} \frac{|Ft_j|}{|D|} \log_2 \frac{|Ft_j|}{|D|}$          (9)
        compute the Gain ratio by use Equation 10
            $\text{Gain Ratio (A)} = \frac{\text{Gain (A)}}{\text{Split Info (A)}}$          (10)
        End For
    4. Select set of features that have the highest gain ratio.
    5. Applied HNB classifier in training dataset to build the NIDS by use Algorithm 2.
    6. Evaluation the proposed system by use three experimental test dataset
       For each test dataset
            Compute accuracy (acc) by use Equation 12
                $acc = \frac{TP + TN}{TP + TN + FP + FN}$          (12)
            Compute the detection rate (DR) by use Equation 13
                $DR = \frac{TP}{TP+FN}$          (13)
            Compute error rate(ER) by use Equation 14
                $ER = \frac{FP+FN}{TP+TN+FP+FN}$          (14)
            Find the confusion matrix
       End for
End
```

Normalization dataset

The first step in the proposed system is applied normalization process to continue feature in dataset to enhance the performance and effectiveness of the system by making the values of attribute within specific range from 0 to 1, in our system will be used Min-max normalization method.

Discretization dataset

As a result of contains continues and discrete feature in KDD Cup 99 and NSL KDD Datasets it is important to convert the continuous attribute to discrete to ensure the efficiency of the system and to solve the problem of appear new value when test dataset which it is not appeared in training dataset.

Feature Selection

Feature selection is one of the most important preprocessing of DM methods that used to remove the unrelated and redundant features in large dataset, and to improve the performance of the system by use the correct feature and reduce the consuming time. In our study, we used gain ratio as a feature selection method.

Training and Testing

The system used HNB Classifier (see Algorithm 2) by select 4000 records in learning phase by select 2169 DOS, 388 probes, 173 R2L, 35 U2R and 1235 normal in both datasets (KDD cup 99 and NSL KDD), while in test phase it will be used 1200 samples to evaluate the work and two other datasets (600,900) samples to validate the performance of the system in KDD Cup 99 Dataset, the selection samples of attack mention in Table 4. While in NSL KDD Dataset the test samples that have been used is 1028 and two other dataset to validate the performance of system with (795 and 566), as mention in Table 5. It is important to note that, the NSL KDD Dataset different from the original KDD Cup 99 Dataset where the samples of attack is less than the KDD Cup 99 Dataset as a result of remove the redundant samples and there is some kind of attack is not mention in NSL KDD Dataset like (warezclient and spy) Which is R2L attack, for that reason the selected test dataset in NSL KDD is different from the selected test in KDD Cup 99.

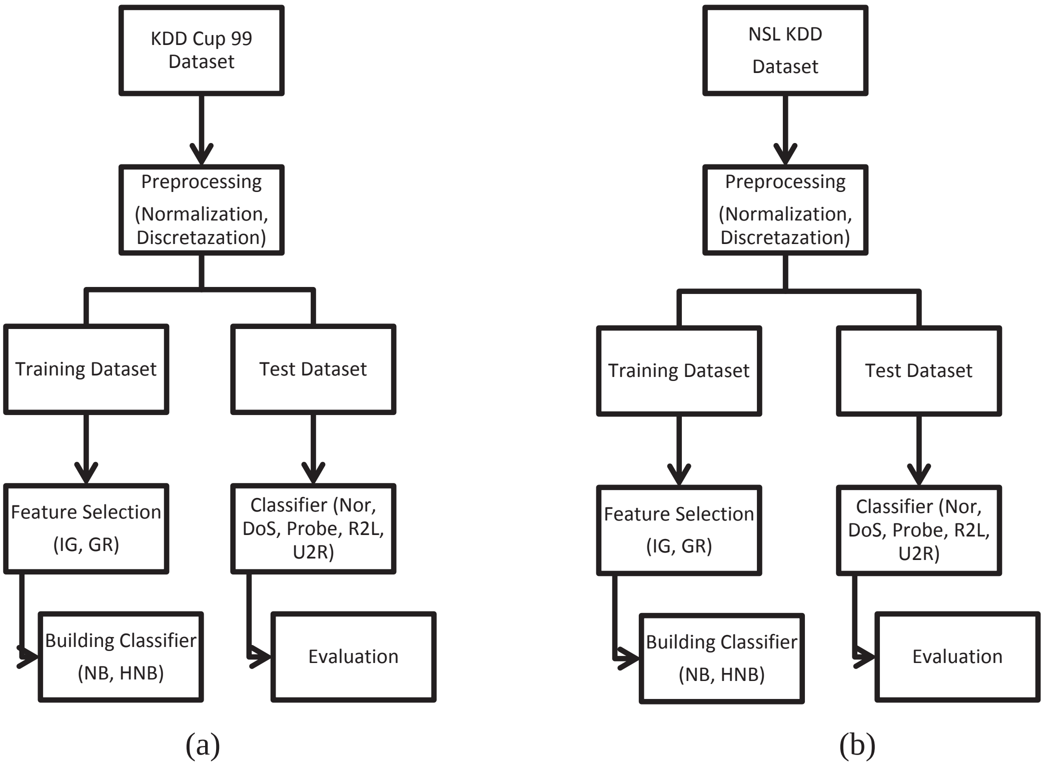
KDD Cup 99 Dataset
Preprocessing (Normalization, Discretazation)
Training Dataset
Test Dataset
Feature Selection (IG, GR)
Classifier (Nor, DoS, Probe, R2L, U2R)
Building Classifier (NB, HNB)
Evaluation
(a)
NSL KDD Dataset
Preprocessing (Normalization, Discretazation)
Training Dataset
Test Dataset
Feature Selection (IG, GR)
Classifier (Nor, DoS, Probe, R2L, U2R)
Building Classifier (NB, HNB)
Evaluation
(b)

Figure 2: (a) Block Diagram of KDD Cup 99 Dataset. (b) Block Diagram of NSL KDD Dataset.

Algorithm 2: Hidden Naïve Bayes Classifier

Input: training and testing dataset after normalization and discrete processes (KDD Cup 99 10% or NSL KDD Dataset)

Output: classification the test dataset

```
Begin
Step1:Training phase
 1) Compute the size of training dataset D
 2) For each class c in training dataset
        Compute P(c) from training dataset
    End for
 3) For each class c
        For each Feature $a_i$, $a_j$ in training dataset
            Compute $P(a_i, a_j, c)$ by divide FrEquation of appeared $(a_i, a_j, c)$ on training D
            Compute $P(a_i, a_j|c)$ by divide FrEquation of appeared $(a_i, a_j|c)$ on FrEquation of class
            Compute $P(a_i|c)$ by divide FrEquation of appeared $(a_i|c)$ on FrEquation of class
            Compute $P(a_j|c)$ by divide FrEquation of appeared $(a_j|c)$ on FrEquation of class
            Apply Equation 5 to find the CMI between two feature:
$I_p(A_i; A_j|C) = \sum_{a_i,a_j,c} P(a_i, a_j, c) log \frac{P(a_i,a_j|c)}{P(a_i|c)*P(a_j|c)}$ (5)
        End for
        For each feature
         compute $Wi = \sum_{j=1,j<>1}^{n} Ip(Ai, Aj|C)$
         compute $W_{ij} = \frac{I_p\ (A_i;A_j|C)}{Wi}$
        End for
    End for
Step 2: Testing phase
 4) For each record in test dataset
        For each value in test dataset
            find probability of vi with c in training dataset
        End for
        Multiply the probability of each record as Equation 2
            $P(E|c) = P(a_1, a_2, \dots, a_n|c) = \prod_{i=1}^{n} P(a_i|c)$   (2)
        Classify the record by Multiply the result of Equation 2 with probability of class and choose the maximum value to classify the record as Equation 1:
        $c(E) = \arg\max_{c \in C} P(c)P(a_1, a_2, \dots, a_n|c)$   (1)
    End For
End
```

Table 4: Test KDD Cup 99 Dataset selected.

Dataset	DOS	Probe	R2L	U2R	normal
600	342	74	23	4	157
900	515	111	36	5	233
1200	680	133	53	8	326

Table 5: Test NSL KDD Dataset selected.

Dataset	DOS	Probe	R2L	U2R	normal
566	326	68	10	6	156
795	434	100	17	11	233
1028	539	122	24	13	330

Experimental Work and Results

The proposed network intrusion detection system is used three test dataset (KDD cup 99 and NSL KDD) to evaluate the system where the records selected randomly and then build the classifier proposed system by use HNB classifier supported by discretization and feature selection method, to evaluate the detection effectiveness of the proposed system we used confusion matrix, accuracy, detection rate and error rate, the confusion matrix is a quality measurement of classifier.

KDD CUP 99 Dataset Evaluations

Table 6 shows the evaluation of classification in three KDD cup 99 test datasets with used 12 best features selected by gain ratio method. The evaluation consists of (Accuracy binary) which is the accuracy of detecting normal and attack, the accuracy of multiclass is the accuracy of detecting normal, DoS, probe, R2L and U2R, detection rate (DR), error rate (ER) and Precision. The accuracy for each class show in Table 7 that demonstrates the accuracy of detecting DoS attack is 100%.

Table 6: Performance measure of KDD cup 99 Dataset.

DS	Acc. multiclass	Acc. binary	DR	ER	Precision
Test1	0.94	0.97	0.96	0.02	100
Test2	0.92	0.97	0.97	0.02	0.98
Test3	0.93	0.96	0.95	0.03	0.99

Table 7: Accuracy for each class in KDD Cup 99 Dataset.

DS	DOS	Probe	R2L	U2R	Normal
Test1	100	0.89	0	0	100
Test2	100	0.83	0	0	0.96
Test3	100	0.87	0	0	0.99

Tables (8, 9, and 10) show the confusion matrix for Tests (1, 2 and 3) of KDD Cup 99 dataset based on select 12 feature by gain ratio which achieve best result in detecting DoS attack.

Table 8: Confusion matrix for test1.

	Normal	DOS	Probe	R2L	U2R
Normal	**157**	**0**	**0**	**0**	**0**
DOS	**0**	**342**	**0**	**0**	**0**
probe	**0**	**8**	**66**	**0**	**0**
R2L	**11**	**12**	**0**	**0**	**0**
U2R	**3**	**1**	**0**	**0**	**0**

Table 9: Confusion matrix for test2.

	Normal	DOS	Probe	R2L	U2R
Normal	**226**	**7**	**0**	**0**	**0**
DOS	**0**	**515**	**0**	**0**	**0**
probe	**0**	**18**	**93**	**0**	**0**
R2L	**17**	**13**	**6**	**0**	**0**
U2R	**0**	**5**	**0**	**0**	**0**

Table 10: Confusion matrix for test3.

	Normal	DOS	Probe	R2L	U2R
Normal	**324**	**2**	**0**	**0**	**0**
DOS	**0**	**680**	**0**	**0**	**0**
probe	**0**	**16**	**117**	**0**	**0**
R2L	**28**	**25**	**0**	**0**	**0**

As shown in the Table 7 the rate of detect R2L as R2L attack and U2R as U2R attack is low, but actually when you look at Tables (Table 8, Table 9, and
Table 10) you can observe that it's detected but another kind of attack.

NSL KDD Dataset Evaluations

The evaluation of classification in three NSL KDD test datasets viewed in Table 11 (accuracy binary, accuracy of multiclass, detection rate (DR) and error rate (ER), Precision), while the accuracy for each class show in Table 12 that demonstrate the accuracy of detecting DoS attack is best when select 10 feature based on gain ratio method**.**

Table 11: Performance measure of NSL KDD Dataset.

DS	Acc. multiclass	Acc. binary	DR	ER	Precision
Test1	0.83	0.92	0.90	0.07	100
Test2	0.82	0.92	0.90	0.06	100
Test3	0.83	0.93	0.90	0.06	100

Table 12: the accuracy for each class in NSL KDD.

DS	DOS	Probe	R2L	U2R	Normal
Test1	0.90	0.29	0	0	100
Test2	0.90	0.29	0	0	100
Test3	0.90	0.28	0	0	100

In Tables (13, 14, and 15), show the confusion matrix for Test (1, 2 and 3) of NSL KDD dataset based on select 10 features by using gain ratio method which achieves best result in detecting DoS attack.

Table 13: Confusion matrix for test1.

	Normal	DOS	Probe	R2L	U2R
Normal	157	0	0	0	0
DOS	30	296	0	0	0
probe	0	48	20	0	0
R2L	0	10	0	0	0
U2R	3	3	0	0	0

Table 14: Confusion matrix for test2.

	Normal	DOS	Probe	R2L	U2R
Normal	233	0	0	0	0
DOS	40	394	0	0	0
probe	0	71	29	0	0
R2L	1	16	0	0	0
U2R	6	5	0	0	0

Table 15: Confusion matrix for test3.

	Normal	DOS	Probe	R2L	U2R
Normal	330	0	0	0	0
DOS	50	489	0	0	0
probe	0	87	35	0	0
R2L	3	21	0	0	0
U2R	7	6	0	0	0

As shown in Table 12 the accuracy of detect normal events is 100% and the accuracy of detect DOS is 90 %, while the accuracy of detect probe as probe attack, R2L as R2L attack and U2R as U2R attack is low rate, but its detect it as a DoS attack and this is the important is to detect attack as any kind of attacks, look at Tables (13, 14, and 15).

Table 16 shows comparison the experimental results between the proposed system and the previous studies [12] [8].

Conclusions

Our research indicates the important to use NIDS in cloud environment to detect the most harmful attack in network which is DoS attack that effect the availability of the resource, The experimental results have revealed that when working with gain ratio and select only 12 features from 41 features in KDD Cup 99 dataset our detection system achieves high accuracy rate, reduce the computation time and reduce the error rate as mention in Table 7, while in NSL KDD it is best to select only 10 feature by used gain ratio method as shown in Table 11. The proposed system show that use KDD Cup 99 dataset in cloud environment is best than NSL KDD in detecting DOS attacks.

Table 165: Comparison between proposed system and previous studies.

Dataset	parameters	Pre1 2012	Pre2 2015	Proposed system		
				Test1	Test2	Test3
KDD Cup 99	Accuracy binary	_	0.9340	0.97	0.97	0.96
	Accuracy of DoS	0.99	_	100	100	100
	Accuracy multiclass	0.9372	_	0.94	0.92	0.93
	Precision	_	_	100	0.98	0.99
	Error_rate	0.06	0.0660	0.02	0.02	0.03
	Detection rate	_	_	0.96	0.97	0.95

NSL KDD	Accuracy binary	–	–	0.92	0.92	0.93
	Accuracy of DoS	–	–	0.90	0.90	0.90
	Accuracy multiclass	–	–	0.83	0.82	0.83
	Precision	_	_	100	100	100
	Error_rate	_	_	0.07	0.06	0.06
	Detection rate	_	_	0.90	0.90	0.90

References

[1] Vidhya. V, "A Review of DOS Attacks in Cloud Computing," *IOSR Journal of Computer Engineering*, vol.16, no.5, 2014.

[2] Singh S. and Srivastava R, "Intrusion Detection Using Data Mining Technique," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol.2, no.4, 2013..

[3] Bandgar M., dhurve K., Jadhav S., Kayastha V. and Parvat T. J., "Intrusion Detection System using Hidden Markov Model (HMM)," *OSR Journal of Computer Engineering (IOSR-JCE)*, vol.10, no.3, 2013.

[4] Madni H. A., Javed M. and M. J. Arshad, "An Overview of Intrusion Detection System (IDS) along with its Commonly Used Techniques and Classifications," *International Journal of Computer Science and Telecommunications*, vol.5, no.2, 2014.

[5] Arora S. and Bawa R. K., "A Review on Intrusion Detection System to Protect Cloud Data," *International Journal of Innovations & Advancement in Computer Science*, vol.3, no.5, 2014.

[6] Abhaya, Kumar K., Jha3 R. and Afroz S., "Data Mining Techniques for Intrusion Detection: A Review," *International Journal of Advanced Research in Computer and Communication Engineering*, vol.3, no.6, 2014.

[7] Kanagalakshmi. R and Raj N. V., "Network Intrusion Detection Using Hidden Naive Bayes Multiclass Classifier Model," *International Journal of Science, Technology & Management*, vol.3, no.12, 2014.

[8] Koc L. and Carswell A. D., "Network Intrusion Detection Using a HNB Binary Classifier," in *UKSIM-AMSS International Conference on Modelling and Simulation*, 2015.

[9] Ghosh P., Debnath C., Metia D. and Dr. Dutta R., "An Efficient Hybrid Multilevel Intrusion Detection System in Cloud Environment," *IOSR Journal of Computer Engineering*, vol.6, no.4, 2014.

[10] Ibrahim H. E., Badr S. M. and Shaheen M. A., "Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems," *International Journal of Computer Applications*, vol.56, no.7, 2012.

[11] Mukherjeea S., Sharmaa N., "Intrusion Detection using Naive Bayes Classifier with Feature Reduction," *Elsevier Ltd.*, 2012.

[12] Koc L., Thomas A. Mazzuchi and Sarkani S., "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier," *Elsevier Ltd.*, 2012.

[13] Padmakumari P., Surendra K., Sowmya M. and Sravya M., "Effective Intrusion Detection System for Cloud Architecture," *ARPN Journal of Engineering and Applied Sciences*, vol.9, no.11, 2014.

[14] Ibrahim L. M., Basheer D. T. and Mahmod M. S., "A Comparison Study For Intrusion Database (KDD99, NSL-KDD) Based On Self Organization Map (SOM) Artificial Neural Network," *Journal of Engineering Science and Technology*, vol.8, no.1, 2013.