

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

Projekt dyplomowy

*Wykrywanie anomalii w ruchu sieciowym z wykorzystaniem algorytmów populacyjnych*

*Anomaly detection in network traffic using population algorithms*

Autor: *Zofia Jankowska*

Kierunek studiów: *Automatyka i Robotyka*

Opiekun pracy: *dr hab. inż. Joanna Kwiecień*

Kraków, 2024

Spis treści

[Wstęp 2](#_Toc141036487)

[Cel i założenia projektu 2](#_Toc141036488)

[Implementacja 2](#_Toc141036489)

[Testy 2](#_Toc141036490)

[Wnioski 2](#_Toc141036491)

[Bibliografia 2](#_Toc141036492)

# Wstęp

Problem anomalii

Problem włamań do sieci

Przegląd metod wykrywania anomalii, w szczególności włamań (GA i PSO)

# Cel i założenia projektu

Celem jest implementacja i porównanie działania algorytmu genetycznego oraz PSO w celu wykrywania anomalii w ruchu sieciowym na podstawie bazy danych KDDCUP’99.

Poszczególne rekordy z bazy będą klasyfikowane jako połączenie normalne lub atak (czy dzielimy na 4 różne główne typy ataków)

Trenowanie algorytmów zostanie wykonane na podstawie zestawu treningowego 10% całej bazy, a następnie przeprowadzona zostanie analiza działania na podstawie całego 100% bazy.

Wykonany zostanie interfejs graficzny przedstawiający przebieg uczenia oraz oraz sprawdzenie efektywności wyznaczonych reguł.

Wykonane zostaną testy statystyczne: test Wilkoxona (rozwinięcie testu t-studenta).

# 

# Implementacja

**Opis bazy danych i preprocessingu**

Opis bazy danych

W ramach preprocessingu – odpowiednia klasyfikacja rekordów :

- Zmiana danych tekstowych na 1 z n.

- Wyznaczenie wartości minimalnych i maksymalnych w każdej kolumnie

- należy jeszcze sprawdzić czy każda kolumna z tych 41 jest rzeczywiście potrzebna (w genetic\_network zastosowano tylko 9, genetic\_intrusion 8)

- zakodowanie odpowiednie adresów IP – rozbicie na 4 wartości?

**Algorytm genetyczny**

- kodowanie jako tablica

- osobniki będą miały zastosowane uogólnienie w postaci dwóch ostatnich wartości oznaczonych xxx.xxx (w genetic\_network jako -1)

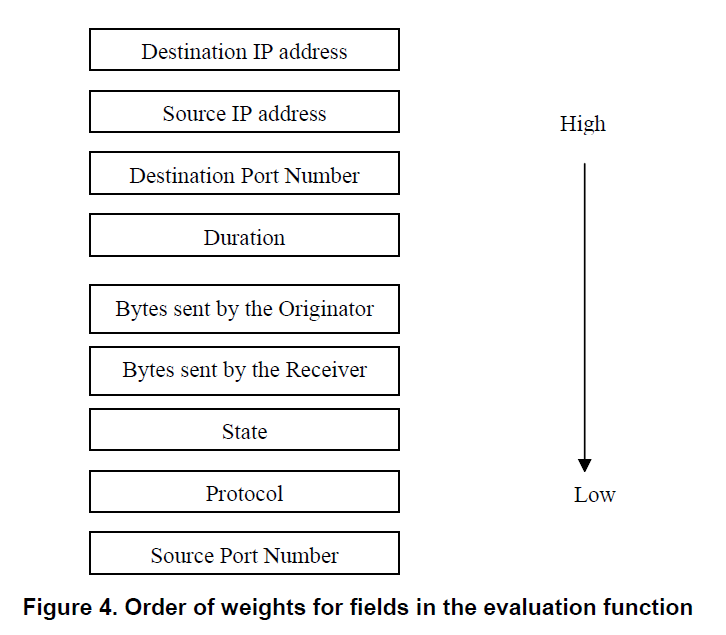
- generowanie losowej populacji początkowej (100-200 osobników)

- krzyżowanie jednopunktowe – po prostu tniemy i zamieniamy (możliwe że w 2 miejscach)

- mutacja – jedno z pól zmienia swoją wartość na wylosowaną z wyznaczonego w trakcie preprocessingu dla zakresu wartości

- funkcja ewaluacyjna – ile dany potomek poprawnie zaklasyfikuje anomalii w zbiorze testowym, ale z wykorzystaniem funkcji F = a/A – b/B, gdzie a i b to poprawnie zaklasyfikowane ataki i nie, a A i B to ilość ataków i nie w zbiorze treningowym. W ten sposób jeśli F > 0 to znaczy że poprawnie sklasyfikowano więcej ataków niż normalsów.

- funkcja ewaluacyjna może też wykorzystywać wagi dla mismatchów pojedynczych pól zgodnie z tabelą wag – genetic\_network



- no i dalej schody, bo podejrzewam że jedna reguła nie przyniesie bardzo widowiskowych efektów, dlatego nie wiem czy nie użyć kilku na raz

- selekcja - w artykule genetic\_intrusion wybierano 60 najlepszych potomków i 40 najlepszych jeszcze z nich jest kopiowane. W artykule effective… metoda selekcji to ruletka z wagami. Im lepsza funkcja ewaluacyjna tym więcej pól na ruletce zajmuje.

- kryterium stopu – funkcja fitness nie rośnie po n iteracjach

**Algorytm PSO**

**Aplikacja**

# Testy

# Wnioski

# Bibliografia

<https://www.kaggle.com/code/abhaymudgal/intrusion-detection-system>

<https://www.ecb.torontomu.ca/~bagheri/papers/cisda.pdf>

<https://www.kaggle.com/datasets/galaxyh/kdd-cup-1999-data>

<https://onedrive.aptiv.com/personal/zofia_jankowska_aptiv_com/Documents/Documents/PracaInz/genetic_algorithm_in_anomaly_detection2.pdf>

<https://onedrive.aptiv.com/personal/zofia_jankowska_aptiv_com/Documents/Documents/PracaInz/Particle_Swarm_Optimisation_for_Outlier_Detection-1.pdf>

<https://onedrive.aptiv.com/personal/zofia_jankowska_aptiv_com/Documents/Documents/PracaInz/Genetic_network.pdf>