Towards the Low False Alarms and High Detection Rate in Intrusions Detection System

Hafiz Muhammad Imran, Azween Bin Abdullah, and Sellappan Palaniappan

***Abstract*—Due to the fast growth of network systems, abundant intrusive approaches have been grown extensively which are escalating many security and solidity threats. Intrusions Detection Systems (IDS) are security programs to decide whether events and activities occurring in the network are intrusive or legitimate. The purpose of IDS is to identify intrusions in network traffic with low false alarms and high detection rate while consuming lesser resources and computational cost. There are plentiful issues in traditional IDS including regular updating, low detection capability to unknown attacks, high false alarms rate, extraordinary resources consumption and many others. Similarly, Intelligent Network IDS have snags of performance efficiency, false positive and false negative while today's advance Neural Network approaches are also facing training/learning overhead, high false alarms and low detection rate. Soft computing is an innovative field to develop intelligent IDS while minimizing the deficiencies in other approaches. The objective of this research is to propose an efficient soft computing approach with low false alarms and high detection rate while maintaining low cost and less time. Our research promising results show that a new proposed system is an improved and applicable representation of an ideal intrusion detection system.**

***Index Terms*—IDS, features selection, NSL-KDD, LDA, GA, SVM Kernels**

I. Introduction

Intrusion Detection Systems are software or hardware security programs that decide whether events and activities occurring in a system or network are intrusive or legitimate based on integrity, confidentiality and the availability of information resources [1]. There are abundant issues in traditional IDS including regular updating, low detection capability to unknown intrusions, non-adapting high false alarm rate, high resources occupation, poor connection and time consuming analysis of attack data requiring excessive human participation, as well as weak defense capability against the common scripting attacks [2]. Similarly, Intelligent Network IDS like Rule based, Hybrid and Graphical approaches have problems of performance efficiency, false positive and false negative while today's advance Neural Network approaches are also facing training/learning overheads, false alarms and low detection rate [3].

Soft computing is an innovative approach to construct a computationally intelligent Intrusion detection system which analogous to the extraordinary ability of the human mind to reason and learn in an environment of uncertainty and imprecision [4].

Selection of suitable dataset is a backbone of efficient intrusion detection approaches. Performance of any IDS depends on the efficiency and accuracy of the dataset. If the training dataset is precise with optimal contents and rich features, efficiency of the trained as well as test system will be improved. Therefore, it is crucial to select an optimal dataset for testing and training the system. There are many standard pre-built simulated datasets like Darpa's KDD Cup 98, 99, Six UCI db and NSL-KDD etc. KDD-Cup 99 is most widely used as a benchmark dataset for training and testing of Intrusion detection systems. KDD-CUP 99 is built based on the data captured in DARPA'98 which has been criticized by McHugh [5], mainly because of the characteristics of the synthetic data. One of the most important deficiencies in the KDD data set is the huge number of redundant records. Analyzing KDD train and test sets, it is found that about 78% and 75% of the records are duplicated in the train and test set, respectively, which causes the learning algorithms to be biased towards the frequent records, and thus prevent them from learning infrequent records which are usually more harmful to networks such as U2R and R2L attacks.

Due to the deficiencies of KDD-Cup which highly affects the performance of evaluated systems and results in a very poor evaluation of intrusion detection approaches, an advanced form of KDD-Cup was proposed namely NSL-KDD which consists of selected records of the complete KDD data set. Important drawbacks of KDD-Cup are fixed in NSL-KDD dataset.

The rest of the paper is organized as follows. In section II, related work to IDS is discussed briefly. In section III, proposed model is discussed and analyzed in depth. Different phases of IDS are described in detail. Conclusion and future work is briefly mentioned in section IV.

II. Related Work

An ensemble approach for features reduction is adopted by [6]. As PCA is not suitable for nonlinear dataset as well as for large dataset, in their work, authors preferred Generalized Discriminant Analysis (GDA) over PCA for features selection. Besides reduction in number of input features, GDA reduces classifiers training time by selection of most discriminant features. It also increases classification accuracy. Anomaly detection approach is used to differentiate between normal data based on normal behavior and attack or intrusive data based on its attack behavior. Self-Organizing Map (SOM) approach and C4.5 decision

Manuscript received April 15, 2013; revised June 28, 2013.

Hafiz Muhammad Imran and Sellappan Palaniappan are with Malaysia University of Science and Technology (MUST), Malaysia (e-mail: engimran_uet@hotmail.com).

Azween Bin Abdullah is with Taylors University, Malaysia.

tree techniques are applied for classification of reduced feature space. Dataset KDDCup-99 is applied in this research and 41 features are reduced to 12 features space by GDA. Experimental results show that GDA outperforms PCA especially for large scale dataset by providing better detection rate, reduced training and testing time. Moreover, C4.5 classifier outperforms SOM for all attack classes.

An efficient intrusion detection system is proposed by Ahmed and his colleagues [7] using features subset selection based on MLP. They used PCA (Principal Component Analysis) and GA for preprocessing and MLP for features classification using KDD-cup dataset. LDA outperforms PCA and PCA is not suitable for large dataset [4], hence their work is limited for small size datasets and results are not more realistic to actual network traffic as there are approved deficiencies in KDD-Cup dataset.

PCA was used for features reduction and Naive Bayes algorithm for classification [8] to generate less false positive alarms ratio and to increase the detection rate efficiently. A total of 41 features of KDD 99 dataset were reduced to 14 features set and 12 major features out of 14 having greater Eigen values were identified by PCA. Brief comparison of different approaches is shown in Table I.

TABLE I: COMPARATIVE ANALYSIS OF EXISTING APPROACES FOR IDS

Author [Year]	Dataset	Architecture	Accuracy Rate
Yu *et al*. [2008]	SNMP MIB TCP,UDP	SVM	97 %
Osareh, and Bita [2008]	KDD	SVM	83 %
Alice Este *et al*. [2009]	CAIDA	SVM	90 %
A. Chandrasekar and V. Vasudevan [2009]	KDD-Cup	PSO-SVM	95%
Lakhina *et al*. [2010]	KDD cup	PCANNA	80.4 %
Ahmad *et al*. [2011]	KDD-Cup	PCA, GA & MLP	99%
Shailendra Singh and Sanjay Silakari [2011]	KDD-Cup	GDA, SOM, C4.5	98%

III. PROPOSED MODEL FOR IDS

There are different interdependent phases in the proposed architecture for the efficient IDS. NSL-KDD is selected during selection of suitable dataset phase. LDA approach is used for features transformation and GA for features optimum subset selection. In the third phase, SVM Kernels are used as classification approach in this research. After Classification, the system is trained and tested according to the standard rules. Fig. 1 is a block diagram for the proposed system.

A. Selection of Suitable Dataset

KDD-Cup is the widely used dataset for training and testing of IDS. There are a total of 41 features which are classified into Basic, Content and Traffic features. KDD-Cup is developed on the basis of DARPA'98 data and this data has been criticized by McHugh [5]. As a result, some of inherited issues also exist in KDD-Cup like redundancy of similar records and complexity level of data behavior. NSL-KDD is an advanced version of KDD-Cup dataset and doesn't suffer from the shortcomings in KDD-Cup. The following are unique features for which we preferred NSL-KDD over KDD-Cup.

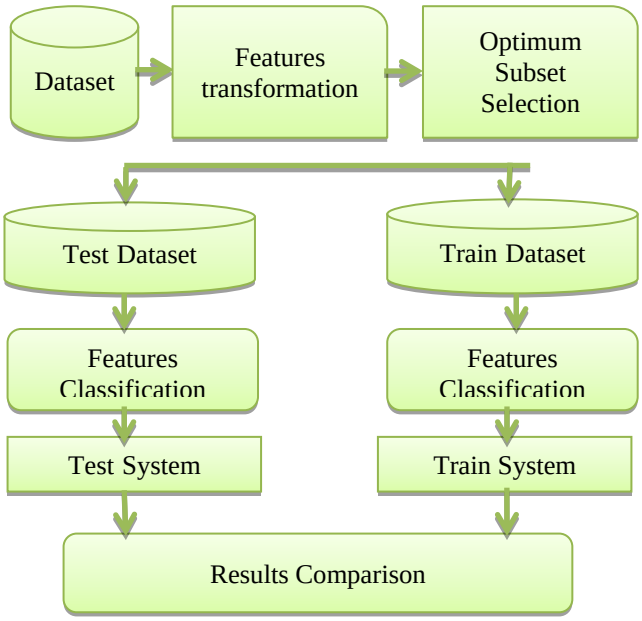

Fig. 1. Block diagram of proposed system for IDS.

1) *No Redundancy of Records:* NSL-KDD doesn't include redundant records in the train set; hence the classifiers would not be biased towards more frequent records.
2) *No Duplication:* No duplicate record in the proposed test sets; therefore, the performance of the learners is not biased.
3) *Less Complexity Level:* The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary widely, which makes it more efficient to have an accurate evaluation of different learning techniques.
4) *Reasonable Records:* The number of records in the train and test sets is reasonable, *which* makes it affordable to run the experiments on the complete set without the need to randomly select a small portion.

NSL-KDD features can be classified into three groups as shown in following Fig. 2.

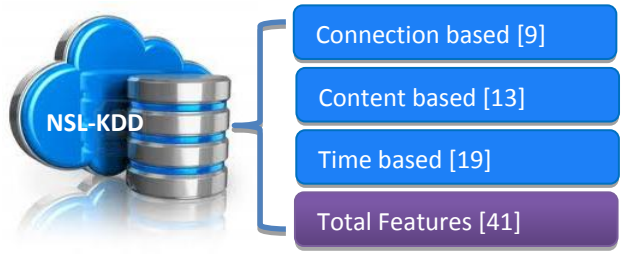

Fig. 2. Feature categories in NSL-KDD dataset.

B. Preprocessing of Raw Dataset

In most of existing intrusion detection approaches, raw features set are given as input directly to classifiers which causes many problems. In some cases, features are transformed and subset of features is given as input to classifier. In this case, there are also some issues regarding subset selection scenario. Some major issues in both above

mentioned approaches involve high false alarms, low detection rate and accuracy, loosing important information and many others. Detailed diagram with related issues is shown in Fig. 3.

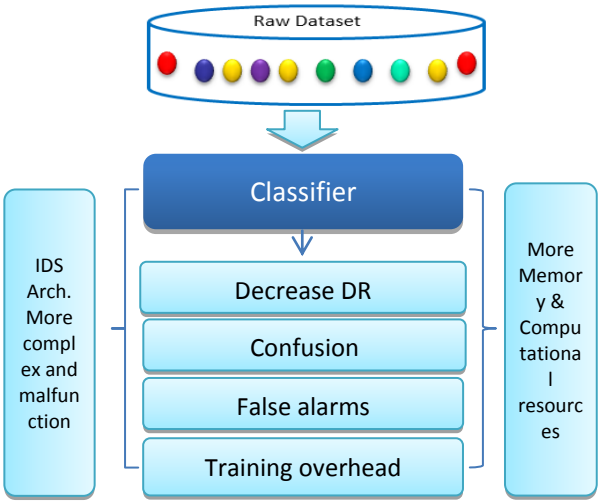

Fig. 3. Issues in existing approaches.

Instead of direct input of raw dataset to selected classifiers, raw dataset is preprocessed in different ways to overcome different issues like training overhead, classifier confusion, false alarms and detection rate ratios. Preprocessing phase is divided into three sub phases as shown in Fig. 4.

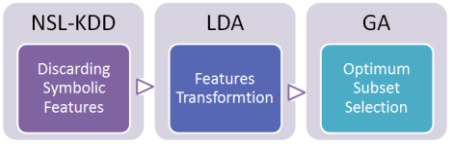

Fig. 4. Flow chart of Pre-processing steps.

C. Discarding Symbolic Feature Vectors

There are three kinds of symbolic features (tcp, ftp_data and SF etc.) in feature space of 41 features. As symbolic values are not of interest to our research, these three feature vectors are discarded to get the following new feature space.

$$F(Xm) = X1, X2 \qquad Xm \text{ where } m = 38 \qquad (1)$$

D. Features Transformation and Organization

In most of existing intrusion detection approaches, raw features set are given as input directly to classifiers which cause many issues out of which some are as following.

1) Using raw data set directly for classifiers guzzles more memory space as well as computational resources during training and testing phases of the system.
2) Detection rate decreases in this case.
3) Classifier may become confused and generate false alarms.
4) Training overhead is increased due to the processing over each input feature even it is not important for the classifier.
5) The architecture of IDS becomes more complex.

In order to avoid the above mentioned issues, LDA approach is adopted to transform original numeric features space into new linear features space. LDA is a high-dimensional data analysis method and suitable for features transformation to facilitate classification [9]. Its steps are shown in Fig. 5. There has been a tendency to use PCA approach for features subset selection or reduction in many different domains like face recognition, image compression as well as intrusion detection [10] but LDA has more benefits over PCA and is preferred over PCA due to the following reasons.

1) LDA outperforms PCA in case of large dataset [4].
2) LDA directly deals with both discrimination within-classes as well as between-classes while PCA does not have any concept of the between-classes structure [1].
3) LDA preserves class discriminatory information as much as possible while performing dimensionality reduction [11].

Suppose $x = (x_1, x_2, x_3, x_4, \ldots\ldots\ldots\ldots\ldots\ldots x_C)$ are $Nx1$ feature vectors where C=38 and each feature vector contains n feature samples. Following are steps adapted in LDA algorithm.

Step 1. Compute the between class scatters using complete feature samples.

$$S_b = \sum_{i=1}^{C} (\alpha_i^j - \alpha_i)(\alpha_i^j - \alpha_i)^T$$

Step 2. Calculate the Total class scatter matrix.

$$S_t = \sum_{i=1}^{C} \sum_{j=1}^{n} (\alpha_i^j - \bar{\alpha})(\alpha_i^j - \bar{\alpha})^T$$

Step 3. Compute Eigenvalues and Eigenvectors using Eigen equation for LDA. $S_b X = \lambda S_t X$

Step 4. Compute the Eigenvectors corresponding to Eigenvalues such that $Eigenvalues: \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots\ldots \lambda_N$ and Eigenvectors: X_1, X_2, X_3 … X_N where N represents dimensionality of feature vectors and N = 38 in our case.

Step 5. Evaluate the contribution of each feature vector.

$$C_j = \sum_{p=1}^{m} |V_{pj}|$$

Step 6. Sort the feature vectors in descending order corresponding to their impact or contribution.

Fig. 5. Algorithm for Linear Discriminant Analysis.

E. Features Optimum Subset Selection

By the implementation of LDA for features transformation, data set is transformed into a new feature space called as linear features space. This new features space may also be used as input to classifier but classifier becomes biased due to architecture complexity as well as training and testing efficiency decreases which increases memory consumption rate and computational cost. Hence, GA (Genetic Algorithm) is applied to select optimal subset of linear features space.

Step 1. Input: Arranged linear space
Step 2. Generate random population of n chromosomes.
Step 3. Evaluate the fitness f(x) of each chromosome x in population.
Step 4. Form a new offspring with a crossover probability.
Step 5. Mutate offspring at each locus.
Step 6. Accept, Replace and Test
Step 7. [Loop] Go to step 2

Fig. 6. Steps in Genetic Algorithm.

There are three basic genetic operators [12] such as selection, crossover, and mutation which guide the genetic process. The genetic search is an iterative process which

iterates for evaluation, selection and recombination of strings. This iteration continues in the population until the termination condition is reached. Detailed algorithm for GA is shown in Fig. 6.

F. Features Classification

After the selection of the optimum features subset, the classifier is designed to train and test the feature using different SVM (Support Vector Machines) Kernels. The proposed approach is implemented with kernel functions by tuning different parameters including the cost parameter C and other kernel parameters. This is done by parameter selection using 5×2 cross validation. Overview of different SVM kernels is shown in Fig. 7. The system is trained and tested with the given set of parameters to evaluate best possible classifier performance on the selected dataset.

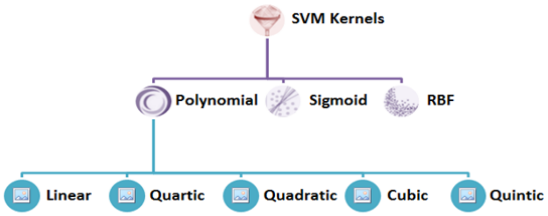

Fig. 7. SVM Kernels.

Fig. 8 shows the different steps taken to classify the network traffic into normal or intrusive using SVM kernels.

Step 1. Input: Features Subset space
Step 2. Classification using best SVM Kernels
Step 3. Selection of best parameters using Loose & Fine Grid Search
Step 4. Performance evaluation using 5x2 CV
Step 5. Train the model
Step 6. Testing and Prediction of the model

Fig. 8. Classification steps.

IV. Experimental Results

Datasets with 11, 15 and 21 feature vectors were selected using GA approach as optimum subset selection from complete dataset with 41 features for training as well as for testing experiments. Different tools including .Net LDA, Neuro Solutions and Matlab were used for this implementation purpose. Table II shows the 11 features selected using GA approach.

TABLE II: Optimum Features Subset Selection

No	Feature Name	Type
1	Duration	Continuous
2	Service	Discrete
3	Count	Continuous
4	dst_bytes	Continuous
5	logged_in	Discrete
6	srv_count	Continuous
7	rv_rerror_rate	Continuous
8	serror_rate	Continuous
9	srv_diff_host_rate	Continuous
10	dst_host_count	Continuous
11	Is_guest_login	Discrete

Network weights are adjusted during training phase. Confusion matrices are used to verify the training process. Weights of the system are frozen after training of the system is completed and system performance is evaluated under testing phase. Testing phase is divided into verification and generalization steps. The objective of verification is to calculate the learning efficiency of trained system while the generalization step is used to measure the generalization ability of the trained system using another dataset besides train dataset. We selected randomly 10,000 feature samples as training dataset from total of 125974 preprocessed linear feature samples while 20% of training data is used as cross validation dataset. Separate dataset of 5,000 is selected randomly from NSL-KDD preprocessed test dataset of 22545 connection records.

We have used several parameters to evaluate the performance of proposed system which include True Positive, True Negative, False Positive, False Negative, Accuracy rate, Detection rate, Sensitivity and Specificity.

Sensitivity: It is the measure of detecting normal patterns accurately.

$$\text{Sensitivity} = (100 \times TP / TP + FN)$$

Specificity: It is the measure of detecting intrusive patterns accurately.

$$\text{Specificity} = (100 \times TN / TN + FP)$$

Three different experiments are conducted using different SVM Kernels. Results in Table III reflect that when optimum subset of features is selected, time consumption rate is relatively reduced and accuracy ratio is increased. Since reduced features space is given as input to classifier, so lesser resources are utilized due to minimum training and learning overheads, hence, computational cost is also minimized. Fig. 9 depicts the performance using different subsets.

TABLE III: Time & Detection Rate Analysis

No.	Features	Not Selected	Time	Detection Rate
1	11	27	45 h	99.3 %
2	15	23	51 h	99 %
3	21	17	55 h	98.7 %

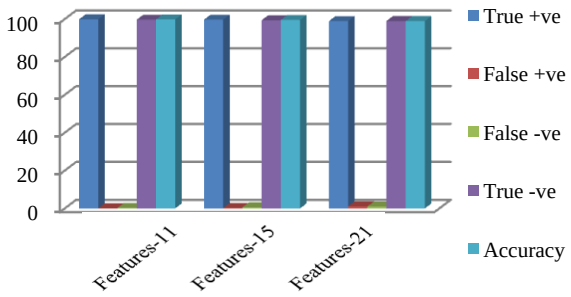

Fig. 9. Performance measurements with different features space.

Sensitivity and specificity results for different SVM kernels and data features combinations are shown in Table IV. Results in table clearly show that RBF kernel performs best for all recipes of features.

Comparison of this research results with some existing approaches is shown in Table V.

TABLE IV: SENSITIVITY & SPECIFICITY ANALYSIS WITH DIFFERENT FEATURES RANGE

Cases	NSL-KDD – 11 Features		15 Features		NSL – 21 Features	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
RBF Kernel	100	99.2	99.1	99.7	98	98.3
Linear Polynomial	99.7	99.3	99	99.3	99.1	98.7
Sigmoid Kernel	98.9	99	99.1	98.6	99.55	98
Quadratic Polynomial	97	97.4	100	95.7	99.1	97.7
Cubic Polynomial	98.45	95.5	90.1	97.1	100	98.1
Quartic Polynomial	99.9	98.75	100	97.5	99.1	98.1
Quintic Polynomial	99	98.5	93.9	98.1	98.34	96.4

TABLE V: RESULTS COMPARISON WITH EXISTING APPROACHES

Author	Year	Approach	Accuracy Rate
Polat and Gunes	2007	LS-SVM	98.53
SterndDobnikar	1996	LDA	96.80
Mehmet Faith Akay	2007	F-Score + SVM	99.51
M. Muthu Rama *et al.*	2008	Nu – SVM	99.38
Abonyi and Szeifert	2003	Supervised Fuzzy Clustering	95.57
Iftikhar Ahmad *et al.*	2011	PCA+GA+MLP	99.0
This Study	2013	LDA+GA+RBF	99.58

V. CONCLUSION AND FUTURE PLAN

Features transformation and selection is generally performed using single approach but in our work, hybrid approach LDA + GA is adopted for features transformation and selection to get better results. LDA is preferred over PCA as it outperforms PCA. Advanced form of KDD-Cup named NSL-KDD is used as standard dataset. Prominent classification approach SVM with different kernels is used to classify network traffic into normal or intrusive. Our work shows that time consumption rate is relatively reduced and accuracy ratio as well as detection rate is increased due to optimum subsets. Since reduced features space is used as classifier input, hence minimum resources are utilized and computational cost is minimized due to minimum training and learning overheads.

Our future plan is to design and develop an efficient intrusion detection system for multi class problems by selecting the optimal subset of features.

REFERENCES

[1] A. Martinez and A. Kak, "PCA versus LDA," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

[2] China Papers Online, study on application of hybrid soft-computing technique to intrusion detection, 2011.

[3] A. N. Toosi and M. Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers," Department of Computer, Ferdowsi, University of Mashhad, Iran, 2007.

[4] K. Delac, M. Grgic, and S. Grgic, *Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set*, University of Zagreb, FER, Unska 3/XII, Croatia, 2006.

[5] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection," *ACM Trans. on Information and System Security*, 2000.

[6] S. Singh, S. Silakari, and R. Patel, "An efficient feature reduction technique for intrusion detection system," *IPCSIT*, vol. 3, 2011.

[7] I. Ahmad, A. B. Abdullah, and Alghamdi, "Intrusion detection using feature subset selction based on MLP," *Scientific Research and Essays*, vol. 6, no. 34, 2011.

[8] S. M. Aqil, M. S. A. Khan, and J. Naeem, "Efficient probabilistic classification methods for NIDS," *IJCSIS*, vol. 8, no. 8, November 2010.

[9] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisher faces: Recognition using class specific linear projection," in *Proc. Fourth Eur. Conf. Computer Vision*, Cambridge, UK, April 1996, vol. 1, no. 1418, pp. 45–58.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. CognNeurosci* 3, pp. 71–86, 1991.

[11] K. Baek, B. Draper, J. R. Beveridge, and K. She, "PCA vs. ICA: A comparison on the FERET data set," in *Proc. the Fourth International Conf. on Computer Vision, Pattern Recognition and Image Processing*, Durham, NC, USA, March 8-14, 2002, pp. 824-827.

[12] A. Chittur, "Model generation for an intrusion detection system using genetic algorithms," High school Honors Thesis accessed in 2006.

Hafiz Muhammad Imran obtained the Ph.D.-Informatics scholar from Malaysia University of Science & Technology, Malaysia. His primary domain is software development and IT management, but his interest in networks thrilled him to do Ph.D. in informatics. His M.S. is in software engineering and B.S. is in computer engineering. He has developed bundle of projects and products out of which TeleNoc suite is most prominent, and he has specially designed to handle managed services contracts in telecom and power sector which includes different modules like GSM Sites Management, CM Alarms, PM Tickets, Fueling Management, Billing, HRMS, AMS, PMS, EAMS, Fleet management and Tracking system.

Hafiz Muhammad Imran is a member of PMI-1439868, he has published his research work in international journals.

Azween Bin Abdullah is a Ph.D. in IT and is an associate dean in Taylors university as well as a research professor in Malaysia University of Science & Technology, Malaysia. He is a fellow of the British Computer Society and a senior member in International Association of Computer Science and Information Technology (IACSIT). He is also a member of Association of Computing Machinery, International Association of Engineers (IAENG), The Society for Modeling and Simulation, Malaysian National Computer Confederation, American Mathematical Society and IEEE Computer Society. He has published in many international journals and conferences.

Sellappan Palaniappan holds a Ph.D. in Interdisciplinary Information Science from the University of Pittsburgh and a master degree in Computer Science from the University of London and is currently the acting provost and the dean of School of Science and Engineering at Malaysia University of Science and Technology (MUST). He has published numerous journals, conference papers and IT books. He has served as an IT consultant for several local and international agencies such as the Asian Development Bank, the United Nations Development Program, the World Bank and the Government of Malaysia.

He was a member of IEEE (USA), Chartered Engineering Council (UK) and British Computer Society (UK), and is currently a member of the Malaysian National Computer Confederation (MNCC).