Mixed-UCT: Zastosowanie metod symulacyjnych do poszukiwania równowagi Stackelberga w grach wielokrokowych

Jan Karwowski

Zakład Sztucznej Inteligencji i Metod Obliczeniowych
Wydział Matematyki i Nauk Informacyjnych PW

5 IV 2017

Plan

1. Security Games
2. Elementy teorii gier
3. Mixed-UCT
4. Wyniki
5. Problemy

Security Games

- Obrońcy i atakujący
- Asymetria graczy
- Różne rodzaje przestrzeni
- Zwykle obrońca ma większą przestrzeń decyzyjną
- Brak jednolitej definicji
- Zazwyczaj równowaga Stackelberga

Cele

- Metoda dobrze działająca z grami wielokrokowymi
- Metoda dobrze skalująca się z rozmiarem gry
- Metoda łatwa do dostosowania do różnych modeli gry

Postać normalna gry o sumie (nie)zerowej

Macierz wypłat gracza P1

P1/P2	a	b	c	d
A	6	4	-2	1
B	0	0	2	3
C	-3	-2	-1	-1

- $U_{P1}(A, b) = 4$
- $U_{P2}(A, b) = -4$

Macierz wypłat gracza P2

P1/P2	a	b	c	d
A	-6	**3**	2	-1
B	0	0	-2	-3
C	-3	2	0	1

- $U_{P2}(A, b) = 3$

Postać normalna gry o sumie (nie)zerowej

Macierz wypłat gracza P1

P1/P2	a	b	c	d
A	6	4	-2	1
B	0	0	2	3
C	-3	-2	-1	-1

- $U_{P1}(A, b) = 4$
- $U_{P2}(A, b) = -4$

Macierz wypłat gracza P2

P1/P2	a	b	c	d
A	-6	**3**	2	-1
B	0	0	-2	-3
C	-3	2	0	1

-
- $U_{P2}(A, b) = 3$

Strategia mieszana

E	H	T
H	1	−1
T	−1	1

O	H	T
H	−1	1
T	1	−1

Równowaga Stackelberga

- Asymetryczni gracze: *Leader*, *Follower*
- *Follower* zna **strategię** *Leadera* w momencie wyboru swojej strategii. (Ale niekoniecznie wykonuje **ruch** po leaderze).
- *Follower* rozstrzyga remisy (swojej wypłaty) na korzyść *Leadera*

Gra

	A1	A2
D1	-15	3
D2	3	-15

	A1	A2
D1	**30**	0
D2	0	**20**

Equilibrium

Obrońca		
Pr.		U_D
0.4	D1	-15
0.6	D2	3

Atakujący			
Pr.		U_A	U_D
1	A1	12	-4.2
0	A2	12	-7.8

Zaburzone equilirbium

Obrońca		
Pr.		U_D
0.39	D1	3
0.61	D2	-15

Atakujący			
Pr.		R_A	U_D
0	A1	11.7	-4
1	A2	12.2	-8

Równowaga Stackelberga

- Asymetryczni gracze: *Leader*, *Follower*
- *Follower* zna **strategię** *Leadera* w momencie wyboru swojej strategii. (Ale niekoniecznie wykonuje **ruch** po leaderze).
- *Follower* rozstrzyga remisy (swojej wypłaty) na korzyść *Leadera*

Gra

	A1	A2
D1	-15	3
D2	3	-15

	A1	A2
D1	**30**	0
D2	0	**20**

Dwupoziomowy problem optymalizacyjny

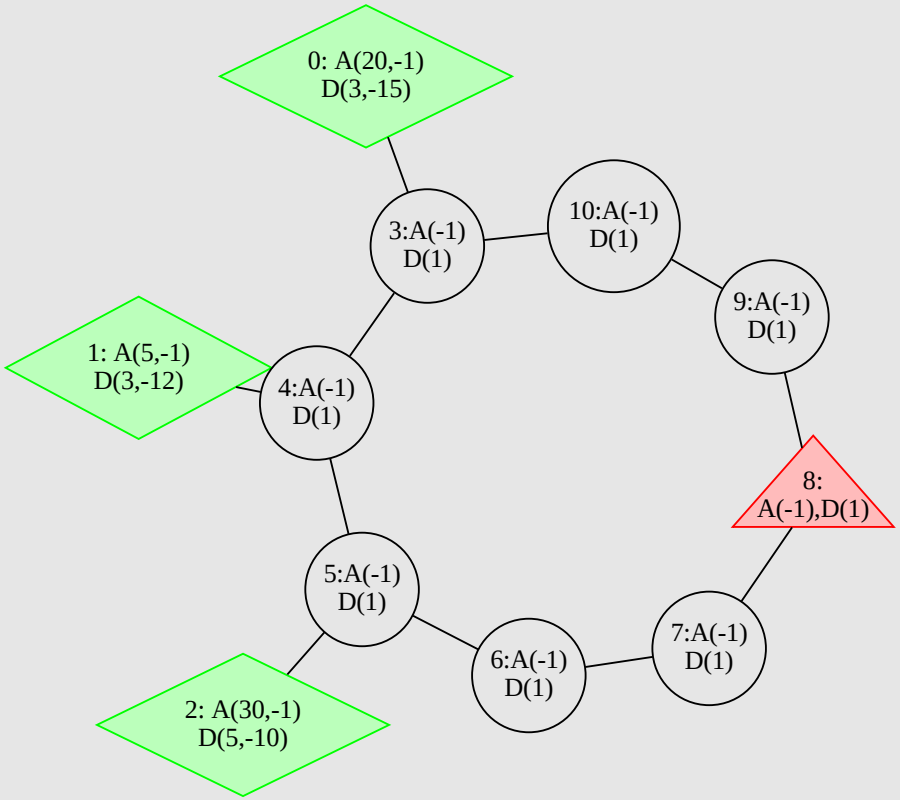
$$\arg\max_{\pi_d \in \Pi_d} U_d(\pi_d, R_a(\pi_d))$$

$R_a(\pi_d) = \arg\max_{\pi_a \in \Pi_a} U_a(\pi_d, \pi_a)$ – funkcja schodkowa

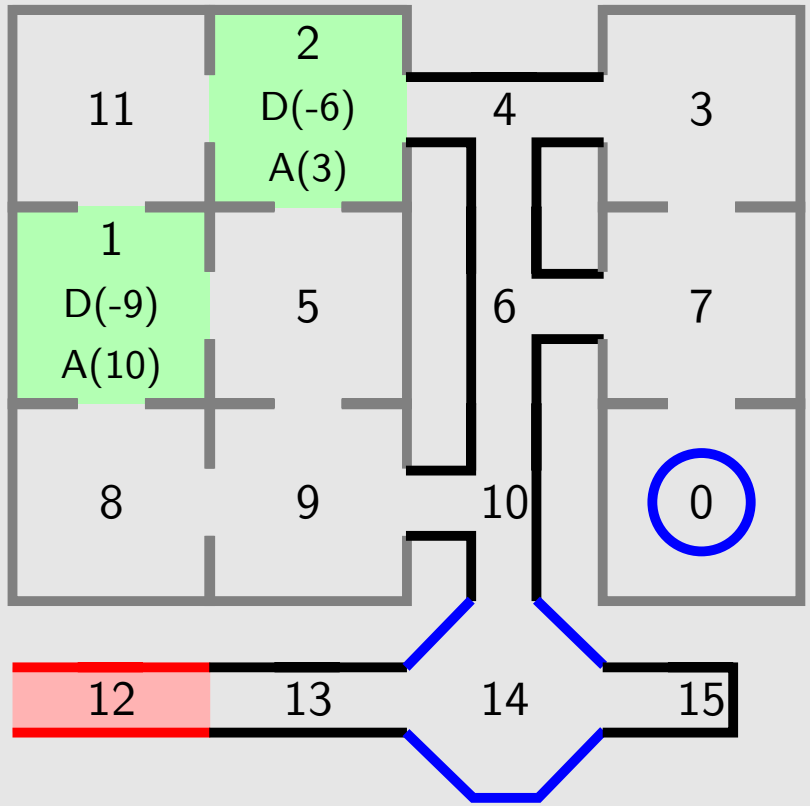
Gra na grafie i generator budynków

- Każdy z graczy operuje jednostkami chodzącymi po grafie
- Obaj wykonują ruch jednocześnie
- Nie widzą siebie nawzajem

Przykładowa gra

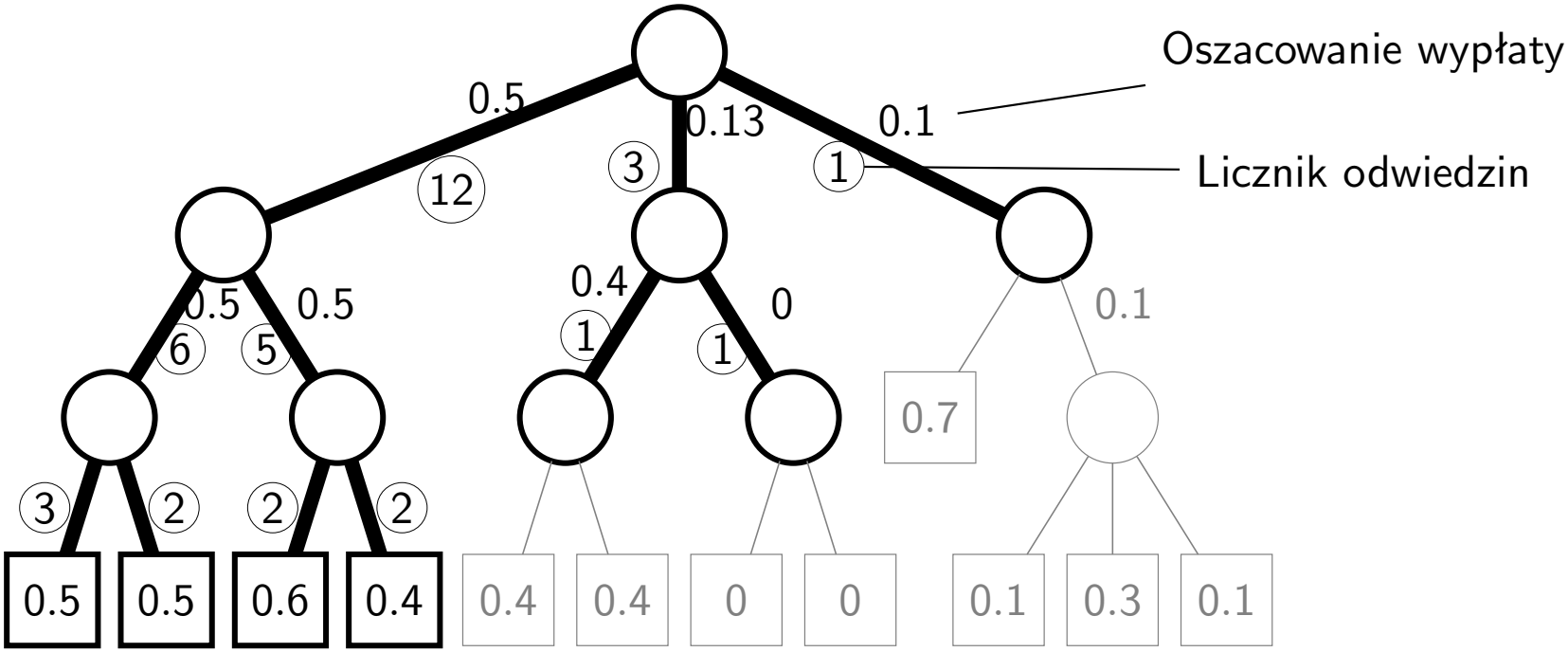


Generator budynków



MCTS i UCT

- Próbkowanie losowe
- Wykorzystanie tylko symulacji gry



Mixed-UCT I

Cele

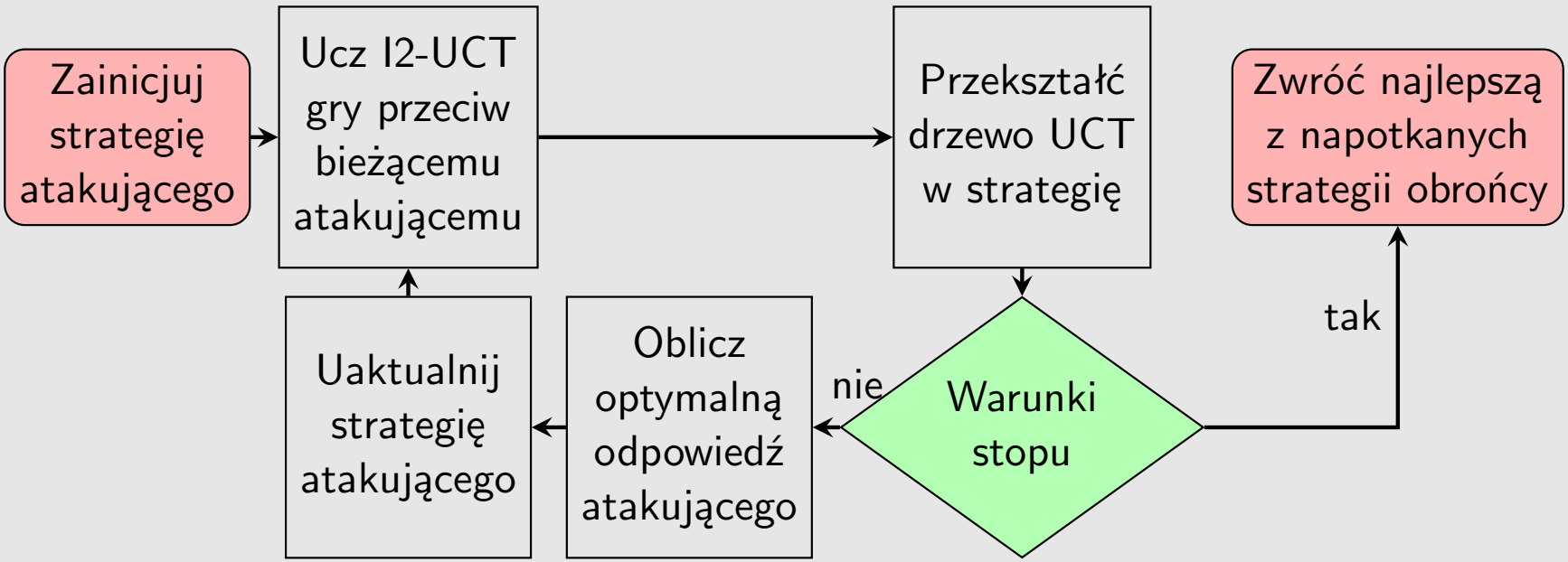
- Znaleźć dobre przybliżenie strategi leadera w stanie równowagowym (w sensie wypłaty)
- Skrócić czas obliczeń w stosunku do metod dokładnych

Problemy

- UCT wymaga gry z pełną informacją
- UCT podaje w wyniku pojedynczy ruch
- Equilibrium Stackelberga zaskakuje

Mixed-UCT II

Zarys



- I2UCT
- Jak wyliczać strategię obrońcy?
- Jak wyliczać odpowiedź atakującego?

Stan równowagowy Stackelberga (raz jeszcze)

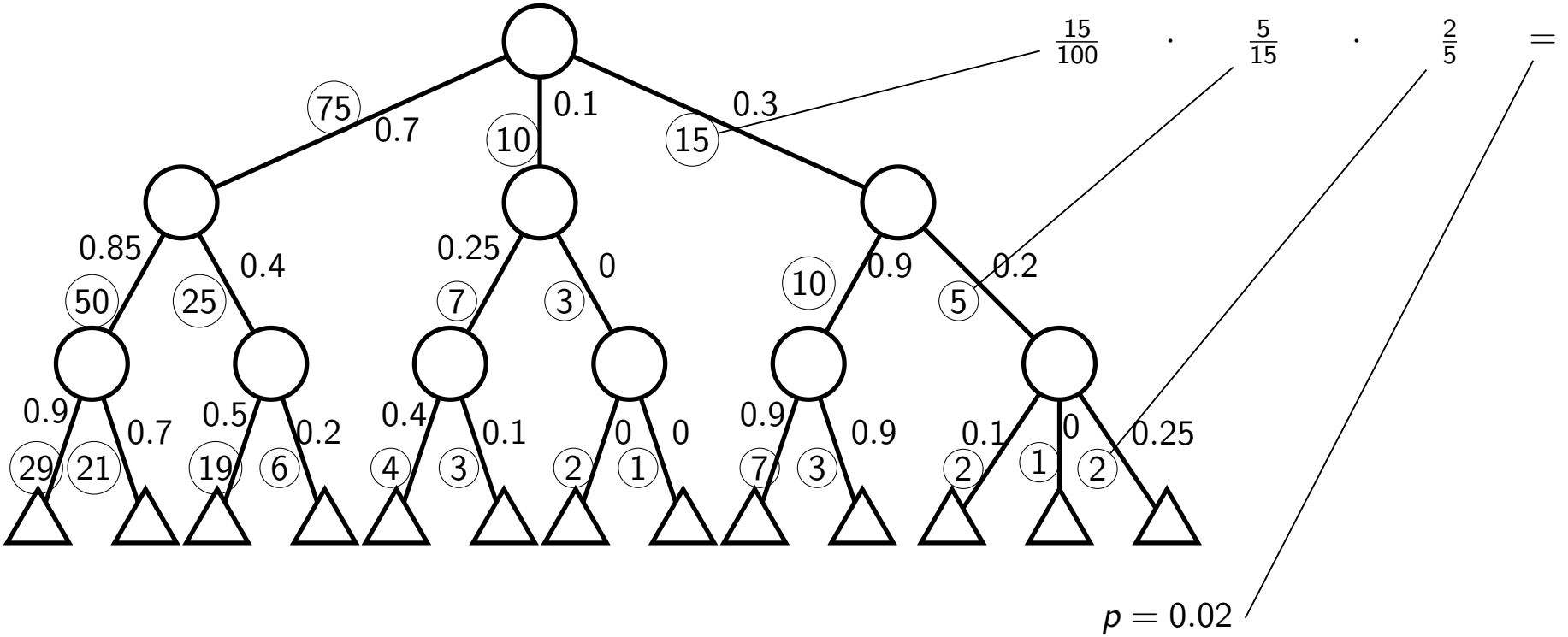
Defender

Pr.	Move	E(R_D)
0.4	D1	-15
0.6	D2	3

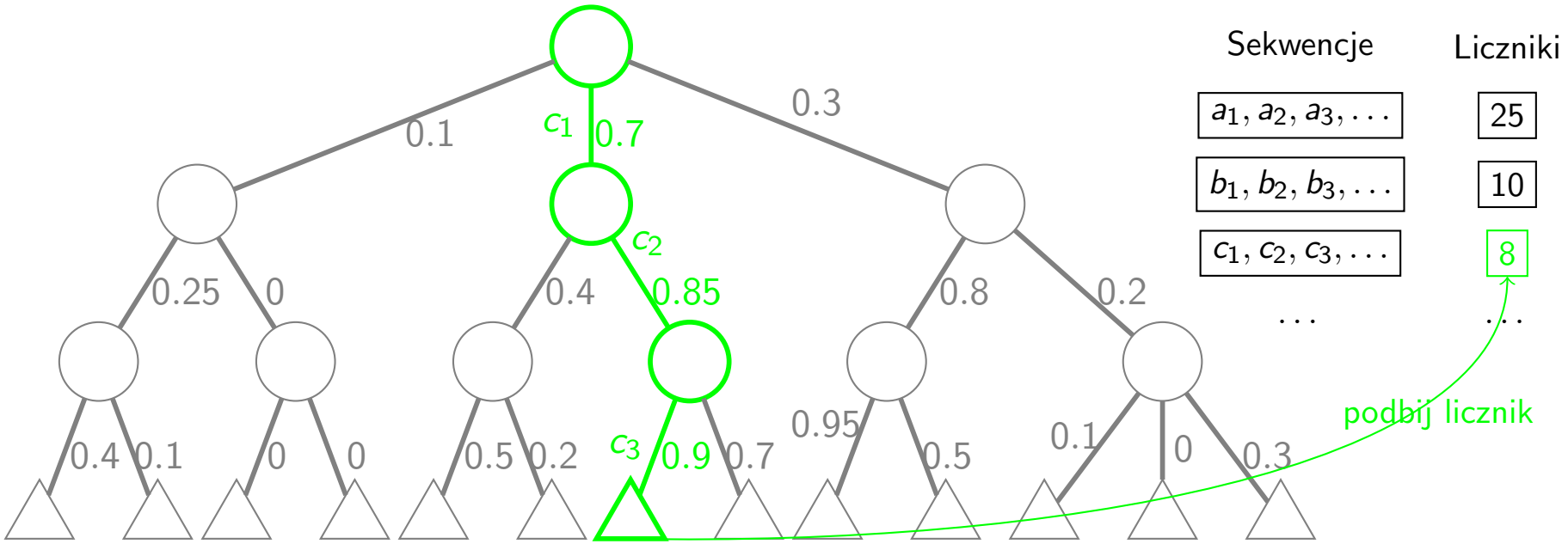
Attacker

Pr.	Move	E(R_A)	E(R_D)
1	A1	12	-4.2
0	A2	12	-7.8

Uzyskiwanie strategii mieszanej v1: FullTree



Uzyskiwanie strategii mieszanej v2: BestPathFreq



Uśrednianie atakującego

Strategia atakującego jest średnią najlepszych odpowiedzi z h poprzednich iteracji.

game	1			10			100			1000			MILP		U
	R	t	sc	R	t	sc	R	t	sc	R	t	sc	R	t	R
1	-1.57	1448	0.81	0.34	1436	0.98	0.48	1492	0.99	0.47	1611	0.99	0.54	28052	-10.46
2	0.03	1049	0.99	0.05	1388	1	0.06	1113	1	0.07	1351	1	0.08	106	-7.21
3	-4.32	2319	0.99	-4.34	1618	0.99	-4.28	1786	1	-4.28	1135	1	-4.27	42946	-13.97
3a	-4.5	2003	1	-4.53	1452	1	-4.51	1230	1	-4.5	1487	1	-4.5	31926	-13.97
3b	2.44	1838	0.96	2.57	2020	1	2.57	1382	1	2.58	1639	1	2.58	227	-0.87
3c	-1.52	1777	0.94	-1.18	1724	0.99	-1.07	1358	1	-1.06	1169	1	-1.06	41624	-9.29
3d	0.84	2129	0.99	0.69	1595	0.96	0.9	1467	1	0.9	1218	1	0.9	37870	-4.61
4	-4.87	1777	1	-4.87	1378	1	-4.87	1995	1	-4.87	1310	1	-4.87	5312	-13.91
4a	-6	1681	1	-6	1638	1	-6	1637	1	-6	1114	1	-6	5949	-13.84
4b	0.79	1788	1	0.79	1266	1	0.79	1639	1	0.79	1061	1	0.79	5546	-0.81
4c	-2.85	1783	1	-2.85	1498	1	-2.85	1494	1	-2.85	1203	1	-2.85	5928	-9.23
4d	0.17	1948	1	0.17	1455	1	0.17	1391	1	0.16	1165	1	0.17	5155	-4.55

MILP: metoda do porównań

MILP

$$
\begin{array}{rl}
 & \max_{q,z,a} \sum_{i \in X} \sum_{j \in Q} R_{ij} z_{ij} \\
\text{s.t.} & \sum_{i \in X} \sum_{j \in Q} z_{ij} = 1 \\
(\forall_{i \in X}) & \sum_{j \in Q} z_{i,j} \leqslant 1 \\
(\forall_{j \in Q}) & q_j \leqslant \sum_{i \in X} z_{ij} \leqslant 1 \\
 & \sum_{j \in Q} q_j = 1 \\
(\forall_{j \in Q}) & 0 \leqslant (a - \sum_{i \in X} C_{ij} (\sum_{h \in Q} z_{ih})) \\
(\forall_{j \in Q}) & (a - \sum_{i \in X} C_{ij} (\sum_{h \in Q} z_{ih})) \leqslant (1 - q_j) M
\end{array}
$$

$$
\begin{array}{l}
z_{ij} \in [0,1] \\
q_j \in \{0,1\} \\
a \in \mathbb{R}
\end{array}
$$

Złożoność

$|X|$ – liczba strategii (sekwencji) obrońcy, $|Q|$ – liczba strategii (sekwencji) atakującego

- $O(|X| \cdot |Q|)$ zmiennych, w tym $|Q|$ binarnych.
- $O(|X| + |Q|)$ ograniczeń

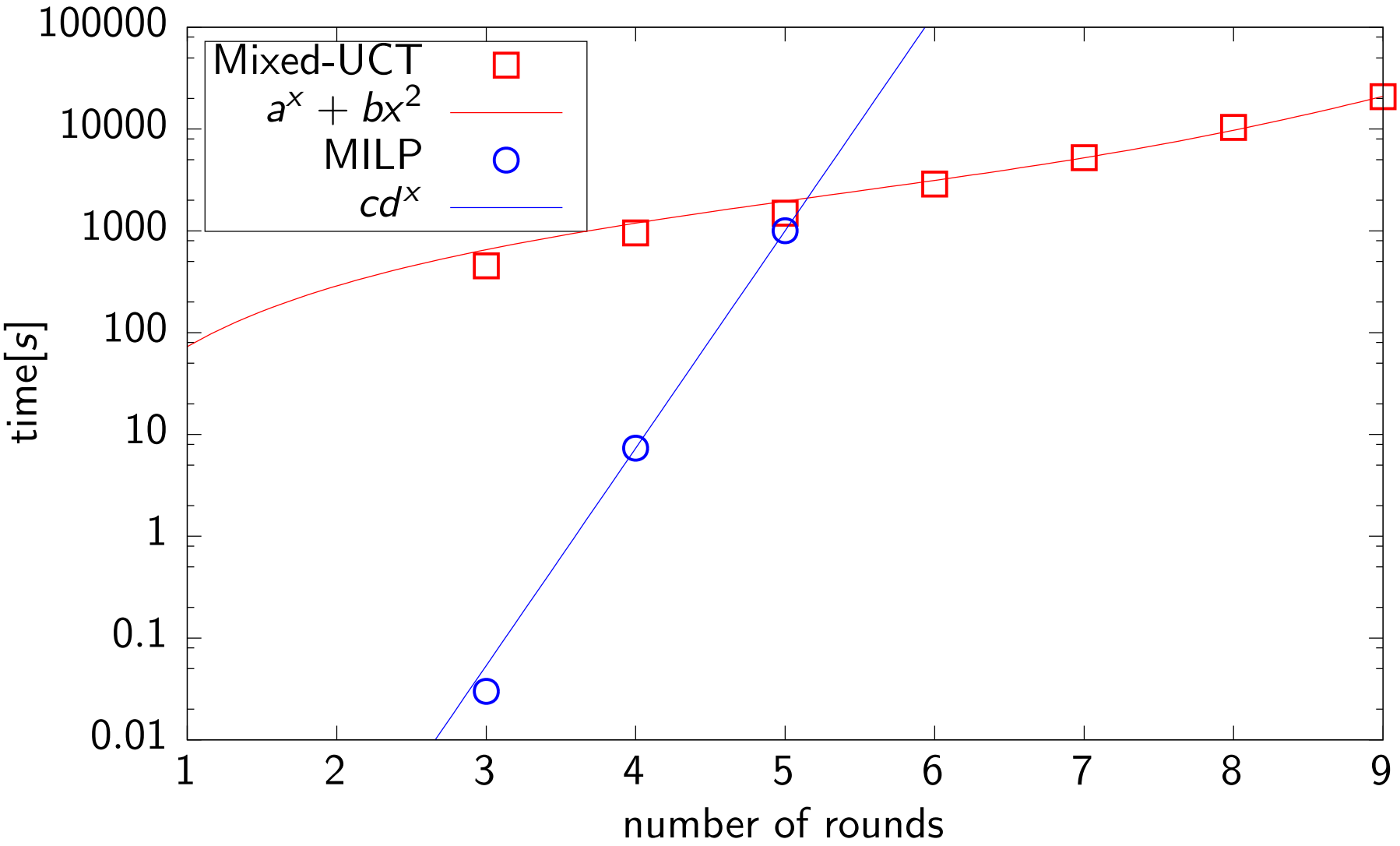
Solvery:

- Gurobi 6.5
- **SCIP 3.1**

FullTree

Game	Mixed-UCT Payoff mean	Mixed-UCT Payoff sd	Mixed-UCT Time [s]	Uniform Payoff	MILP Payoff	MILP Time [s]	Score
1	0.03	0	1437	-2.79	0.03	2350	100%
17	0.26	0.02	1705	-7.95	0.37	409	99%
2	0	0	5258	-15.08	0	593	100%
20	1.28	0.01	2199	-1.28	1.29	2705	100%
24	1.13	0.17	2431	0.08	1.55	692	71%
3	0.46	0.05	2111	-8.55	0.5	1684	100%
35	0.47	0.03	1117	-0.89	0.5	34	98%
39	0.04	0.01	1376	-17.06	0.09	204	100%
41	-2.89	0.01	1661	-5.07	-2.85	3212	98%
42	0	0	2149	-16.53	0	1861	100%
43	0	0	1236	-17.4	0	361	100%
56	1.13	0.06	4410	0.52	1.6	1107	56%
59	0.55	0.01	1632	-7.39	0.62	631	99%
64	0.08	0.01	1793	-11.81	0.16	541	99%
7	-0.79	0.05	1744	-9.7	-0.76	2464	100%
74	1.39	0.02	3379	-0.09	1.47	88	95%
78	0.34	0.07	1379	-9.28	0.5	2463	98%
82	0	0	1819	-10.66	0	899	100%
85	-0.98	0.02	3833	-1.79	-0.89	1531	90%
87	0.77	0	3395	-1.66	0.8	8	99%
89	0.02	0.05	1151	-13.78	0.21	172	99%
91	-5.65	0.09	2096	-5.97	-5.62	118	91%
96	0	0	1409	-10.15	0.19	82	98%
mean	-0.1	0.03	2178	-7.8	-0.02	1019	96%

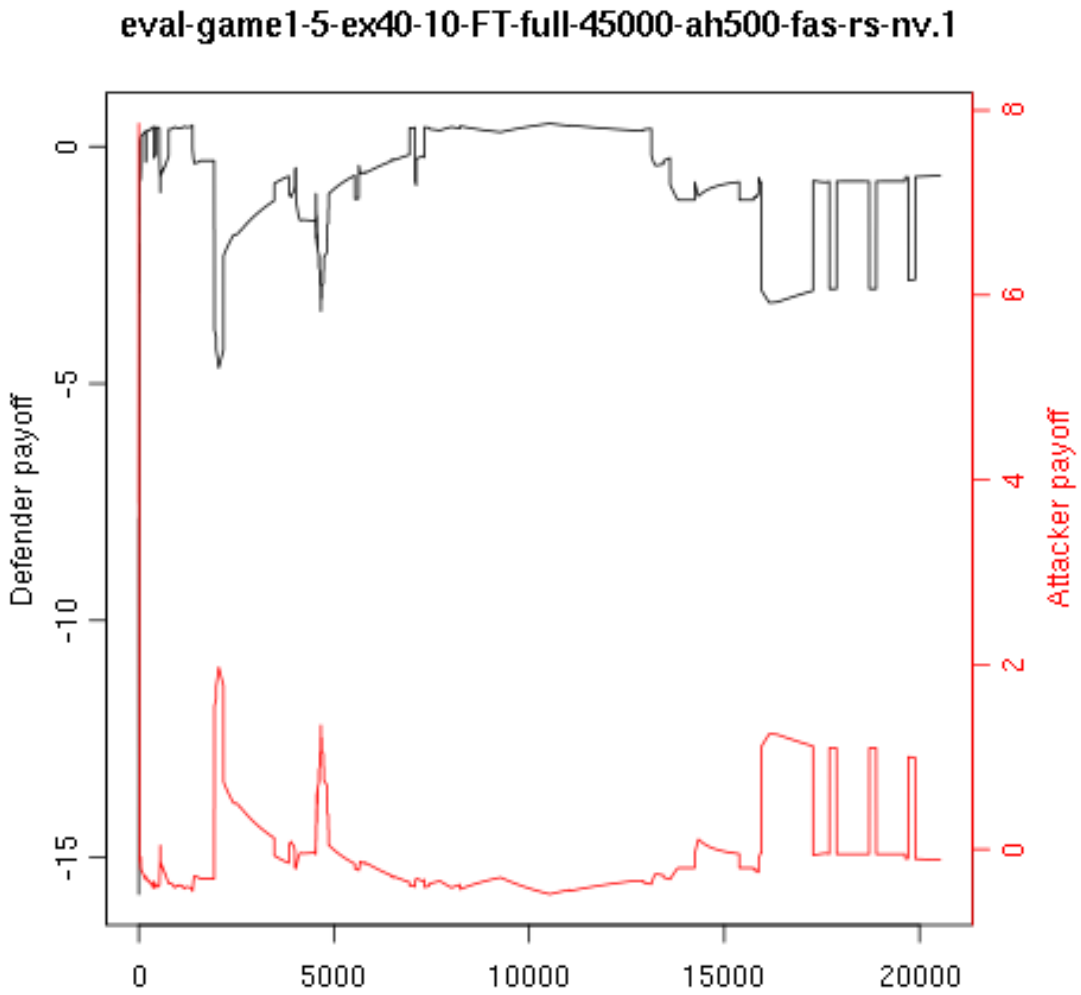
Czas obliczeń



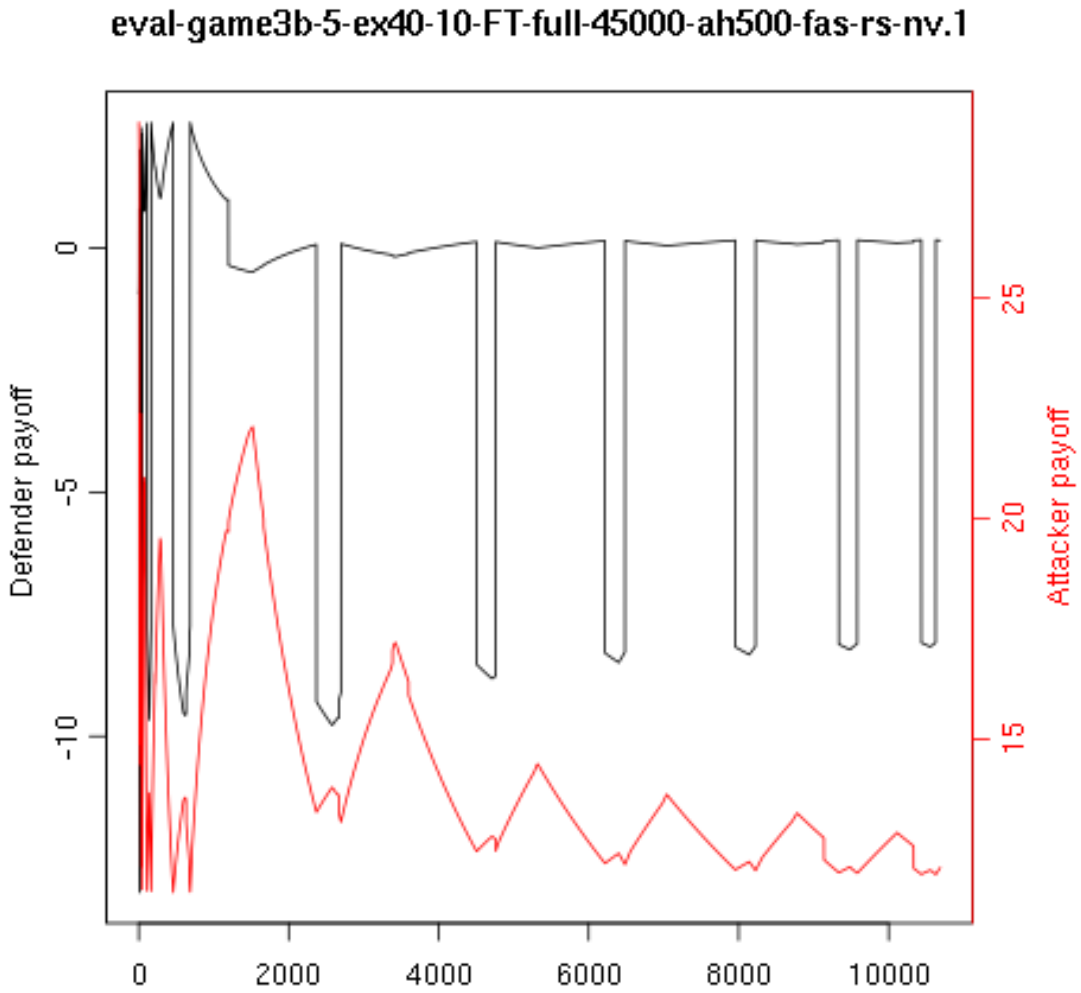
BestPath

Game	Mixed-UCT Payoff		Mixed-UCT Time	Uniform Payoff	MILP Payoff	Score
	mean	sd				
1	-0.2	0.08	1150.77	-2.79	0.03	0.92
17	0.33	0	1060.36	-7.95	0.01	1.04
2	0	0	1723.67	-15.08	0	1
20	1	0.06	1050.01	-1.28	1.29	0.89
24	0.92	0	1873.81	0.08	1.55	0.57
3	0.5	0	1002.34	-8.55	0.5	1
31	0	0	1129.1	-7.02	0	1
35	0.5	0	1582.27	-0.89	0.5	1
39	0.03	0	2100.06	-17.06	0.09	1
41	-2.88	0	2239.92	-5.07	-2.85	0.99
42	0	0	1440.14	-16.53	0	1
43	0	0	1447.3	-17.4	0	1
56	1.01	0.02	1676.38	0.52	1.6	0.45
59	0.6	0.01	1320.73	-7.39	0.62	1
64	0.12	0.01	1698.35	-11.81	0.16	1
7	-1	0.08	1052.87	-9.7	-0.77	0.97
78	0.31	0.01	1207.51	-9.28	0.5	0.98
82	0	0	2497.62	-10.66	0	1
85	-0.95	0.02	1456.07	-1.79	-0.89	0.93
87	0.7	0.02	1201.91	-1.66	0.8	0.96
89	0.21	0	1291.88	-13.78	0.21	1
91	-5.62	0	1851.86	-5.97	-5.62	1
96	0.13	0.04	1876.05	-10.15	0.19	0.99
98	0	0	1101.88	-13.02	-0.3	1.02
mean	-0.14	0.01	1485.22	-7.77	-0.1	0.94

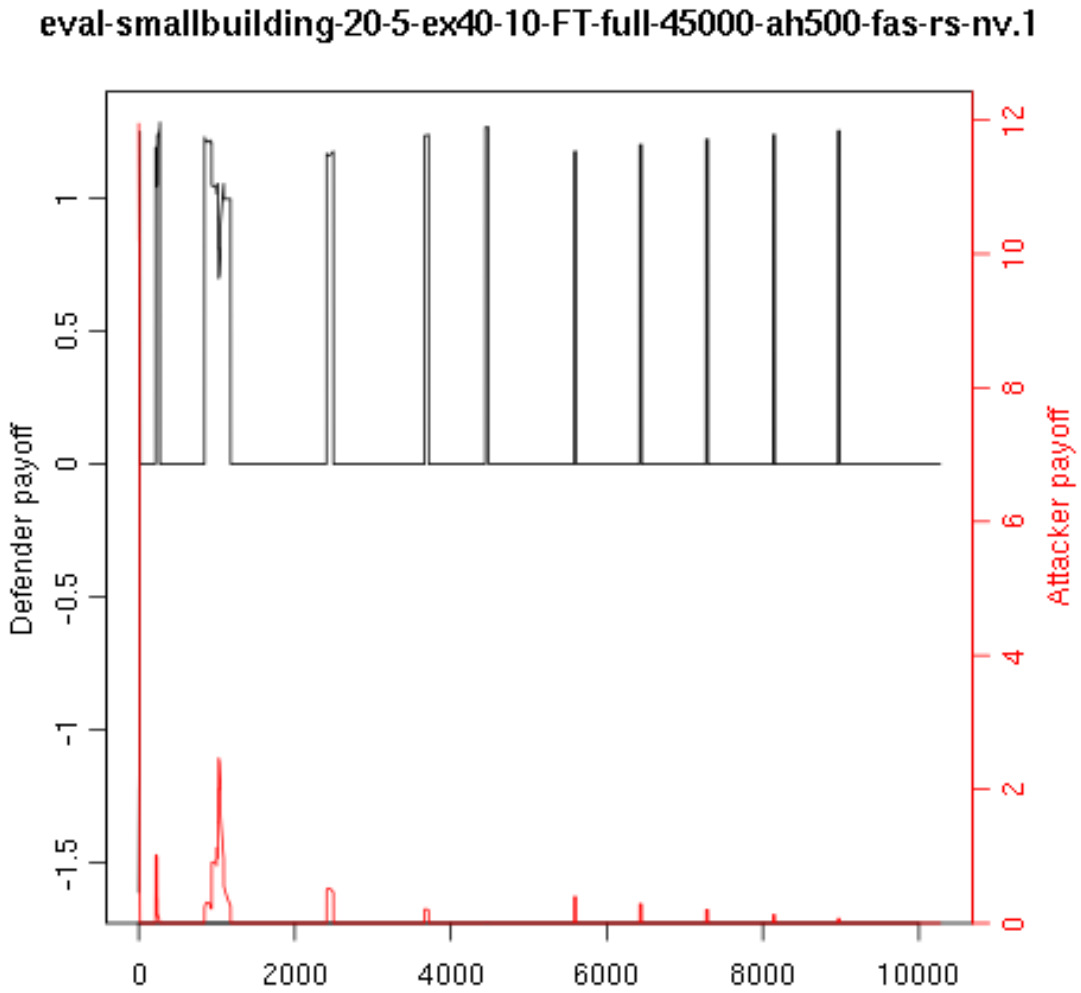
Przebiegi I


eval-game1-5-ex40-10-FT-full-45000-ah500-fas-rs-nv.1
Defender payoff
Attacker payoff
0
-5
-10
-15
8
6
4
2
0
0
5000
10000
15000
20000

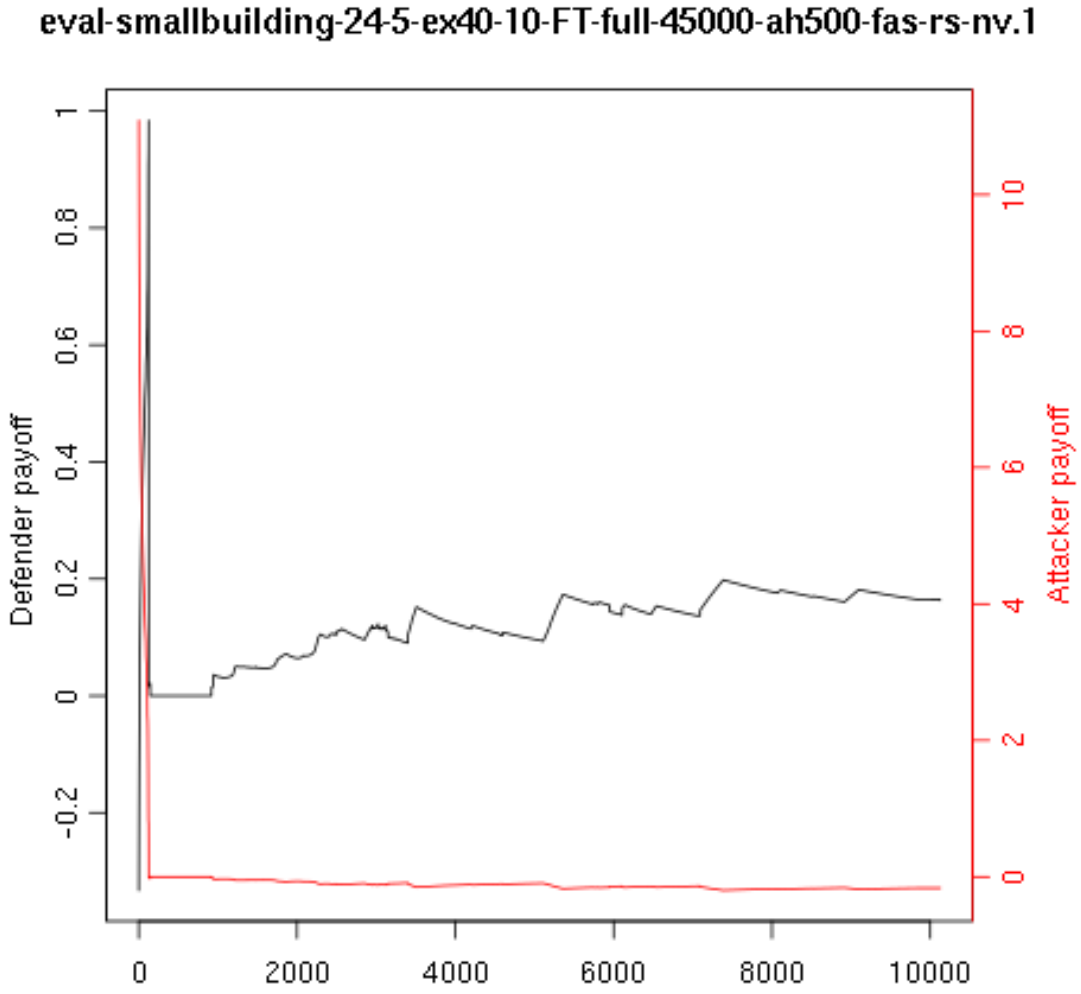
Przebiegi II



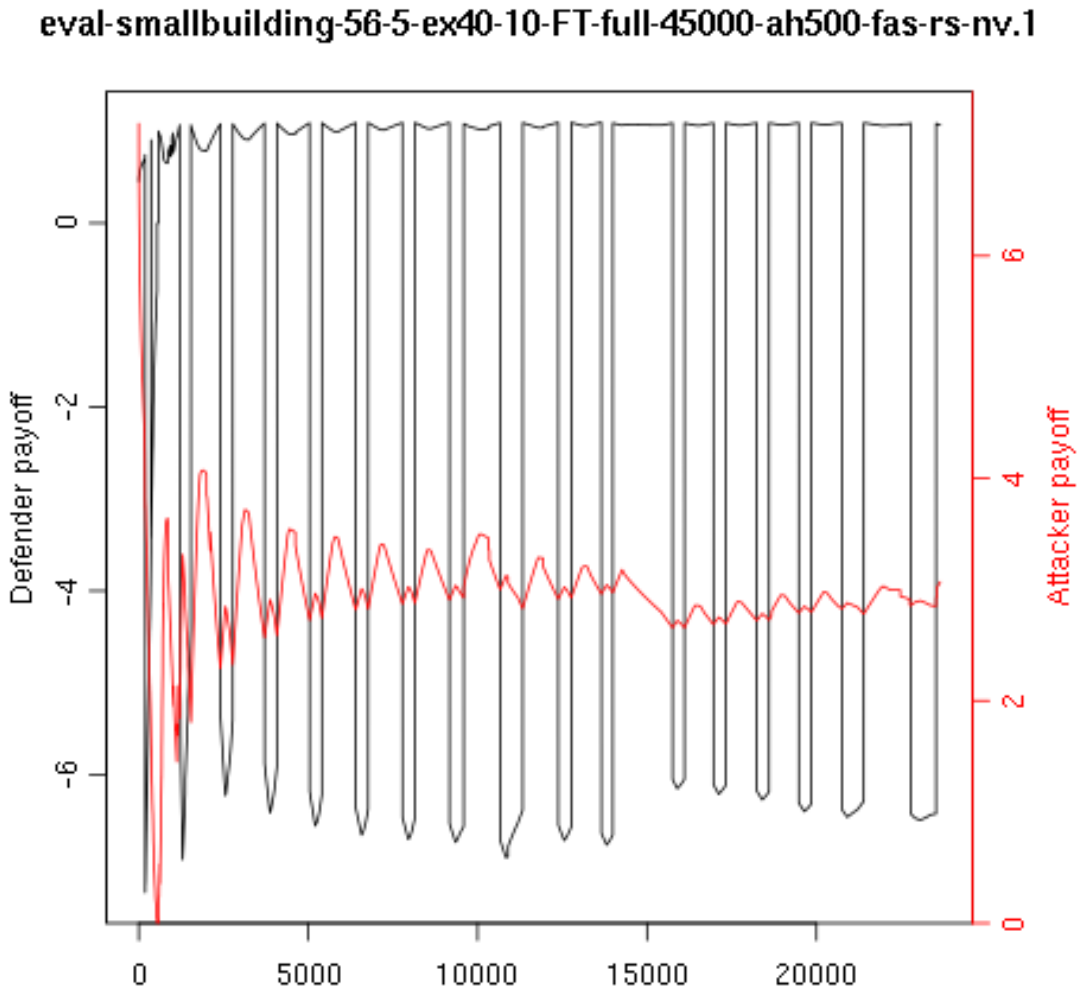
Przebiegi III



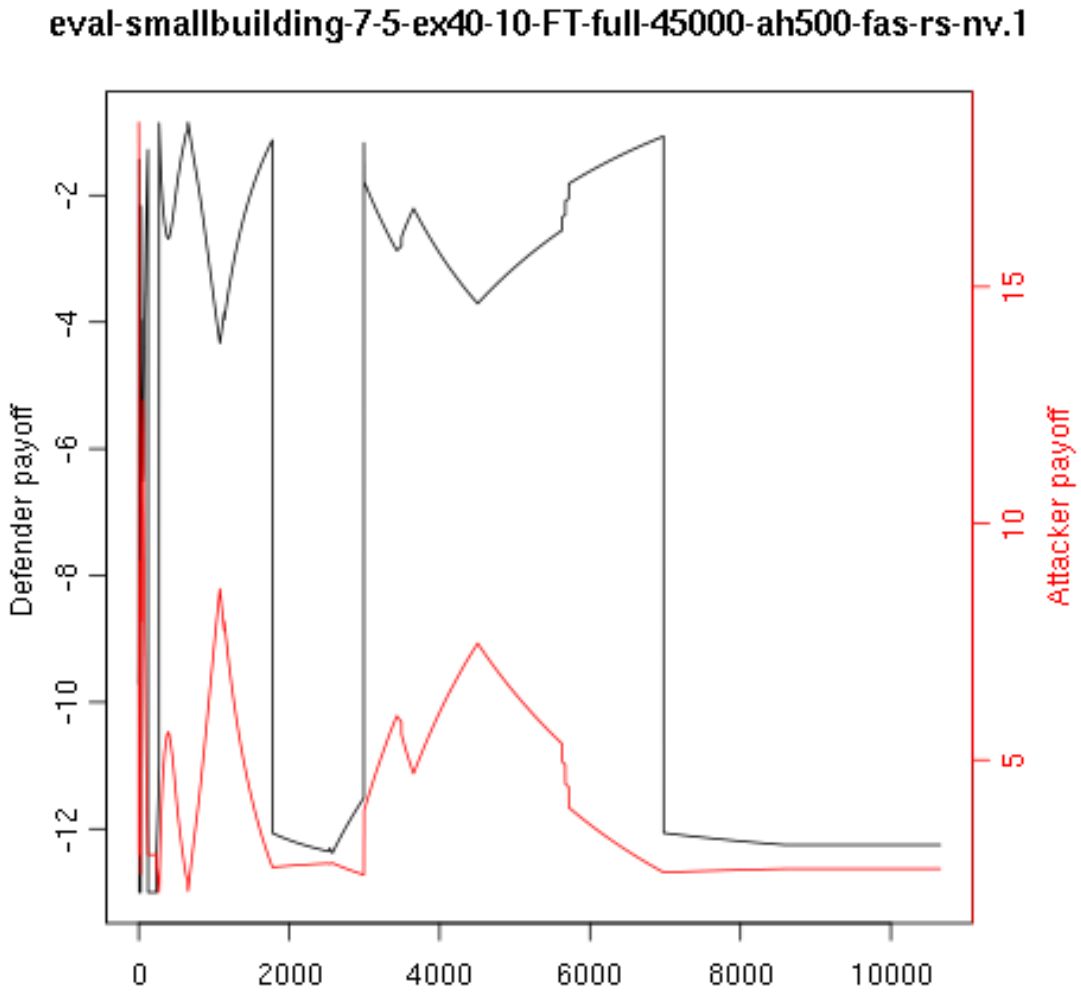
Przebiegi IV



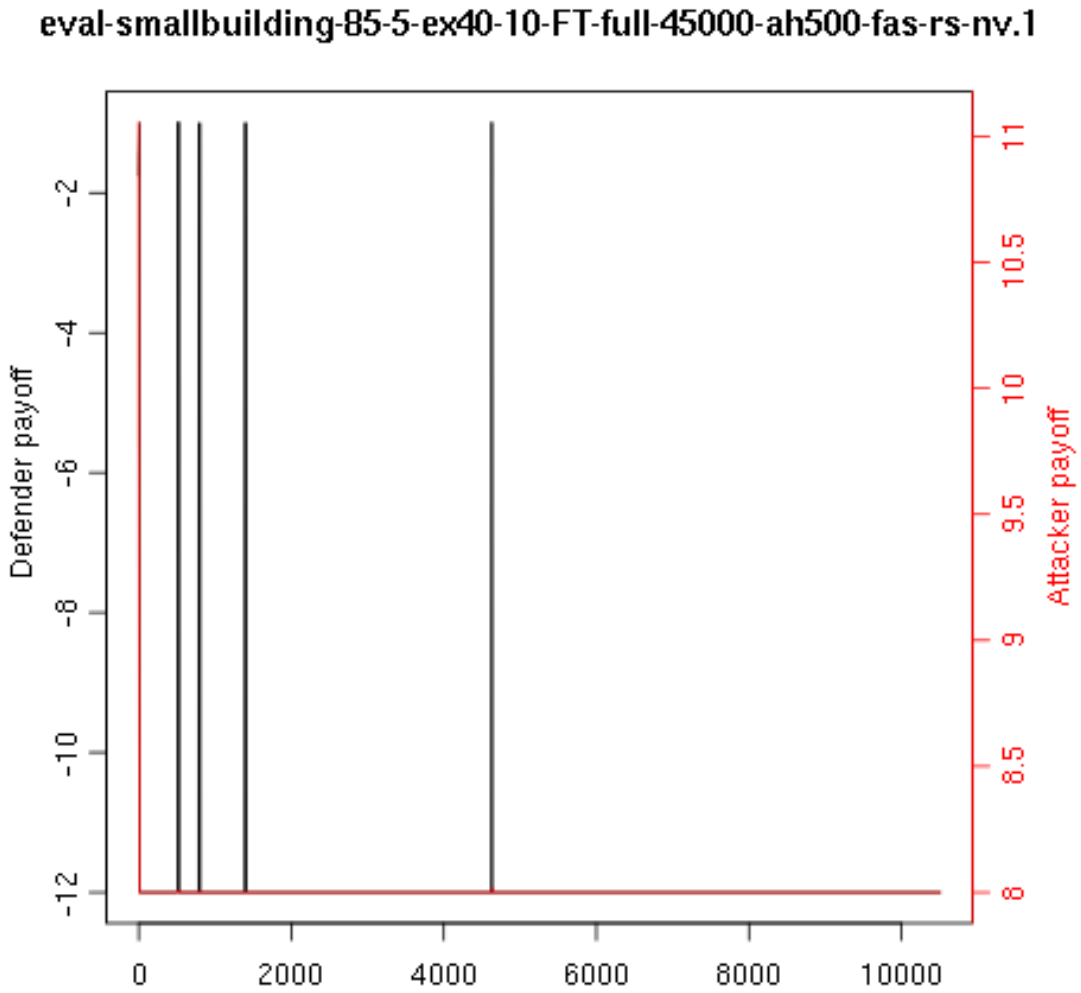
Przebiegi V


eval-smallbuilding-56-5-ex40-10-FT-full-45000-ah500-fas-rs-nv.1
Defender payoff
0
-2
-4
-6
Attacker payoff
6
4
2
0
0
5000
10000
15000
20000

Przebiegi VI



Przebiegi VII

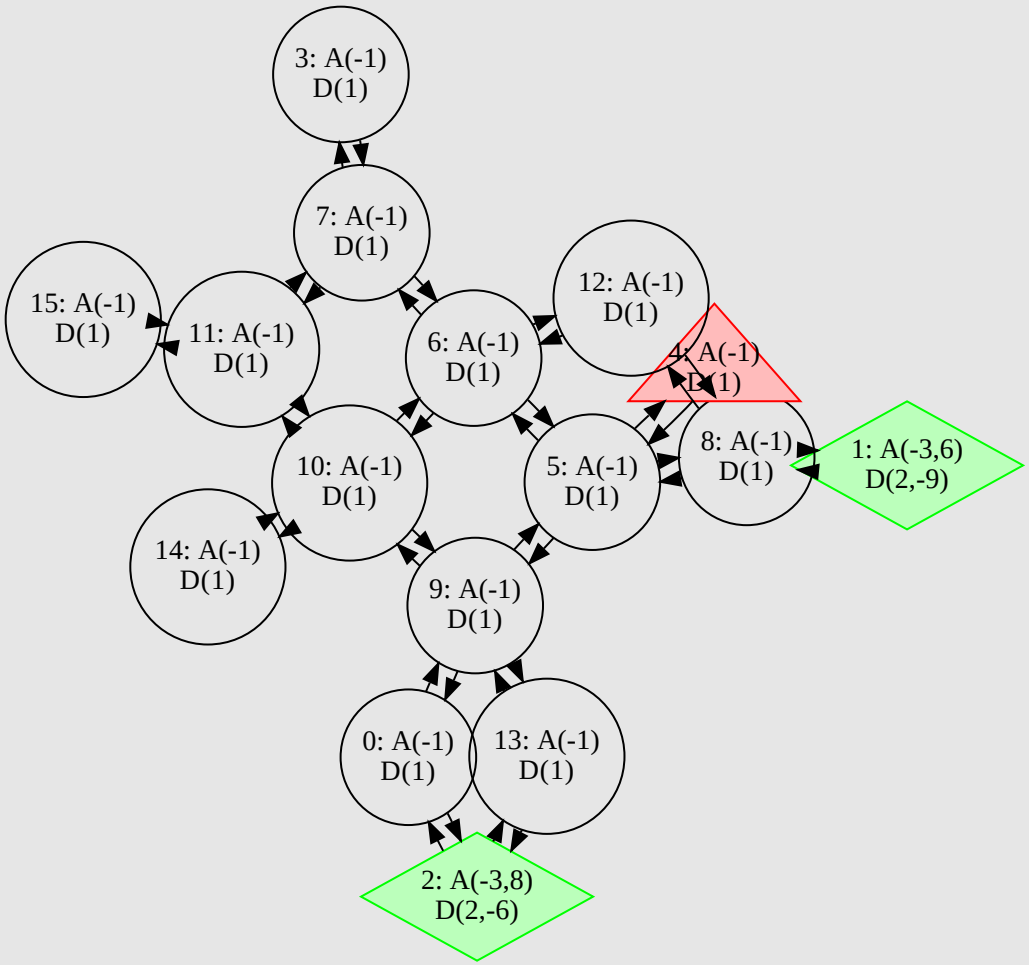


Wnioski

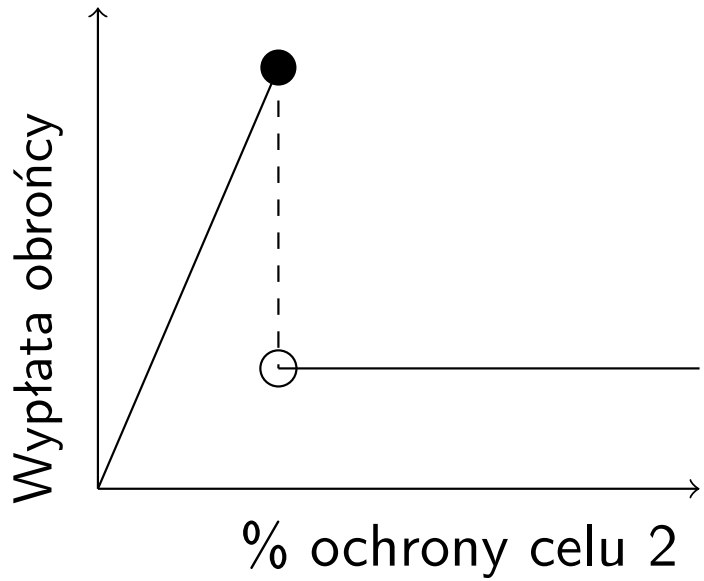
- Działa dobrze w większości przypadków
- Skaluje się lepiej niż MILP

Minima lokalne

Gra



- Metoda prawie deterministyczna
- Podąża w kierunku najlepszej wypłaty przy danym przeciwniku



Zatrzymywanie obliczeń

- Kryterium stopu?
- Restarty?
- Zmienne parametry uczenia?

*

(Koniec)