IoT Intrusion Detection By kNN:

Multi-Objective Optimization

William Eng

IoT Cyber Security

- People's Republic of China-Linked Actors Compromise Routers and IoT Devices for Botnet Operations
- Intrusion Detection
  - AI Use Case
    - Attack Accuracy
      - Type
    - False alarm
    - Detection rate

Multi-Objective Optimization

- kNN
  - Change k
  - Optimize accuracy and F1-score

Swarm-Based Algorithm

- Particle Swarm Optimization (PSO)
- Whale Optimization Algorithm (WOA)

Runtime type	Python 3
Hardware accelerator	CPU
Initial k Nearest Neighbor	5
Population	30
Epoch	3
Filter Out	Analysis, Backdoor, Shellcode, Worms
Optimize For	k, f1

Libraries

MEALPY

- Open Source Python Library

Scikit-Learn

- kNN model
- Metrics
  - Classification, Confusion

Pandas

- Load Data

Data

UNSW-NB15

- Intelligent Security Group at UNSW Canberra
- 100GB network capture data
- Initial Training Subset
  - Exclude Analysis, Backdoor, Shellcode, Worms
  - 10,000 of each Attack, 56,000 Normal
- Final Training Subset
  - Exclude Analysis, Backdoor, Shellcode, Worms
  - 2,000 of each Attack, 10,000 Normal
- 49 Features
  - Standard Scaling

Attack Categories	**Training Dataset Total Count**	**Testing Dataset Total Count**
Analysis	2000	37000
Backdoor	1746	18871
DoS	12264	11132
Exploits	33393	6062
Fuzzers	18184	4089
Generic	40000	3496
Normal	56000	677
Reconnaissance	10491	583
Shellcode	1133	378
Worms	130	44
Grand Total	**175341**	**82332**

Baseline kNN Model Attack Cat Classification

	Precision	Recall	F1-score	Support
DoS	0.19	0.47	0.27	4089
Exploits	0.62	0.48	0.54	11132
Fuzzers	0.17	0.66	0.27	6062
Generic	0.99	0.52	0.68	18871
Normal	0.88	0.48	0.62	37000
Reconnaissance	0.24	0.52	0.33	3496
Accuracy			0.51	80650
Macro avg	0.51	0.52	0.45	80650
Weighted avg	0.71	0.51	0.57	80650

Baseline kNN Model Attack Cat Confusion

	DoS	Exploits	Fuzzers	Generic	Normal	Reconnaissance
DoS	1933	634	813	5	440	264
Exploits	2625	5361	1685	10	630	821
Fuzzers	981	78	4030	8	402	563
Generic	2426	297	5699	9768	386	295
Normal	1989	1975	11167	89	17870	3910
Reconnaissance	222	249	594	0	606	1825

Baseline kNN Model Attack Label

	Precision	Recall	F1-score	Support
Normal	0.88	0.48	0.62	37000
Attack	0.68	0.95	0.79	43650
Accuracy			0.77	82332
Macro avg	0.76	0.76	0.76	82332
Weighted avg	0.77	0.77	0.77	82332

	Normal	Attack
Normal	27291	9709
Attack	9448	35884

Optimization Results (Attack Category)

PSO

- k, 23.18

WOA

- k, 22.52

Both F1 and accuracy improved slightly

	Baseline F1-score	PSO F1-score
DoS	0.27	0.32
Exploits	0.54	0.55
Fuzzers	0.27	0.28
Generic	0.68	0.71
Normal	0.62	0.64
Reconnaissance	0.33	0.40
Accuracy	0.51	0.53
Macro avg	0.45	0.48
Weighted avg	0.57	0.59

Optimization Results (Attack Label)

- F1 score improved
- Accuracy worsened

	Baseline F1-score	PSO F1-score
Normal	0.62	0.65
Attack	0.79	0.82
Accuracy	0.77	0.76
Macro avg	0.76	0.73
Weighted avg	0.77	0.74

Conclusion

Multi-Objective Optimization

- Marginally improved F1 scores and accuracy for Attack Category
- Marginally improved F1 scores for Attack Label

Future Possible Work

- Improve objective function
    - Fitness Score, Search Space
- Try three objectives
- Move to a more complicated machine learning model
- Increase Epoch

Reference

Federal Bureau of Investigation, Cyber National Mission Force, & National Security Agency. (2024, September 18). *People's Republic of China-Linked Actors Compromise Routers and IoT Devices for Botnet Operations*. Department of Defense. Retrieved November 3, 2024, from https://media.defense.gov/2024/Sep/18/2003547016/-1/-1/0/CSA-PRC-LINKED-ACTORS-BOTNET.PDF

Kareem, S. S., Mostafa, R. R., Hashim, F. A., & El-Bakry, H. M. (2022, February 11). An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection. *Sensors*, *22*(4), 1396. https://www.mdpi.com/1424-8220/22/4/1396

Mastafa, N. (2016). The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data se. *Information Security Journal: A Global Perspective*, 1-14. https://www.tandfonline.com/doi/abs/10.1080/19393555.2015.1125974

Moustafa, N., Creech, G., & Slay, J. (2017). Big Data Analytics for Intrusion Detection System: Statistical Decision-Making Using Finite Dirichlet Mixture Models. In I. Palomares Carrascosa, H. K. Kalutarage, & Y. Huang (Eds.), *Data Analytics and Decision Support for Cybersecurity: Trends, Methodologies and Applications*. Springer International Publishing.

Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*. 10.1109/MilCIS.2015.7348942

Moustafa, N., & Slay, J. (2021, June 2). *The UNSW-NB15 Dataset*. The UNSW-NB15 Dataset | UNSW Research. Retrieved November 3, 2024, from https://research.unsw.edu.au/projects/unsw-nb15-dataset

Moustafa, N., Slay, J., & Creech, G. (2019, December 01). Novel Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation on Large-Scale Networks. *IEEE Transactions on Big Data*, *5*(4), 481-494. https://ieeexplore.ieee.org/abstract/document/7948715

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. *Journal of Big Data*, *11*(105). https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-00957-y

Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2020, November 18). NetFlow Datasets for Machine Learning-based Network Intrusion Detection Systems. *BDTA 2020*, 1-16. https://arxiv.org/abs/2011.09144

scikit-learn. (2024, September). *scikit-learn*. scikit-learn: Machine Learning in Python. https://scikit-learn.org/stable/

Sharma, S., Kumar, V., & Dutta, K. (2024). Multi-objective optimization algorithms for intrusion detection in IoT networks: A systematic review. *Internet of Things and Cyber-Physical Systems*, *4*, 258-267. https://doi.org/10.1016/j.iotcps.2024.01.003

Thieu, N. V., & Mirjalili, S. (2023). MEALPY: An open-source library for latest meta-heuristic algorithms in Python. *Journal of Systems Architecture*, *139*(2023). https://www.sciencedirect.com/science/article/abs/pii/S1383762123000504?via%3Dihub