Applied Probability Trust (27 March 2019)

ONLINE SUPLEMENT TO "AN ASYMPTOTICALLY OPTIMAL HEURISTIC FOR GENERAL NON-STATIONARY FINITE-HORIZON RESTLESS MULTI-ARMED MULTI-ACTION BANDITS"

GABRIEL ZAYAS-CABÁN,* *University of Wisconsin-Madison*

STEFANUS JASIN,** *University of Michigan*

GUIHUA WANG,** *University of Michigan*

* Postal address: Mechanical Engineering Building, 1513 University Ave, Room 3011 Madison, WI 53706-1691, zayascaban@wisc.edu

** Postal address: Stephen M. Ross School of Business, University of Michigan, 701 Tappan St, Ann Arbor, MI 48109, jasin@umich.edu, guihuaw@umich.edu

APPENDIX 1: NUMERICAL SIMULATIONS AND RESULTS

1.1. Parameters Used in Numerical Experiments in Section 5

TABLE 1: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1	Experiment 2
Number of states ($\lvert\mathbb{J}\rvert$)	2	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	1	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 2 & 1 & 1 & 1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	2	5
Penalty threshold (m)	1	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}8.14 & 8.81\\ 6.72 & 1.04\end{bmatrix}$	$\begin{bmatrix}1.33 & 3.29 & 3.42 & 0.44 & 6.22\\ 5.48 & 5.45 & 3.79 & 5.42 & 2.90\\ 1.57 & 1.44 & 4.98 & 7.18 & 1.45\\ 5.01 & 4.16 & 4.33 & 1.72 & 5.86\\ 9.99 & 6.88 & 0.19 & 3.18 & 0.56\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0}=$	$\begin{bmatrix}0.96 & 0.04\\ 0.51 & 0.49\end{bmatrix}$	$\begin{bmatrix}0.19 & 0.17 & 0.01 & 0.63 & 0.01\\ 0.13 & 0.07 & 0.22 & 0.26 & 0.32\\ 0.17 & 0.17 & 0.24 & 0.18 & 0.24\\ 0.25 & 0.21 & 0.13 & 0.39 & 0.02\\ 0.05 & 0.30 & 0.18 & 0.30 & 0.17\end{bmatrix}$
$P_{j\in J,k\in J}^{1}=$	$\begin{bmatrix}0.15 & 0.85\\ 0.46 & 0.54\end{bmatrix}$	$\begin{bmatrix}0.15 & 0.21 & 0.27 & 0.16 & 0.22\\ 0.36 & 0.26 & 0.22 & 0.02 & 0.14\\ 0.16 & 0.25 & 0.25 & 0.21 & 0.13\\ 0.22 & 0.27 & 0.29 & 0.20 & 0.02\\ 0.15 & 0.29 & 0.26 & 0.28 & 0.01\end{bmatrix}$
$P_{j\in J,k\in J}^{2}=$	$\begin{bmatrix}0.78 & 0.22\\ 0.29 & 0.71\end{bmatrix}$	$\begin{bmatrix}0.02 & 0.26 & 0.35 & 0.28 & 0.10\\ 0.27 & 0.12 & 0.06 & 0.24 & 0.30\\ 0.13 & 0.34 & 0.14 & 0.20 & 0.20\\ 0.12 & 0.27 & 0.30 & 0.09 & 0.21\\ 0.13 & 0.27 & 0.05 & 0.25 & 0.30\end{bmatrix}$
$P_{j\in J,k\in J}^{3}=$		$\begin{bmatrix}0.32 & 0.14 & 0.15 & 0.04 & 0.35\\ 0.05 & 0.76 & 0.07 & 0.11 & 0.01\\ 0.34 & 0.23 & 0.11 & 0.01 & 0.32\\ 0.18 & 0.18 & 0.23 & 0.20 & 0.21\\ 0.28 & 0.31 & 0.02 & 0.05 & 0.35\end{bmatrix}$
$P_{j\in J,k\in J}^{4}=$		$\begin{bmatrix}0.16 & 0.21 & 0.16 & 0.24 & 0.23\\ 0.13 & 0.04 & 0.12 & 0.36 & 0.35\\ 0.34 & 0.33 & 0.02 & 0.27 & 0.03\\ 0.44 & 0.39 & 0.08 & 0.05 & 0.04\\ 0.12 & 0.19 & 0.24 & 0.06 & 0.39\end{bmatrix}$
$P_{j\in J,k\in J}^{5}=$		$\begin{bmatrix}0.25 & 0.21 & 0.09 & 0.21 & 0.24\\ 0.20 & 0.31 & 0.06 & 0.22 & 0.21\\ 0.05 & 0.03 & 0.17 & 0.31 & 0.43\\ 0.21 & 0.39 & 0.24 & 0.10 & 0.05\\ 0.31 & 0.15 & 0.37 & 0.15 & 0.02\end{bmatrix}$

1.2. Additional Simulations of Experiment 1

TABLE 2: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1a	Experiment 1b
Number of states ($\lvert\mathbb{J}\rvert$)	2	2
Undesirable state ($\lvert\mathbb{U}\rvert$)	1	1
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	2	2
Penalty threshold (m)	1	1
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}6.09 & 1.90\\ 9.22 & 9.57\end{bmatrix}$	$\begin{bmatrix}6.54 & 4.78\\ 3.13 & 6.04\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0} =$	$\begin{bmatrix}0.13 & 0.87\\ 0.68 & 0.32\end{bmatrix}$	$\begin{bmatrix}0.29 & 0.71\\ 0.44 & 0.56\end{bmatrix}$
$P_{j\in J,k\in J}^{1} =$	$\begin{bmatrix}0.46 & 0.54\\ 0.54 & 0.46\end{bmatrix}$	$\begin{bmatrix}0.15 & 0.85\\ 0.32 & 0.68\end{bmatrix}$
$P_{j\in J,k\in J}^{2} =$	$\begin{bmatrix}0.72 & 0.28\\ 0.00 & 1.00\end{bmatrix}$	$\begin{bmatrix}0.61 & 0.39\\ 0.25 & 0.75\end{bmatrix}$

TABLE 3: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1c	Experiment 1d
Number of states ($\lvert\mathbb{J}\rvert$)	2	2
Undesirable state ($\lvert\mathbb{U}\rvert$)	1	1
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	2	2
Penalty threshold (m)	1	1
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}8.03 & 5.48\\ 4.20 & 3.03\end{bmatrix}$	$\begin{bmatrix}7.75 & 1.12\\ 8.39 & 8.05\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0} =$	$\begin{bmatrix}0.44 & 0.56\\ 0.46 & 0.54\end{bmatrix}$	$\begin{bmatrix}0.51 & 0.49\\ 0.96 & 0.04\end{bmatrix}$
$P_{j\in J,k\in J}^{1} =$	$\begin{bmatrix}0.5 & 0.5\\ 0.49 & 0.51\end{bmatrix}$	$\begin{bmatrix}0.3 & 0.7\\ 0.29 & 0.71\end{bmatrix}$
$P_{j\in J,k\in J}^{2} =$	$\begin{bmatrix}0.47 & 0.53\\ 0.33 & 0.67\end{bmatrix}$	$\begin{bmatrix}0.49 & 0.51\\ 0.01 & 0.99\end{bmatrix}$

TABLE 4: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1e	Experiment 1f
Number of states ($\|\mathbb{J}\|$)	2	2
Undesirable state ($\|\mathbb{U}\|$)	1	1
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\end{bmatrix}$
Number of actions ($\|\mathbb{A}\|$)	2	2
Penalty threshold (m)	1	1
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}8.14 & 8.81\\ 6.72 & 1.04\end{bmatrix}$	$\begin{bmatrix}7.89 & 5.51\\ 9.85 & 7.66\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0} =$	$\begin{bmatrix}0.96 & 0.04\\ 0.51 & 0.49\end{bmatrix}$	$\begin{bmatrix}0.66 & 0.34\\ 0.68 & 0.32\end{bmatrix}$
$P_{j\in J,k\in J}^{1} =$	$\begin{bmatrix}0.15 & 0.85\\ 0.46 & 0.54\end{bmatrix}$	$\begin{bmatrix}0.24 & 0.76\\ 0.5 & 0.5\end{bmatrix}$
$P_{j\in J,k\in J}^{2} =$	$\begin{bmatrix}0.78 & 0.22\\ 0.29 & 0.71\end{bmatrix}$	$\begin{bmatrix}0.97 & 0.03\\ 0.32 & 0.68\end{bmatrix}$

TABLE 5: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1g	Experiment 1h
Number of states ($\|\mathbb{J}\|$)	2	2
Undesirable state ($\|\mathbb{U}\|$)	1	1
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\end{bmatrix}$
Number of actions ($\|\mathbb{A}\|$)	2	2
Penalty threshold (m)	1	1
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}3.29 & 3.42\\ 0.44 & 6.22\end{bmatrix}$	$\begin{bmatrix}4.33 & 1.72\\ 5.86 & 9.99\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0} =$	$\begin{bmatrix}0.5 & 0.5\\ 0.41 & 0.59\end{bmatrix}$	$\begin{bmatrix}0.97 & 0.03\\ 0.85 & 0.15\end{bmatrix}$
$P_{j\in J,k\in J}^{1} =$	$\begin{bmatrix}0.65 & 0.35\\ 0.22 & 0.78\end{bmatrix}$	$\begin{bmatrix}0.52 & 0.48\\ 0.01 & 0.99\end{bmatrix}$
$P_{j\in J,k\in J}^{2} =$	$\begin{bmatrix}0.83 & 0.17\\ 0.55 & 0.45\end{bmatrix}$	$\begin{bmatrix}0.02 & 0.98\\ 0.24 & 0.76\end{bmatrix}$

1.3. Additional Simulations of Experiment 2

TABLE 6: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 1i	Experiment 1j
Number of states ($\lvert\mathbb{J}\rvert$)	2	2
Undesirable state ($\lvert\mathbb{U}\rvert$)	1	1
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2 & 1\end{bmatrix}$	$\begin{bmatrix}2 & 1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	2	2
Penalty threshold (m)	1	1
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}8.03 & 9.80\\ 5.01 & 5.05\end{bmatrix}$	$\begin{bmatrix}4.85 & 5.43\\ 7.62 & 9.78\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0} =$	$\begin{bmatrix}0.58 & 0.42\\ 0.59 & 0.41\end{bmatrix}$	$\begin{bmatrix}0.41 & 0.59\\ 0.58 & 0.42\end{bmatrix}$
$P_{j\in J,k\in J}^{1} =$	$\begin{bmatrix}0.62 & 0.38\\ 0.95 & 0.05\end{bmatrix}$	$\begin{bmatrix}0.92 & 0.08\\ 0.3 & 0.7\end{bmatrix}$
$P_{j\in J,k\in J}^{2} =$	$\begin{bmatrix}0.15 & 0.85\\ 0.37 & 0.63\end{bmatrix}$	$\begin{bmatrix}0.49 & 0.51\\ 0.62 & 0.38\end{bmatrix}$

TABLE 7: Percentage loss for 2 states/2 actions

θ	$T = 10$	$T = 30$	$T = 50$	$T = 100$	$T = 10$	$T = 30$	$T = 50$	$T = 100$
		Experiment 1a				Experiment 1b		
1	81.03	98.95	97.33	90.46	171.45	172.15	181.86	172.15
5	47.74	48.01	44.14	45.29	63.15	67.59	69.40	69.26
10	26.78	31.95	33.68	30.33	30.31	34.82	35.10	30.59
20	22.85	21.19	23.94	22.14	15.52	19.83	15.97	10.52
40	15.65	15.61	15.24	15.81	8.72	12.66	5.80	7.90
60	11.03	12.91	12.81	12.48	-1.05	3.71	1.79	2.18
80	11.74	11.59	11.17	10.00	-0.65	2.93	4.55	0.87
100	10.64	10.15	10.51	9.64	2.24	2.81	3.62	3.19
		Experiment 1c				Experiment 1d		
1	55.72	53.97	50.76	51.05	78.71	74.62	82.76	78.25
5	6.61	14.54	11.34	8.30	24.09	33.36	30.90	28.36
10	4.69	3.99	3.93	3.67	17.53	18.50	19.31	18.89
20	-0.21	3.48	-0.36	3.05	9.24	13.87	12.12	13.92
40	2.59	2.44	-0.58	0.01	9.97	8.81	7.58	7.23
60	-2.52	-0.73	-0.49	-0.80	4.72	6.33	6.39	5.82
80	-0.31	0.88	0.32	-1.20	5.09	5.90	5.54	5.14
100	1.12	0.48	0.19	1.02	6.23	5.79	5.59	6.00
		Experiment 1e				Experiment 1f		
1	6.79	4.77	6.13	5.08	5.38	2.19	6.08	3.38
5	2.15	2.88	2.51	2.38	-1.07	1.69	1.61	-0.41
10	1.78	1.52	1.82	1.80	-1.36	0.69	0.19	-0.72
20	1.33	1.32	1.05	0.98	-0.46	0.89	0.69	0.48
40	1.11	0.63	1.00	1.34	0.58	0.70	-0.62	0.02
60	0.53	0.64	0.80	0.59	-0.92	-0.34	-0.09	0.28
80	0.80	0.69	0.91	0.58	-0.15	-0.06	-0.61	-0.31
100	0.66	0.53	0.49	0.81	0.27	0.43	0.31	-0.01
		Experiment 1g				Experiment 1h		
1	40.43	49.38	56.37	51.62	12.42	12.53	13.91	12.75
5	9.21	14.36	13.91	15.81	6.79	6.55	7.01	5.74
10	3.33	4.03	4.17	1.52	4.27	4.82	4.16	4.45
20	-0.72	2.55	2.44	-0.09	3.15	3.33	3.52	3.31
40	1.32	2.00	-1.21	2.75	2.23	2.24	2.37	2.60
60	-1.92	-0.08	-0.58	-0.06	1.86	1.84	1.80	1.69
80	-0.53	0.85	0.19	-0.07	1.75	1.73	1.60	1.58
100	0.97	-0.18	0.40	-0.93	1.47	1.47	1.44	1.50
		Experiment 2i				Experiment 2j		
1	23.96	21.85	26.18	24.06	22.11	26.55	31.78	21.71
5	5.33	8.31	4.56	3.58	2.20	7.04	6.19	2.81
10	3.37	3.63	3.88	0.81	-0.21	2.26	3.15	-0.23
20	0.24	1.85	1.28	0.99	-0.90	2.17	0.85	2.01
40	1.70	1.53	-0.16	0.05	1.47	1.42	-1.32	-0.22
60	-0.83	0.29	0.12	0.60	-1.32	-0.14	-0.48	-0.46
80	0.41	0.25	0.36	-0.06	-0.49	-0.31	-0.87	-0.54
100	0.70	1.29	0.82	0.78	1.06	0.89	0.28	0.47

TABLE 8: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 2a	Experiment 2b
Number of states ($\lvert\mathbb{J}\rvert$)	5	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	2	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	5	5
Penalty threshold (m)	2	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}2.21&6.56&6.73&6.60&1.78\\4.80&5.88&8.70&7.35&0.43\\2.04&1.97&4.09&8.46&7.51\\2.32&6.80&9.20&7.84&4.61\\4.98&7.30&2.41&3.65&1.85\end{bmatrix}$	$\begin{bmatrix}0.12&9.35&4.27&2.58&2.62\\1.61&9.98&4.66&0.74&7.08\\8.96&3.94&3.49&8.20&3.41\\4.42&0.52&0.91&0.26&2.01\\9.88&0.14&4.38&7.17&3.73\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0}=$	$\begin{bmatrix}0.06&0.07&0.35&0.19&0.33\\0.22&0.20&0.32&0.06&0.20\\0.23&0.05&0.32&0.09&0.31\\0.15&0.26&0.09&0.26&0.25\\0.04&0.35&0.13&0.23&0.25\end{bmatrix}$	$\begin{bmatrix}0.28&0.14&0.06&0.32&0.20\\0.21&0.20&0.19&0.32&0.08\\0.16&0.32&0.00&0.16&0.36\\0.33&0.03&0.21&0.11&0.32\\0.31&0.14&0.32&0.09&0.14\end{bmatrix}$
$P_{j\in J,k\in J}^{1}=$	$\begin{bmatrix}0.14&0.32&0.08&0.15&0.30\\0.38&0.12&0.19&0.21&0.10\\0.08&0.41&0.20&0.01&0.30\\0.63&0.07&0.02&0.19&0.08\\0.03&0.22&0.23&0.28&0.23\end{bmatrix}$	$\begin{bmatrix}0.28&0.30&0.15&0.25&0.02\\0.13&0.22&0.17&0.06&0.41\\0.53&0.35&0.05&0.05&0.02\\0.06&0.34&0.02&0.09&0.49\\0.17&0.13&0.37&0.24&0.09\end{bmatrix}$
$P_{j\in J,k\in J}^{2}=$	$\begin{bmatrix}0.21&0.07&0.17&0.37&0.18\\0.00&0.25&0.14&0.42&0.18\\0.21&0.20&0.22&0.23&0.14\\0.26&0.09&0.11&0.15&0.38\\0.32&0.06&0.26&0.30&0.05\end{bmatrix}$	$\begin{bmatrix}0.03&0.11&0.41&0.25&0.21\\0.09&0.23&0.19&0.24&0.26\\0.05&0.10&0.08&0.37&0.40\\0.15&0.19&0.51&0.01&0.13\\0.41&0.15&0.18&0.05&0.20\end{bmatrix}$
$P_{j\in J,k\in J}^{3}=$	$\begin{bmatrix}0.07&0.26&0.02&0.48&0.17\\0.27&0.02&0.22&0.31&0.17\\0.14&0.35&0.04&0.04&0.43\\0.08&0.21&0.31&0.09&0.32\\0.24&0.13&0.18&0.23&0.21\end{bmatrix}$	$\begin{bmatrix}0.28&0.36&0.25&0.06&0.05\\0.26&0.19&0.24&0.20&0.11\\0.19&0.06&0.27&0.24&0.24\\0.13&0.17&0.10&0.29&0.30\\0.20&0.37&0.31&0.05&0.07\end{bmatrix}$
$P_{j\in J,k\in J}^{4}=$	$\begin{bmatrix}0.21&0.29&0.09&0.37&0.04\\0.11&0.12&0.29&0.29&0.18\\0.03&0.12&0.04&0.43&0.38\\0.12&0.18&0.28&0.37&0.05\\0.10&0.19&0.26&0.22&0.23\end{bmatrix}$	$\begin{bmatrix}0.20&0.26&0.32&0.13&0.10\\0.55&0.32&0.03&0.09&0.01\\0.21&0.00&0.32&0.28&0.18\\0.19&0.37&0.05&0.25&0.14\\0.44&0.24&0.03&0.22&0.06\end{bmatrix}$
$P_{j\in J,k\in J}^{5}=$	$\begin{bmatrix}0.46&0.14&0.04&0.25&0.11\\0.20&0.14&0.28&0.08&0.30\\0.25&0.12&0.21&0.24&0.18\\0.24&0.28&0.14&0.07&0.27\\0.13&0.22&0.11&0.40&0.14\end{bmatrix}$	$\begin{bmatrix}0.07&0.32&0.36&0.13&0.13\\0.28&0.19&0.13&0.27&0.13\\0.21&0.22&0.19&0.11&0.27\\0.30&0.06&0.28&0.31&0.05\\0.26&0.24&0.29&0.18&0.03\end{bmatrix}$

TABLE 9: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 2c	Experiment 2d
Number of states ($\lvert\mathbb{J}\rvert$)	5	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	2	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	5	5
Penalty threshold (m)	2	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}4.04&0.14&0.87&8.90&0.41\\7.16&5.52&6.79&8.97&4.81\\0.65&8.00&0.05&7.75&7.56\\1.02&7.45&6.86&2.25&5.48\\8.77&0.69&2.05&5.89&1.40\end{bmatrix}$	$\begin{bmatrix}6.67&1.96&6.18&6.07&2.43\\9.69&1.01&4.31&8.80&5.53\\6.47&4.94&6.80&7.28&2.38\\6.61&2.61&4.94&7.29&9.52\\9.69&3.12&2.95&0.95&1.07\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0}=$	$\begin{bmatrix}0.22&0.30&0.06&0.18&0.25\\0.14&0.32&0.03&0.27&0.24\\0.05&0.30&0.32&0.22&0.11\\0.17&0.38&0.30&0.02&0.14\\0.21&0.15&0.11&0.39&0.15\end{bmatrix}$	$\begin{bmatrix}0.27&0.18&0.21&0.00&0.34\\0.15&0.28&0.17&0.28&0.12\\0.14&0.22&0.27&0.29&0.07\\0.28&0.15&0.25&0.05&0.27\\0.34&0.17&0.22&0.22&0.05\end{bmatrix}$
$P_{j\in J,k\in J}^{1}=$	$\begin{bmatrix}0.22&0.01&0.46&0.20&0.11\\0.26&0.13&0.21&0.17&0.24\\0.09&0.11&0.30&0.15&0.36\\0.20&0.09&0.22&0.26&0.24\\0.25&0.02&0.34&0.15&0.25\end{bmatrix}$	$\begin{bmatrix}0.20&0.31&0.06&0.21&0.22\\0.04&0.03&0.36&0.31&0.27\\0.10&0.14&0.18&0.31&0.27\\0.23&0.30&0.20&0.12&0.15\\0.29&0.20&0.10&0.19&0.22\end{bmatrix}$
$P_{j\in J,k\in J}^{2}=$	$\begin{bmatrix}0.14&0.09&0.15&0.20&0.42\\0.10&0.28&0.18&0.18&0.26\\0.16&0.19&0.19&0.13&0.33\\0.16&0.21&0.26&0.08&0.29\\0.35&0.02&0.23&0.03&0.37\end{bmatrix}$	$\begin{bmatrix}0.07&0.34&0.01&0.19&0.40\\0.13&0.10&0.21&0.14&0.41\\0.25&0.01&0.25&0.25&0.23\\0.15&0.14&0.02&0.36&0.32\\0.23&0.20&0.14&0.24&0.19\end{bmatrix}$
$P_{j\in J,k\in J}^{3}=$	$\begin{bmatrix}0.22&0.02&0.08&0.32&0.35\\0.16&0.20&0.22&0.20&0.22\\0.28&0.01&0.32&0.09&0.29\\0.10&0.20&0.21&0.18&0.31\\0.31&0.19&0.15&0.12&0.23\end{bmatrix}$	$\begin{bmatrix}0.24&0.27&0.06&0.32&0.11\\0.11&0.28&0.46&0.11&0.03\\0.26&0.29&0.02&0.30&0.12\\0.04&0.32&0.11&0.36&0.18\\0.13&0.35&0.26&0.23&0.02\end{bmatrix}$
$P_{j\in J,k\in J}^{4}=$	$\begin{bmatrix}0.34&0.02&0.25&0.01&0.39\\0.12&0.10&0.09&0.43&0.26\\0.12&0.19&0.32&0.05&0.31\\0.20&0.20&0.27&0.06&0.28\\0.28&0.02&0.33&0.34&0.03\end{bmatrix}$	$\begin{bmatrix}0.23&0.20&0.29&0.09&0.19\\0.06&0.14&0.18&0.28&0.34\\0.36&0.32&0.07&0.09&0.16\\0.27&0.00&0.36&0.09&0.27\\0.22&0.31&0.26&0.02&0.19\end{bmatrix}$
$P_{j\in J,k\in J}^{5}=$	$\begin{bmatrix}0.27&0.32&0.04&0.11&0.26\\0.31&0.08&0.07&0.24&0.30\\0.06&0.29&0.01&0.37&0.28\\0.24&0.01&0.18&0.22&0.36\\0.46&0.30&0.06&0.08&0.09\end{bmatrix}$	$\begin{bmatrix}0.45&0.08&0.13&0.20&0.13\\0.16&0.23&0.19&0.23&0.19\\0.16&0.18&0.32&0.13&0.21\\0.22&0.16&0.19&0.34&0.09\\0.25&0.26&0.06&0.22&0.21\end{bmatrix}$

TABLE 10: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 2e	Experiment 2f
Number of states ($\lvert\mathbb{J}\rvert$)	5	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	2	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	5	5
Penalty threshold (m)	2	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}5.72&7.27&6.88&6.48&2.53\\0.13&5.83&1.00&1.34&3.69\\9.12&2.41&7.26&7.87&8.52\\2.36&0.16&2.85&7.55&4.77\\1.39&1.37&7.56&2.38&2.77\end{bmatrix}$	$\begin{bmatrix}4.29&9.59&3.05&4.71&5.07\\0.06&3.72&1.92&2.54&4.43\\5.34&4.68&2.08&6.16&9.94\\0.47&8.54&9.70&2.63&5.50\\7.49&7.35&7.64&4.38&0.12\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0}=$	$\begin{bmatrix}0.33&0.18&0.10&0.30&0.09\\0.30&0.41&0.02&0.15&0.13\\0.27&0.14&0.40&0.19&0.00\\0.03&0.05&0.32&0.20&0.40\\0.16&0.38&0.20&0.23&0.03\end{bmatrix}$	$\begin{bmatrix}0.27&0.23&0.23&0.06&0.21\\0.32&0.06&0.13&0.28&0.21\\0.16&0.10&0.26&0.21&0.28\\0.11&0.31&0.28&0.25&0.05\\0.11&0.20&0.24&0.17&0.28\end{bmatrix}$
$P_{j\in J,k\in J}^{1}=$	$\begin{bmatrix}0.01&0.25&0.20&0.30&0.24\\0.22&0.01&0.25&0.27&0.26\\0.28&0.03&0.31&0.27&0.11\\0.32&0.12&0.24&0.13&0.20\\0.32&0.33&0.30&0.03&0.02\end{bmatrix}$	$\begin{bmatrix}0.37&0.08&0.21&0.24&0.11\\0.14&0.22&0.32&0.05&0.27\\0.34&0.16&0.11&0.22&0.18\\0.11&0.18&0.01&0.08&0.62\\0.35&0.04&0.17&0.26&0.18\end{bmatrix}$
$P_{j\in J,k\in J}^{2}=$	$\begin{bmatrix}0.23&0.35&0.25&0.02&0.15\\0.32&0.21&0.11&0.31&0.05\\0.25&0.27&0.19&0.09&0.20\\0.27&0.09&0.14&0.29&0.21\\0.33&0.08&0.05&0.41&0.13\end{bmatrix}$	$\begin{bmatrix}0.30&0.08&0.30&0.20&0.12\\0.18&0.13&0.42&0.00&0.27\\0.14&0.29&0.05&0.32&0.20\\0.38&0.14&0.14&0.12&0.23\\0.19&0.24&0.23&0.14&0.20\end{bmatrix}$
$P_{j\in J,k\in J}^{3}=$	$\begin{bmatrix}0.21&0.03&0.21&0.16&0.39\\0.15&0.12&0.03&0.68&0.01\\0.44&0.00&0.38&0.04&0.13\\0.17&0.25&0.12&0.19&0.27\\0.15&0.32&0.02&0.30&0.20\end{bmatrix}$	$\begin{bmatrix}0.22&0.15&0.34&0.17&0.12\\0.09&0.02&0.31&0.21&0.36\\0.13&0.11&0.12&0.32&0.32\\0.28&0.37&0.32&0.00&0.02\\0.09&0.37&0.07&0.33&0.13\end{bmatrix}$
$P_{j\in J,k\in J}^{4}=$	$\begin{bmatrix}0.47&0.01&0.11&0.16&0.25\\0.35&0.28&0.24&0.11&0.03\\0.11&0.22&0.29&0.21&0.18\\0.01&0.20&0.38&0.16&0.25\\0.22&0.20&0.25&0.02&0.31\end{bmatrix}$	$\begin{bmatrix}0.23&0.14&0.04&0.51&0.07\\0.58&0.00&0.05&0.30&0.07\\0.12&0.49&0.18&0.16&0.05\\0.18&0.27&0.02&0.21&0.31\\0.28&0.21&0.22&0.21&0.08\end{bmatrix}$
$P_{j\in J,k\in J}^{5}=$	$\begin{bmatrix}0.32&0.25&0.06&0.11&0.26\\0.04&0.22&0.18&0.20&0.36\\0.21&0.17&0.04&0.39&0.19\\0.34&0.21&0.02&0.07&0.37\\0.35&0.05&0.04&0.28&0.28\end{bmatrix}$	$\begin{bmatrix}0.07&0.13&0.06&0.39&0.34\\0.41&0.08&0.33&0.05&0.15\\0.14&0.10&0.01&0.57&0.18\\0.34&0.09&0.10&0.19&0.28\\0.03&0.27&0.26&0.22&0.23\end{bmatrix}$

Table 11: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 2g	Experiment 2h
Number of states ($\lvert\mathbb{J}\rvert$)	5	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	2	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	5	5
Penalty threshold (m)	2	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}8.19&7.64&0.09&9.28&4.89\\3.15&1.35&2.54&5.04&0.75\\7.55&6.60&8.77&3.58&0.83\\1.41&8.07&4.19&0.60&5.52\\5.50&3.29&0.75&5.95&1.47\end{bmatrix}$	$\begin{bmatrix}1.94&1.53&9.37&6.44&6.84\\6.65&5.95&8.55&7.95&4.37\\4.30&8.26&5.89&7.20&7.25\\4.34&1.93&4.82&5.35&5.60\\2.72&0.81&7.66&7.52&1.25\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P^0_{j\in J,k\in J}=$	$\begin{bmatrix}0.25&0.31&0.06&0.06&0.33\\0.35&0.12&0.34&0.18&0.01\\0.01&0.19&0.40&0.25&0.14\\0.33&0.48&0.02&0.02&0.16\\0.10&0.23&0.33&0.29&0.04\end{bmatrix}$	$\begin{bmatrix}0.21&0.12&0.09&0.38&0.20\\0.02&0.29&0.28&0.12&0.29\\0.36&0.22&0.12&0.19&0.10\\0.14&0.19&0.15&0.14&0.39\\0.31&0.26&0.09&0.17&0.18\end{bmatrix}$
$P^1_{j\in J,k\in J}=$	$\begin{bmatrix}0.05&0.40&0.03&0.27&0.25\\0.20&0.29&0.32&0.12&0.07\\0.28&0.30&0.21&0.18&0.03\\0.28&0.10&0.11&0.19&0.32\\0.16&0.16&0.19&0.26&0.22\end{bmatrix}$	$\begin{bmatrix}0.18&0.11&0.31&0.24&0.15\\0.32&0.23&0.15&0.04&0.26\\0.16&0.29&0.24&0.17&0.14\\0.24&0.20&0.41&0.15&0.00\\0.31&0.23&0.08&0.16&0.22\end{bmatrix}$
$P^2_{j\in J,k\in J}=$	$\begin{bmatrix}0.24&0.21&0.27&0.21&0.07\\0.26&0.09&0.11&0.00&0.54\\0.21&0.27&0.23&0.27&0.02\\0.16&0.14&0.17&0.12&0.43\\0.10&0.26&0.15&0.39&0.11\end{bmatrix}$	$\begin{bmatrix}0.19&0.38&0.18&0.07&0.19\\0.16&0.29&0.16&0.28&0.11\\0.18&0.05&0.04&0.58&0.15\\0.02&0.10&0.24&0.32&0.31\\0.16&0.20&0.16&0.23&0.25\end{bmatrix}$
$P^3_{j\in J,k\in J}=$	$\begin{bmatrix}0.00&0.16&0.44&0.12&0.27\\0.16&0.31&0.16&0.21&0.16\\0.37&0.31&0.09&0.04&0.19\\0.17&0.19&0.07&0.25&0.32\\0.20&0.19&0.18&0.20&0.23\end{bmatrix}$	$\begin{bmatrix}0.11&0.41&0.34&0.07&0.07\\0.02&0.40&0.14&0.41&0.03\\0.28&0.06&0.15&0.41&0.10\\0.13&0.00&0.30&0.29&0.28\\0.21&0.23&0.31&0.05&0.20\end{bmatrix}$
$P^4_{j\in J,k\in J}=$	$\begin{bmatrix}0.23&0.23&0.21&0.22&0.11\\0.32&0.04&0.27&0.13&0.23\\0.16&0.30&0.16&0.30&0.09\\0.19&0.22&0.14&0.13&0.33\\0.14&0.08&0.38&0.00&0.40\end{bmatrix}$	$\begin{bmatrix}0.07&0.09&0.10&0.30&0.45\\0.20&0.19&0.23&0.17&0.22\\0.17&0.28&0.12&0.17&0.25\\0.30&0.24&0.16&0.07&0.23\\0.49&0.10&0.19&0.12&0.10\end{bmatrix}$
$P^5_{j\in J,k\in J}=$	$\begin{bmatrix}0.21&0.03&0.43&0.13&0.19\\0.04&0.06&0.51&0.25&0.15\\0.20&0.04&0.18&0.25&0.33\\0.36&0.09&0.43&0.01&0.11\\0.11&0.01&0.47&0.10&0.31\end{bmatrix}$	$\begin{bmatrix}0.26&0.02&0.35&0.33&0.04\\0.29&0.24&0.18&0.02&0.27\\0.27&0.05&0.05&0.40&0.23\\0.22&0.21&0.21&0.11&0.25\\0.37&0.28&0.02&0.06&0.27\end{bmatrix}$

TABLE 12: Parameters Used in Numerical Experiments

Parameter (Notation)	Experiment 2i	Experiment 2j
Number of states ($\lvert\mathbb{J}\rvert$)	5	5
Undesirable state ($\lvert\mathbb{U}\rvert$)	2	2
Number of arms ($n_{j\in J,1}$)	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$	$\begin{bmatrix}2&2&1&1&1\end{bmatrix}$
Number of actions ($\lvert\mathbb{A}\rvert$)	5	5
Penalty threshold (m)	2	2
Penalty cost (ϕ)	20	20
Cost matrix ($c_{j\in J}^{a\in A}$)	$\begin{bmatrix}9.98&6.67&3.01&8.40&7.51\\7.98&4.72&6.66&0.33&0.35\\4.26&7.27&1.50&9.82&4.19\\6.19&6.06&8.91&5.10&0.39\\9.11&1.55&1.11&7.01&7.01\end{bmatrix}$	$\begin{bmatrix}2.57&3.48&1.84&1.92&5.53\\2.53&0.37&0.21&3.38&8.91\\0.85&9.70&2.48&5.92&1.67\\1.55&6.25&6.25&6.36&4.33\\1.70&1.80&4.62&2.23&6.29\end{bmatrix}$
Transition matrix ($P_{j\in J,k\in J}^{a\in A}$)		
$P_{j\in J,k\in J}^{0}=$	$\begin{bmatrix}0.06&0.21&0.28&0.27&0.18\\0.24&0.26&0.27&0.09&0.14\\0.33&0.23&0.20&0.15&0.09\\0.32&0.14&0.02&0.52&0.01\\0.23&0.14&0.25&0.12&0.26\end{bmatrix}$	$\begin{bmatrix}0.34&0.10&0.15&0.19&0.22\\0.19&0.23&0.15&0.24&0.20\\0.19&0.33&0.12&0.16&0.19\\0.30&0.07&0.28&0.06&0.29\\0.19&0.76&0.03&0.01&0.02\end{bmatrix}$
$P_{j\in J,k\in J}^{1}=$	$\begin{bmatrix}0.30&0.37&0.06&0.16&0.11\\0.34&0.14&0.18&0.33&0.01\\0.13&0.20&0.21&0.31&0.15\\0.18&0.26&0.28&0.03&0.25\\0.22&0.29&0.22&0.17&0.10\end{bmatrix}$	$\begin{bmatrix}0.23&0.15&0.17&0.42&0.03\\0.29&0.23&0.12&0.11&0.25\\0.33&0.30&0.30&0.04&0.02\\0.19&0.28&0.30&0.04&0.19\\0.00&0.35&0.07&0.09&0.49\end{bmatrix}$
$P_{j\in J,k\in J}^{2}=$	$\begin{bmatrix}0.22&0.14&0.27&0.34&0.03\\0.23&0.16&0.17&0.05&0.38\\0.37&0.29&0.05&0.17&0.12\\0.29&0.05&0.24&0.23&0.20\\0.21&0.24&0.04&0.35&0.16\end{bmatrix}$	$\begin{bmatrix}0.37&0.03&0.14&0.13&0.34\\0.19&0.39&0.14&0.25&0.02\\0.13&0.33&0.17&0.35&0.02\\0.53&0.35&0.04&0.00&0.08\\0.17&0.30&0.27&0.12&0.14\end{bmatrix}$
$P_{j\in J,k\in J}^{3}=$	$\begin{bmatrix}0.25&0.20&0.01&0.32&0.22\\0.11&0.12&0.22&0.12&0.43\\0.39&0.08&0.14&0.25&0.14\\0.24&0.36&0.30&0.07&0.04\\0.18&0.03&0.26&0.18&0.34\end{bmatrix}$	$\begin{bmatrix}0.32&0.07&0.26&0.13&0.22\\0.24&0.19&0.12&0.25&0.19\\0.30&0.02&0.25&0.17&0.26\\0.46&0.05&0.05&0.24&0.20\\0.29&0.36&0.11&0.16&0.09\end{bmatrix}$
$P_{j\in J,k\in J}^{4}=$	$\begin{bmatrix}0.21&0.06&0.30&0.25&0.18\\0.21&0.28&0.22&0.05&0.24\\0.23&0.20&0.12&0.20&0.24\\0.09&0.19&0.46&0.13&0.12\\0.07&0.32&0.01&0.35&0.25\end{bmatrix}$	$\begin{bmatrix}0.12&0.33&0.09&0.30&0.15\\0.36&0.03&0.26&0.25&0.09\\0.26&0.16&0.22&0.27&0.09\\0.04&0.35&0.01&0.24&0.35\\0.27&0.25&0.18&0.19&0.10\end{bmatrix}$
$P_{j\in J,k\in J}^{5}=$	$\begin{bmatrix}0.17&0.33&0.16&0.18&0.16\\0.32&0.23&0.15&0.14&0.16\\0.30&0.24&0.03&0.16&0.26\\0.00&0.43&0.05&0.34&0.17\\0.09&0.18&0.09&0.05&0.59\end{bmatrix}$	$\begin{bmatrix}0.18&0.17&0.29&0.33&0.04\\0.26&0.19&0.21&0.06&0.28\\0.23&0.25&0.02&0.29&0.21\\0.19&0.37&0.10&0.23&0.10\\0.04&0.37&0.14&0.34&0.12\end{bmatrix}$

TABLE 13: Percentage loss for 5 states/5 actions

θ	$T = 10$	$T = 30$	$T = 50$	$T = 100$	$T = 10$	$T = 30$	$T = 50$	$T = 100$
		Experiment 2a				Experiment 2b		
1	138.93	144.36	134.02	120.42	36.92	35.09	34.92	37.08
5	50.40	50.48	52.98	50.52	10.11	9.70	8.03	6.48
10	33.31	40.43	34.59	41.22	3.42	3.92	2.74	6.36
20	21.81	22.29	23.41	19.73	3.76	2.31	2.55	2.35
40	12.36	11.73	12.93	11.33	0.67	1.46	2.23	0.83
60	6.27	7.56	9.79	5.83	1.01	1.92	2.57	0.69
80	7.21	4.22	6.16	3.51	1.81	0.57	1.65	0.50
100	2.41	3.75	3.30	3.17	0.84	0.21	0.52	0.51
		Experiment 2c				Experiment 2d		
1	21.33	26.26	24.67	22.64	35.78	31.53	34.80	31.13
5	4.95	4.28	6.27	4.22	5.80	8.09	7.70	10.53
10	3.71	5.40	1.58	5.35	4.23	6.36	4.44	5.56
20	4.09	3.10	2.62	2.41	3.37	3.43	1.50	2.62
40	1.67	1.23	1.71	0.61	1.03	1.88	2.04	1.24
60	1.61	1.62	1.92	0.50	1.31	2.08	1.00	1.76
80	1.23	0.89	1.07	0.72	0.94	1.49	0.67	0.69
100	1.50	0.98	0.97	0.96	1.79	0.37	0.86	1.50
		Experiment 2e				Experiment 2f		
1	57.35	55.95	56.67	55.44	68.67	70.18	64.63	66.75
5	19.17	18.36	17.27	20.96	19.88	11.11	11.75	21.34
10	12.07	15.89	10.50	12.78	10.56	5.98	5.39	6.20
20	8.92	7.65	5.10	6.16	3.47	2.39	3.43	2.44
40	2.69	4.12	4.73	3.48	1.72	2.04	1.54	1.43
60	3.34	4.11	3.68	4.29	0.96	2.64	2.50	-0.41
80	1.90	1.97	3.92	0.85	2.21	1.04	0.21	0.16
100	3.15	1.40	2.97	3.01	2.15	-0.12	1.30	0.43
		Experiment 2g				Experiment 2h		
1	36.59	41.49	35.41	34.65	8.66	16.20	18.90	17.61
5	13.64	9.95	9.24	9.92	1.11	0.64	3.51	1.52
10	6.59	9.91	4.08	4.16	-0.34	1.25	-1.32	0.10
20	3.51	5.26	2.53	3.19	-0.34	-0.51	-1.66	-0.55
40	2.67	2.57	3.25	5.18	-1.04	0.58	0.02	-0.16
60	2.43	2.75	1.65	1.35	-0.96	0.15	-0.14	0.11
80	1.26	1.75	0.85	1.17	0.53	-0.44	-0.41	-0.44
100	2.39	1.06	2.75	1.61	-0.64	-0.63	0.90	0.38
		Experiment 2i				Experiment 2j		
1	16.31	21.80	21.12	15.75	16.20	19.32	12.11	14.42
5	2.57	1.83	2.04	1.00	4.41	3.52	3.75	2.41
10	1.15	3.17	2.74	3.01	2.99	3.65	3.18	2.88
20	0.12	1.10	-0.25	0.65	1.75	2.66	1.35	2.64
40	-0.02	0.89	0.76	-0.13	0.78	1.05	1.34	1.27
60	0.93	0.40	1.01	0.17	0.70	1.03	1.58	0.66
80	-0.06	0.24	0.00	0.23	0.87	0.70	0.79	1.13
100	0.82	0.42	0.02	-0.20	1.43	0.72	1.02	0.47