

Algorytmy detekcji infekcji programów komputerowych inspirowane biologicznymi mechanizmami immunologicznymi

mgr inż. Patryk Widuliński

*Wydział Elektroniki i Informatyki
Politechnika Koszalińska*



Plan prezentacji



1. Cel pracy
2. Teza pracy
3. Wstęp do tematyki
4. Algorytm negatywnej selekcji
5. Metoda losowa
6. Metoda szablonów
7. Badania eksperymentalne
8. Wnioski
9. Podsumowanie

Cel pracy



- Analiza istniejących i opracowanie nowych algorytmów detekcji infekcji programów komputerowych, inspirowanych mechanizmami obronnymi organizmów żywych

Teza pracy



- Możliwa jest realizacja algorytmów detekcji infekcji programów komputerowych o wybranych właściwościach porównywalnych lub lepszych od znanych z literatury rozwiązań wykorzystujących sztuczne systemy immunologiczne

Wstęp do tematyki



- Systemy wykrywania intruzów (SWI) jako sposób detekcji nieuprawnionego/niepożądanego dostępu do danych
- Sztuczne systemy immunologiczne (SSI) – algorytmy inspirowane biologią
- Połączenie SWI i SSI
- Zagrożenia

Algorytm negatywnej selekcji



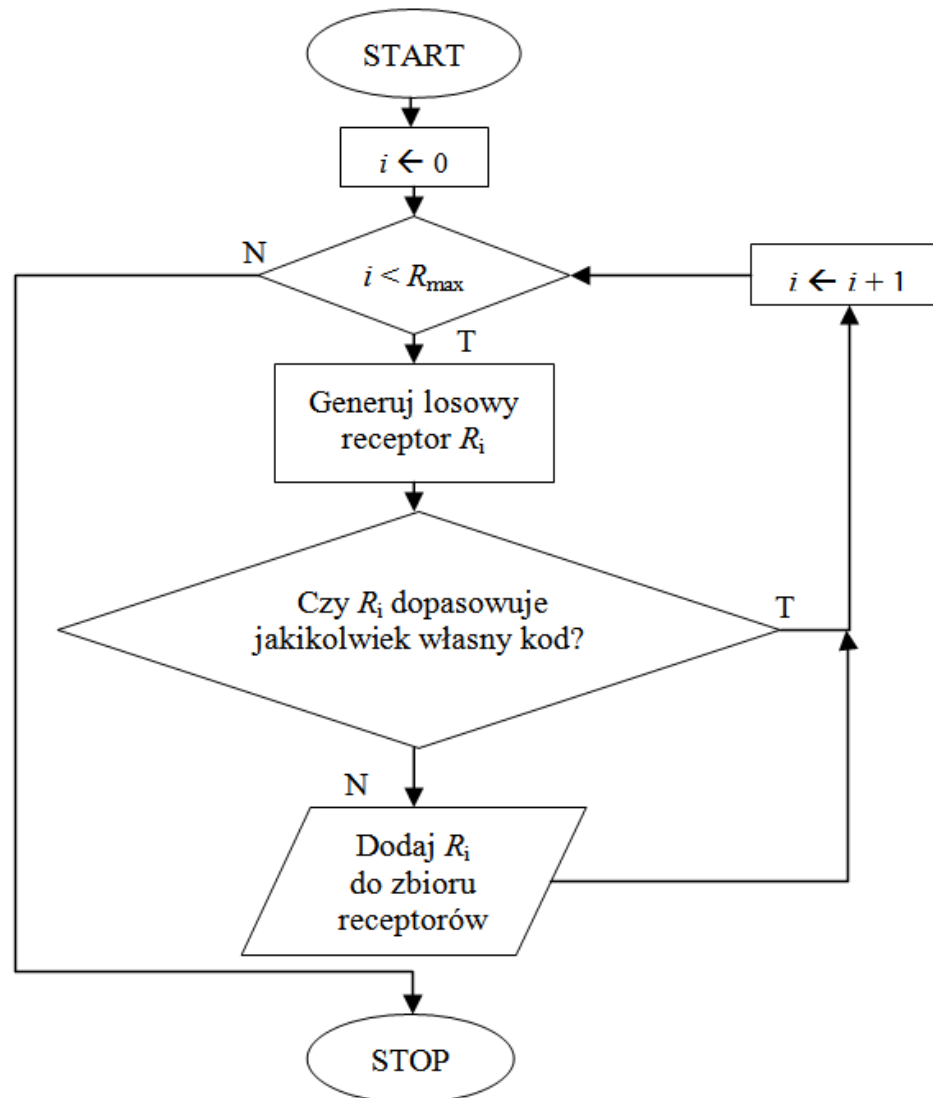
- Algorytm inspirowany odpornością nabytą układu immunologicznego ssaków
- Dyskryminacja ciągów binarnych względem ciągów własnych i obcych
- Proces wyboru ciągów wykrywających obce struktury – receptorów

Metoda losowa



- Generacja receptorów: losowa
- Detekcja anomalii w programie z wykorzystaniem okna

Metoda losowa



Metoda losowa



$$\max(R_{\max}) = 2^l, \quad (1)$$

$$f = \text{ceil}(\text{div}(l, 8)), \quad (2)$$

01001011	11011001	00000101	11101111
4B	D9	05	EF

$$l = 28, f = 4$$

Metoda losowa



$$\max(R_{\max}) = 2^l, \quad (1)$$

$$f = \text{ceil}(\text{div}(l, 8)), \quad (2)$$

11100001	10011010
E1	9A

$$l = 16, f = 2$$

Metoda losowa



$$\max(R_{\max}) = 2^l, \quad (1)$$

$$f = \text{ceil}(\text{div}(l, 8)), \quad (2)$$

Odczytany fragment kodu	10101000	1101100100000101	01101100
Receptor	01001011	1101100100000101	11101111
Odczytany fragment kodu	A8	D9 05	6C
Receptor	4B	D9 05	EF

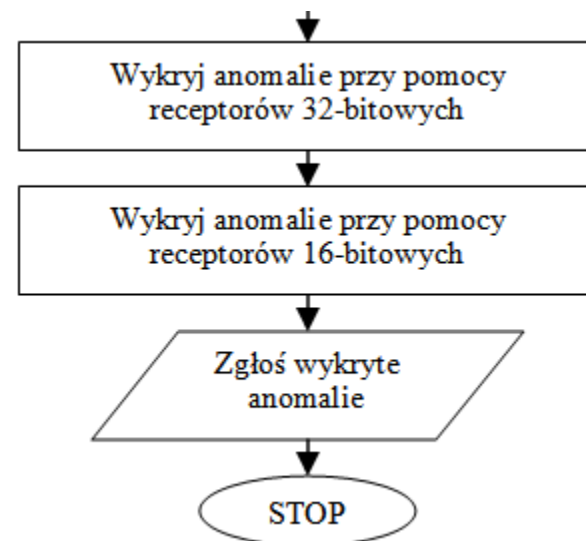
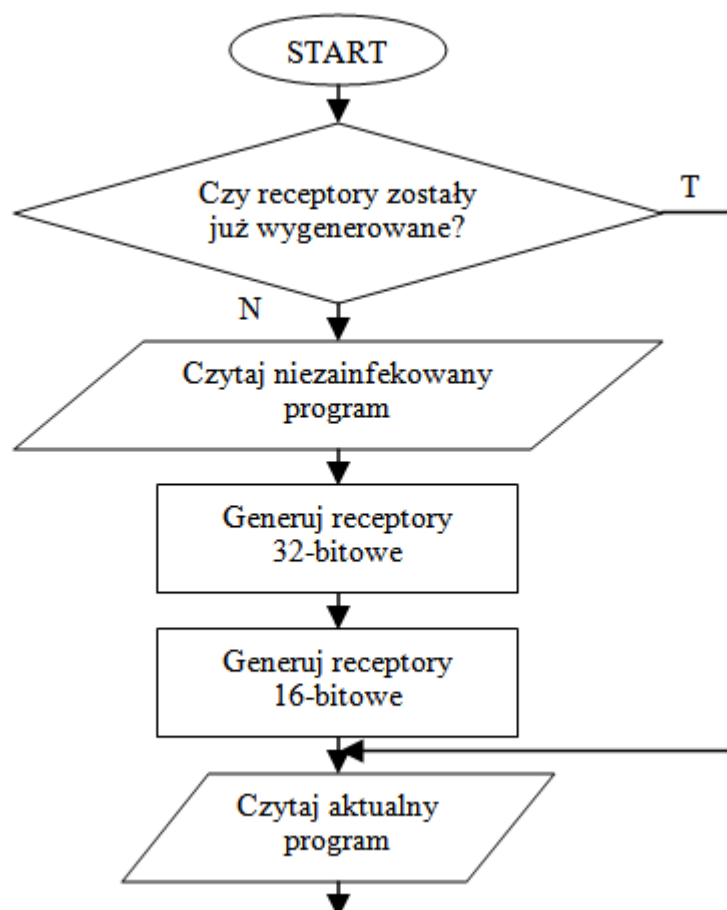
$$l = 32, m = 16, \quad k = 8$$

Metoda szablonów



- Generacja receptorów z wykorzystaniem specjalnych ciągów binarnych zwanych szablonami
- Detekcja anomalii w programie z wykorzystaniem okna

Metoda szablonów



Metoda szablonów



$$S = \{101110, 101101, 101100, 101011, 101010\}.$$

Tabela 1. Tablica szablonów T dla $l = 6$ i $m = 4$

i	m ustalonych bitów	$T[i,1]$	$T[i,2]$	$T[i,3]$
1	0000	0000**	*0000*	**0000
2	0001	0001**	*0001*	**0001
3	0010	0010**	*0010*	**0010
4	0011	0011**	*0011*	**0011
...				
14	1101	1101**	*1101*	**1101
15	1110	1110**	*1110*	**1110
16	1111	1111**	*1111*	**1111

Tabela 2. Tablica T_a dla $l = 6$ i $m = 4$

i	m ustalonych bitów	$T_a[i,1]$	$T_a[i,2]$	$T_a[i,3]$
1	0000	1	1	1
2	0001	1	1	1
3	0010	1	1	1
4	0011	1	1	1
5	0100	1	1	1
6	0101	1	0	1
7	0110	1	0	1
8	0111	1	0	1
9	1000	1	1	1
10	1001	1	1	1
11	1010	0	1	0
12	1011	0	1	0
13	1100	1	1	0
14	1101	1	1	0
15	1110	1	1	0
16	1111	1	1	1

Metoda szablonów



R = {000000, 110001, 000010, 110011, 000100, 110101, 000110, 110111,
001000, 111001, 010000, 010001, 010010, 010011, 010100, 010101,
010110, 010111, 011000, 011001, 111111}

Metoda szablonów



Komórki programu pod adresami X , $X+1$
niezmodyfikowane przez szkodliwe oprogramowanie:

...	Adres X	Adres $X+1$	...
...	01 23 45 67	89 AB CD EF	...

Komórki programu pod adresami X , $X+1$
zmodyfikowane przez szkodliwe oprogramowanie:

...	Adres X	Adres $X+1$	...
...	01 23 45 6A	A9 AB CD EF	...

Badania eksperymentalne



- Algorytmy zaimplementowano w języku C#
- Plik *test.exe*, rozmiar 6584 B
- Badane metody: losowa, szablonów 32-bitowych, szablonów 32+16, sekcji
- Anomalie wprowadzane losowo, o rosnącej długości
- 100 prób dla każdego rozmiaru anomalii

Badania eksperymentalne



Tabela 3. Metoda losowa, $R_{max} = 1000$, $R_n = 927$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	1	1046	10,00%
2	1	1030	20,00%
3	1	1041	40,00%
4	1	1020	20,00%
5	1	1025	30,00%
6	3	1037	20,00%
7	2	1050	20,00%
8	2	1045	20,00%
Średnia	1,5	1036,75	22,50%

$l = 32$, $m = 16$

Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

Badania eksperymentalne



Tabela 4. Metoda losowa, $R_{max} = 5000$, $R_n = 4693$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	1	5396	30,00%
2	2	5321	50,00%
3	2	5313	40,00%
4	3	5311	70,00%
5	4	5298	90,00%
6	4	5264	60,00%
7	4	5348	80,00%
8	5	5211	60,00%
Średnia	3,125	5307,75	60,00%

$l = 32$, $m = 16$

Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

Badania eksperymentalne



Tabela 5. Metoda losowa, $R_{max} = 10000$, $R_n = 9427$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	2	10072	40,00%
2	3	10017	40,00%
3	2	10020	90,00%
4	3	10072	60,00%
5	4	10063	90,00%
6	3	9952	100,00%
7	3	10216	100,00%
8	5	9995	100,00%
Średnia	3,125	10050,875	77,50%

$l = 32$, $m = 16$

Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

Badania eksperymentalne



Tabela 6. Metoda S-32, $R_n = 32$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	4	12	30,00%
2	3	22	50,00%
3	8	8	20,00%
4	2	13	30,00%
5	14	21	50,00%
6	6	26	60,00%
7	6	26	60,00%
8	8	35	80,00%
Średnia	6,375	20,375	47,50%

$l = 32, m = 16$

Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

Badania eksperymentalne



Tabela 7. Metoda RMK, $R_n = 46$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	4	12	30,00%
2	3	22	50,00%
3	11	8	40,00%
4	11	13	70,00%
5	14	21	50,00%
6	6	26	60,00%
7	6	26	60,00%
8	11	8	100,00%
Średnia	8,25	17	57,50%

$l = 32, m = 16$

$l = 16, m = 8$

Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

Badania eksperymentalne



Tabela 8. Metoda sekcji, $R_n = 780$

<i>Rozmiar anomalii [B]</i>	<i>NRA</i>	<i>CWavg</i>	<i>W</i>
1	5	10	60,00%
2	7	17	100,00%
3	9	16	90,00%
4	11	20	100,00%
5	13	20	100,00%
6	15	17	100,00%
7	24	16	100,00%
8	20	19	100,00%
Średnia	13	16,875	93,75%

$l = 32, m = 16$

4 sekcje

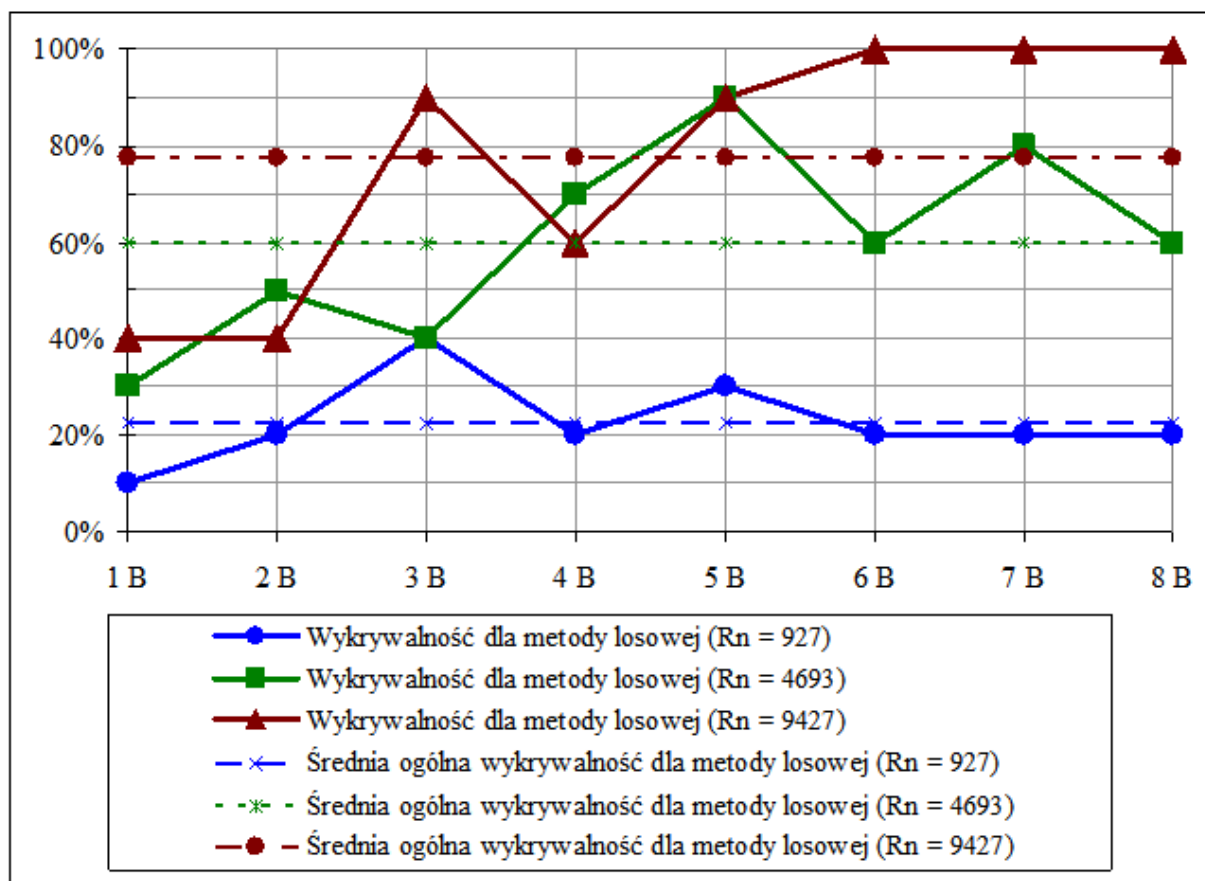
Legenda:

NRA – największa liczba
aktywowanych receptorów
dla 100 prób

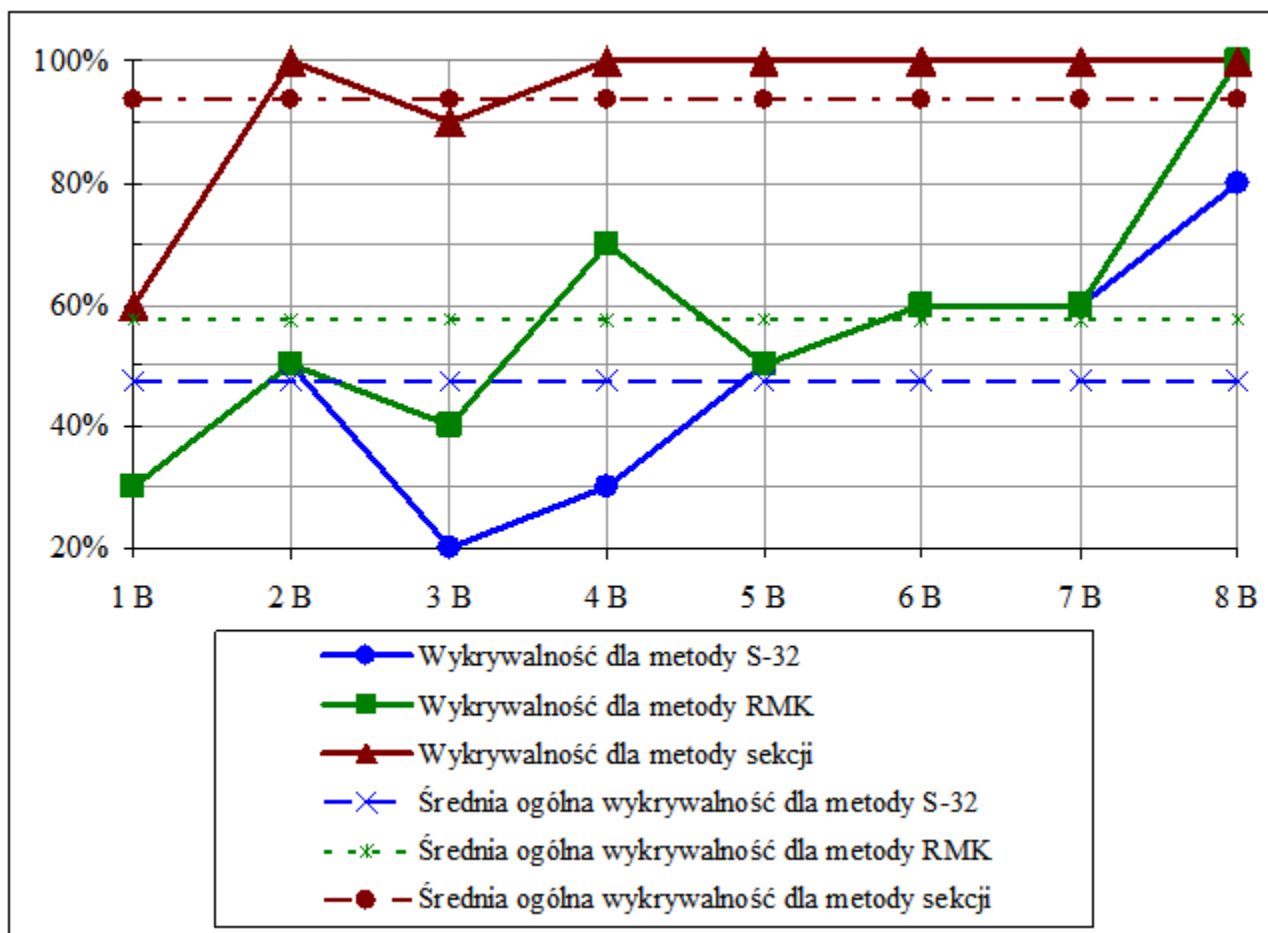
CWavg – średni czas wykrycia
anomalii [ms]

W – wykrywalność anomalii [%]

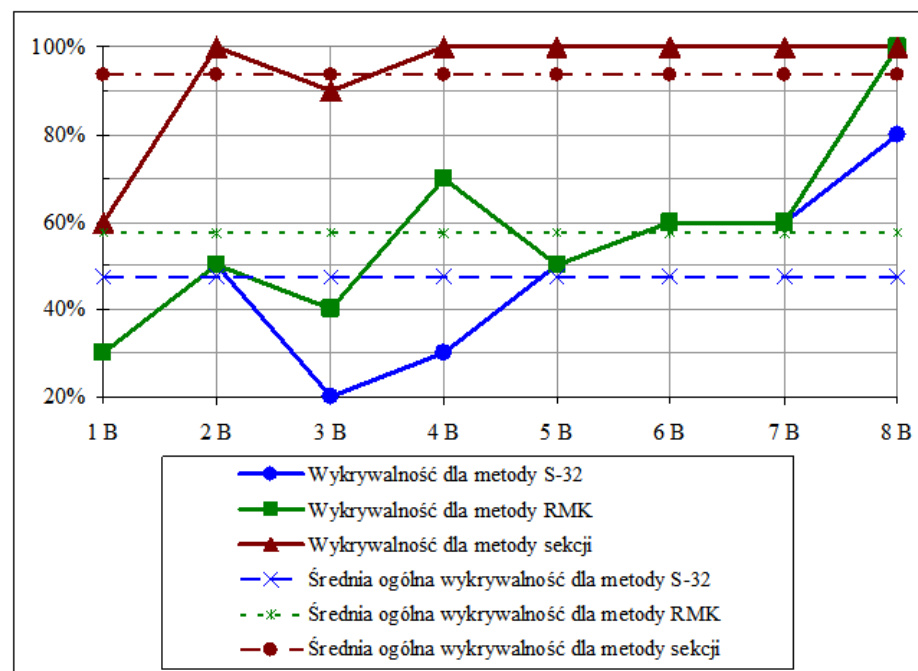
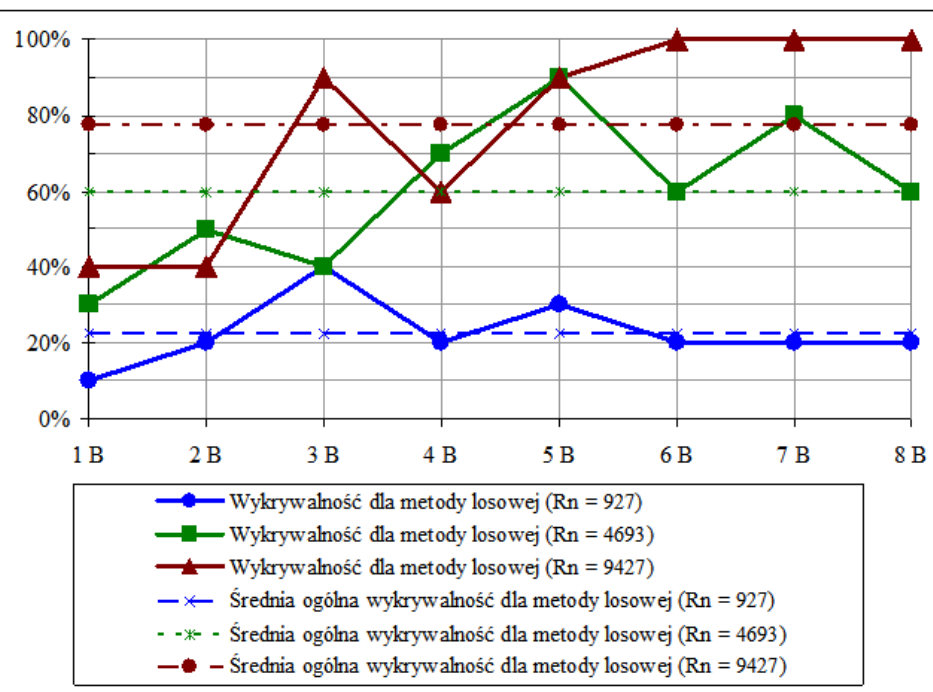
Badania eksperymentalne



Badania eksperymentalne



Badania eksperymentalne



Wnioski



- Mniejsze anomalie są trudniejsze w wykryciu
- Generacja losowa jest znacznie szybsza, ale wykrywanie jest wolniejsze i mniej skuteczne
- Generacja szablonowa jest powolna, ale wykrywanie jest znacznie szybsze i bardziej skuteczne

Podsumowanie



- Zastosowanie sztucznych systemów immunologicznych pozwala na znajdowanie modyfikacji w programach w systemie operacyjnym
- Kolejny krok:
 - Badania ze zmiennymi: liczbą bitów receptora l i progiem aktywacji m

Publikacje



1. Marciniak J., Wawryn K., Widuliński P., *An Artificial Immune Negative Selection Algorithm to Control Water Temperature in the Outlet of the Chamber*, w: *2018 International Conference on Signals and Electronic Systems, ICSES 2018 – Proceedings*, 2018, s. 236-241. (Materiały konferencyjne indeksowane w Web of Science i Scopus)
2. Wawryn K., Widuliński P., *A Human Immunity Inspired Algorithm to Detect Infections in a Computer Program*, *MIXDES 2019*. (Artykuł zgłoszony)
3. Widuliński P., Wawryn K., *Detekcja anomalii w plikach za pomocą wybranych algorytmów inspirowanych systemami immunologicznymi*, „ZN WEil”, Koszalin 2019.

Dziękuję za uwagę