DEFENCE AND SECURITY SYMPOSIUM

IISRI INSTITUTE FOR INTELLIGENT SYSTEMS RESEARCH AND INNOVATION

DEAKIN UNIVERSITY

Optimising Network Intrusion Detection Systems with an Ensemble Multi-objective Harris' Hawks Optimiser

By: Kelvin Choo

INTRODUCTION

Machine Learning (ML)-based Network Intrusion Detection Systems (NIDSs) have proven to become a crucial technique for detecting malicious network activities from cyber criminals [1,2,3]. This research leverages a decision tree to differentiate and classify between normal network activities and invasions. The model is trained with the UNSW-NB15 dataset. While each data sample comprises many features, not all are discriminative in the classification task. An ensemble multi-objective Harris' hawk optimiser is designed and developed to optimise the model with multiple objectives, viz., minimising the number of features, maximising sensitivity, and maximising specificity.

OBJECTIVES

- To improve the network anomalies detection rate utilising a decision tree algorithm
- To reduce the model complexity and training time.
- To optimise the number of features used in training by selecting a compact feature set

METHODOLOGY

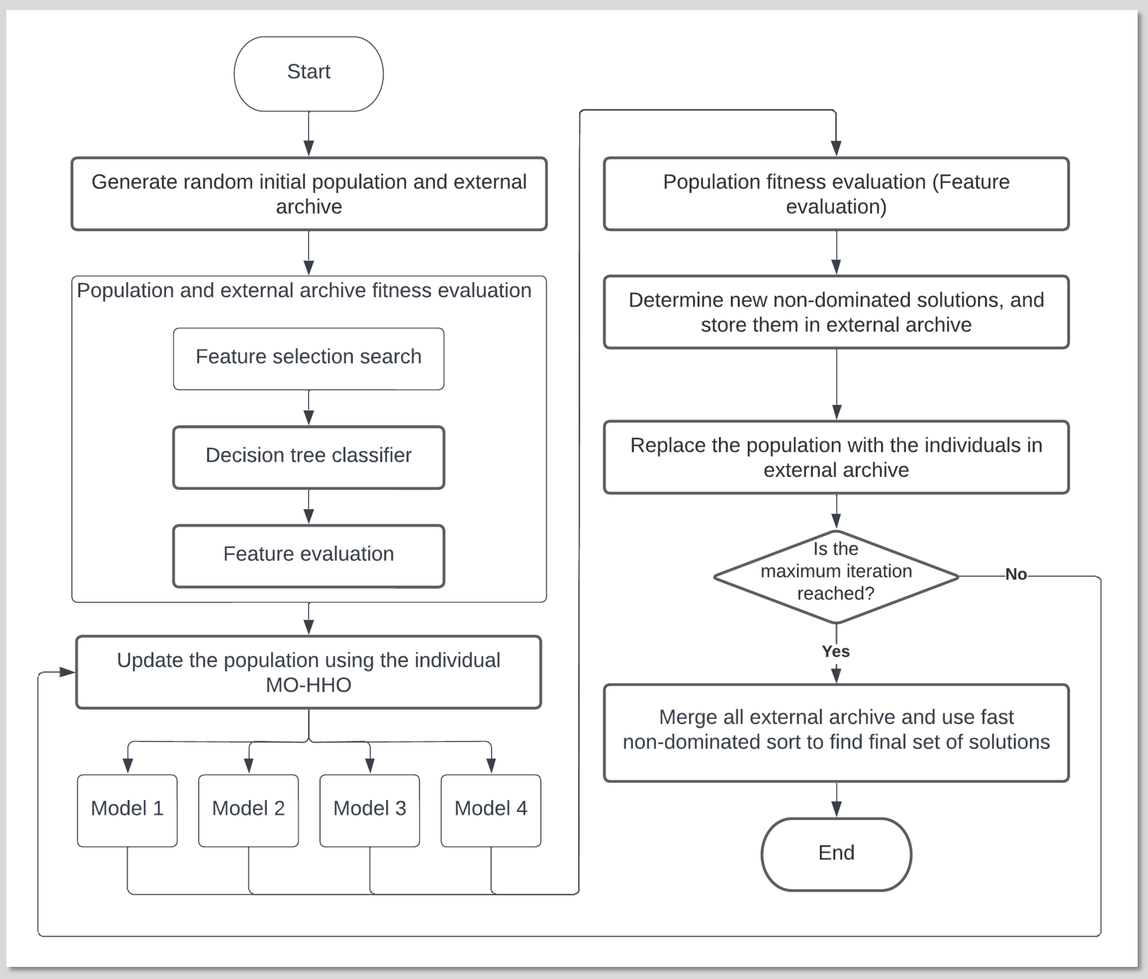

Figure 1: Flowchart of the feature selection with EMO-HHO.

- Dataset: UNSW-NB15 (containing 196 features)
- Feature selection: a wrapper method
- Machine learning : decision tree classifier
  - Settings: Criterion = Entropy, Max depth = 8, Min. samples leaf = 9
- Optimiser: Ensemble Multi-objective Harris' Hawk (EMO-HHO) comprising four enhanced MO-HHO variants
  - Population size: 20; Max iteration: 100
- Objective function:
  - Minimise the number of features
  - Maximise sensitivity
  - Maximise specificity

RESULTS

Table 1: Best solution obtained by each MO-HHO model and its ensemble EMO-HHO model for each objective function.

	Objective 1 (Number of features)	Objective 2 (Sensitivity %)	Objective 3 (Specificity %)
Benchmark	196	95.40%	85.61%
The best solution with minimum features			
MO-HHO1	54	91.15%	81.04%
MO-HHO2	38	99.21%	72.81%
MO-HHO3	**19**	**96.21%**	**77.09%**
MO-HHO4	39	99.70%	64.67%
EMO-HHO	**19**	**96.21%**	**77.09%**
The best solution with the highest sensitivity			
MO-HHO1	**74**	**99.95%**	**71.01%**
MO-HHO2	47	99.78%	71.41%
MO-HHO3	58	99.75%	72.33%
MO-HHO4	44	99.87%	66.69%
EMO-HHO	**74**	**99.95%**	**71.01%**
The best solution with the highest specificity			
MO-HHO1	65	91.76%	93.69%
MO-HHO2	69	92.50%	93.52%
MO-HHO3	**61**	**88.45%**	**95.81%**
MO-HHO4	63	91.02%	94.87%
EMO-HHO	**61**	**88.45%**	**95.81%**

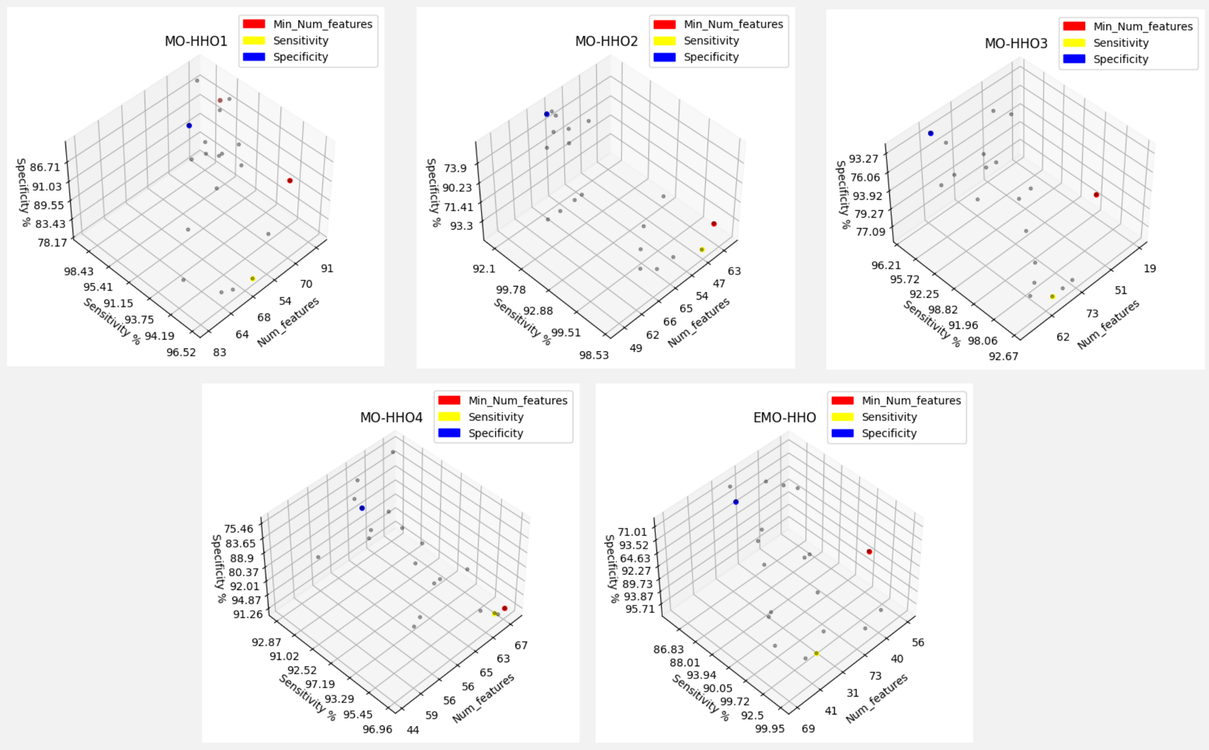

Figure 2: Scatter plot of the solutions generated by EMO-HHO.

DISCUSSION

- MO-HHO3 and EMO-HHO can generate better solutions in objectives 1 and 3, which have fewer features and the highest specificity rate.
- MO-HHO1 and EMO-HHO produce the best solution in objective 2
- However, MO-HHO2 yields the best overall solution (highlighted in yellow) with a better balance between sensitivity and specificity scores.
- The feature set selected by the proposed model performs better than those from the benchmark in terms of sensitivity and specificity with fewer number of features.

CONCLUSION

- The proposed model is able to yield a better network anomaly detection rate than those from the benchmark with fewer features used in model training.
- Users can choose the most preferred solutions with respect to each objective for implementation.
- A generic framework has been developed in which the decision tree classifier can be replaced with various machine learning algorithms such as K-mean, support vector machines, and random forest.

REFERENCES

[1] Moustafa, Nour, and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)." Military Communications and Information Systems Conference (MilCIS), 2015. IEEE, 2015.

[2] Sarhan, Mohanad, Siamak Layeghy, Nour Moustafa, and Marius Portmann. NetFlow Datasets for Machine Learning-Based Network Intrusion Detection Systems. In Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings (p. 117). Springer Nature.

[3] Moustafa, Nour, Gideon Creech, and Jill Slay. "Big data analytics for intrusion detection system: Statistical decision-making using finite dirichlet mixture models." Data analytics and decision support for cybersecurity. Springer, Cham, 2017. 127-156.