Stopping Criteria for Value Iteration on Stochastic Games with Quantitative Objectives

Jan Kretinsky, Tobias Meggendorfer, Maximilian Weininger

TECHNISCHE UNIVERSITÄT MÜNCHEN

Institute of Science and Technology Austria

Talk in one slide

- **Probabilistic systems**: Best algorithm (usually) is **Value Iteration (VI)**
- But: Requires a **stopping criterion**
  For **Stochastic Games (SG)** with most **infinite-horizon, quantitative objectives** there is **none!**

- This paper: **Uniform** solution for **large class of quantitative objectives** (including total reward, mean payoff, …)

Talk in one slide

- **Probabilistic systems**: Best algorithm (usually) is **Value Iteration (VI)**
- But: Requires a **stopping criterion** for correctness
  For **Stochastic Games (SG)** with most **infinite-horizon, quantitative objectives** there is **none!**

1-player SG: separate papers giving stopping criteria for each objective [BCC+14, HM14, BKL+17, ACD+17].

- This paper: **Uniform** solution for **large class of quantitative objectives** (including total reward, mean payoff, …)

Talk in one slide

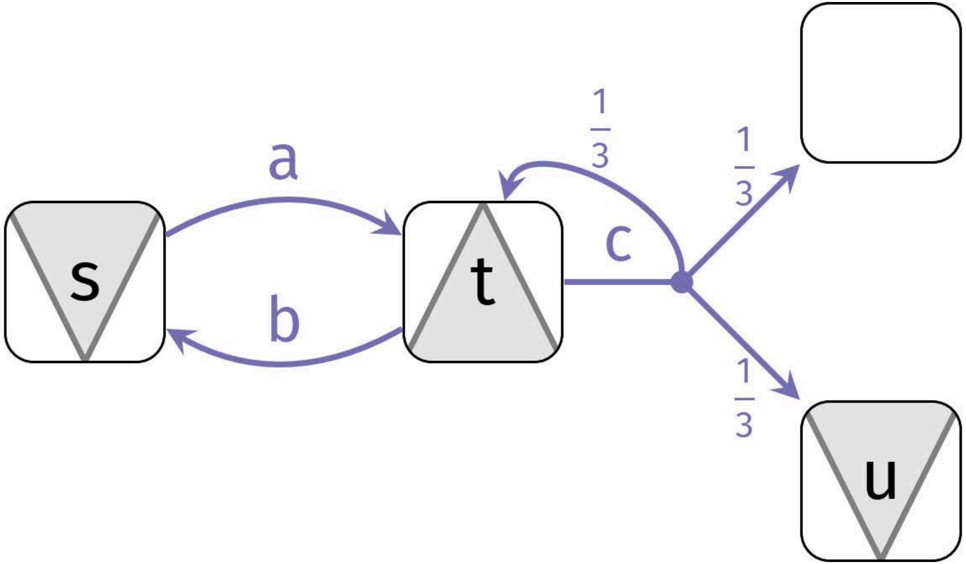
- **Probabilistic systems**: Best algorithm (usually) is **Value Iteration (VI)**
- But: Requires a **stopping criterion** for correctness
  For **Stochastic Games (SG)** with most **infinite-horizon, quantitative objectives** there is **none!**

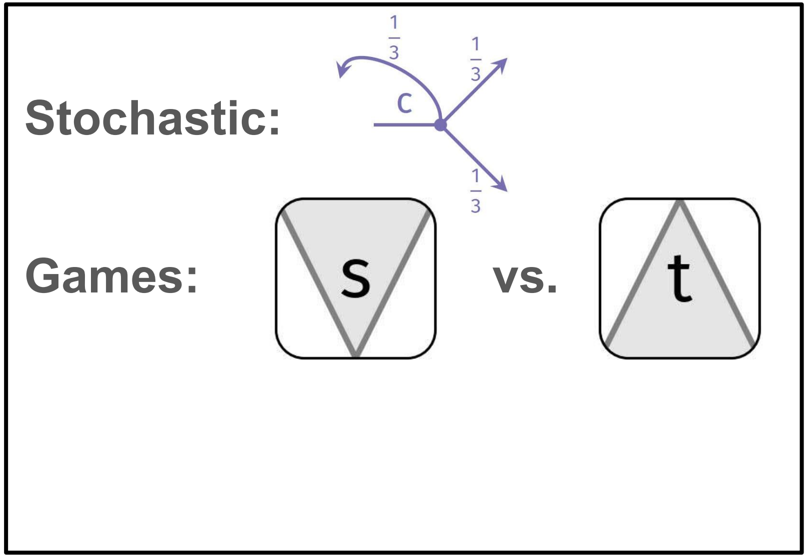
1-player SG: separate papers giving stopping criteria for each objective [BCC+14, HM14, BKL+17, ACD+17].

- This paper: **Uniform** solution for **large class of quantitative objectives** (including total reward, mean payoff, …)

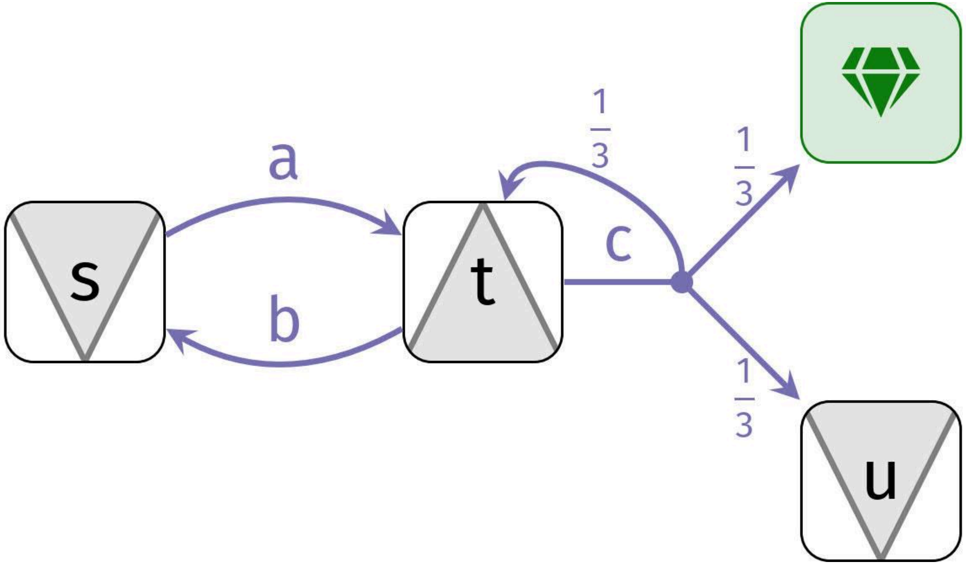
Unifies all previous ones and is more broadly applicable.

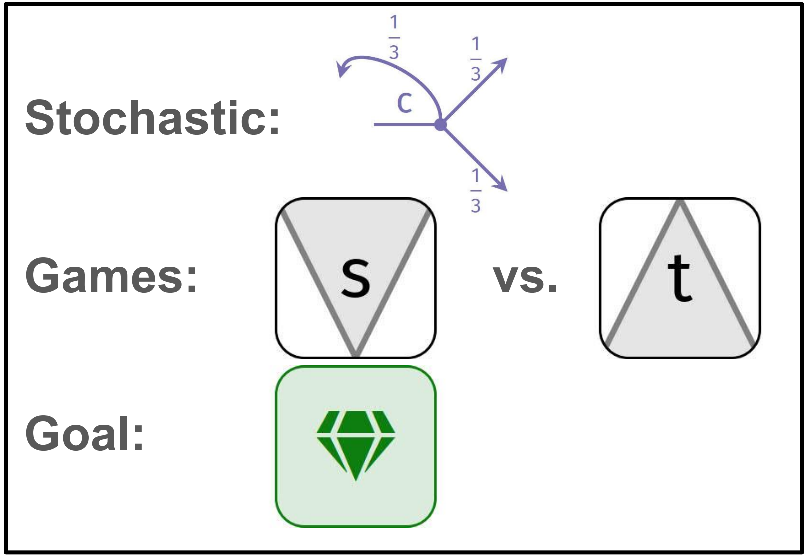
Stochastic Games



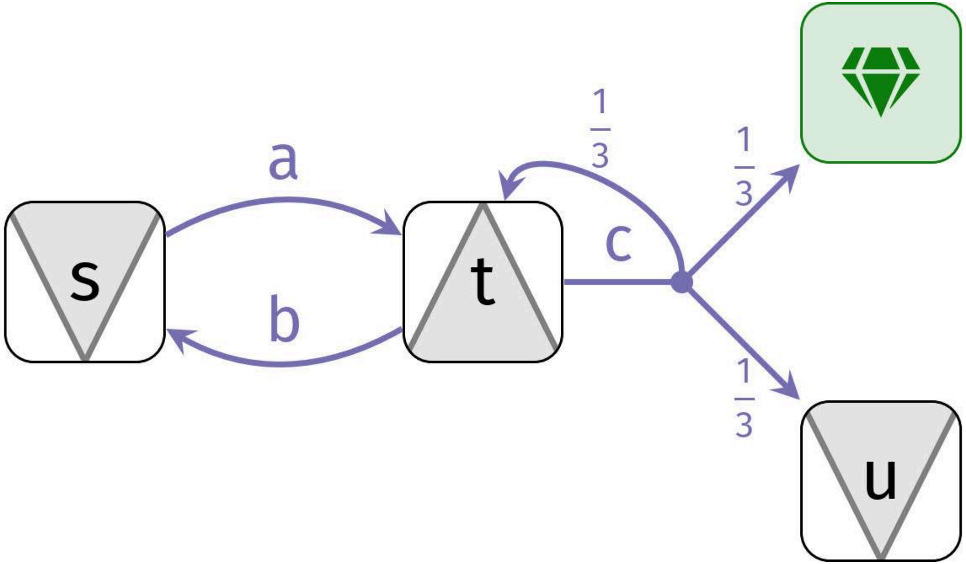


Stochastic Games



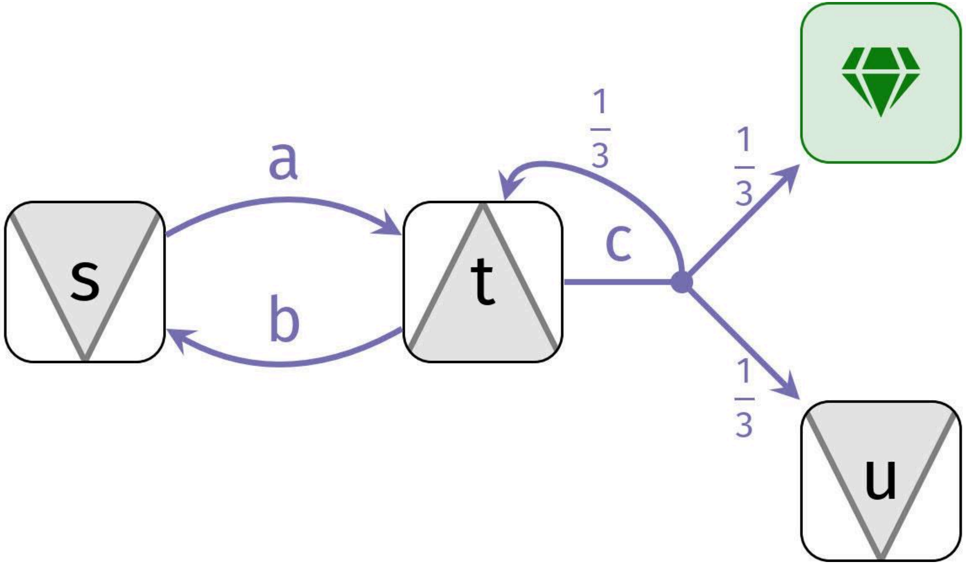


Stochastic Games and Value Iteration



Iteration	L(s)	L(t)
0	0	0
1		
2		
…		

Stochastic Games and Value Iteration

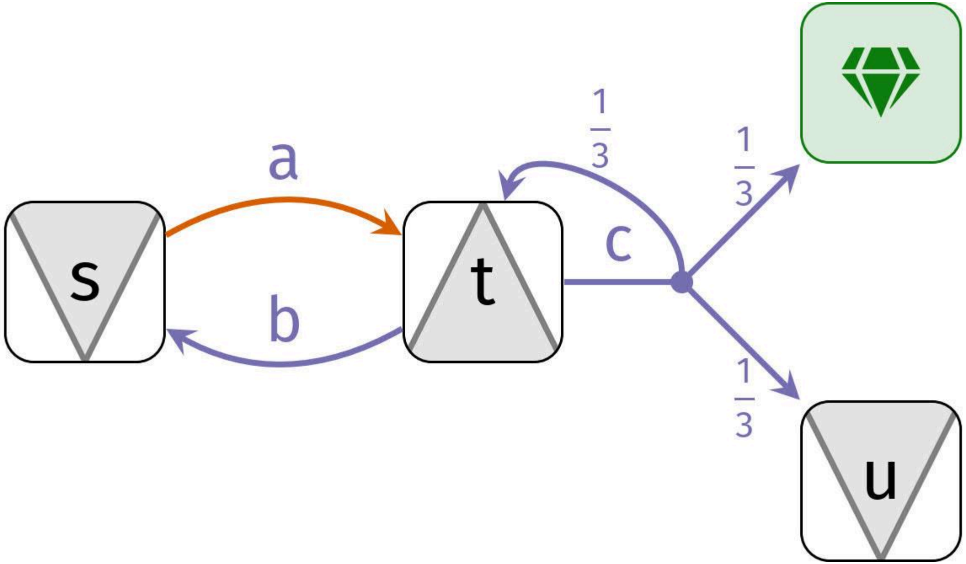


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s,a)$$

Iteration	L(s)	L(t)
0	0	0
1		
2		
…		

Stochastic Games and Value Iteration

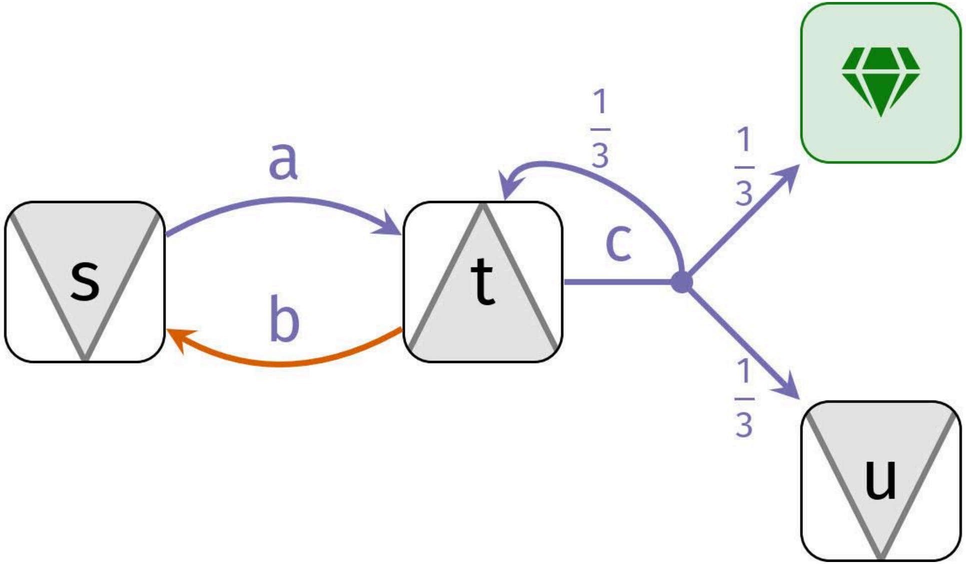


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s,a)$$

Iteration	L(s)	L(t)
0	0	0
1	0	
2		
…		

Stochastic Games and Value Iteration

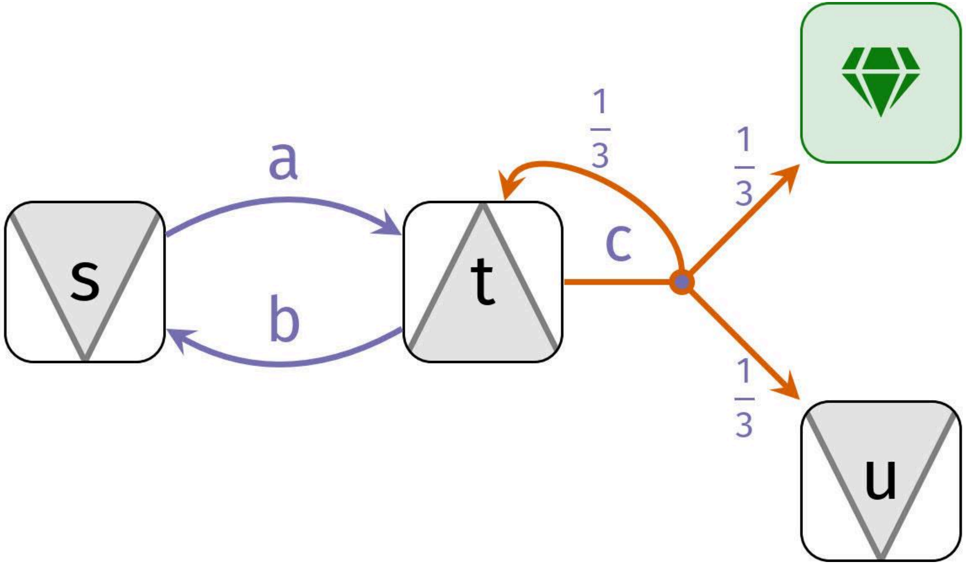


Bellman update:

$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s,a)$$

Iteration	L(s)	L(t)
0	0	0
1	0	
2		
…		

Stochastic Games and Value Iteration

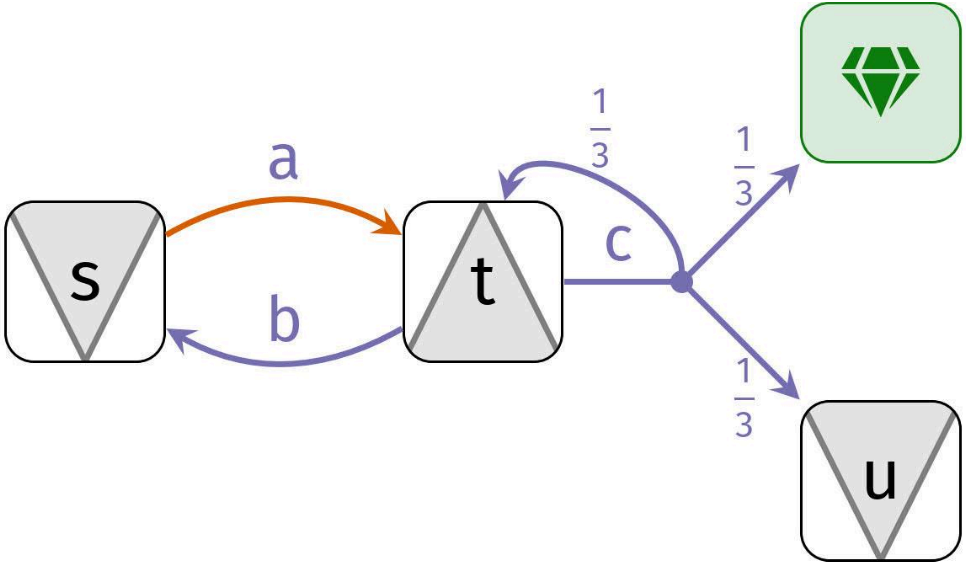


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$

Iteration	L(s)	L(t)
0	0	0
1	0	1/3
2		
…		

Stochastic Games and Value Iteration



Bellman update:

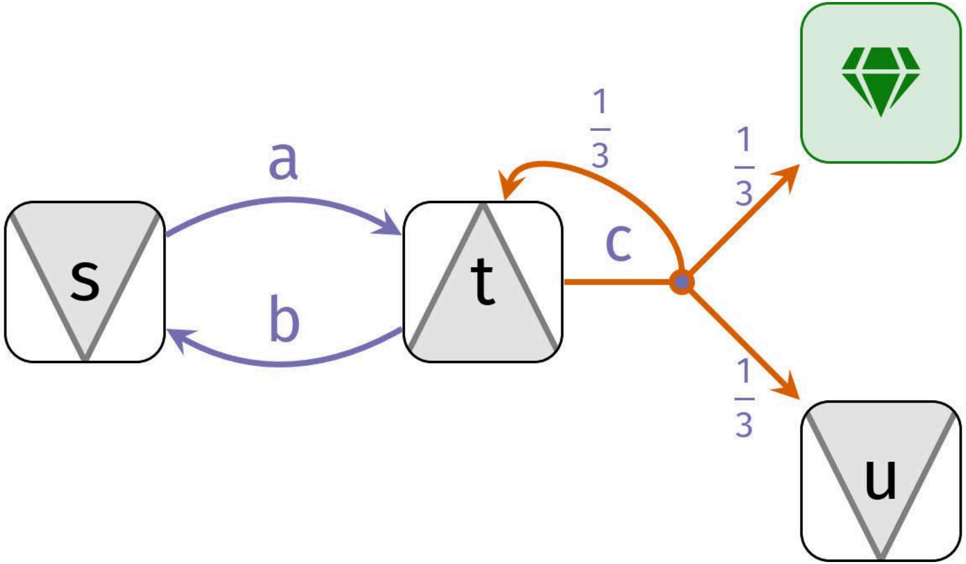
$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s, a)$$

Iteration	L(s)	L(t)
0	0	0
1	0	1/3
2	1/3	
…		

Stochastic Games and Value Iteration

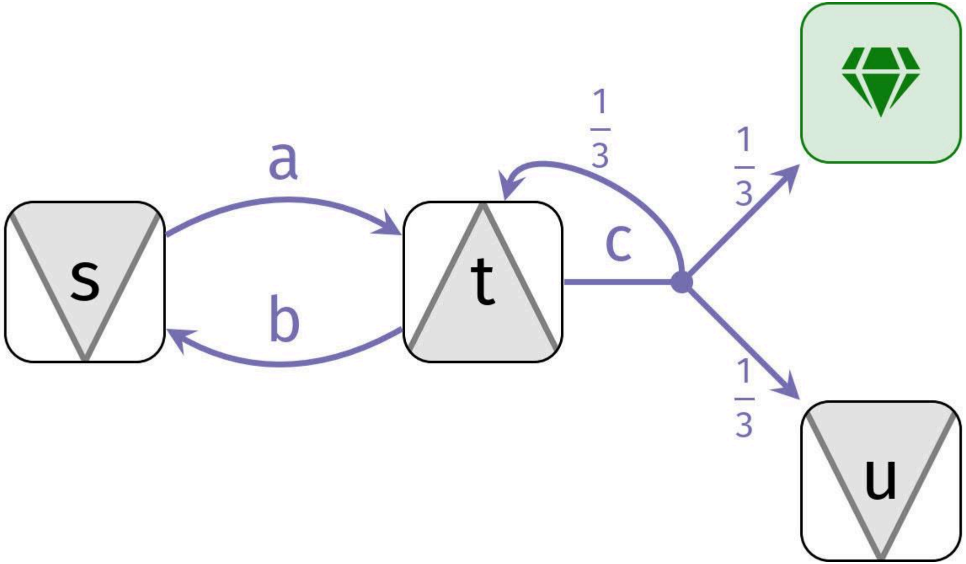
Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$



Iteration	L(s)	L(t)
0	0	0
1	0	1/3
2	1/3	4/9
…		

Stochastic Games and Value Iteration

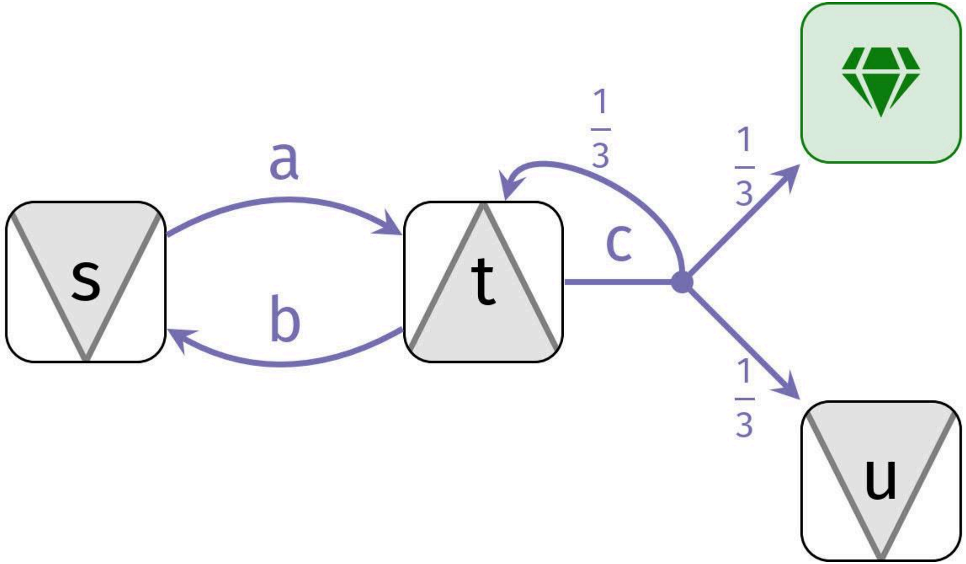


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s,a)$$

Iteration	L(s)	L(t)
0	0	0
1	0	1/3
2	1/3	4/9
…	…	…

Stochastic Games and Value Iteration



Bellman update:

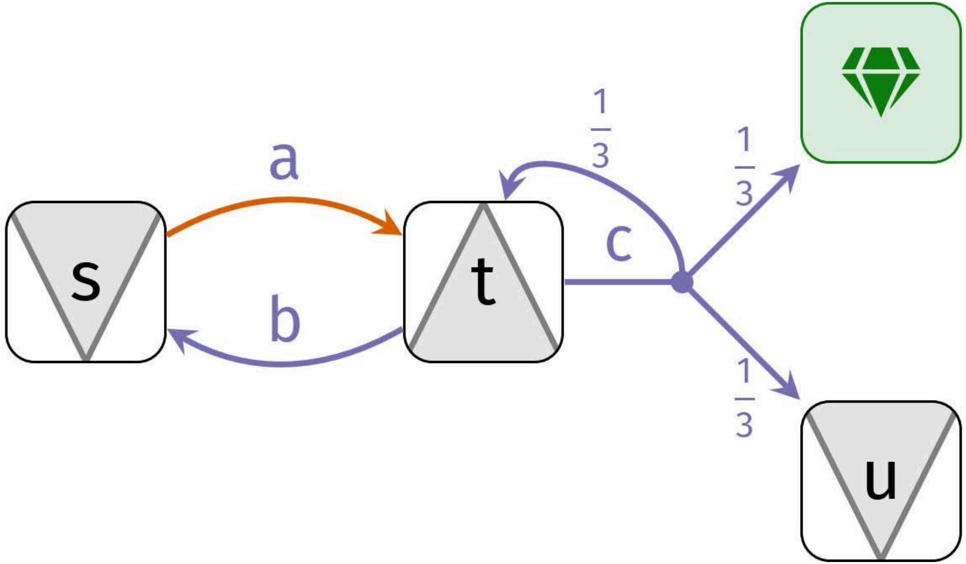
$$x_i(s) = \mathbf{opt}_a\ x_{i-1}(s,a)$$

Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3			
2	1/3	4/9			
…	…	…			

Stochastic Games and Value Iteration

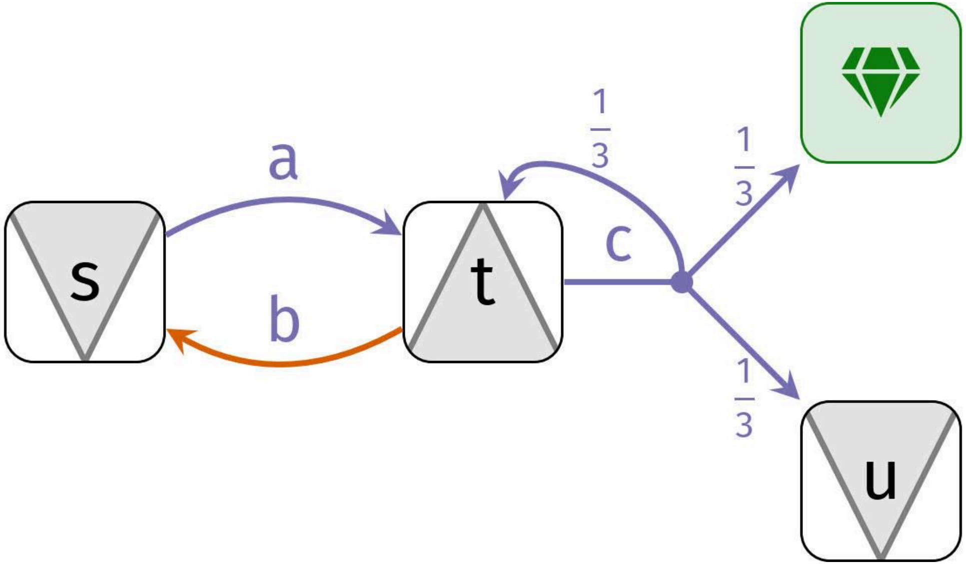
Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$



Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3		1	
2	1/3	4/9			
…	…	…			

Stochastic Games and Value Iteration



Bellman update:

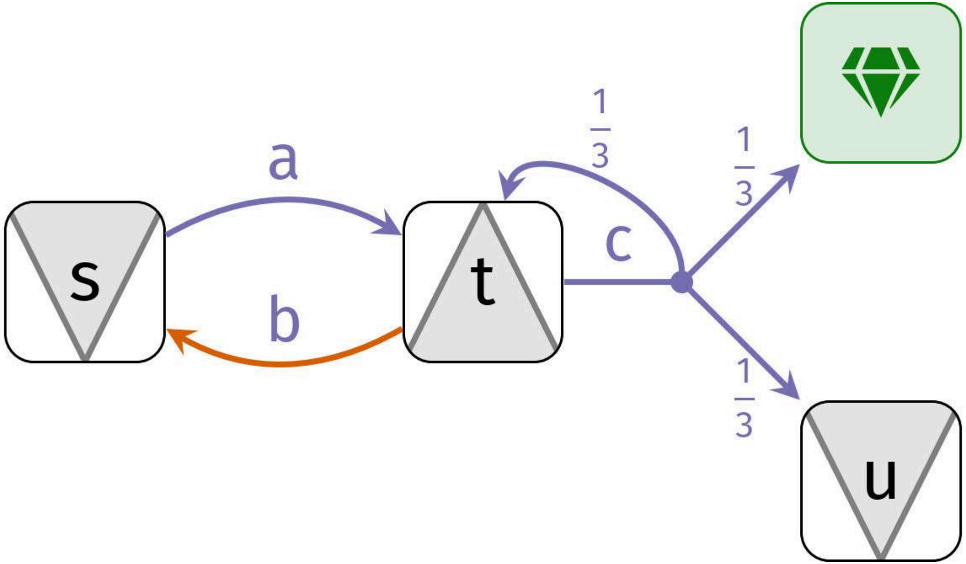
$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s,a)$$

Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3		1	1
2	1/3	4/9			
…	…	…			

Stochastic Games and Value Iteration

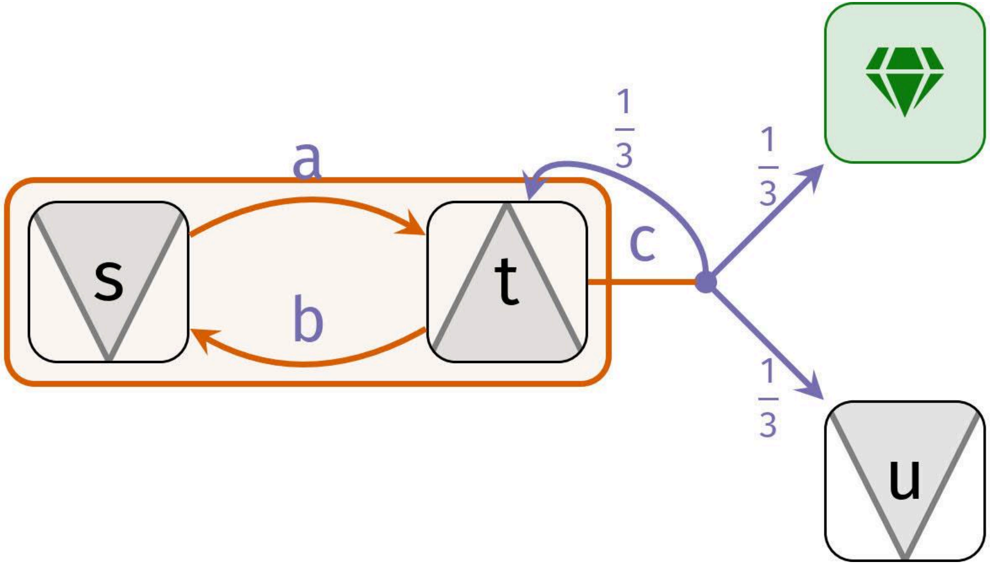
Bellman update:

$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s, a)$$



Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3		1	1
2	1/3	4/9		1	1
…	…	…		…	…

Stochastic Games and Value Iteration

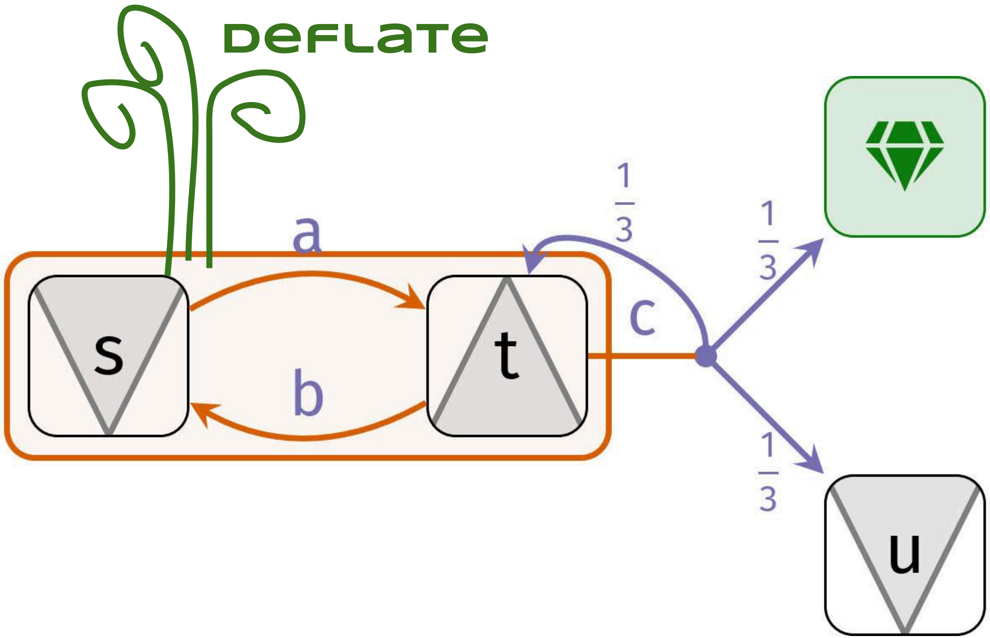


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$

Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3		1	1
2	1/3	4/9		1	1
…	…	…		…	…

Stochastic Games and Value Iteration

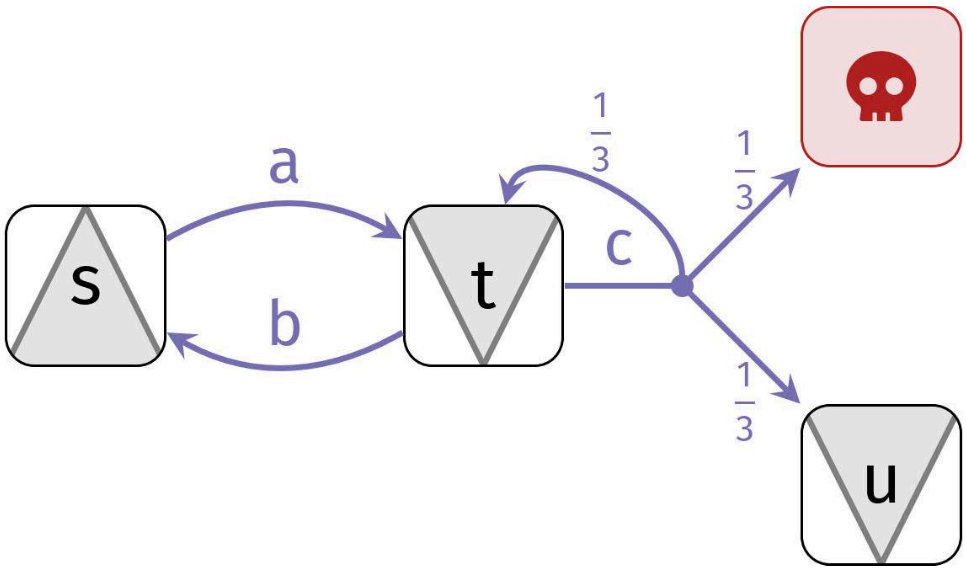


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s,a)$$

Iteration	L(s)	L(t)		U(s)	U(t)
0	0	0		1	1
1	0	1/3		1	**2/3**
2	1/3	4/9		**2/3**	**5/9**
…	…	…		…	…

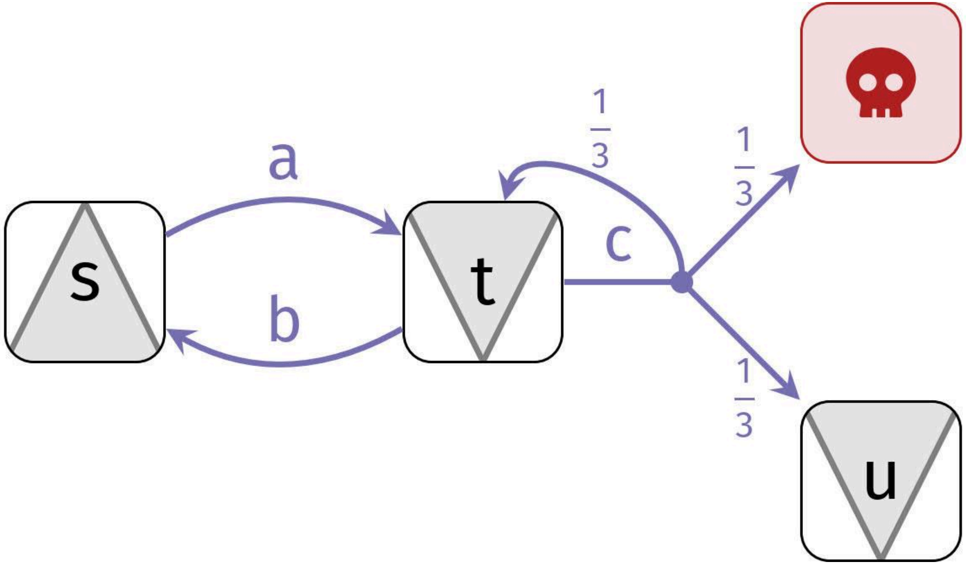
Stochastic Games and Value Iteration



Stochastic Games and Value Iteration

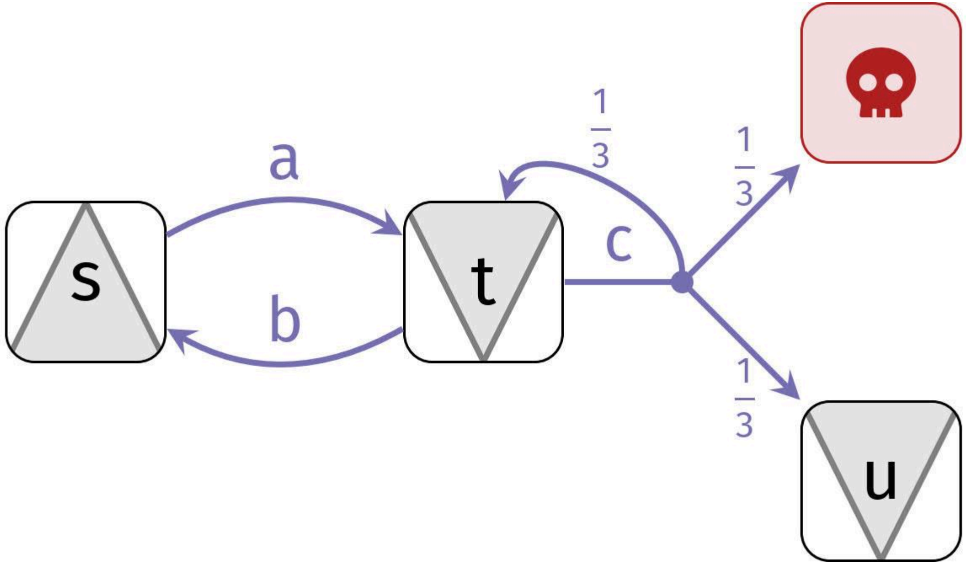
Bellman update:

$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s,a)$$



Iteration	L(s)	L(t)	L(u)
0	0	0	1
1			
2			
…	…	…	

Stochastic Games and Value Iteration

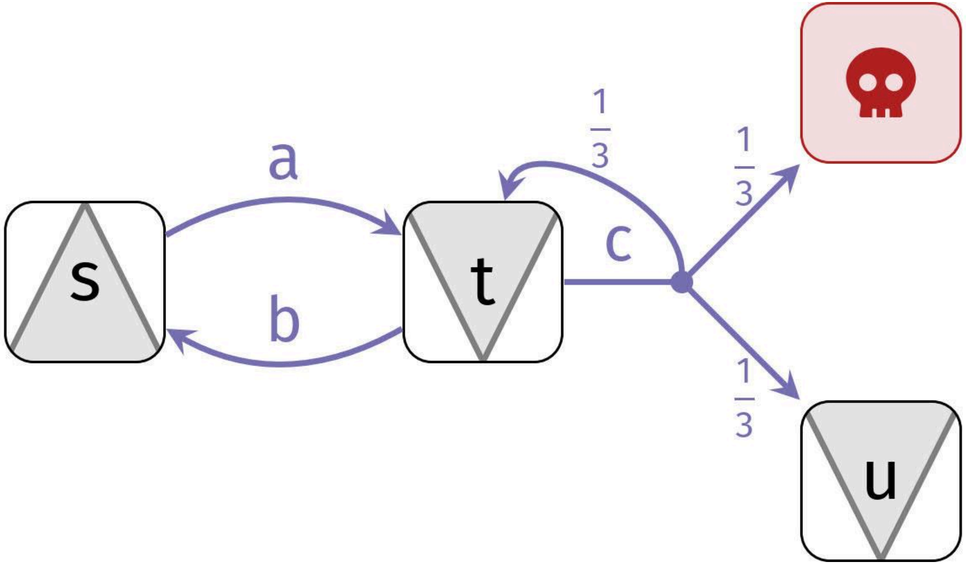


Bellman update:

$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s,a)$$

Iteration	L(s)	L(t)	L(u)
0	0	0	1
1			1
2			1
…	…	…	…

Stochastic Games and Value Iteration

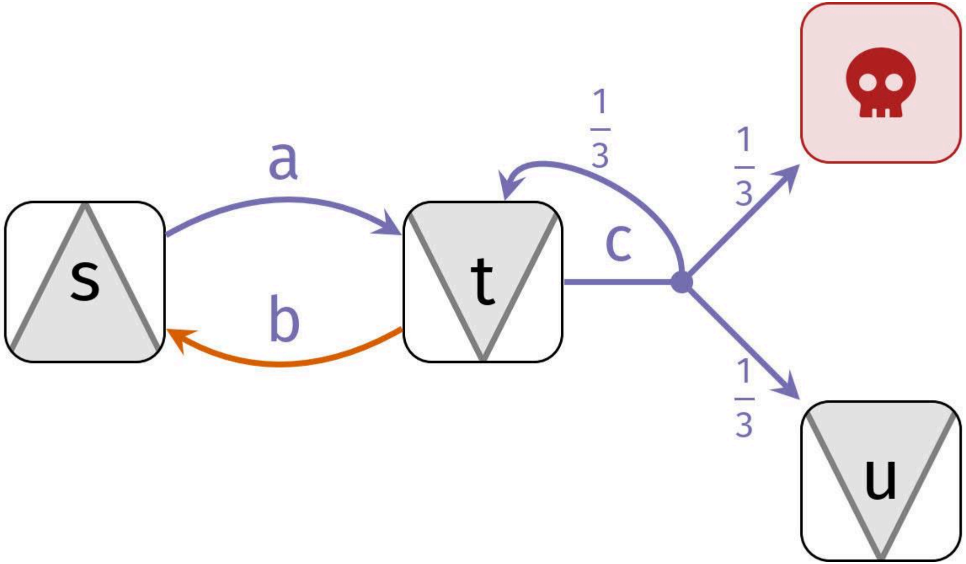


Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$

Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0		1
2			1
…	…	…	…

Stochastic Games and Value Iteration



Bellman update:

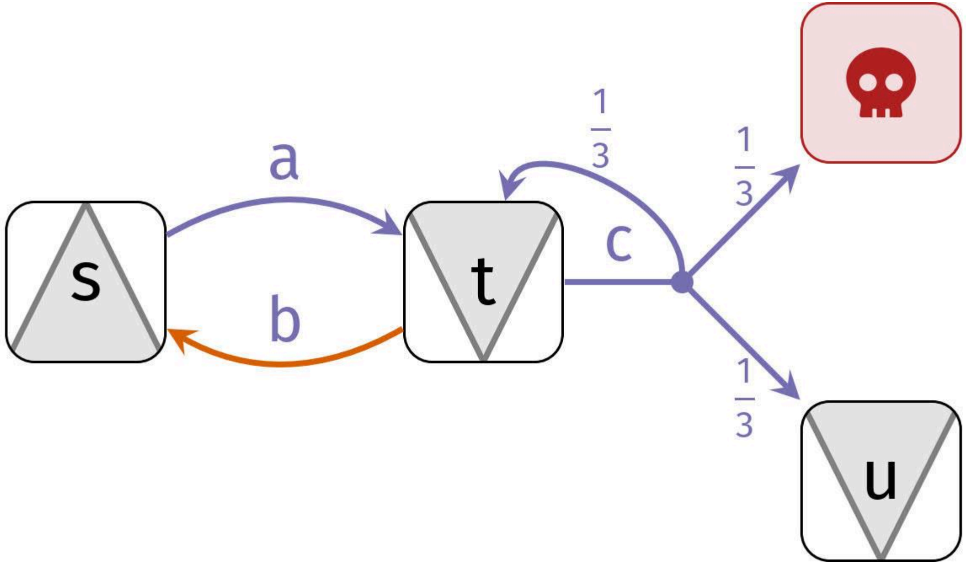
$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$

Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0	0	1
2			1
…	…	…	…

Stochastic Games and Value Iteration

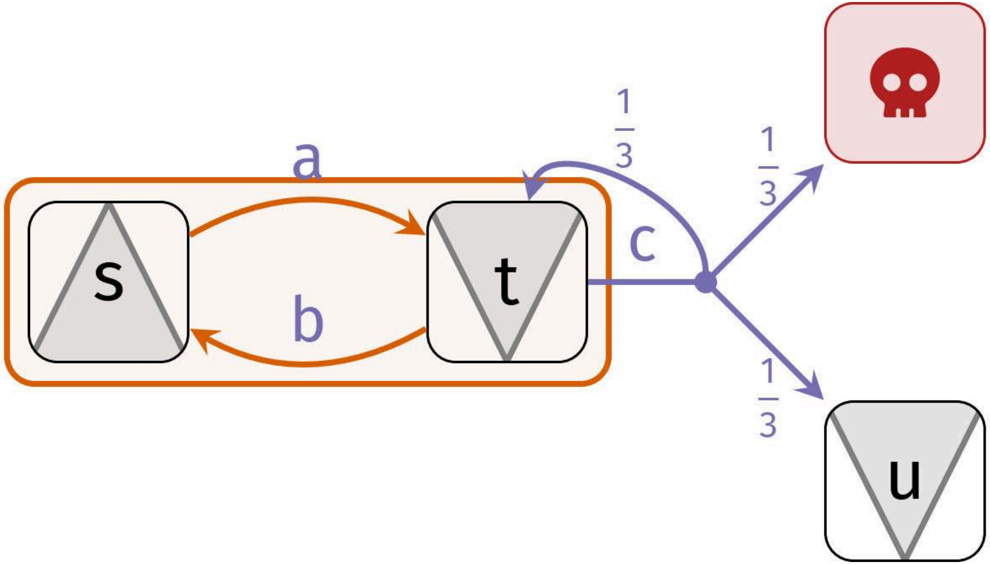
Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$



Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0	0	1
2	0	0	1
…	…	…	…

Stochastic Games and Value Iteration



Bellman update:

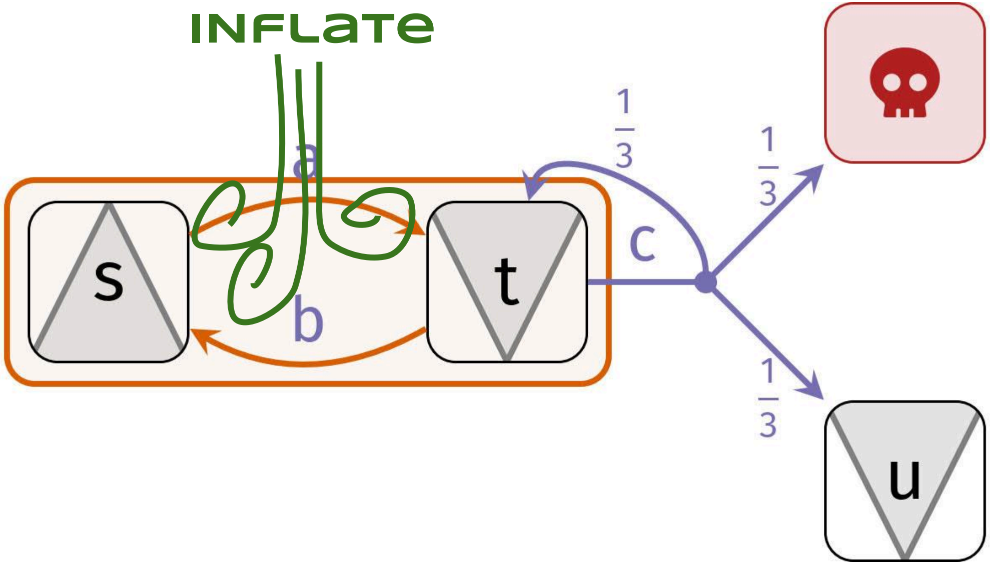
$$x_i(s) = \mathbf{opt}_a \, x_{i-1}(s,a)$$

Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0	0	1
2	0	0	1
…	…	…	…

Stochastic Games and Value Iteration

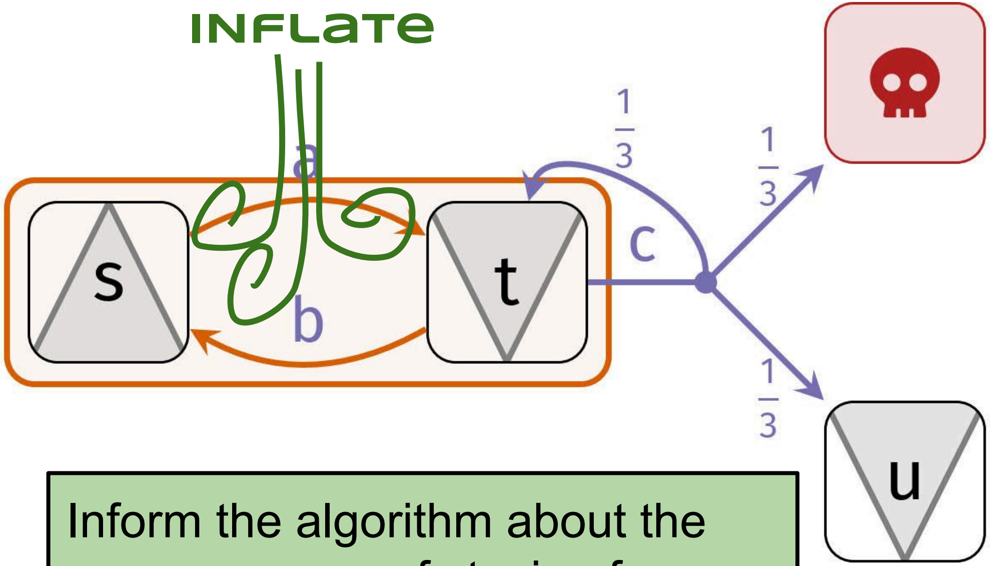
Bellman update:

$$x_i(s) = \mathbf{opt}_a \; x_{i-1}(s, a)$$



Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0	**1/3**	1
2	**1/3**	**4/9**	1
…	…	…	…

Stochastic Games and Value Iteration



Inform the algorithm about the consequences of staying forever.

Bellman update:

$$x_i(s) = \mathbf{opt}_a\, x_{i-1}(s,a)$$

Iteration	L(s)	L(t)	L(u)
0	0	0	1
1	0	**1/3**	1
2	**1/3**	**4/9**	1
…	…	…	…

Should I stay or should I go now?

Should I stay or should I go now?

Reachability:

stay=0

Safety:

stay=1

Should I stay or should I go now?

Reachability:

stay=0 max(stay,exit) = exit

Safety:

stay=1

Should I stay or should I go now?

Reachability:

stay=0 max(stay,exit) = exit

Safety:

stay=1

min(stay,exit) = exit

Should I stay or should I go now?

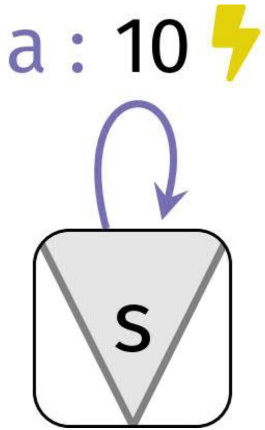
Reachability:

stay=0	max(stay,exit) = exit
min(stay,exit) = 0

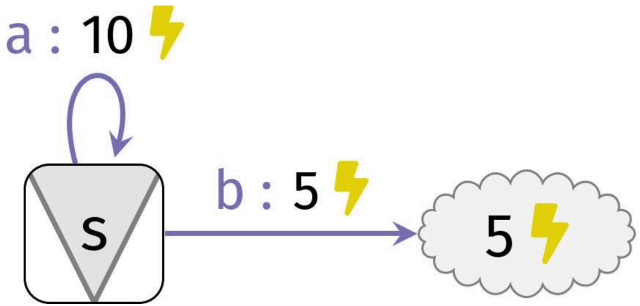
Safety:

stay=1	max(stay,exit) = 1
min(stay,exit) = exit

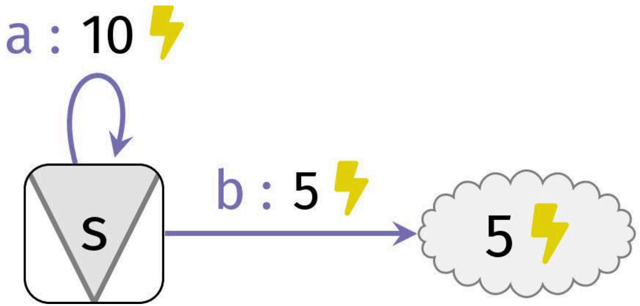
Deflating and Inflating for Mean Payoff



Deflating and Inflating for Mean Payoff

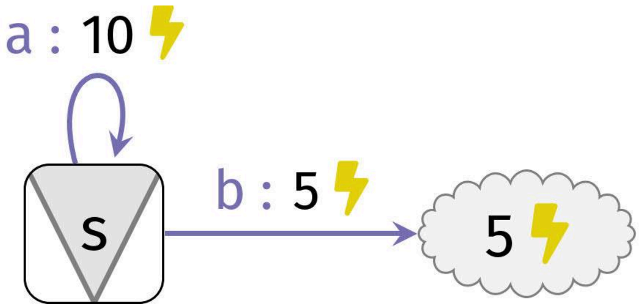


Deflating and Inflating for Mean Payoff

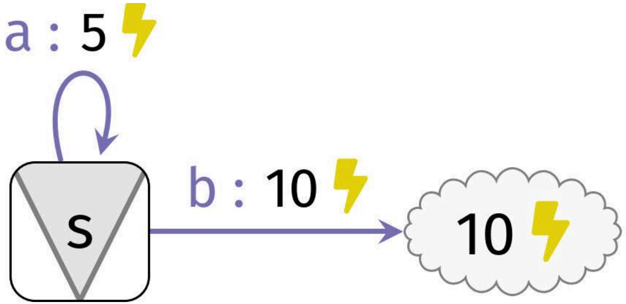


Inflate from 0 to **exit** value 5

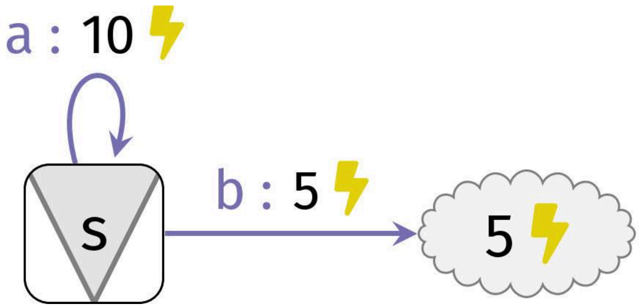
Deflating and Inflating for Mean Payoff



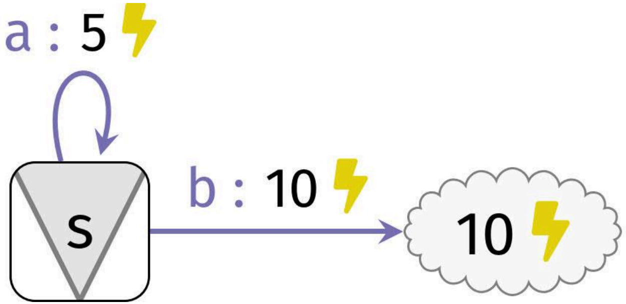
Inflate from 0 to **exit** value 5



Deflating and Inflating for Mean Payoff

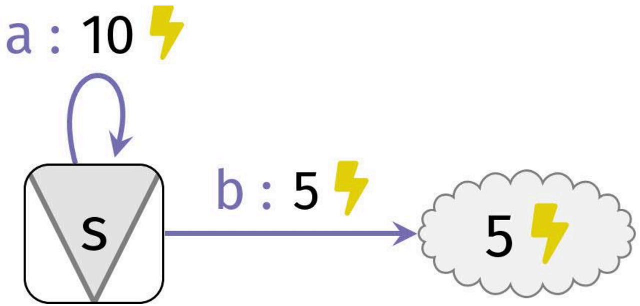


Inflate from 0 to **exit** value 5



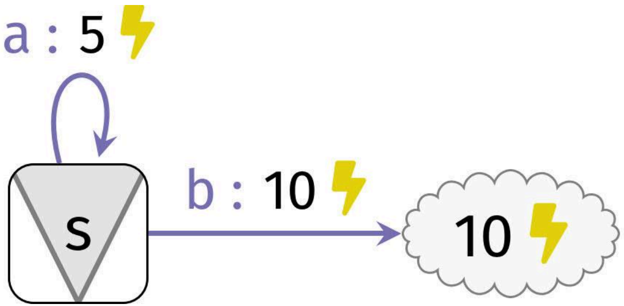
Inflate from 0 to **stay** value 5

Deflating and Inflating for Mean Payoff



Inflate from 0 to **exit** value 5

And dually for Maximizer states



Inflate from 0 to **stay** value 5

Where can I stay and go?

- Idea: If opponent thinks staying here is good,
  what happens if we really do stay?

Where can I stay and go?

- Idea: If opponent thinks staying here is good,
  what happens if we really do stay?

- Thus: Fix opponent’s strategy and analyse remaining cycles.

Where can I stay and go?

- Idea: If opponent thinks staying here is good,
  what happens if we really do stay?

- Thus: Fix opponent’s strategy and analyse remaining cycles

SOUND

Where can I stay and go?

- Idea: If opponent thinks staying here is good,
  what happens if we really do stay?

- Thus: Fix opponent’s strategy and analyse remaining cycles

SOUND

COMPLETE

Where can I stay and go?

- Idea: If opponent thinks staying here is good,
  what happens if we really do stay?

- Thus: Fix opponent’s strategy and analyse remaining cycles

SOUND

COMPLETE

OBJECTIVE-INDEPENDENT

Conclusion

- Given: **Stochastic Games** with **quantitative objectives** (including reachability, safety, mean payoff, expected total reward, …),
- Goal: Solving them **quickly** and with **precision-guarantees**
- Approach: **Value Iteration** with our **new stopping criterion**

Idea: Inform the algorithm about the consequences of staying forever:
Should I stay or should I go now?

Unifies previous work [BCC+14, HM14, BKL+17, ACD+17, KKKW18, PTHH20]
in an **elegant, objective-independent way**