ISSN 2278-3091
Volume 8, No.5, September - October 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse43852019.pdf
https://doi.org/10.30534/ijatcse/2019/43852019

An Implementation of Novel Feature Subset Selection Algorithm for IDS in Mobile Networks

N Chandra Sekhar Reddy[1], Dr. Purna Chandra Rao Vemuri[2], Dr. A Govardhan[3]
[1,2]Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India. naguchinni@gmail.com
[3]Rector and Professor, Department of Computer Science and Engineering, JNTUCEH Hyderabad, India.
[3]govardhan_cse@jntuh.ac.in

ABSTRACT

Intrusion identification and prevention in networks and network related sources has been considered as an all time challenge for researchers and network scientists. Research towards this is conducted using many data mining algorithms and ML techniques. Wide usage of Internet through hand held devices like mobiles and laptops which tracks users every step like location, activities, transfers etc., may be misused through various intrusions. The traditional security measures like firewall, authentication and cryptography used as first line of defence are not sufficient provided the full landscape of security challenges in the heterogeneous and distributed networks of the real world. Therefore, there is need for Intrusion Detection Systems (IDS) that are to be integrated with anti-virus and other security mechanisms to ensure end-to-end security. The existing IDSs suffer from the problem of massive amounts of network traffic that lead to unsatisfactory results or inaccurate intrusion results or computational difficulties that form barrier in providing timely protection to a communication network. It is important to have intelligent security systems that can segregate relevant features from irrelevant ones, discard irrelevant features, and remove redundant features besides discovering Indicative features to enhance the scalability and efficiency of IDS. This is the challenging problem that is addressed in this work which is aimed at proposing and implementing a comprehensive intrusion detection framework which not only takes care of detecting intrusions but also reduces feature space more elegantly and intelligently.

Key words: Feature Selection, Feature Extraction IDS, Mobile Intrusion Dataset, Indicative Features, SVM Classifier, Subsets.

1. INTRODUCTION

Every Intrusion technique is dependent on the variety of observing information that is given as contribution to the recognition algorithm, with respect to the sort of the genuine recognition strategy utilized. Intrusion detection techniques are classified into 2 types which are Host_based and Network_based IDS. Network_based intrusions are used to capture the information on network like data packets, system traffic and network traffic. In network traffic, there is a possibility of cyber-attacks which involves to theft sensitive data such as user credentials, pass keys, credit-card details, banking details etc. and cyber-attacks have been viewed in different forms such as port scan, wireless packets injection, code injection, session hijacking, exploiting on vulnerability, unauthorized access and some of the cyber-attacks leads to damage the resources in network such includes denial of service, ping of death. In LAN Network and Wireless Sensor Network (WSN), resources are vulnerable to cyber-attacks because constrained assets of the data centres and their open and distributed nature in network. Moreover, network traffic, data packets spreading can be done regularly, nodes in network can be organized arbitrarily so an attacker can be effortlessly injected or damage the packets and system in network environment. A cyber attacker can compromise systems, mobile devices, network resources, spy packets, infuse fake information, modify the data integration. Denial of service, ping of death, probe attacks are viewed as most unsafe and risky assaults that impends network security[12].

Since the process of avoidance or mitigating security attacks cannot be constantly effective, an Intrusion Detection System is expected to distinguish known and obscure attacks and notify mobile network hubs about them. IDS permit distinguishing suspicious or strange exercises and trigger a caution when an interruption happens. The execution of IDSs for WSNs are more troublesome than different frameworks since network hubs are normally intended to be minor and inexpensive, and they don't have enough equipment assets. Furthermore, there is no particular dataset that contains ordinary profiles and assaults in network analyzers that can be utilized to identify an invader signature. Thinking about the above difficulties, there are two conditions while outlining IDS for WSNs: The IDS must be of high level of exactness in distinguishing an invader that incorporates obscure assaults, and it should be lightweight to guarantee least overhead on the framework of WSNs. We have applied the existing combinations on a new intrusion dataset and

developed an approach for feature selection which give better results for new mobile intrusions.

1.1 Dataset Description

The dataset below represents the features of Network Intrusion dataset gathered from an organization based on many user's handling of various applications. This dataset can be used for classifying intrusions similar to benchmark datasets like KDD [7], NSL-KDD. The features mentioned in below Table:1 describes the various parameters taken into consideration based on the path to various intrusions [9]. The MNID dataset(table 1) is categorized into 4 ways of intrusion attacks. Location 4, Network 14, Personal 9, System 25.

1.2 Pre-processing of Dataset

Pre-processing is performed by normalization of the discrete attributes into continuous ones by Min-Max technique on both datasets. Pre-processing is performed on the dataset and then the data is classified as train set and test set.

Table 1: MNID Dataset Features

Feature identity	Feature_name	Type	Feature Description
A	AccessCoarseLocation	Location	This permission permits a mobile application to use estimated location.
B	AccessFineLocation	Location	This permission permits a mobile application to use accurate location.
C	AccessLocationExtraCommands	Location	This permission permits a mobile application to access extra location provider commands.
D	ControlLocationUpdates	Location	This permission permits a mobile application to enable or disable location update notifications.
E	InstallLocationProvider	Location	This permission permits a mobile application to mount a location provider to Manager.
F	AccessNetworkState	Network	This permission permits a mobile application to use info about n/w and its underlying areas.
G	AccessWifiState	Network	This permission permits a mobile application to now info about WIFI n/w's
H	BindPrintService	Network	This permission is required by a printer service such that only a network system can be linked to it.
I	ReadExternalStorage	System	This permission permits a mobile application to read data from external source.
J	ReadInputState	System	This permission permits a mobile application to get the current status of switches and network keys
K	BroadcastSMS	Network	This permission permits a mobile application to receive notification about broadcast SMS.
L	BroadcastWapPush	Network	This permission permits a mobile application to broad cast a WAP Push notification.
M	ChangeConfiguration	Network	This permission permits a mobile application to change the current configurations of network.
N	ChangeNetworkState	Network	This permission permits a mobile application to modify network connection status.
O	ChangeWifiMulticastState	Network	This permission permits a mobile applications to into a wireless multicast mode.
P	ChangeWifiState	Network	This permission permits a mobile application to modify connectivity of Wifi state.
Q	SetTime	Location	This permission permits a mobile application to adjust the time of system.
R	RecieveMMS	System	This permission permits a mobile application to check regularly about inward MMS messages.
S	RecieveSMS	System	This permission permits a mobile application to collect SMS information.

T	RecordAudio	System	This permission permits a mobile application to record audio
U	MountFormatFilesystems	System	This permission permits a mobile application to format file system for removable storage.
V	ModifyPhoneState	System	This permission permits a mobile application to modify telephone state like power on, off etc., It doesn't include placing calls.
W	PersistentActivity	Location	This permission permits a mobile application to perform activities.
X	Internet	Network	This permission permits a mobile application to unlock network sockets.
Y	BindVpnService	Network	It is required for use of VPN to ensure only system can link to it.
Z	SetTimeZone	Location	This permission permits a mobile application to locate system time zone.
AA	BindRemoteViews	Network	This permission is required by remote view service to make sure only system will be linked to it.
AB	ReadCallLog	Personal	This permission permits a mobile application to read caller log data of user.
AC	ReadContacts	Personal	This permission permits a mobile application to access users contact data.
AD	ReadLogs	Personal	This permission permits a mobile application to access low level log files in system.
AE	ReadPhoneState	Personal	This permission permits a mobile application to use mobile state, cellular n/w info, ongoing calls status, and contacts list details which are in device.
AF	ReadSms	Personal	This permission permits a mobile application to read SMS messages in mobile.
AG	MediaContentControl	System	This permission permits a mobile application to know about current playing and control playback.
AH	ModifyAudioSettings	System	This permission permits a mobile application to change audio settings.
AI	MountUnmountFilesystems	System	This permission permits a mobile application to mount and unmount File system for removable storage.
AJ	WriteSettings	System	This permission permits a mobile application to read and write system settings.
AK	WriteSecureSettings	System	This permission permits a mobile application to use read or write system settings.
AL	WriteSyncSettings	System	This permission permits a mobile application to write the synchronization settings.
AM	WriteCallLog	System	This permission permits a mobile application to write and doesn't allow to read call log data of user.
AN	WriteContacts	System	This permission permits a mobile application to write contacts data of user.
AO	UseFingerprint	System	This permission permits a mobile application to access finger print option of mobile phone.
AP	GetAccounts	Personal	This permission permits a mobile application to provide list of accounts from accounts service.
AQ	BindTelecomConnectionService	Network	This permission is required by telecom connection service to make sure only system can link it.
AR	SendSms	System	This permission permits a mobile application to send SMS from mobile to mobile.
AS	SendRespondViaMessage	System	This permission permits a mobile application to send a request to other applications. It gives response via a message during incoming calls like I'm in meeting, etc.,

AT	Bluetooth	System	This permission permits a mobile application to connect or disconnect and pair bluetooth devices.
AU	BluetoothAdmin	System	This permission permits a mobile application to search and connect to new blueetooth devices.
AV	CallPhone	System	This permission permits a mobile application to instigate a mobile phone call without using dialer interface to place a call.
AW	Camera	System	This permission permits a mobile application to use device camera facilities.
AX	CaptureAudioOutput	System	This permission permits a mobile application to capture output of audio.
AY	CaptureVideoOutput	System	This permission permits a mobile application to capture output of video.
AZ	GetAccountsPrivileged	Personal	This permission permits an access to acconutd in the accounts service.

2. RELATED WORKS

With the advancements in technology and vulnerabilities in software systems, today's smart devices are prone for different unknown attacks. Even though firewall is used in many organizations as first line of defence which can scrutinize the inflow and outflow it cannot generate an alert whenever it detects an attack. To categorize any attack accurately and fight against them we are in need of an effective intrusion detection and prevention mechanisms which are very crucial in this internet world. To defend the network traffic across many known and unknown attacks an exemplary solution is a good Intrusion detection which has an ability to even identify unknown attacks based on previous features.

The main reason we develop any Intrusion detection mechanism is to classify the data from network traffic as genuine or intruded. Any network traffic is concluded as intruder when the behaviour slews away from its standard pattern with abnormalities. The efficiency and effectiveness of network Intrusion detection is reliant upon the mechanism or algorithm which we use for classifying any network data. The parameters like time consumption by any algorithm and accuracy plays a major criteria for the algorithm selection. As a common practice Machine learning methods use a sample dataset for the development of a model to decide the effectiveness by predicting and decision making of the developed model[11].

In this paper further it is described sections wise as in section 2 implementation of existing approach of PCA and SVM, In section 3 we have clearly explained our new approach of IFSS for making subsets. In section 4 the observations and results are clearly explained and in section 5 we have explained the conclusions and further enhancements.[1]

Intrusion Detection Systems (IDS) are developed to identify unauthorized effort to access or manipulate the computer systems. IDS collects network data to identify different kinds of malware and attacks against services and applications. [2]

Data mining techniques have been used recently in the development of intrusion detection models to minimize information overloading. These models extract the meaningful information in search of patterns and relationships from the data collected, thereby improving decision making. Data mining methods such as k-nn, Logistic regression, Neural networks, naive bayes, and support vector machine are used for classification and pattern recognition as they have improved the performance of the models that deploy such algorithms. In classification, the features of newly present objects are examined and are assigned to one of the existing set of classes. Classifier models gain knowledge from the training data and identify the class label for the new instances. Many supervised learning models are used to solve classification problems.

2.1 SVM Classifier

Support Vector Machine supervised classifier is among the efficient techniques used as the generalization capability is higher even when the sample training data is small. In our work an existing data mining classifier SVM is used for forming a detection model for any intrusion detection. SVM is considered as the most preferable supervised machine learning technique for classifying intrusions in network traffic data. The datasets which are used to measure the detection rate of Support Vector Machine algorithm is the KDDCup‘99 dataset and NSL KDD [8]. In the recent years, many hybrid intelligent systems have been proposed to improve the accuracy in comparison to individual techniques [5]. SVM is an approach for classifying datasets based on Statistical Learning. SVM works on the thought of hyperplane classification also called as linear seperability [6].

2.2 Principal Component Analysis

IDS mechanism is used to detect un authorised behaviour in network and its systems. Many data mining classification methods have been implemented in collaboration with feature extraction and reduction techniques to get better results in short time with reduced features [3].PCA is most widely statistical approach in data mining for dimensionality

reduction and to identify data points with highest possible variance. Principal Component Analysis is widely used to determine the patterns in data with high dimension [4].

2.3 PCA Approach:

The steps to PCA approach are summarized as below:

Step I: Find the covariance matrix of the normalized d- dimensional dataset.

Step II: From the results obtained from Step I find Eigen vectors and Eigen values.

Step III: Arrange the results of Eigen values in descending order.

Step IV: From the 'n' largest Eigen values select the 'n' Eigen vectors. (n indicates number of dimensions)

Step V: Build a matrix from selected Eigen vectors

Step VI: Convert the Original dataset to form a new n-dimensional feature space [10].

3. RESEARCH METHOD

This section explains about the architecture of the algorithm developed to identify intrusions in a new dataset whose features are explained in section II. Government organizations and many private associations are working and putting lots of efforts to address the problem of Intrusion detection, but still there is a need for efficient algorithm which can protect the internet applications which are distributed.

As showin in Figure 1 in our architecture a new dataset with 52 features has been selected which is an intrusion dataset as described earlier in this paper. The flow represents the feature selection process and how dataset is partitioned into training and test sets, also indicating the evaluation process to find out the parameters required to calculate accuracy.

3.1 IFSS-SVM framework:

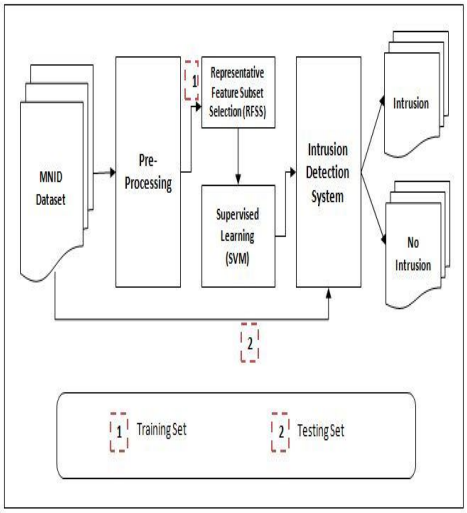

Figure 1: Indicative feature subset selection phase of IFSS-SVM framework

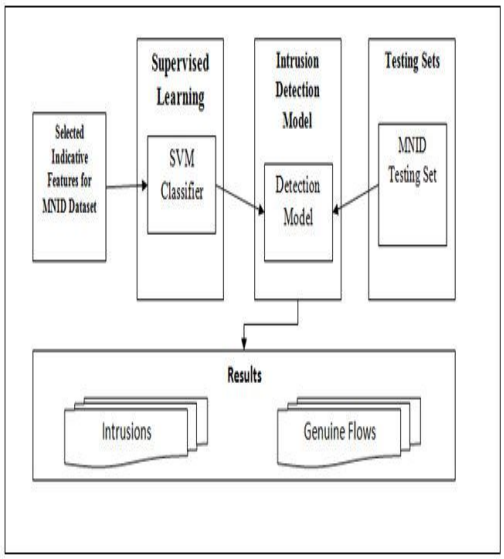

Figure 2: Testing Phase of Intrusion detection framework

In Figure 2 the Indicative features of MNID Dataset are sent for classification using SVM and using the intrusion detection model the training and testing datasets are classified using Indicative feature subsets and SVM Classification. After the successful completion of the classification through the process, the results will be generated as represented in Figure 2. The two types of classifications among the dataset are genuine flows and intrusions.

Algorithm: Indicative Feature Subset Algorithm

Inputs: Dataset ***D,*** attack categories C

Outputs: Chosen Feature Subset ***FS***

01 Initialize feature vector ***F***

02 Initialize threshold vector TR

03 Initialize Indicative feature vector InF

04 Extract features into F from D

05 For each c in C

06 Compute P(X)

07 Compute entropy e

08 Compute gain g

09 Computer symmetric uncertainty su

10 End For

11 For each c in C

12 Compute threshold tr using su of c and su of overall features

13 Add tr to TR

14 End For

15 For each c in C

16 For each feature f in c
17 Computer entropy e
18 Compute gain g
19 Compute e symmetric uncertainty su
20 If su satisfies tr of c in C Then
21 Add f to InF
22 End If
23 End For
24 End For
25 Return InF

4. RESULTS AND ANALYSIS

The methodology explained in above section is implemented step by step and the practical implementation for the same approach is executed for analyzing the results generated. Initially the dataset is analyzed using machine learning classification mechanisms and the results were not satisfactory. Hence the Novel IFSS mechanism is applied on the same dataset combined with SVM Classifier. The results obtained are finally compared. The below table 2 represents the obtained Indicative features and its subsets

Table 2: Obtained feature subsets from IFSS

S.No	Indicative Feature	Subset Features	Type	Description
1	AB	{AC,AD}	Personal	AB is indicative feature for AC,AD as the threshold value of AC and AD are in the range of AB they are classified as subsets.
2	AE	{AF}	Personal	AE is indicative feature for AF as threshold value of AF is in the range of AE.
3	AP	{AZ}	Personal	AP is indicative feature for AZ as threshold value of AZ is in the range of AP.
4	A	{B,C,D,E}	Location	The features B,C,D,E are indicated by feature A as the ranges are near and falls into this subsets.
5	Q	{W,Z}	Location	W,Z Features are subsets of Indicative feature Q.
6	F	{G}	Network	F is indicating G as superset of G.
7	H	H	Network	Here H is the only unique feature which is not indicating any other features.
8	K	{L,M}	Network	L, M fall into subset of K as the threshold ranges are near to K.
9	N	{O,P}	Network	O,P aresubsets of Indicative feature N. N is considered as Indicative feature in first come first serve basis. As N is arrived first it is considered as Indiative feature.
10	X	{Y, AA, AQ}	Network	Features Y,AA,AQ are subsets of Indicative feature X.
11	I	{J, R, S, T}	System	Features J,R,S,T are subsets of Indicative feature I.
12	U	{V, AG, AH, AI}	System	U is the Indicative feature for the subset features V,AG,AH,AI.
13	AJ	{AK, AL, AM, AN}	System	AJ is the indicative feature representating AK,AL,AM,AN.
14	AO	{AR, AS, AT, AU, AV, AW, AX, AY}	System	AO is an indicative feature for group of features AR,AS,AT,AU,AV,AW, AX,AY. As the threshold range values of subsets are in the range of AO they are grouped as one subset

4.1 Performance Evaluation

True Positive (TP) is the number of actual attacks classified as attacks,

True Negative (TN) is the number of actual normal records classified as normal ones,

False Positive (FP) is the number of actual normal records classified as attacks, and

False Negative (FN) is the number of actual attacks classified as normal records[13].

The above metrics are used in construction of a confusion matrix table 3.

Table 3: Confusion Matrix used for evaluation

	Ground Truth (correct prediction)	Ground Truth (incorrect prediction)
Result of IFSS-SVM (correct prediction)	True Positive (TP)	False Positive (FP)
Result of IFSS-SVM (incorrect prediction)	False Negative (FN)	True Negative (TN)

$$\text{Precision} = \left(\frac{TP}{TP+FP}\right)*100 \quad \text{--- Eq (1)}$$

$$\text{Recall} = \left(\frac{TP}{TP+FN}\right)*100 \quad \text{--- Eq (2)}$$

$$\text{F-Measure} = 2 * \left(\frac{precision*recall}{precision+recall}\right)*100 \quad \text{--- Eq (3)}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{--- Eq (4)}$$

Precision value Eq:1 and Recall Eq:2 value directly affects the performance of the system. F Mearsure Eq:3 is calculated using values obtained from precision and recall.

As shown in fig:3 the results of the existing feature extraction mechanism and the developed mechanism are compared and using the evaluation parameters Accuracy Eq:4 has been calculated. From the Table 4 and Fig:3 it is very clear that the performance of IFSS-SVM is better when compared with state of art, which makes our approach a better feature selection mechanism for new types of attacks than existing approach in terms of accuracy. The table 4 represents the accuracy values obtained for the two approaches.

Table:4 Evaluation metrics

Algorithm	**Accuracy**	**Precision**	**Recall**	**F-Measure**
J48	97.65	93.65	92.65	96.90
Random Forest	98.11	96.11	95.11	95.60
SVM	98.6	96.6	92.6	96.37
IFSS & J48	99.12	95.12	94.12	98.37
IFSS & Random Forest	99.4	97.4	96.4	96.89
IFSS & SVM	99.65	97.65	93.65	97.42

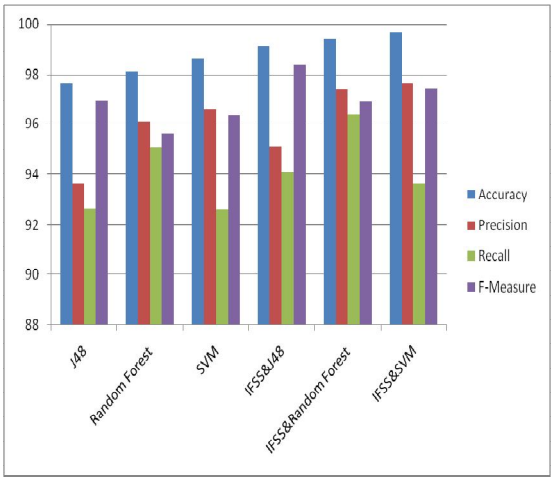

Figure 3: Performance of IDS

Figure 3 shows clearly the performance of IFSS and SVM gives better resuts when compared with other state of art.

5.CONCLUSION

The work presented in this paper is mainly focused on Mobile intrusions which are very dangerous to users confidential data. The method IFSS explained through this research is a novel approach developed with an intetion to detect new type of intrusion attacks affecting users mobile information. There are many approaches concentrating on various intrusions but lacking performance for new type of Intrusion attacks. Hence we herewith introduced a new IFSS-SVM Framework with a better results compared to State of Art.

As internet usage is streaming very fast the data transfer across network and the types attacks need have increased which are to be classified. For doing this task we have taken a real time data mobile intrusion dataset from an organization which works on application oriented network data. We have successfully applied the existing approaches and checked the performance evaluation of the new dataset. This approach gave good accuracy and overall system performance was good. We have also developed a new feature subset selection mechanism which can dynamically reduce the redundant features and check the accuracy of classifying the data. There is a much scope for new approaches which can be combined with various feature extraction and classification algorithms. The current methodology can also be tested on benchmark datasets which is our

future work. The accuracy obtained on existing benchmark datasets will be used as a evaluation parameter for the developed algorithm.

ACKNOWLEDGEMENT

I sincerely wish to take this oppurtunity to thank my guides (co-authors) for their continuous support in carrying out my research work. I thank my college management for providing resources in completing the work. I wish to thank cartel organization for providing their support in data collection.

REFERENCES

1. Praneeth NSKH, Naveen Varma M, Roshan Ramakrishna Naik, "Principle Component Analysis based Intrusion Detection System Using Support Vector Machine," IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India.
2. SumaiyaThaseen Ikram and Aswani Kumar Cherukuri, "Improving Accuracy of Intrusion Detection Model Using PCA and Optimized SVM," *Journal of Computing and Information Technology*, Vol. 24, No. 2, June 2016, 133–148 133, doi: 10.20532/cit.2016.100270.
3. NoreenKausar, BrahimBelhaouari Samir, SuziahBtSulaiman, Iftikhar Ahmad, Muhammad Hussain, *An Approach towards Intrusion Detection using PCA Feature Subsets and SVM*, 2012 International Conference on Computer & Information Science (ICCIS) IEEE.
4. N. Chandra Sekhar Reddy, Purna Chandra Rao Vemuri, A. Govardhan, "Evaluation of PCA and K-means Algorithm for Efficient Intrusion Detection," *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 12, Number 12 (2017) pp. 3370-3376
5. SumaiyaThaseen, Ch.Aswani Kumar, "Intrusion Detection Model Using fusion of PCA andoptimized SVM," *International Conference on Contemporary Computing and Informatics (IC3I) 2014 IEEE*.
6. Heba F. Eid, Ashraf Darwish, Aboul Ella Hassanien, and Ajith Abraham, "Principle Components Analysis and Support Vector Machine based Intrusion Detection System," *10th International Conference on Intelligent Systems Design and Applications*, 2010 IEEE.
7. PreechaSomwang and Woraphon Lilakiatsakun, "Computer Network Security Based On Support Vector Machine Approach," 11th International Conference on Control, Automation and Systems, Oct. 26-29, 2011 in KINTEX, Gyeonggi-do, Korea.
8. N Chandra Sekhar Reddy, Dr. Purna Chandra Rao Vemuri, Dr. A Govardhan,Ch. Vijay, "An Empirical Study On Feature Extraction Techniques For Intrusion Detection System," *Journal of Advanced Research in Dynamical and Control Systems* Vol. 9. Sp– 12 / 2017.
9. N. Chandra Sekhar Reddy, B.Rama, B. Madhuravani, Raman Dugyala, N. Rajasekhar, "Mobile Network Services From The Presence Cloud," *International Journal of Pure and Applied Mathematics*, Volume 119 No. 14 2018, 95-100.
10. Aburomman, A. A., &Reaz, M. B. I. (2017), "Ensemble of binary SVM classifiers based on PCA and LDA feature extraction for intrusion detection," In *Proceedings of 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2016* (pp. 636-640). [7867287] Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/IMCEC.2016. 7867287
11. N. Chandra Sekhar Reddy, Dr. Purna Chandra Rao, Dr. G Govardhan, " An Intrusion Detection System for Secure Distributed Local Action Detection and Retransmission of Packets " International Journal of Soft Computing 12(1): 45-49, 2017.
12. Krishan Kumar, Gulshan Kumar, Jabarweer Singh, An Effective Combination Technique for Artificial intelligence based Ensembles for Intrusion detection, International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE), Vol.2 , No.5, Pages :29-36 (2013) Special Issue of ICCECT 2013.
13. Mohammed A. Ambusaidi, Xiangjian He*, Priyadarsi Nanda, and Zhiyuan Tan, Building an intrusion detection system using a filter-based feature selection algorithm, IEEE Transactions On Computers, Vol., No November 2014