

An Intelligent Data Mining Approach to Detect Intrusions in Social Programming Networks

Hayder Hussein Azeez^{1*}, Mustafa Musa Jaber²

[*Hayder.hussein@stu.edu.iq](mailto:Hayder.hussein@stu.edu.iq)

¹Nassryia Technical Institute, Southern Technical University, Thi Qar, Iraq

**²Department of Medical Instruments Engineering Techniques, Al-Farahidi University,
Baghdad 10022, Iraq**

Abstract: Data mining it's one of the successful techniques for intrusion detection in the fast growing internet world. Moreover, they are extensively used to develop intrusion detection systems that classify the users based on the behavior in various contexts. Soft computing techniques are used in data mining for making effective decisions by the applications of neural networks, fuzzy logics and genetic algorithms. Moreover, Fuzzy Rule-Based Classification Systems (FRBCSs) are productive and they are recognized well in the machine learning area, since they can present an interpretable decision model for the end user. The use of fuzzy logic algorithms helps to develop an efficient classification model that can detect the network intrusions more effectively. In this work, new feature selection algorithms namely Fuzzy based intelligent non-dominated particle sorting optimization (F-INDPS) are proposed to select The optimal number of features from dataset. These features contribute in classification process for forming rules and to apply them for effective decision making. An experimental result of the proposed system increases the security through the intrusion detection system and Total completion time has been reduced. Moreover, the proposed system increased the throughput and overall performance of the networks by enhancing the packet delivery ratio and by reducing the delay in communication and thereby provided more reliable and secured communication in social programming networks.

Keywords: Data mining, Fuzzy, network intrusions, machine learning throughput, packet delivery ratio.

1. Introduction:

Intrusion Detection System (IDS) is playing an significant it's part in security of Social Programming Networks since it is used for detecting the known and unknown attackers[1]. The attack detection in Social Networks is a difficult task while it is related with the IDS in fixed wired and wireless networks, because of the difficulty to fulfill the wireless and mobility requirements of Social Networks [2,3]. In this scenario, the characteristics of the ad-hoc network create new complexities during the implementation [4]. Moreover, few more additional challenges are also present in the detection of attacks in Social Networks which include lack of monitoring facilities for the mobile nodes in the network [5]. Moreover, the trust based routing protocol proposed in this research work needs the cooperation from all participating nodes for finding suitable and secure routes in Social Networks[6]. Therefore, it is necessary to analyze the use of datasets and pre-processing of datasets in order to develop efficient IDS[7]. Recently, Selection and classification techniques Features are useful for detecting intruders effectively in social network [8]. Classification methods used to categorize the node as normal and abnormal are based on certain conditions [9]. Many of soft computing techniques introduced in the past for effective classification of a dataset. Moreover, the existing classification techniques are not able to classify the dataset successfully to detect the intruders with all kinds of attacks [10]. For this purposing, two new feature selection techniques are introduced in this research work and they are useful for improving the classification accuracy. Specifically, Fuzzy based intelligent non-dominated particle sorting optimization (F-INDPS) is proposed in this research work for effective pre-processing of intrusion detection dataset and it is capable of selecting optimal number of features to be sent to the classifier. In this work, solution for IDS with an objective of minimizing energy consumption, throughput and overall performance of the networks by enhancing the packet delivery ratio and by reducing the delay in communication is proposed using highly efficient Fuzzy with non-dominated particle sorting optimization which is based on Pareto centered Metaheuristics. In the recent past, Data mining of Pareto-fronts were used to represent the plot of the objective functions in IDS problems of various networks whose non-dominated vectors are in the Data mining of pareto optimal set. The research has been organized by: section 2 presented the literature wiping carried out in the space of scientific workflow of IDS , feature selection based on data mining concepts security issues in CIoT framework. Section 3 provides the overall system architecture of the proposed system and performance of scientific workflow application using F-INDPS for minimizing IDS in social networking which reduces security issues between dependent tasks. Section 4 provides empirical analysis and discussion in detail of the proposed and other existing works. Section 5 presents the conclusions drawn from this work and suggests some potential future work.

2. Literature Survey:

Feature selection techniques were introduced by many of researchers who are work in the area of medical data mining, image processing and intrusion detection [11]. In all these applications, feature selection is used to choose ideal number of highlights from the datasets consisting of more number of features [12]. There are many research doing that have been carried out in this trend by various researchers ago. Among them,

[13] proposed the first additional feature selection algorithm which was used to find the relevant features which are suitable to solve the problem with optimal number of features. In their work, the number of features and patterns are huge in their dataset[14]. Therefore, the proposed and effective method for feature selection can help in dimensionality reduction. They also designed an incremental probabilistic algorithm as an alternative approach to the exhaustive and heuristic methods. [15] Proposed an incremental feature selection (IFS) algorithm which is dynamic in nature for taking decisions over the different functionalities. Their calculation is able to dodge the re-calculation process as opposed to retraining the dynamic choice framework as new one to compute the element of data mining subset without any preparation.

[16] A new evolutionary-incremental framework for feature selection has been developed. Their framework is applied to a normal evolutionary algorithm called a genetic algorithm. Their framework introduced some generic modifications on ordinary evaluation algorithms which are To be compatible with the variable length for data mining solutions. Moreover, their framework has been deployed for performing classification operation with two new operators namely addition and deletion operators which change the length of solution randomly.

[17] Proposed a new unsupervised feature selection method for effective of decision making over the dataset. [18] Proposed a new model for choose which feature is It was developed with the group of Incremental Mixed Variable Ant Colony Optimization (IACOMV) and Support Vector Machine (SVM) to solving model effectively check problems. [19,20] proposed a wrapper based element choice methodology that consolidates two machine learning calculations with an enhanced variant of the Binary Bat Algorithm for performing successful classification. In spite of the presence of all these works, most of the works are not suitable for feature selection in intrusion detection applications. Moreover, the existing feature selection algorithms are more complex and hence they consume lot of time for convergence as well consumes more energy with less throughput and overall performance for the intrusion detection. Therefore, new feature selection algorithms namely F-INDPS selection algorithm are proposed in this research to choose the optimal number of features.

3. FUZZY BASED INTELLIGENT NON-DOMINATED PARTICLE SORTING OPTIMIZATION SYSTEM

The overall structure of the proposal classification system as shown in Figure.1. It consists of seven major parts components namely NSL-KDD99 cup dataset, Wireless Network Trace data, User Interface Module, Intrusion Detection System, Decision Manager, Temporal Information Manager and Knowledge base.

- a. **NSL-KDD'99 Cup Dataset:** It is a standard bench mark data set which contains millions of records with 41 attributes.
- b. **Wireless Network Trace Data:** It consists of 10000 records with 21 attributes which are collected from wireless network server.

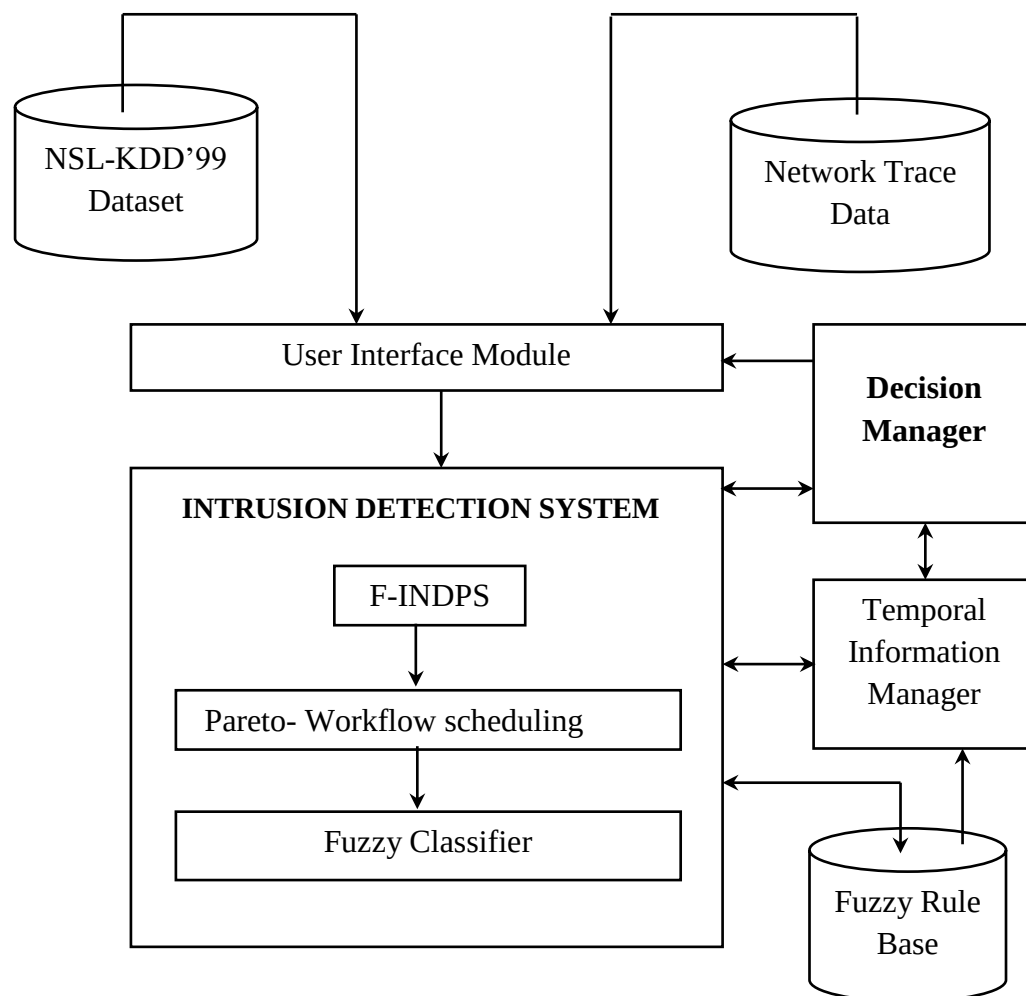


Figure.1. Architecture of F-INDPS classification system

- c. **User Interface Module:** It collects necessary data from these two dataset and also decides what data have to collect for evaluating the proposed system. Collected data's are transferred to the intrusion detection system for further processes over the network data and received the recommended (resulted) data from decision manager.
- d. **Intrusion Detection System:** It consists from of two sub-systems namely the pre-processing subsystem and the classification subsystem. The pre-processing subsystem performs pre-processing using two newly proposed pre-processing algorithms namely

F-INDPS Algorithm for effective pre-processing of the datasets. Fuzzy Classification Algorithm which is proposed in this research for performing effective classification of the intrusion detection dataset based on temporal fuzzy rules. The collected records are involved pre-processing first, fuzzy operations are carried out and then they are forwarded to the classification subsystem for applying fuzzy rules with temporal constraint for taking final decision over the network data regarding intrusions for helping temporal information manager and fuzzy rule base.

- e. **Temporal Information Manager:** It is helpful for classification intrusion detection process of this proposed system. It provides the necessary details about time constraints in the proposed system.
- f. **Fuzzy Rule Base data mining:** It contains the vague rules needed for decision-making on the dataset using a parallel approach to applying genetic processes and forming ambiguous time rules.
- g. **Decision Manager:** This decision manager assists in making the final decision about which network data is to move from the intrusion detection system with the help of Temporal Information manager and fuzzy rule base.

In the proposed work, non-dominated solutions generated by workflow scheduling algorithm are used to construct the Data mining of Pareto-fronts and fuzzy rules with a newly proposed fitness function being used to perform resource and schedule optimization. data mining of pareto optimal set denotes the notion of solution dominance in multi-objective optimization problems. X is the solution which is considered to control additional result Y, On the off chance that X is superior to Y in somewhere around one of the destinations and superior to or equivalent to in different goals. Formally, Data mining of pareto optimality is characterized as, a result $yy \in YY$ such that $fn(yy) < paretof(x^\alpha)$ as shown in the Eq(1)

Pareto optimal set is defined as follows

$$PP^* = \{yy \in YY \mid \forall y \in YY: fn(xx) \leq fn(yy)\} \quad (1)$$

The primary objective of multi-target advancement issue is finding however many solutions as could be expected under the circumstances which are near the Pareto-front. This issue has been defeated in this proposed work by applying fuzzy standards with the triangular membership function defined in Eq (2) for a lower limit 'lt', upper limit 'v' and 'a' value 'mm' which lies between c and d.

$$\mu_c(yy) = \begin{cases} 0 & yy \leq lt \\ \frac{yy - lt}{mm - lt} & lt < yy \leq mm \\ \frac{v - yy}{v - mm} & mm < yy < v \\ 0 & yy \geq v \end{cases} \quad (2)$$

The resultant data mining of Pareto-fronts has been used in this work to provide an optimal configuration of cloud to the scientific users. Moreover, the proposed F-INDPS based multi-objective optimization is useful for workflow scheduling which helps to reduce IDS and the proposed model has been tested by applying the proposed algorithm on various scientific

workflow applications with different sizes. In addition, the quality of the Fuzzy and Pareto-front generated from the proposed F-INDPS has been tested using different Pareto-front analysis metrics and the newly proposed fitness function.

$$I(Data) = - \sum_{j=1}^m \left[\frac{fn(C_j, Data)}{|Data|} \right] \log_2 \left[\frac{fn(C_j, Data)}{|Data|} \right] \quad (3)$$

$$I(Fn) = - \sum_{i=1}^n \left[\frac{|Fn_i|}{|Fn|} \right] \times I(Fn_i) \quad (4)$$

$$InGR(A_i) = \left[\frac{I(Data) - I(Fn)}{I(Data) + I(Fn)} \right] \times I(Fn_i) \quad (5)$$

As Shown in the Eq(4) &(5) &(6) Here, dataset, C_i are the fragments made from the dataset, Fn is the feature set with Fn_i as the feature and $InGR$ indicate the information gain ratio. The word info indicates the information gain values. Now, the steps of the non-dominated particle sorting optimization are as follows:

Algorithm.1. Non-Dominated Particle Sorting Optimization Algorithm

Step 1: $D = C1, C2, \dots, Cn$.

Where $C1, C2, \dots, Cn \rightarrow$ Subsets

Step 2: Agent computes $InGR(A_i) = \left[\frac{I(Data) - I(Fn)}{I(Data) + I(Fn)} \right] \times I(Fn_i)$

Step 3: $Check = \text{Max}(InGR)$

Step 4: $Partition = \text{neighbour nodes}$.

Step 5: fitness values (FV), Threshold (Th), Selected Feature set SS

If $FV(A_i) \geq Th$ & $FV(A_j) \geq Th$ then

$SS = SS \cup (A_i) \cup (A_j);$

Step 6: If $FV(A_i) \geq Th$ & $FV(A_j) < Th$ then

$SS = SS \cup (A_i);$

Step 7 : If $FV(A_i) < Th$ & $FV(A_j) \geq Th$ then

$SS = SS \cup (A_j);$

Step 8: Repeat this process until all attributes are tested with the rules used in this procedure.

Step 9: Return SS

Step 10: If No attributes are selected then change the Threshold with $Th = Th * 0.9$ and go to step 5.

Algorithm.2.Fitness Function Computing in F-INDPS

```

1  Function Fitness Function(PSO_List)
2  for every ready task  $t \in T$  do
3    find  $t.length, t.memory, t.input\_data, t.output\_data$ 
4  end for
5  for each virtual machine  $\in R$  do
6    find  $vm.cpu\_cores, vm.memory, vm.storage,$ 
        $vm.bandwidth, vm.mips, vm.available\_time$ 
7  end for
8  for each particle  $p \in S$  do
9    Calculate  $p.Total\ Completion\ Time$ 
10   Calculate energy consumption
11   Calculate  $p.consistency$ 
12     Find Best_Value using equation 5.13
13   Best_Value, High_Value, Medium_Value, Fitness_Value(BV, HV, MV, FV)
       Low_Value(LV)
       if  $BV > HV$  then
          $FV = BV$ 
       else if  $BV > MV$  then
          $FV = 2 * BV$ 
       else // if  $BV > LV$ 
          $FV = 3 * BV$ 
       end if
14 end for
15 return Fitness_value
16 end procedure

```

From the Algorithm.2.Fuzzy logic helps to perform reasoning under uncertainty. It incorporates a basic rule-based approach on well-formed formulas and the rules are represented by IF x AND y THEN z. A fuzzy inference systems applies rules which are stored in a knowledge base against facts present in the application to perform inference. Table.1 shows the fuzzy rules used in the proposed algorithm for finding the best solution from a set of workflow scheduling solutions with optimal configuration from the centre in which the scheduling activities are carried out to avoid intrusion detection. The rules in the table .1. are interpreted as IF... THEN rules, for example, the last row of the table represents the rule:

IF MEMORY_SIZE IS HIGH AND ENERGY_CONSUMPTION IS LESS AND AVAILABLE_TIME IS SMALL THEN SOLUTION_TYPE IS EXCELLENT

Similarly, the first row of the Table, 1 represents the rule:

IF MEMORY_SIZE IS LOW AND ENERGY_CONSUMPTION IS HIGH AND AVAILABLE_TIME IS LARGE THEN SOLUTION_TYPE IS POOR

In this way, the fuzzy inference system developed in this work applies the fuzzy rules using forward chaining inference mechanism to perform deductive inference in order to make efficient decisions on configurations and the scheduling methods. In memory size, Low memory size indicates a memory size up to 32 GB. Medium size of memory indicates a memory upto 1 TB. Finally, High memory indicates the availability of memory upto 1PB through virtual machine allocation. In the case of Energy Consumption, Less energy indicates the energy upto 1KW. Average energy consumption indicates upto 1000 KW and High energy consumption indicates upto 100 MW.

On account of Time to Complete the Workflow Task, little demonstrates an opportunity to finish the task upto 60 minutes, Medium shows an opportunity to finish the task upto 2 hours, High shows an opportunity to finish the task upto 3 hours. Hence, the proposed algorithm aims at optimizing these parameters in the IDS.

Table.1.Fuzzy rules to find the best solution

Memory_size	Energy_consumption	Time to Complete the Workflow Task	Solution_type
Low	High	Large	Poor
Low	High	Medium	Fair
Low	High	Small	Very Fair
Low	Average	Large	Fair
Low	Average	Medium	Very Fair
Low	Average	Small	Good
Low	Less	Large	Very Fair
Low	Less	Medium	Good
Low	Less	Small	Very Good
Medium	High	Large	Fair
Medium	High	Medium	Very Fair
Medium	High	Small	Good
Medium	Average	Large	Very Fair
Medium	Average	Medium	Good
Medium	Average	Small	Very Good
Medium	Less	Large	Good
Medium	Less	Medium	Very Good
Medium	Less	Small	Best
High	High	Large	Very Fair
High	High	Medium	Good
High	High	Small	Very Good
High	Average	Large	Good
High	Average	Medium	Very Good
High	Average	Small	Best
High	Less	Large	Very Good
High	Less	Medium	Best
High	Less	Small	Excellent

4. Results and Discussion

In this section, the description of parameter settings of this work for implementing the proposed F-INDPS and the performance is evaluated using standard network dataset called KDD 1999 cup dataset for test the performance of the proposed system. The various parameters has been taken in to consideration have been conducted by using this dataset with its feature is shown in the Table.2.

Table.2.Parameter values used in F-NSPSO workflow scheduling

S.No.	Parameter	Values
1	Number of particles[S] and its duration rage	200
2	Number of iterations(Which includes Error rate, failed logins)	50-200
3	Inertia Weight[W]	0.4 to 1.0
4	Random Variables [C1,C2]	2.0
5	Dimension of Particles	Number of tasks in the workflow

a. Total Completion Time(TCT) and Energy Consumption (EC)

To verify the proposed algorithm, initially the results of TCT and energy consumption are compared with IACOMV, SVM, and IFS algorithm. In F-INDPS based implementation, minimum energy consumption by operating the processor at low voltage for five different numbers of tasks such as 100, 200, 300, 400 and 500. Figure 2. Shows the energy consumption level of F-INDPS. In the case of Energy Consumption, Less energy indicates the energy upto 1KW. Average energy consumption indicates upto 1000 KW and High energy consumption indicates upto 100 MW.

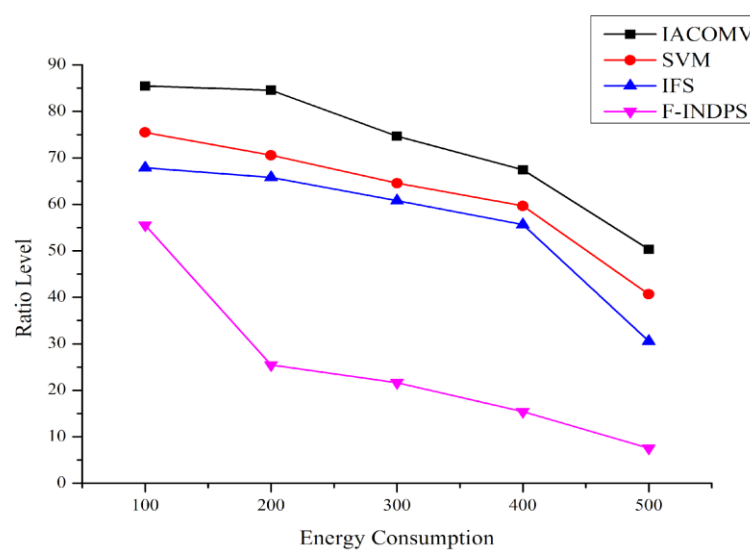


Figure 2. Energy Consumption level of F-INDPS

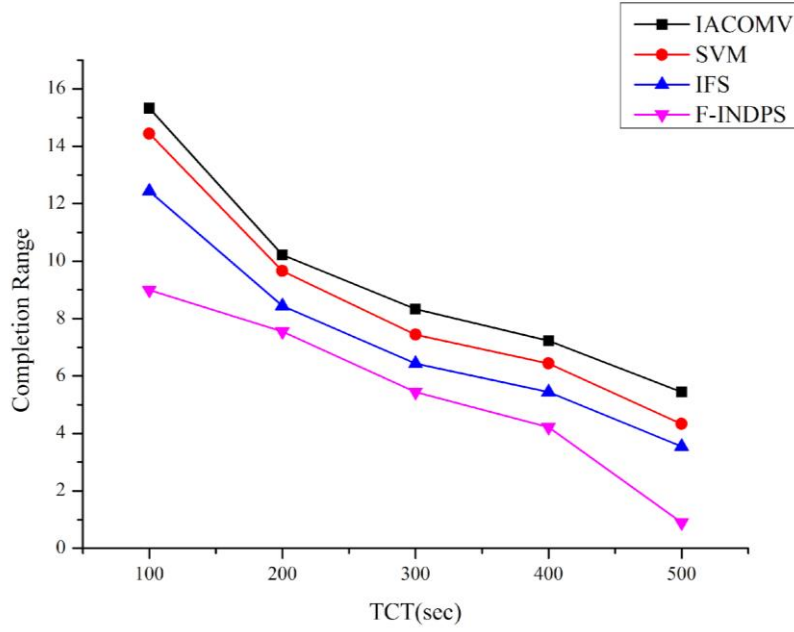


Figure 3. TCT Consumption level of F-INDPS

As shown in the Figure.3. On account of Time to Complete the Workflow Task, little demonstrates an opportunity to finish the task upto 60 minutes, Medium shows an opportunity to finish the task upto 2 hours, High shows an opportunity to finish the task upto 3 hours. This is due to the use of effective optimization process which is carried out by incorporating the appropriate fuzzy membership function.

b. Distance Coverage using Spacing(S)

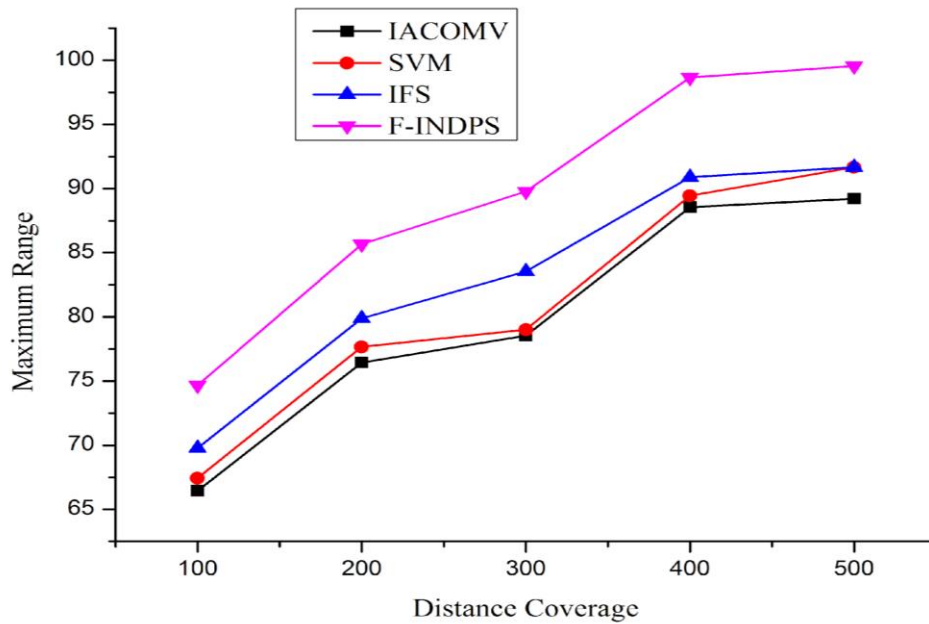


Figure 4. Distance Coverage using Spacing for F-INDPS

As shown in the Figure.4.This metric is used to measure the distribution of particles during intrusion in the obtained data mining of Pareto-front [21]. It is calculated based on the average distance of each point from its nearest neighbour. An ideal value of spacing is Eq(6) and the F-INDPS has good distance coverage due to pareto optimality.

$$s = \frac{1}{n-1} \sqrt{\sum_{i=1}^n (\underline{d} - d_i)^2} \quad (6)$$

Where $d_i = \min_j (\sum_{k=1}^m |f_m^i - f_m^j|)$; $i,j=1..n$, m represents the number of objectives, n represents number of non-dominated solutions and $\underline{d}$ is computed as the mean of d_i

c. Performance based on Ratio of Non-dominated Individuals (RNI)

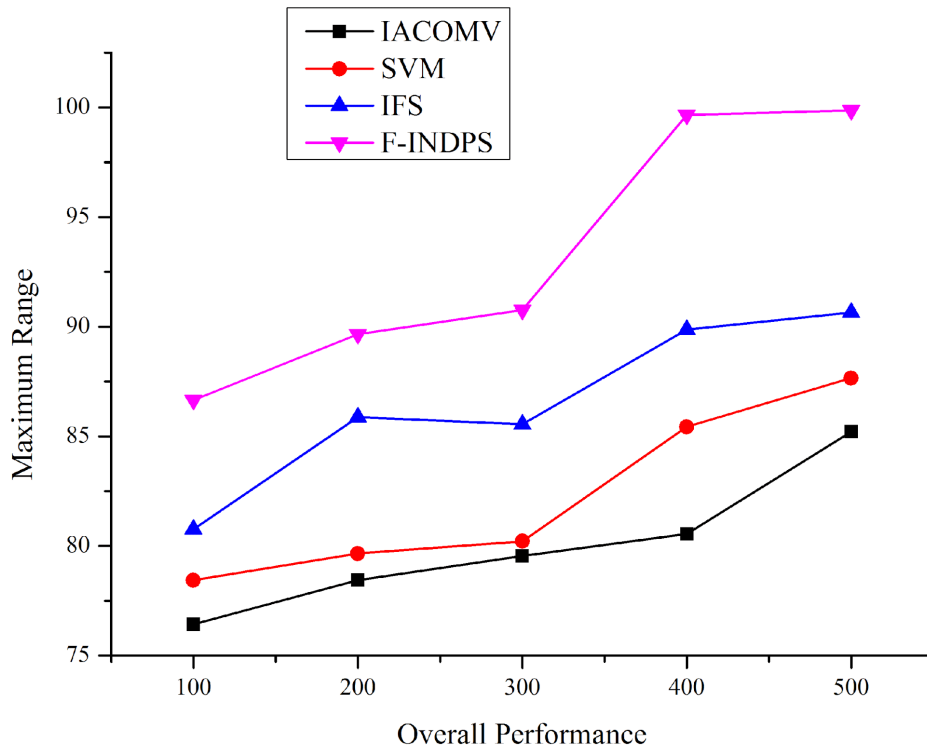


Figure 5. Performance based on Ratio of Non-dominated Individuals for F-INDPS

As shown in the Figure.5.This metric measures the cardinality of the obtained non dominated solutions [22] It is always desirable to have a F-INDPS large number of non-dominated solutions than IACOMV, SVM, IFS algorithm and is calculated using the following Eq(7)

$$RNI = \frac{\text{Number of Non-dominated solutions}}{\text{Total number of solutions}} \quad (7)$$

d. Consistency using Maximum Spread(MS)

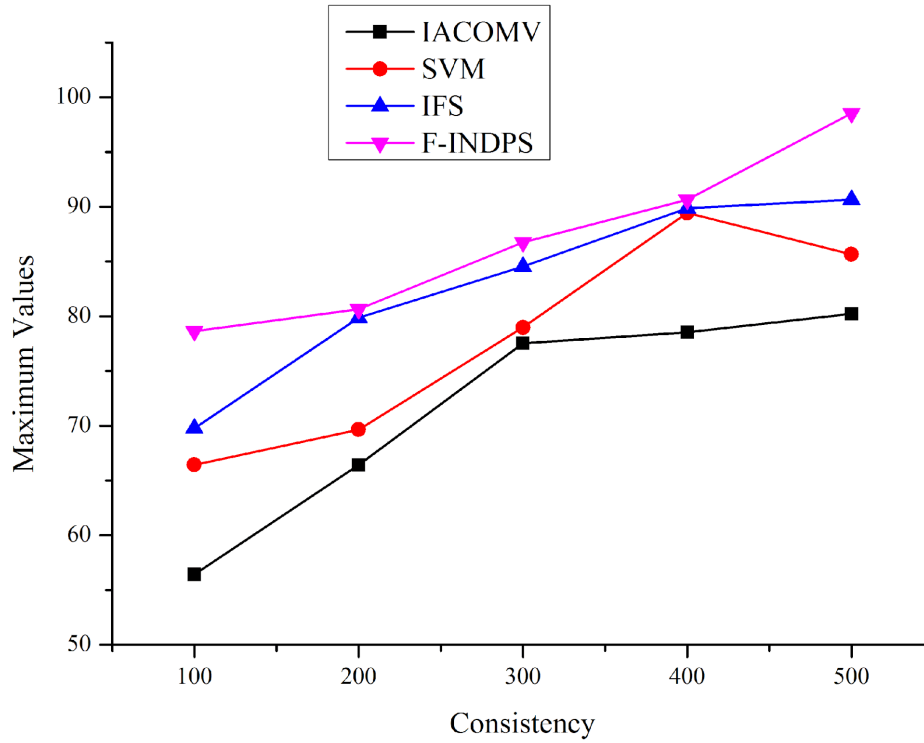


Figure 6. Consistency using Maximum Spread for F-INDPS

It measures the quality of the generated Pareto- front by analysing how much search space is covered [23-25] and is calculated by finding Euclidian distance between the maximum and minimum value of the objective function(OF) as given below in the Eq(8) with high consistency ratio [26] than IACOMV, SVM, and IFS approach as shown in the Figure.6.

$$MS = \sqrt{\sum_{i=1}^n |f_i^{max} - f_i^{min}|} \quad (8)$$

where f_i^{max} represents the maximum value and f_i^{min} represents the minimum value with i^{th} OF. It helps in evaluating the diversity of the search space.

The proposed work has been evaluated on different metrics [27] and the results have indicated that the proposed work has given good results which are more suitable for social programming network for intrusion detection [28].

5. Conclusion

The F-INDPS algorithm uses a fitness function to minimize the energy consumption, total completion time and paid idle time of the resources. For large scale scientific applications number of non-dominated solutions will be more, for deciding the quality of the solution memory utilization is also considered as one more objective which helps to reduce intrusion in the systems. In order to quickly decide the solution 16 Fuzzy rules have been developed to classify the solution and also these rules facilitate easy decision making. Simulations have been conducted to evaluate the performance of F-INDPS compared to IACOMV, SVM, IFS. Performance analysis of F-INDPS algorithm shows that the quality of the non-dominated solutions is good. In future extension cluster based scheduling solution for IDS optimization for data intensive applications to minimize data transfer overhead will be presented.

References:

1. Dutta, S., Mawla, T., & Rabbi, M. F. (2019). A Comparison Study of Temporal Signature Mining Over Traditional Data Mining Techniques to Detect Network Intrusion. In *Emerging Technologies in Data Mining and Information Security* (pp. 757-763). Springer, Singapore.
2. Mawla, T., Dutta, S., & Rabbi, M. F. (2019). Temporal Signature Mining for Network Intrusion Detection Using TEMR. In *Emerging Technologies in Data Mining and Information Security* (pp. 645-655). Springer, Singapore.
3. Wei, T. E., Lee, H. M., Jeng, A. B., Lamba, H., & Faloutsos, C. (2019). WebHound: a data-driven intrusion detection from real-world web access logs. *Soft Computing*, 1-19.
4. P. Mohamed Shakeel; Tarek E. El. Tobely; Haytham Al-Feel; Gunasekaran Manogaran; S. Baskar., "Neural Network Based Brain Tumor Detection Using Wireless Infrared Imaging Sensor", IEEE Access, 2019, Page(s): 1
5. Shakeel, P.M., Tolba, A., Al-Makhadmeh, Zafer Al-Makhadmeh, Mustafa Musa Jaber, "Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks", Neural Computing and Applications, 2019, pp1-14. <https://doi.org/10.1007/s00521-018-03972-2>
6. Neelima, D., Karthik, J., John, K. A., Gowthami, S., & Nayak, J. (2019). Soft Computing-Based Intrusion Detection Approaches: An Analytical Study. In *Soft Computing in Data Analytics* (pp. 635-651). Springer, Singapore.
7. Saxena, A., Saxena, K., & Goyal, J. (2019). Hybrid Technique Based on DBSCAN for Selection of Improved Features for Intrusion Detection System. In *Emerging Trends in Expert Applications and Security* (pp. 365-377). Springer, Singapore.
8. Tan, X., Su, S., Huang, Z., Guo, X., Zuo, Z., Sun, X., & Li, L. (2019). Wireless Sensor Networks Intrusion Detection Based on SMOTE and the Random Forest Algorithm. *Sensors*, 19(1), 203.
9. Baskar, S., Dhulipala, V.R.S., Shakeel, P.M., Sridhar, K. P., Kumar, R. Hybrid fuzzy based spearman rank correlation for cranial nerve palsy detection in MIIoT environment. Health Technology. (2019). <https://doi.org/10.1007/s12553-019-00294-8>
10. Ashraf, M., Guohua, G., Wang, X., & Ahmad, F. (2018, December). Detection and Resolving WSN issues with the use of Data Mining & Intelligent Algorithms. In *2018 International Conference on Frontiers of Information Technology (FIT)* (pp. 253-258). IEEE.
11. Arshad, I., Dutta, C., Choudhury, T., & Thakral, A. (2018, June). Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques. In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 163-168). IEEE.
12. Yang, Y., Chen, D., Wang, H., & Wang, X. (2018). Incremental perspective for feature selection based on fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, 26(3), 1257-1273.

13. Dorigo, M., & Stützle, T. (2019). Ant colony optimization: overview and recent advances. In *Handbook of metaheuristics*(pp. 311-351). Springer, Cham.
14. Scholkopf, B., Smola, A. J., & Bach, F. (2018). Learning with kernels: Support vector machines, regularization, optimization, and beyond.
15. Li, X., Cui, Z., Sun, L., Lu, G., & Narayan, D. (2019). Research on iterative repair algorithm of Hyperchaotic image based on support vector machine. *Discrete & Continuous Dynamical Systems-S*, 12(4&5), 1199-1218.
16. Kannan, N., Sivasubramanian, S., Kaliappan, M., Vimal, S., & Suresh, A. (2018). Predictive big data analytic on demonetization data using support vector machine. *Cluster Computing*, 1-12.
17. Mathew, J., Pang, C. K., Luo, M., & Leong, W. H. (2018). Classification of imbalanced data by oversampling in kernel space of support vector machines. *IEEE transactions on neural networks and learning systems*, 29(9), 4065-4076.
18. Maione, C., de Oliveira Souza, V. C., Togni, L. R., da Costa, J. L., Campiglia, A. D., Barbosa, F., & Barbosa, R. M. (2018). Establishing chemical profiling for ecstasy tablets based on trace element levels and support vector machine. *Neural Computing and Applications*, 30(3), 947-955.
19. Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2019). Sentence Level Sentiment Identification and Calculation from News Articles Using Machine Learning Techniques. In *Computing, Communication and Signal Processing* (pp. 371-376). Springer, Singapore.
20. Qi, C., Chen, Q., Fourie, A., Tang, X., Zhang, Q., Dong, X., & Feng, Y. (2019). Constitutive modelling of cemented paste backfill: A data-mining approach. *Construction and Building Materials*, 197, 262-270.
21. Nisioti, A., Mylonas, A., Yoo, P. D., & Katos, V. (2018). From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods. *IEEE Communications Surveys & Tutorials*, 20(4), 3369-3388.
22. Selvakumar S, Inbarani H, Shakeel PM. A Hybrid Personalized Tag Recommendations for Social E-Learning System. *International Journal of Control Theory and Applications*. 2016;9(2):1187-99.
23. Nelson-Barber, S., & Johnson, Z. (2019). Raising the standard for testing research-based interventions in Indigenous learning communities. *International Review of Education*, 1-19.
24. Elhag, S., Fernández, A., Alshomrani, S., & Herrera, F. (2019). Evolutionary Fuzzy Systems: A Case Study for Intrusion Detection Systems. In *Evolutionary and Swarm Intelligence Algorithms* (pp. 169-190). Springer, Cham.
25. Manogaran G, Shakeel PM, Hassanein AS, Priyan MK, Gokulnath C. Machine-Learning Approach Based Gamma Distribution for Brain Abnormalities Detection and Data Sample Imbalance Analysis. *IEEE Access*. 2018 Nov 9.DOI 10.1109/ACCESS.2018.2878276
26. Kozak, J. (2019). Evolutionary Computing Techniques in Data Mining. In *Decision Tree and Ensemble Learning Based on Ant Colony Optimization* (pp. 29-44). Springer, Cham.
27. Shah, S. A. R., Issac, B., & Jacob, S. M. (2018). Intelligent Intrusion Detection System Through Combined and Optimized Machine Learning. *International Journal of Computational Intelligence and Applications*, 1850007.
28. Metiaf, A., Hong, W. Q., & Aljeroudi, Y. (2019). Searching with Direction Awareness: Multi-Objective Genetic Algorithm Based on Angle Quantization and Crowding Distance MOGA-AQCD. *IEEE Access*.