Introduction to *Monte Carlo Tree Search*

By Xi Huang @SIST, ShanghaiTech

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

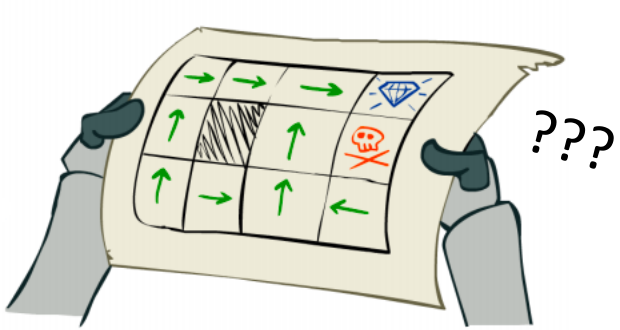
Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10

Markov Decision Process

Policy π : for each state s, $\pi(\cdot\,|s)$ as a mapping from A to [0, 1]



A Markov Decision Process is defined by a tuple $< S, A, P, R, \gamma >$, where

☆ S is a (finite) set of states

☆ A is a finite set of actions

☆ P is a state transition probability matrix

$$P^{a}_{s,s'} = P[S_{t+1} = s' | S_t = s, A_t = a]$$

☆ R is a reward function

$$R^{a}_{s} = E[R_{t+1} | S_t = s, A_t = a]$$

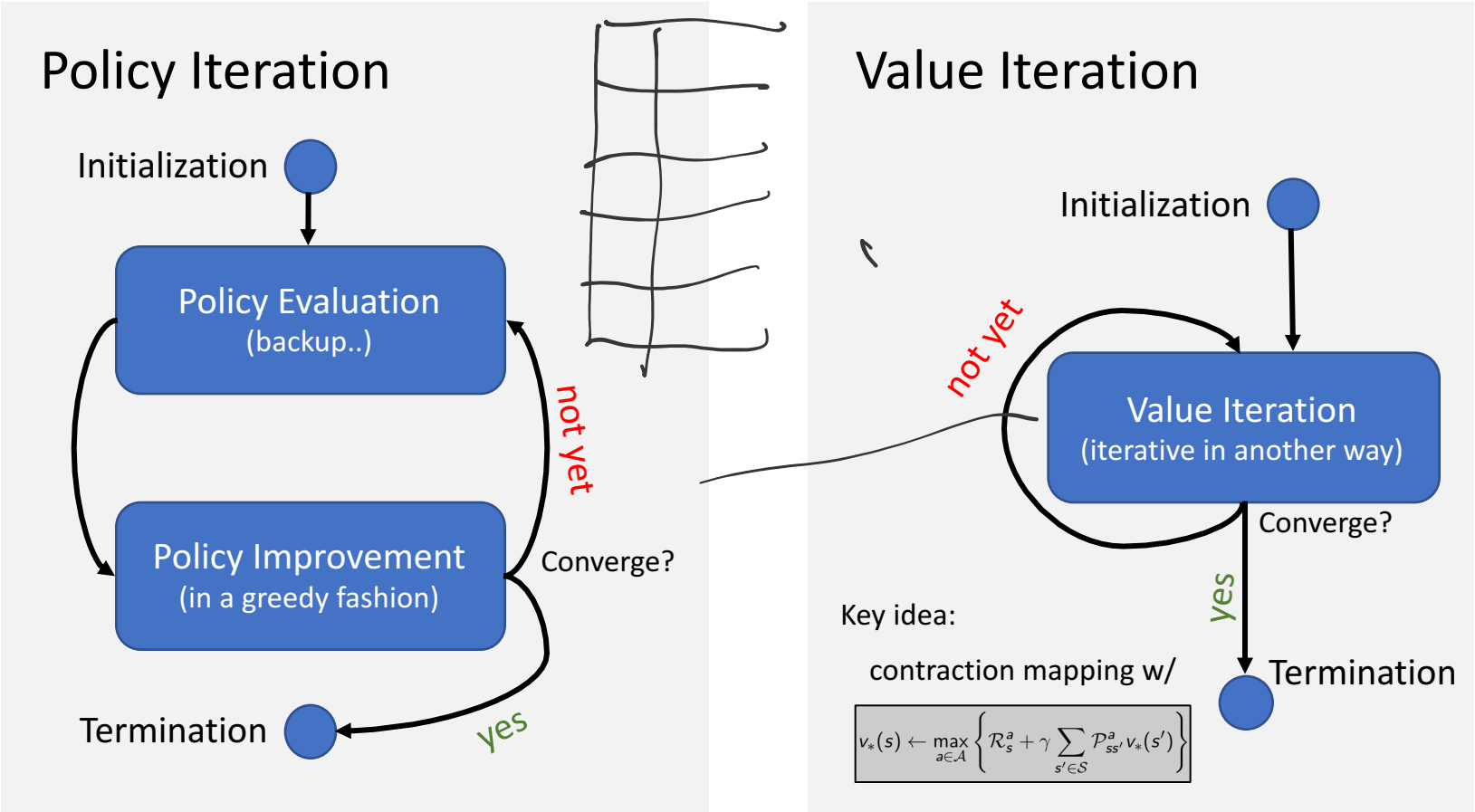
☆ γ is a discount factor such that $\gamma \in [0, 1]$

Solving MDP via DP

Tabular Methods

☆ State-value function: $v_\pi(s) = E[R_{t+1} + \gamma R_{t+2} + \cdots | S_t = s]$

☆ Action-value function: $q_\pi(s, a) = E[R_{t+1} + \gamma R_{t+2} + \cdots | S_t = s, A_t = a]$



Limitations of Pure DP Algs

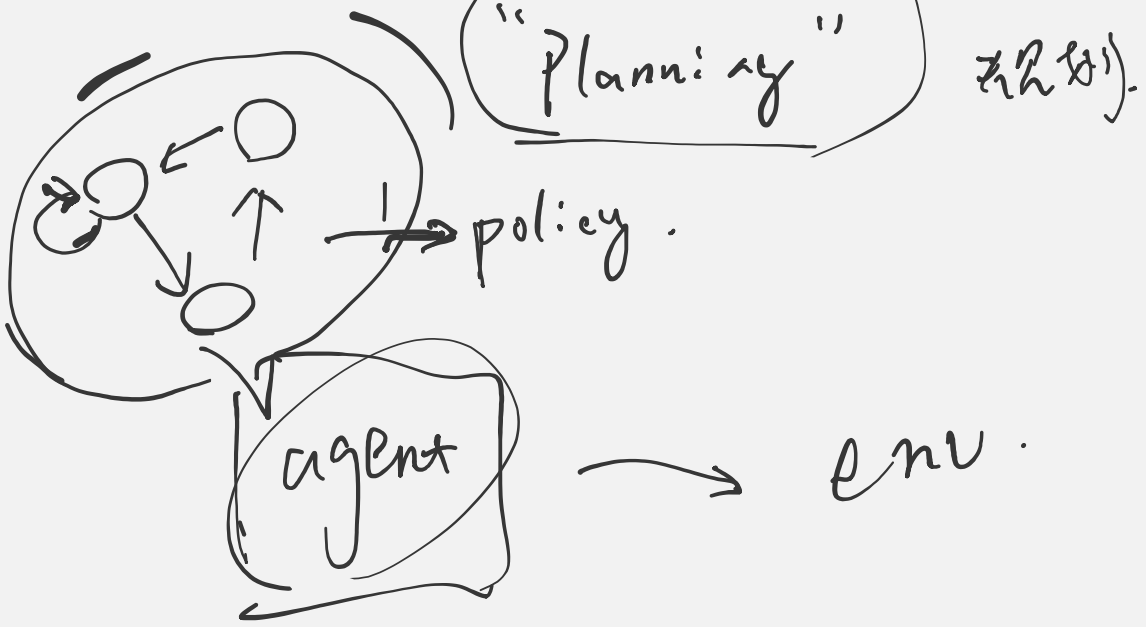
☆ Efficiency (should be addressed; better if independent of state space size)

- Computational complexity polynomial in the state space size (*e.g.,* ~10^{170})
- "Curse of dimensionality" vs. **[resource scarcity & time limit of decision-making]**

Limitations of Pure DP Algs

☆ Availability of dynamics (leverage those given right on the spot)

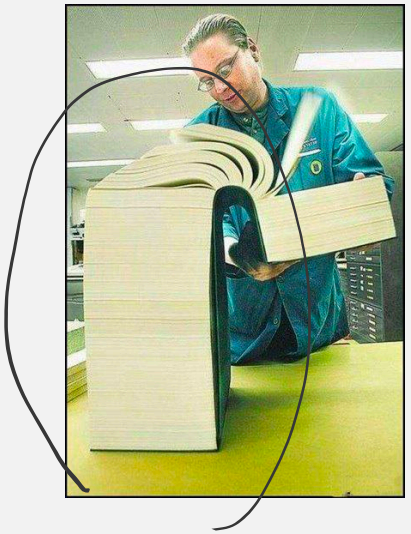
- Full knowledge of environment dynamics (required to be *Markovian*) should be explicitly given for decision making
- Usually incomplete or acquired through (simulated) interaction w/ the env.



Limitations of Pure DP Algs

☆ Necessity (why not do it in a lazy fashion)

- Only a small portion of states are frequently encountered, not ALL of them, so..



DOES the agent really need

a "THICK manual" beforehand?

Limitations of Pure DP Algs

☆ Efficiency (should be addressed; better if independent of state space size)

- Computational complexity polynomial in the state space size (*e.g.*, ~10^{170})
- "Curse of dimensionality" vs. **[resource scarcity & time limit of decision-making]**

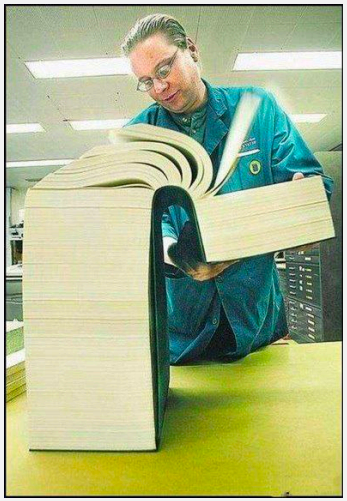
☆ Availability of dynamics (leverage those given right on the spot)

- Full knowledge of environment dynamics (required to be *Markovian*) should be explicitly given for decision making
- Usually incomplete or acquired through (simulated) interaction w/ the env.

☆ Necessity (do it in a lazy fashion)

- Only a small portion of states are frequently encountered, not ALL of them, so..

DOES it really need a "THICK manual" beforehand?



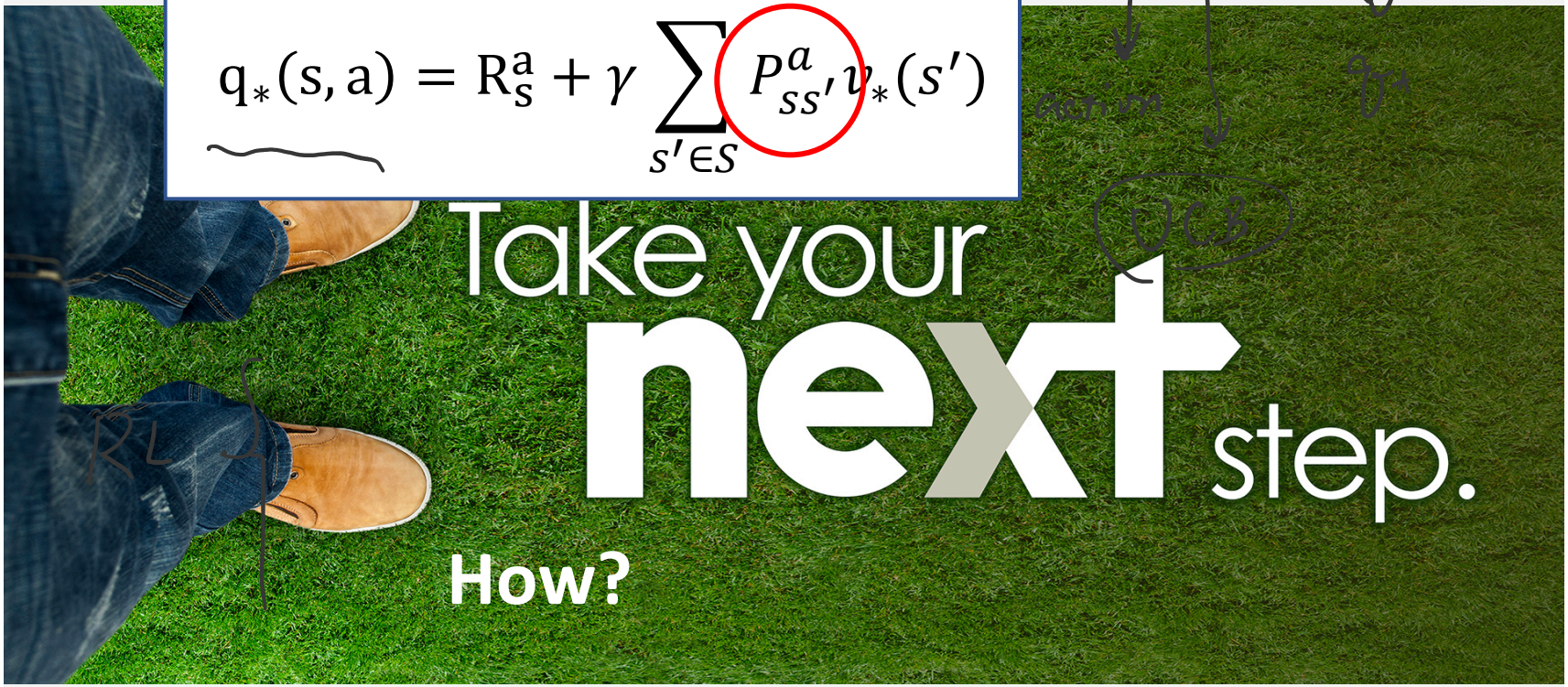
Key Idea of Monte Carlo Tree Search

☆ How the decision is made

$$v_*(s) = \max_{a \in A} q_*(s, a)$$

1. How to approximate?
   Monte Carlo Simulation
2. Exploration-exploitation?
   UCB method

$$q_*(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v_*(s')$$



Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

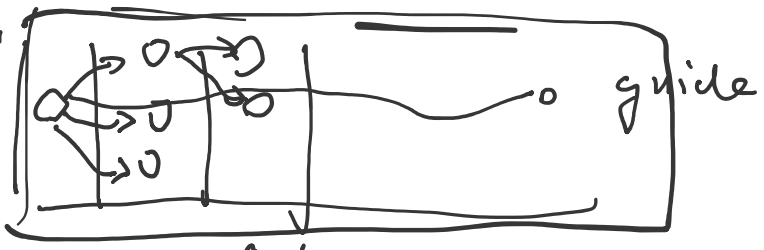
A Close Look at MCTS

Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10

History of MCTS



Timeline of RL

1987 — 1992 — 1997 — 2002 — 2005 — 2006 — 2008

MAB

2002

Prof. Auer *et al.* developed a series of UCB methods w/ log regrets (theoretically opt.)

2005

H. S. Chang *et al.* integrates bandit learning into simulation-based tree search for the approximate computing of optimal value functions in MDP

Prof. Remi Coulom

MCTS

Success in
- MoGo
- Zen
- AlphaGo
…

Timeline of game-tree search

1987

Dr. Bruce Abramson in his PhD thesis, CU

1992

Dr. Bernd Brügmann firstly applied in Go

1997

Deep Blue vs. Kasparov

2006

Dr. K. Levente & Prof. C. Szepesvári introduced UCB method to MCTS *a.k.a.* UCT

2016

Source: https://medium.com/point-nine-news/

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

Remarks

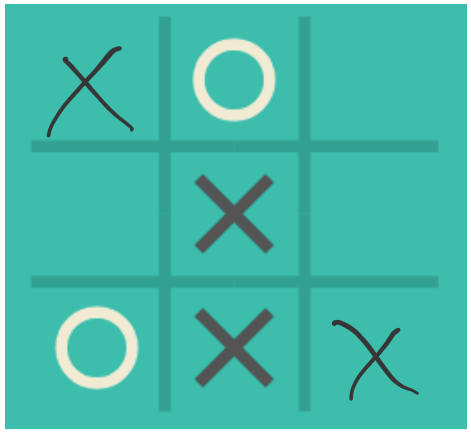
MCTS for Combinatorial Optimization Problems

MCMC HW Q10

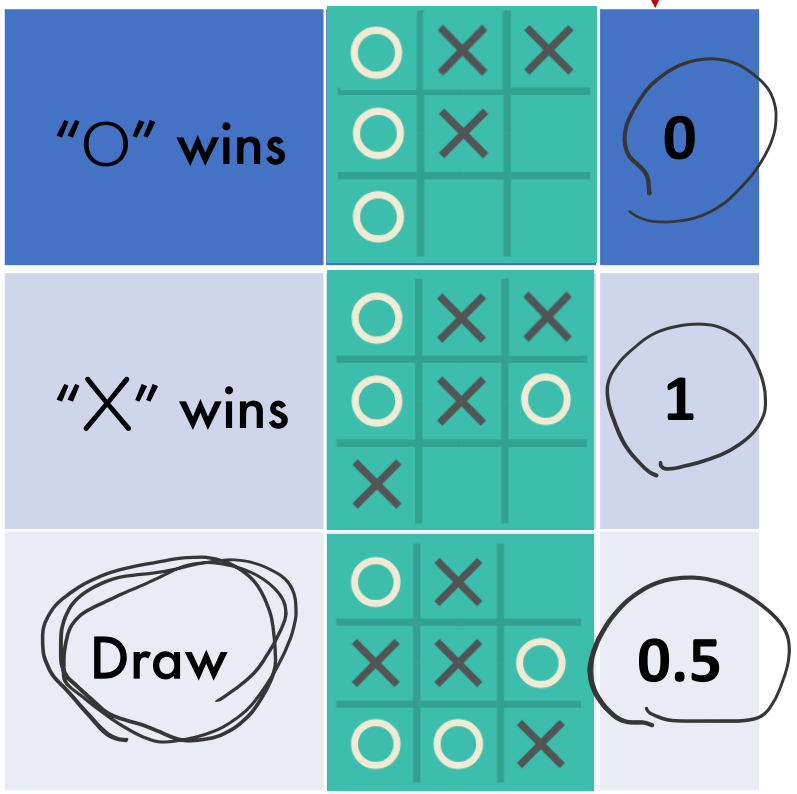
Monte Carlo Tree Search (MDP)

Sequential Decision-Making Problems as Tree-Search problems

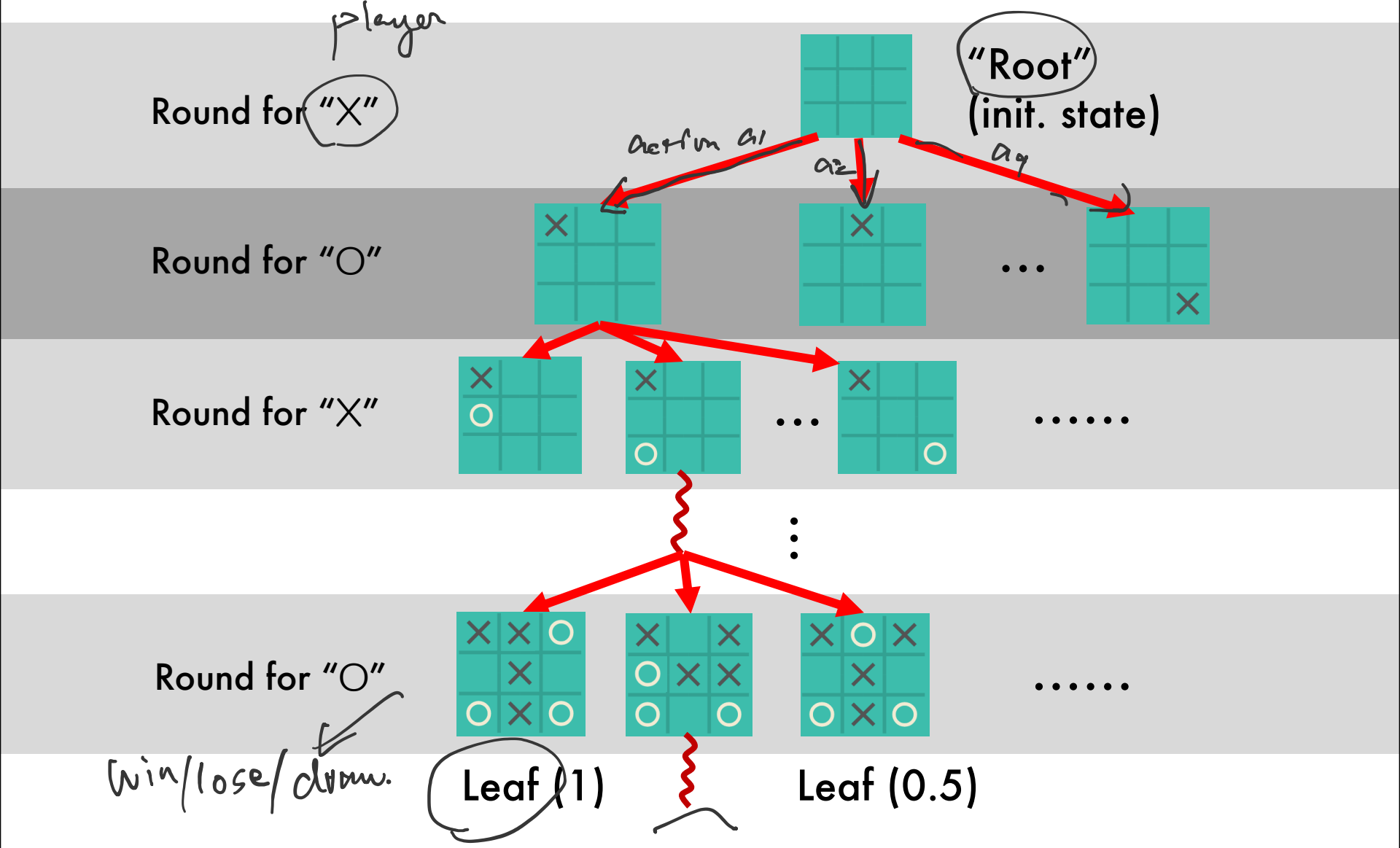
- Sequential Decision-Making Problems



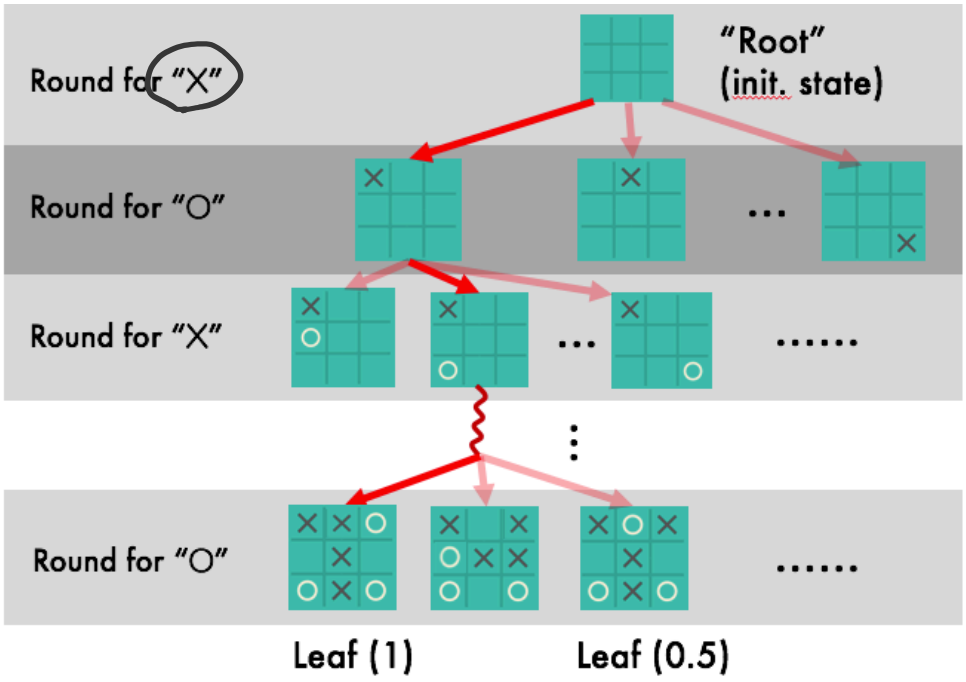
Tic-Tac-Toe



Game-play Tree for "Tic-Tac-Toe"

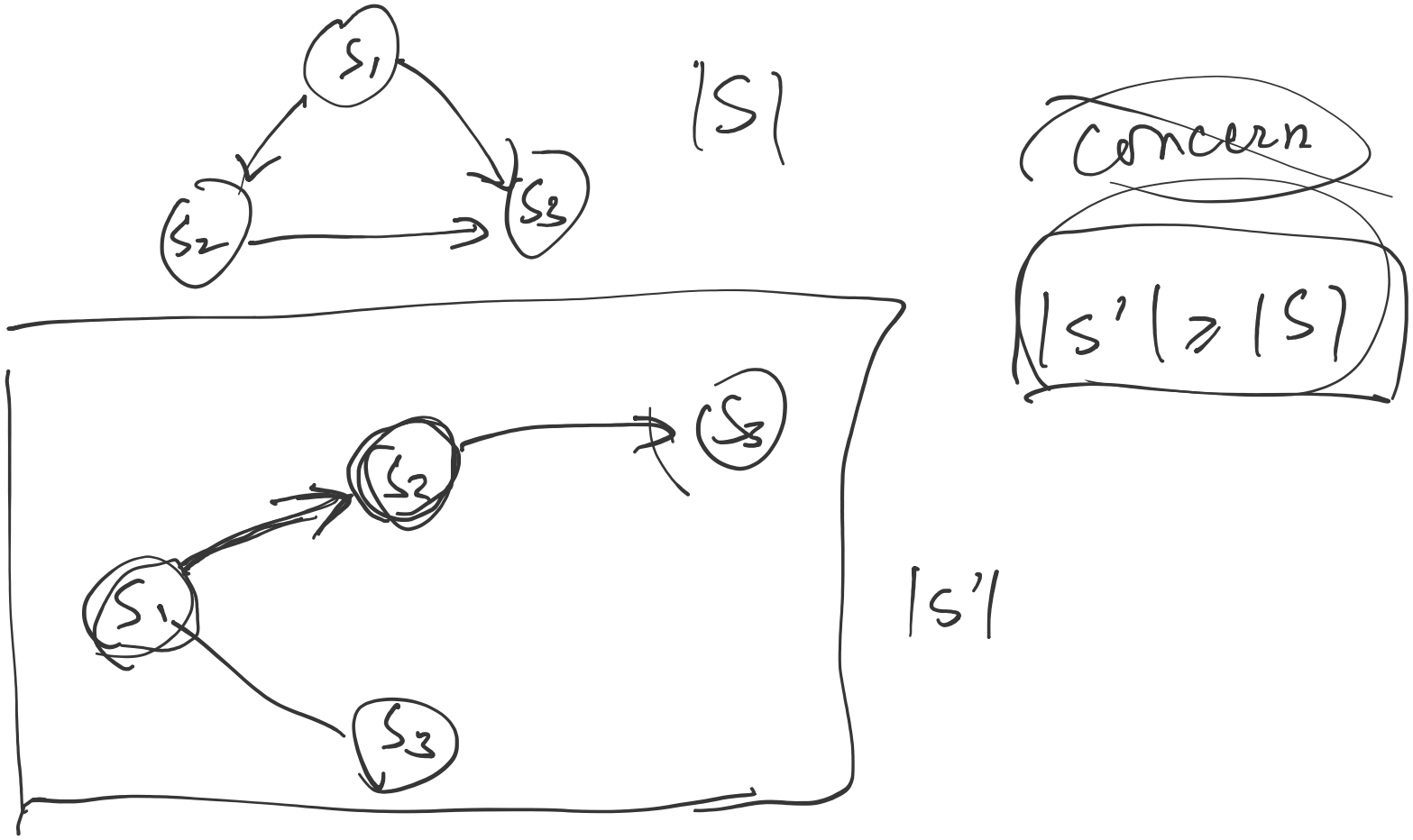


Game-play Tree for “Tic-Tac-Toe”

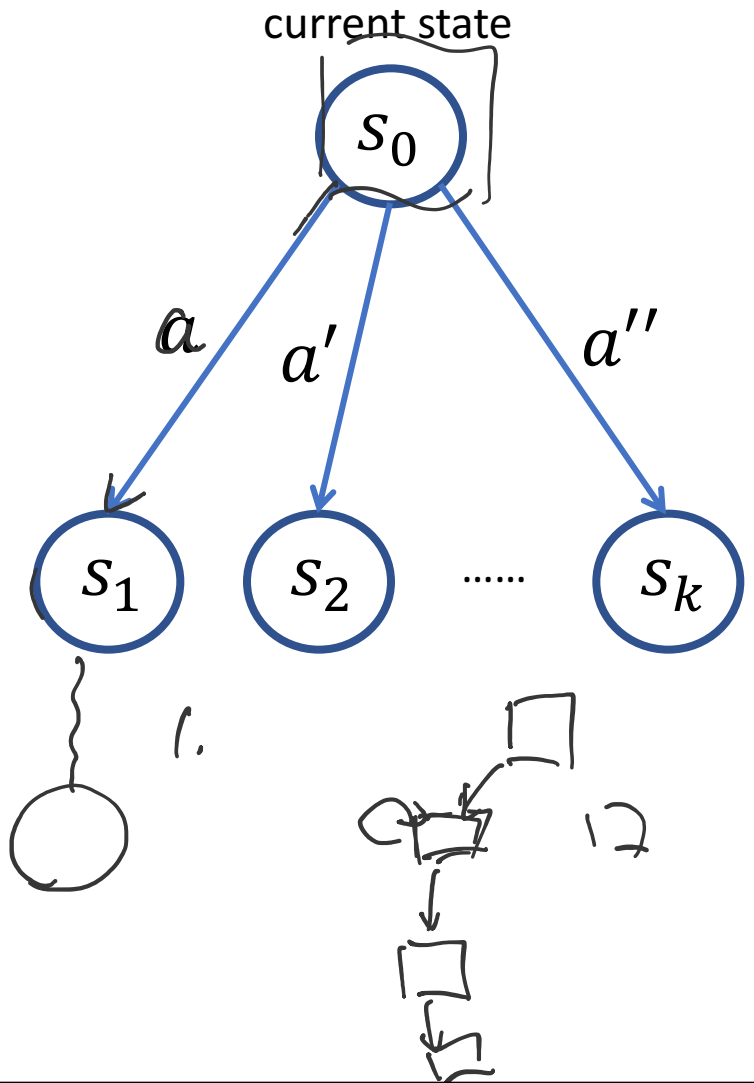


Optimal Policy: Find a path that leads to the best possible outcome (*a.k.a. planning*)

What about MDP?



Simplified Assumptions for Illustration



$$\max_{a \in A} q_*(s_0, a)$$

$$q_*(s_0, a) = R^a_{s_0} + \gamma \sum_{s' \in S} P^a_{s_0 s'} v_*(s')$$

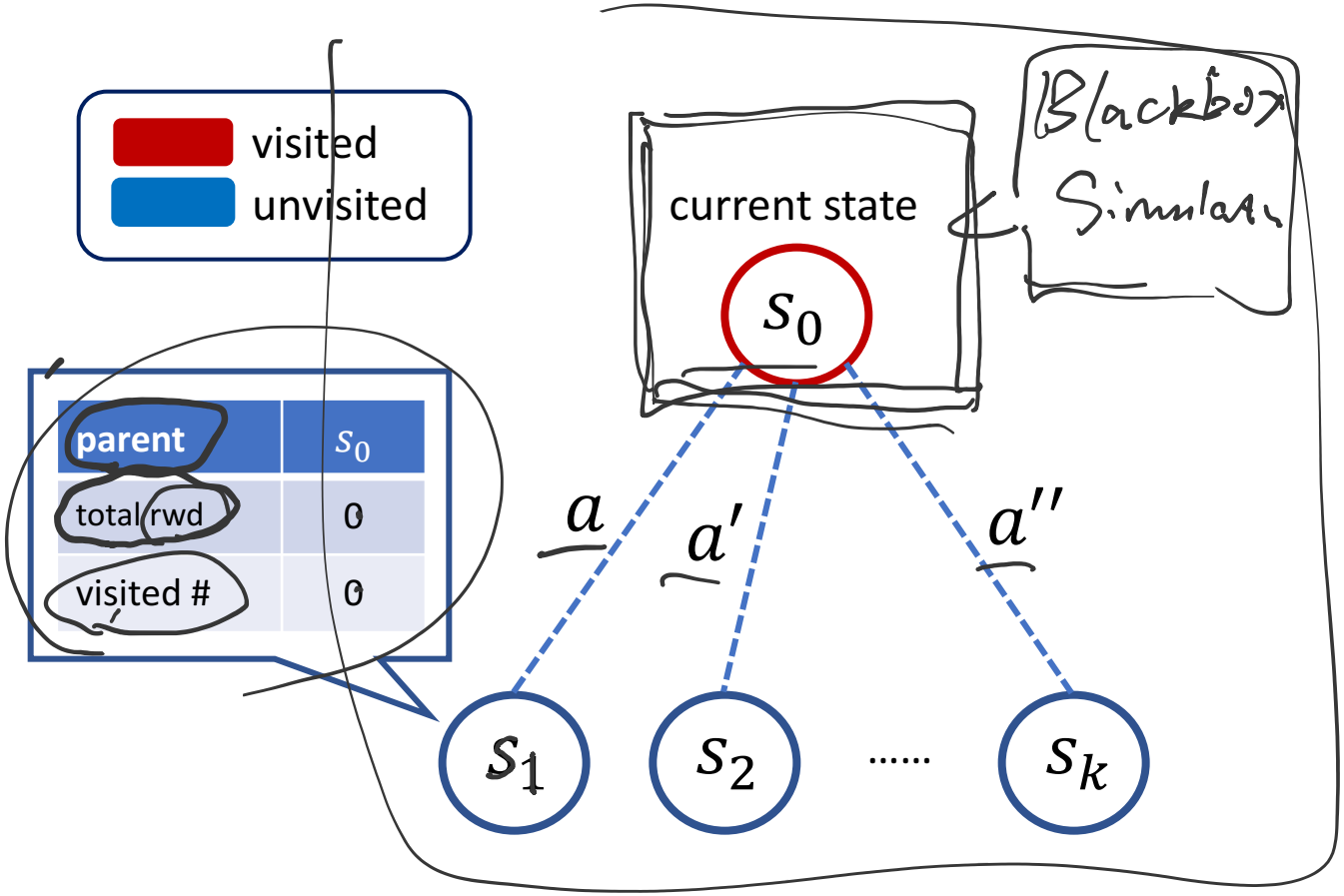
Reward

0

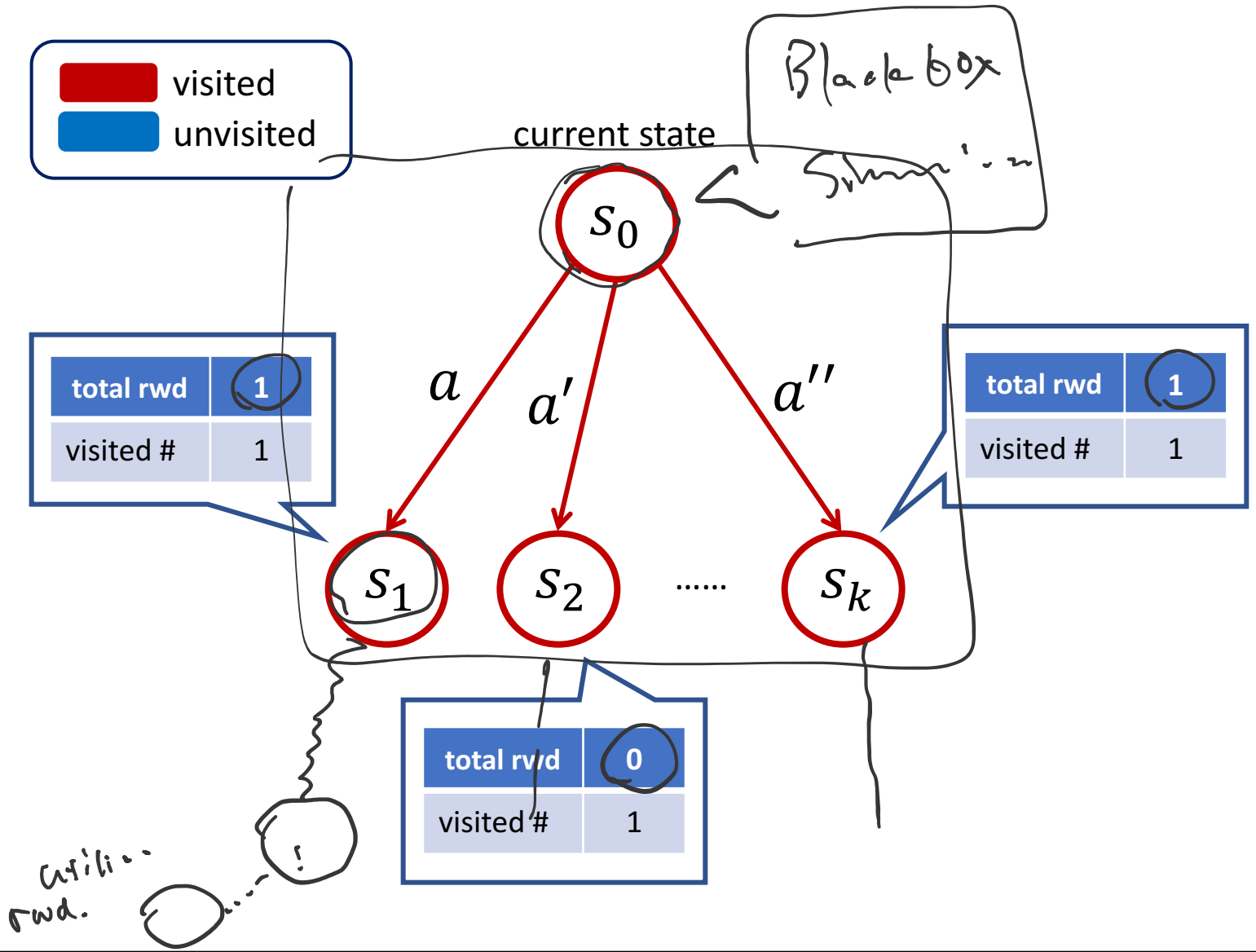
0

1

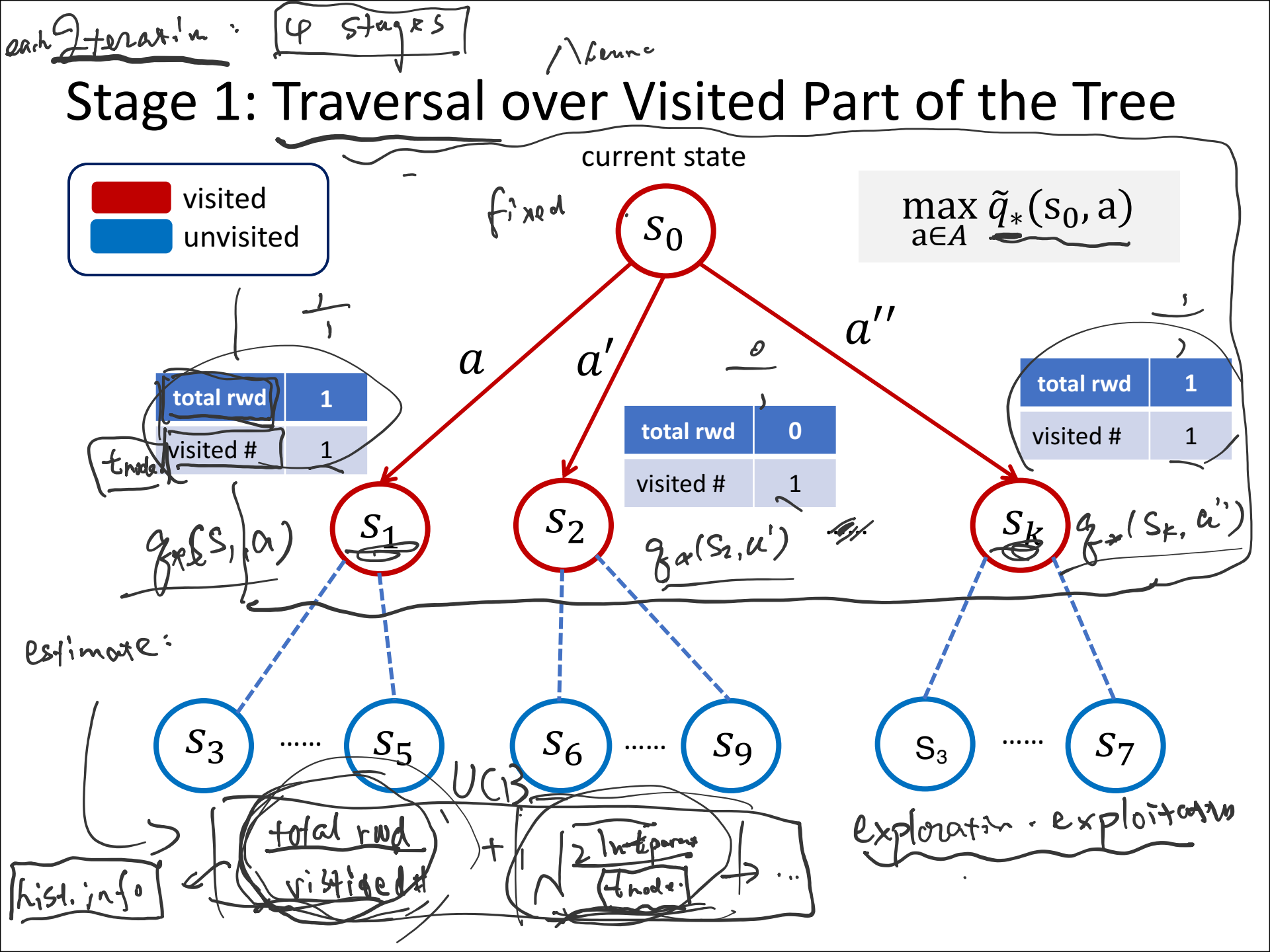
MCTS: Initial trials



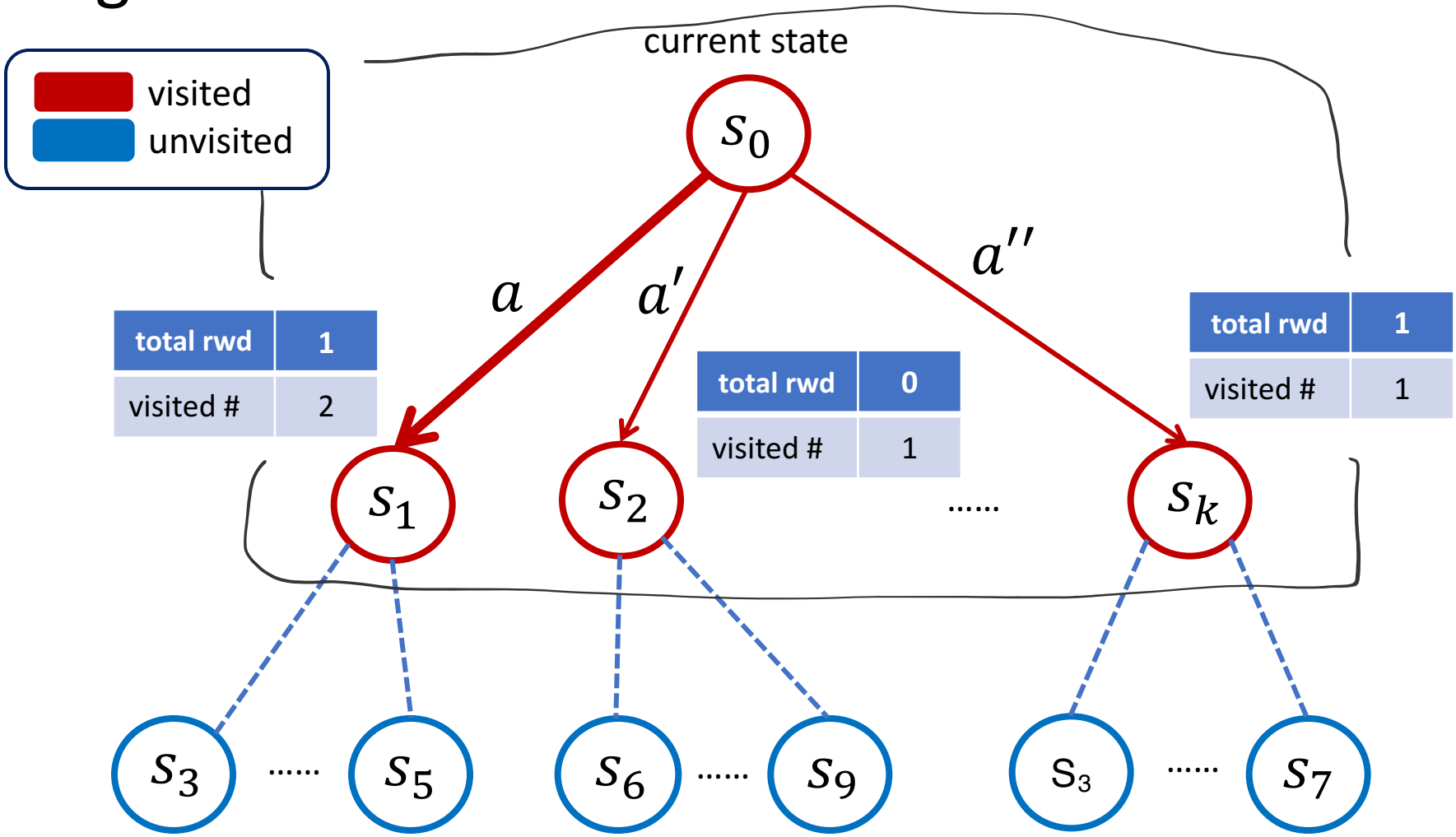
MCTS: Initial trials



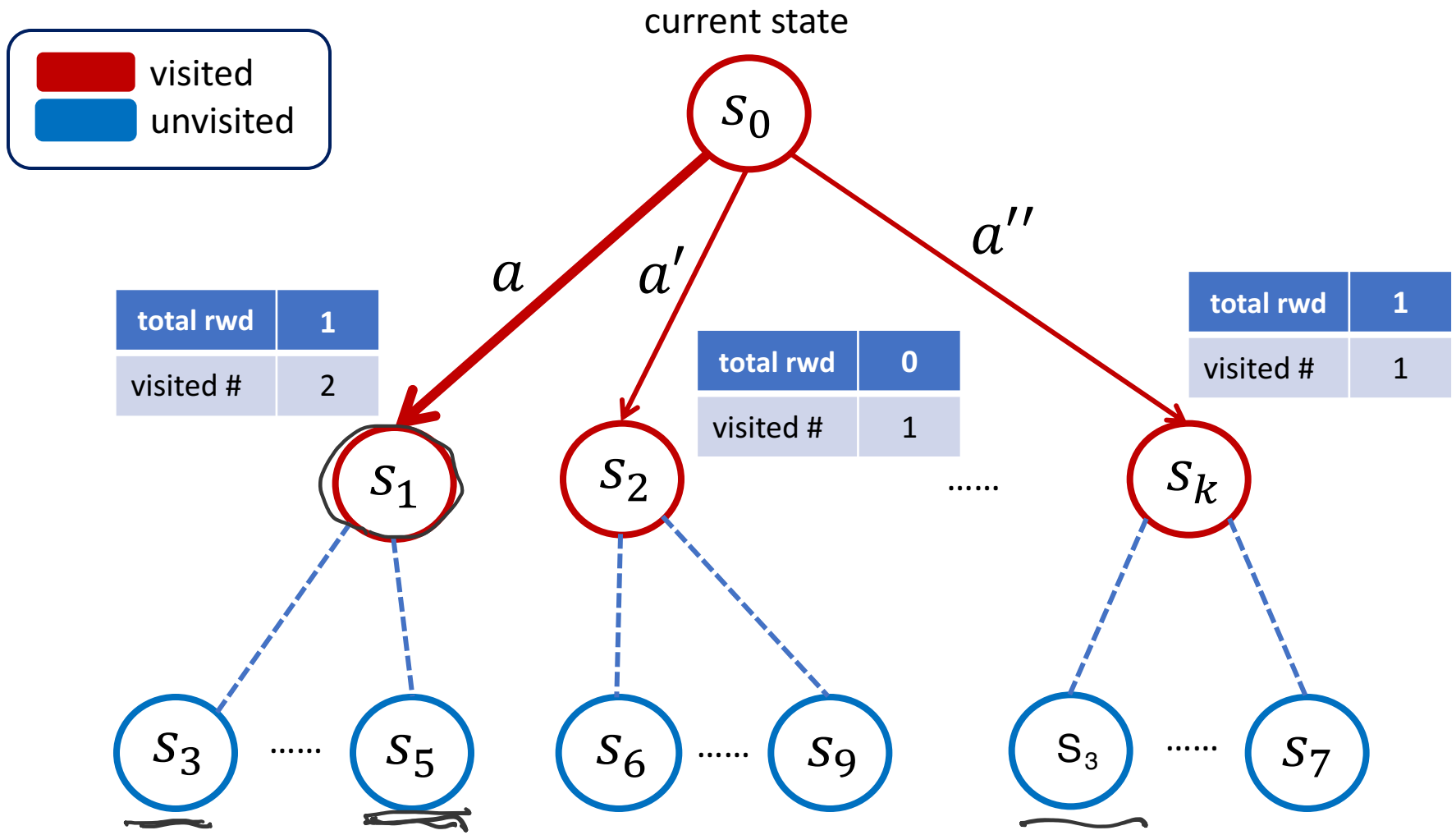
Stage 1: Traversal over Visited Part of the Tree



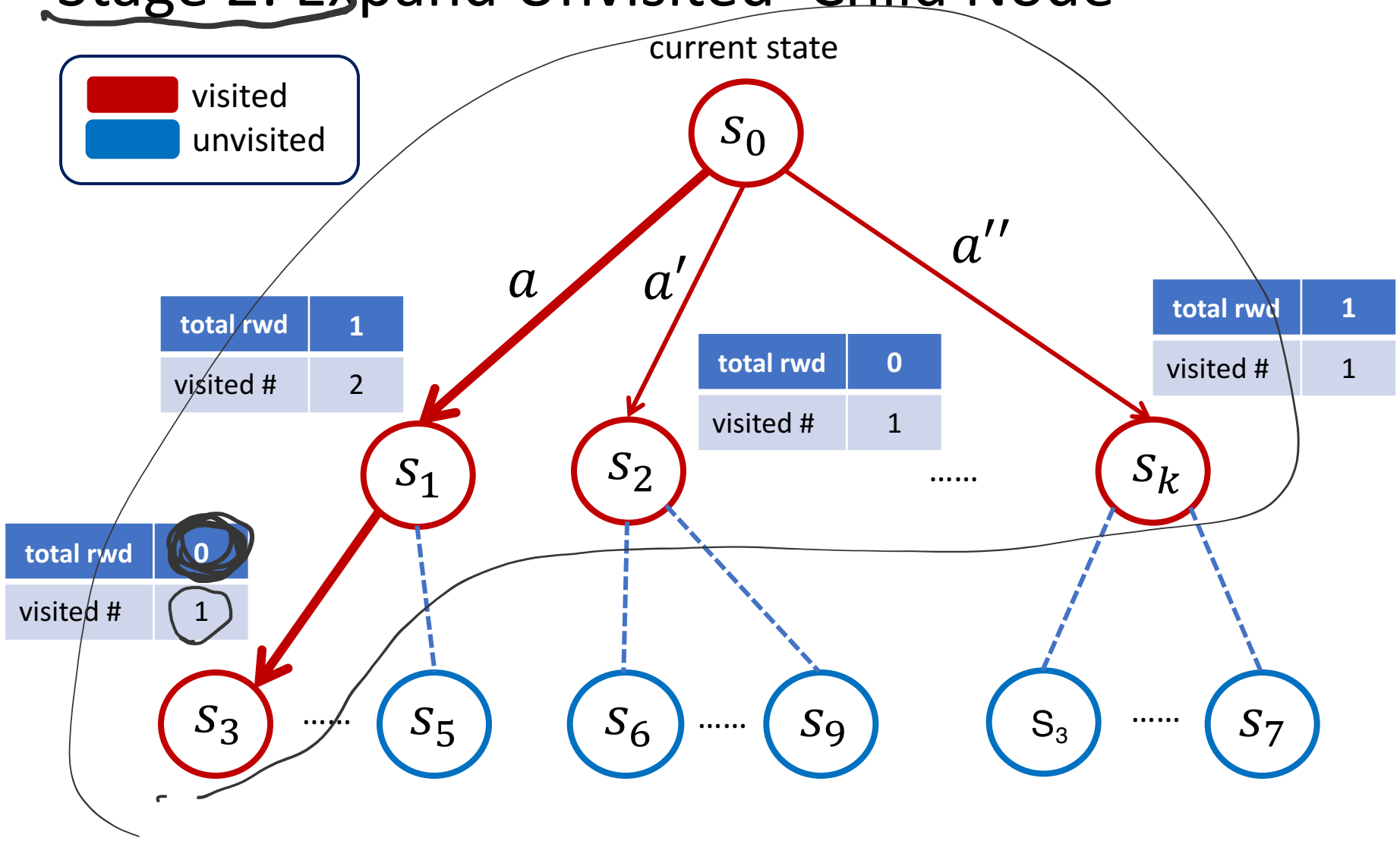
Stage 1: Traversal over Visited Part of the Tree



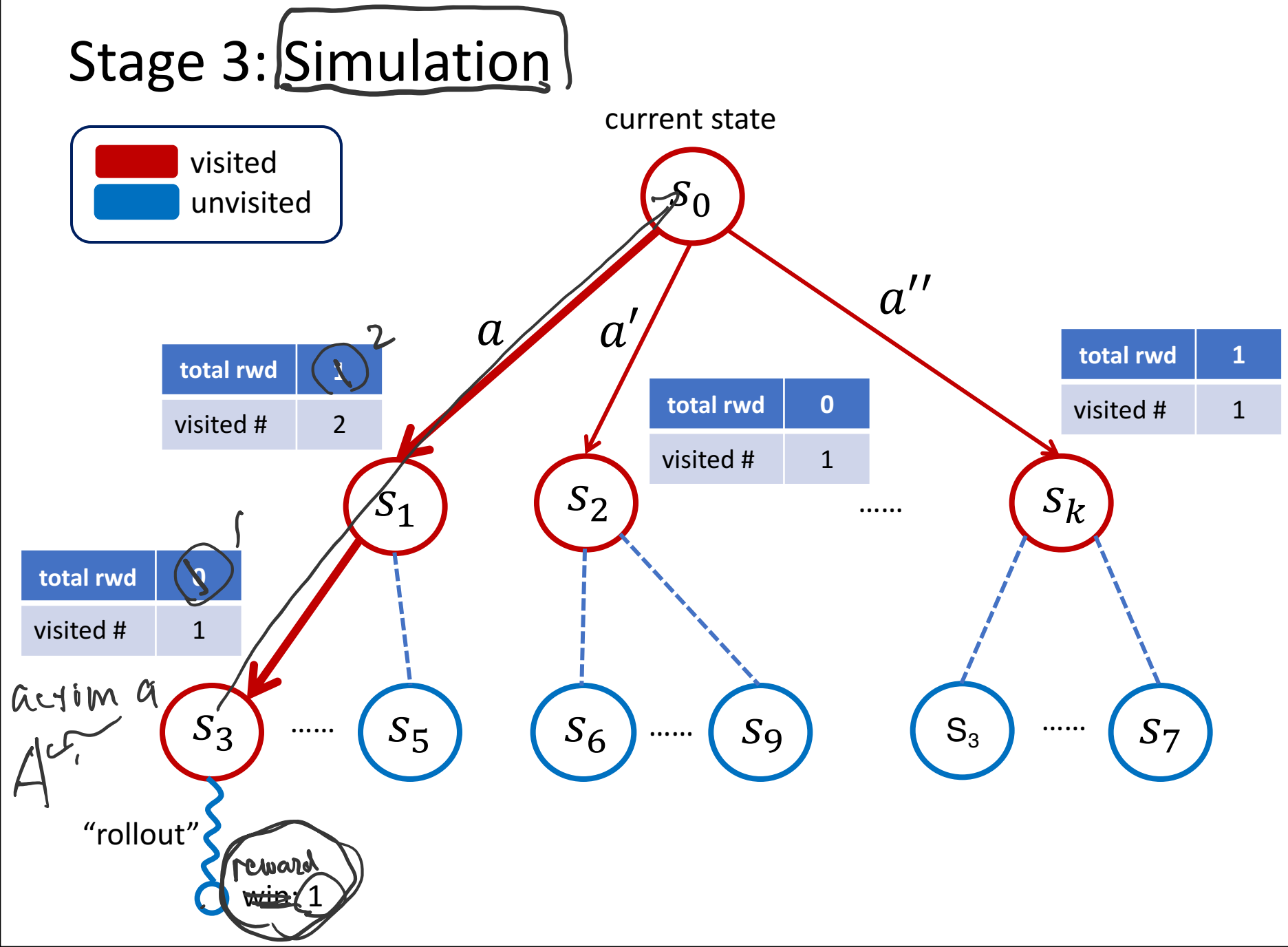
Stage 2: Expand Unvisited 'Child Node'



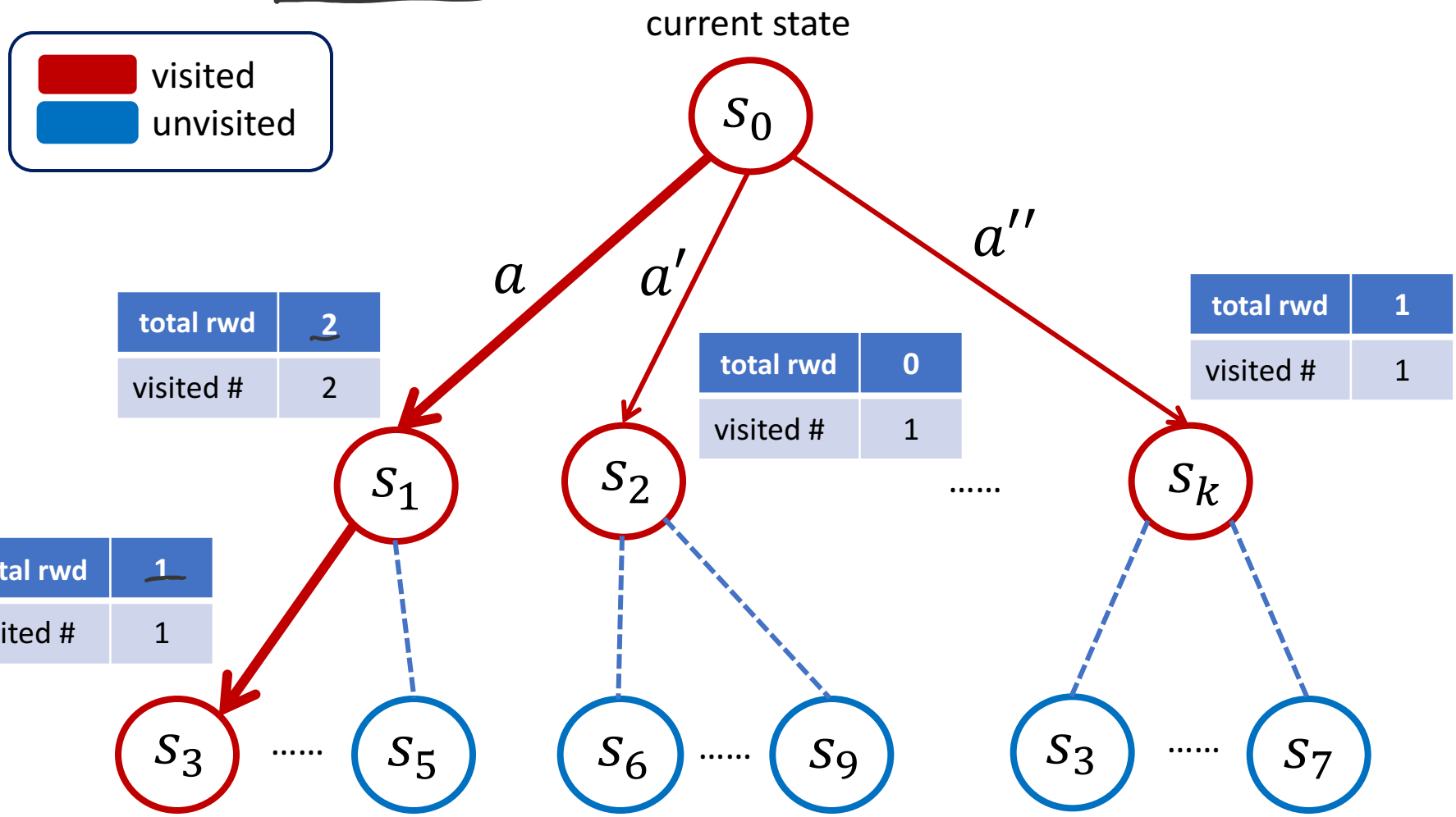
Stage 2: Expand Unvisited 'Child Node'



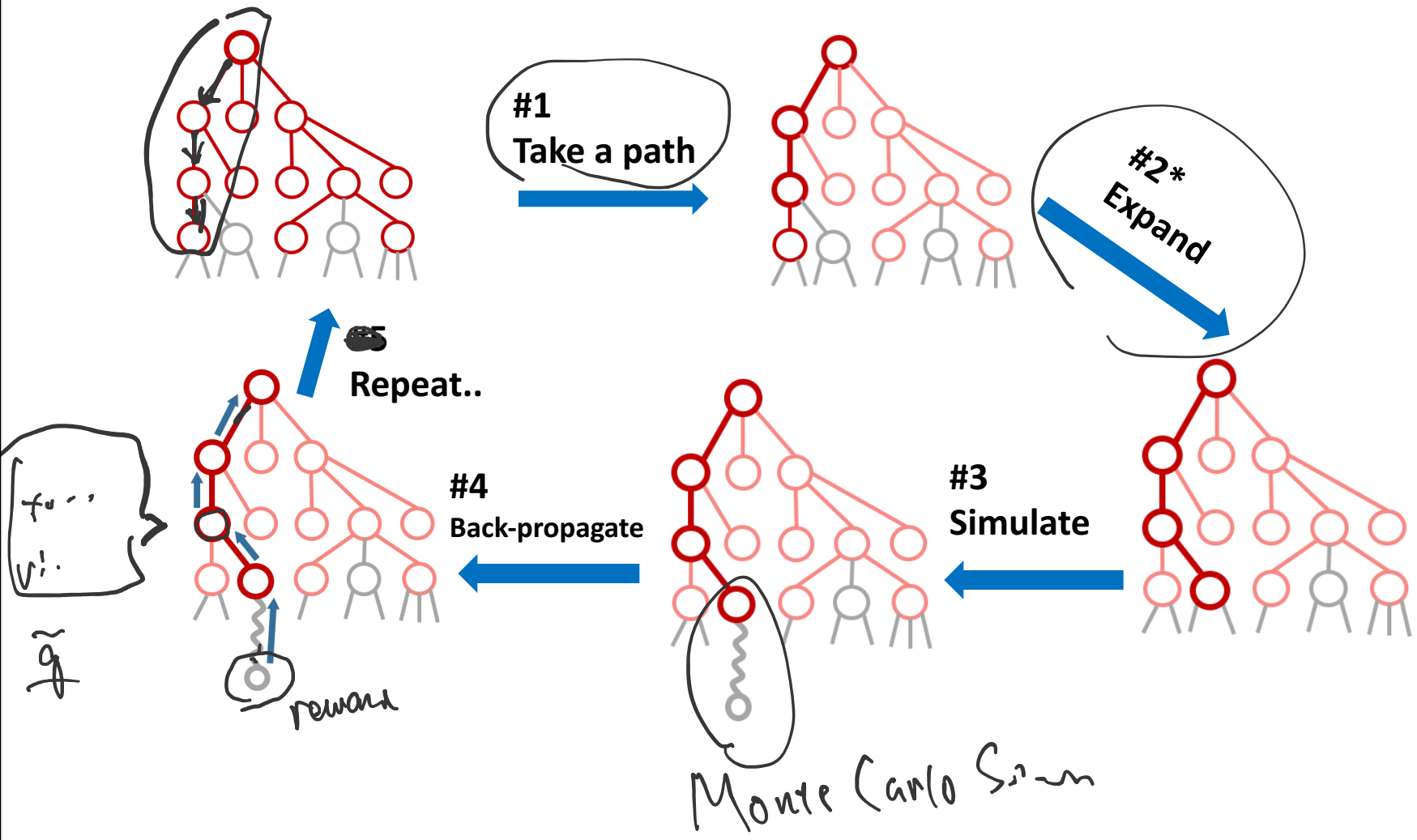
Stage 3: Simulation



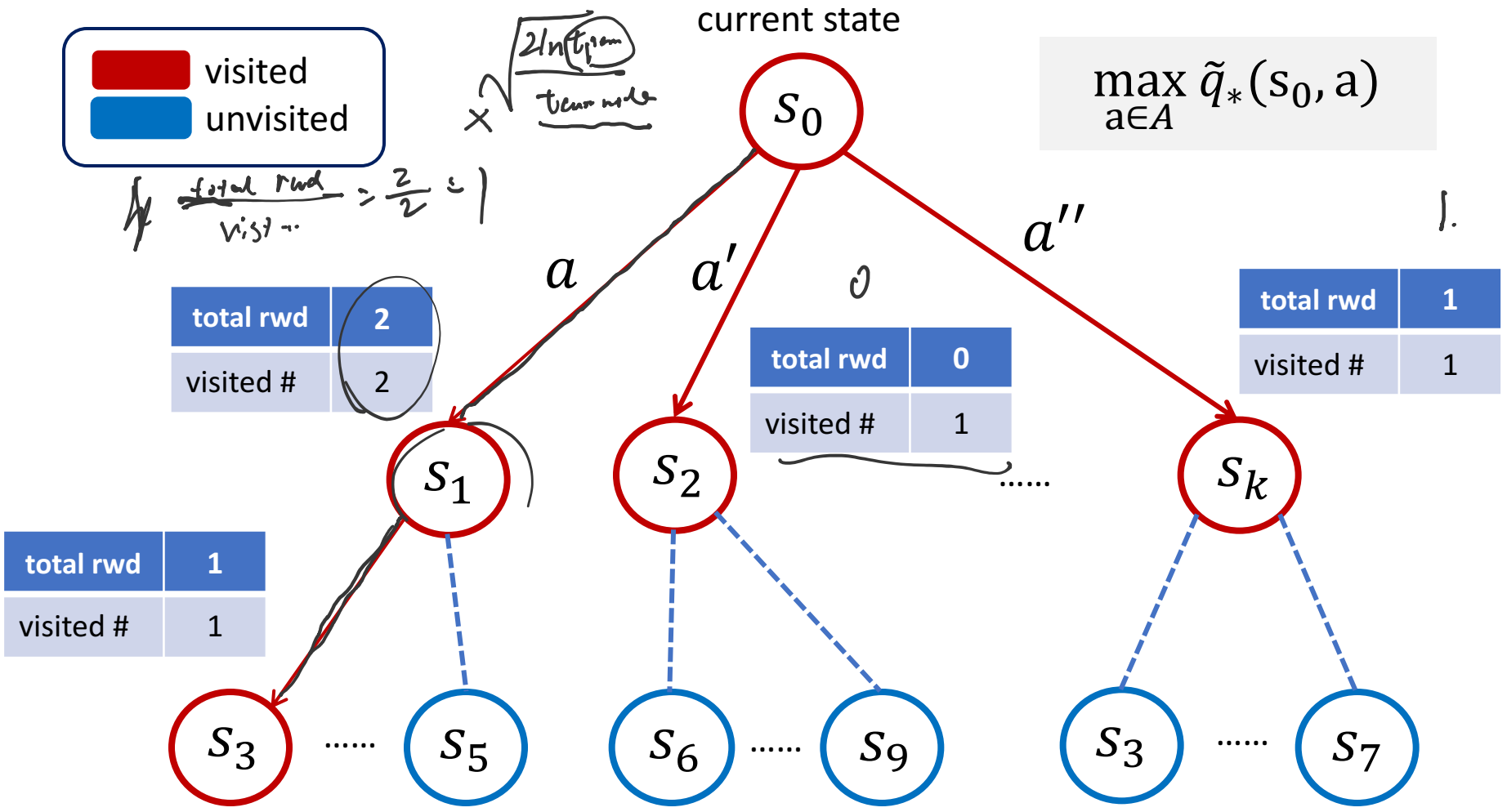
Stage 4: Backpropagation



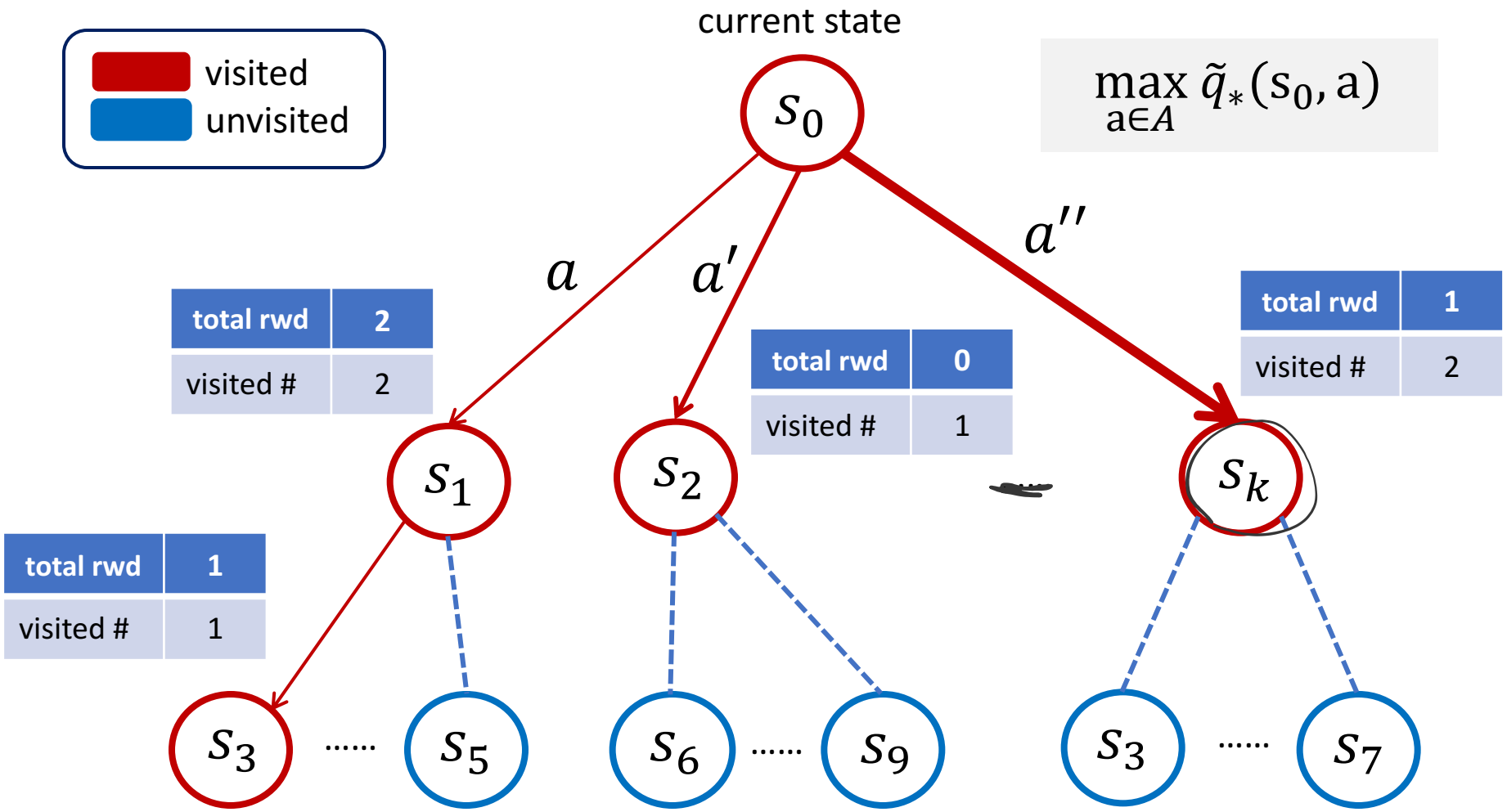
We just finished one iteration of MCTS!



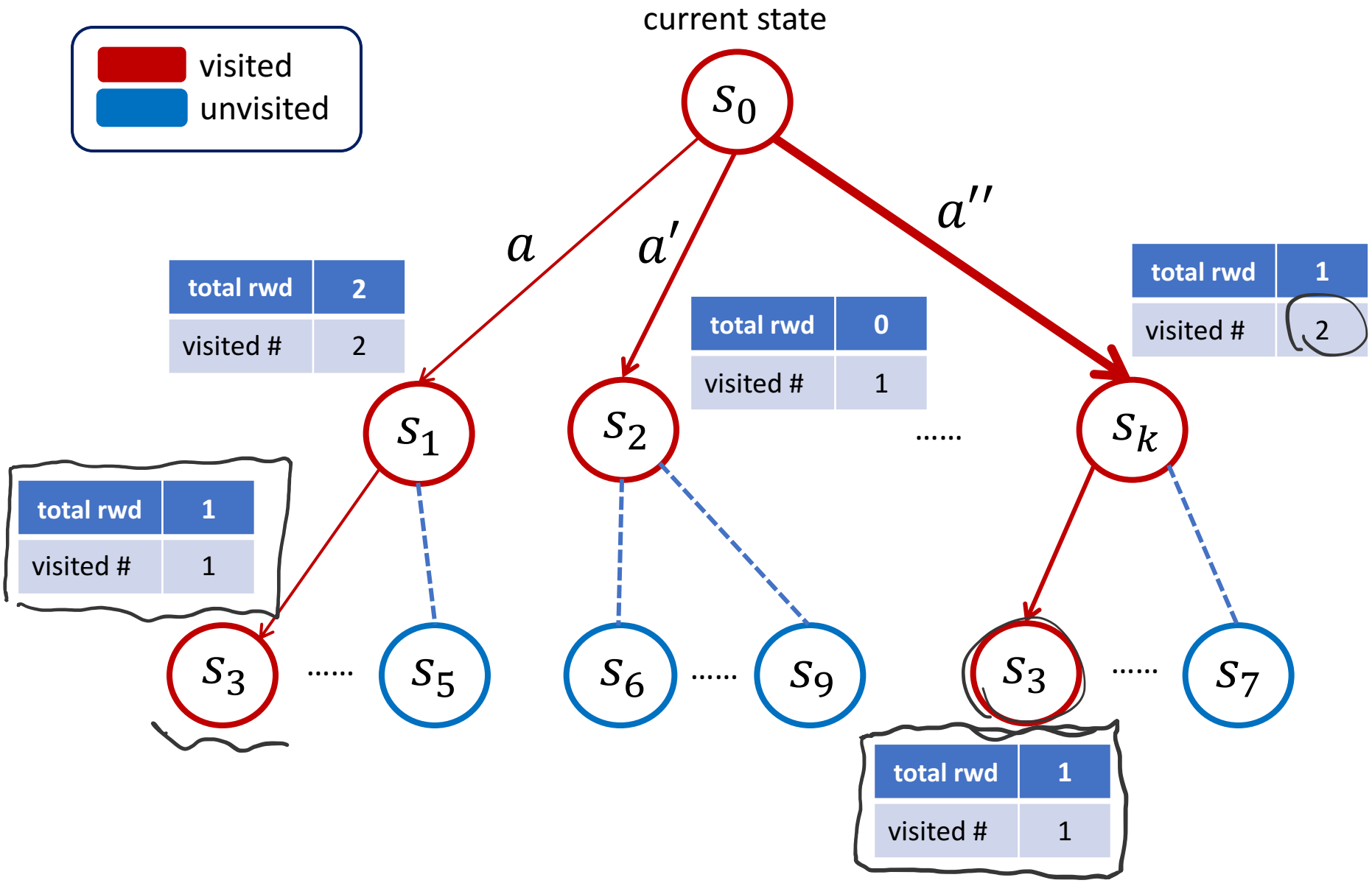
Stage 1: Traversal over Visited Part of the Tree



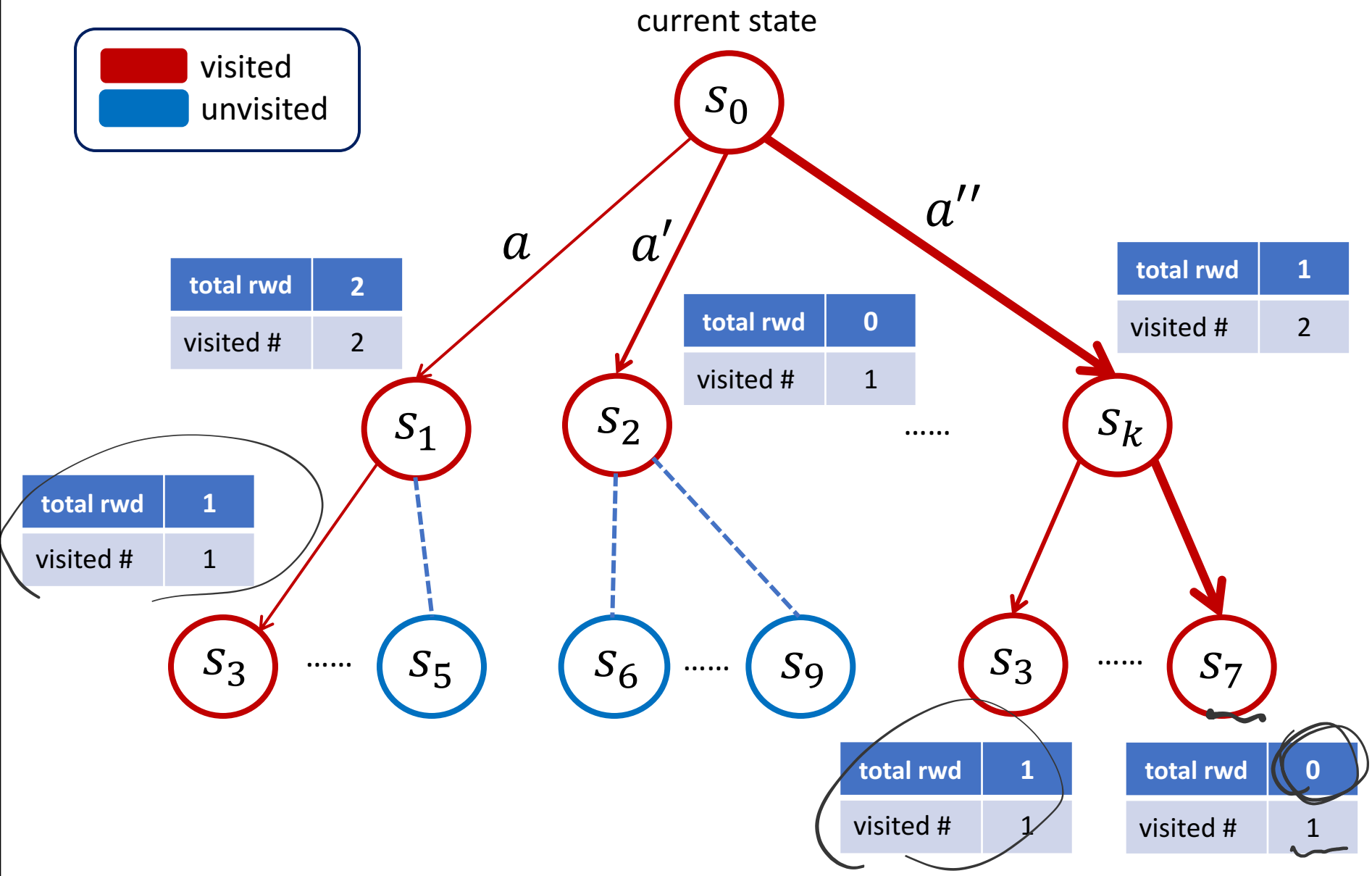
Stage 1: Traversal over Visited Part of the Tree



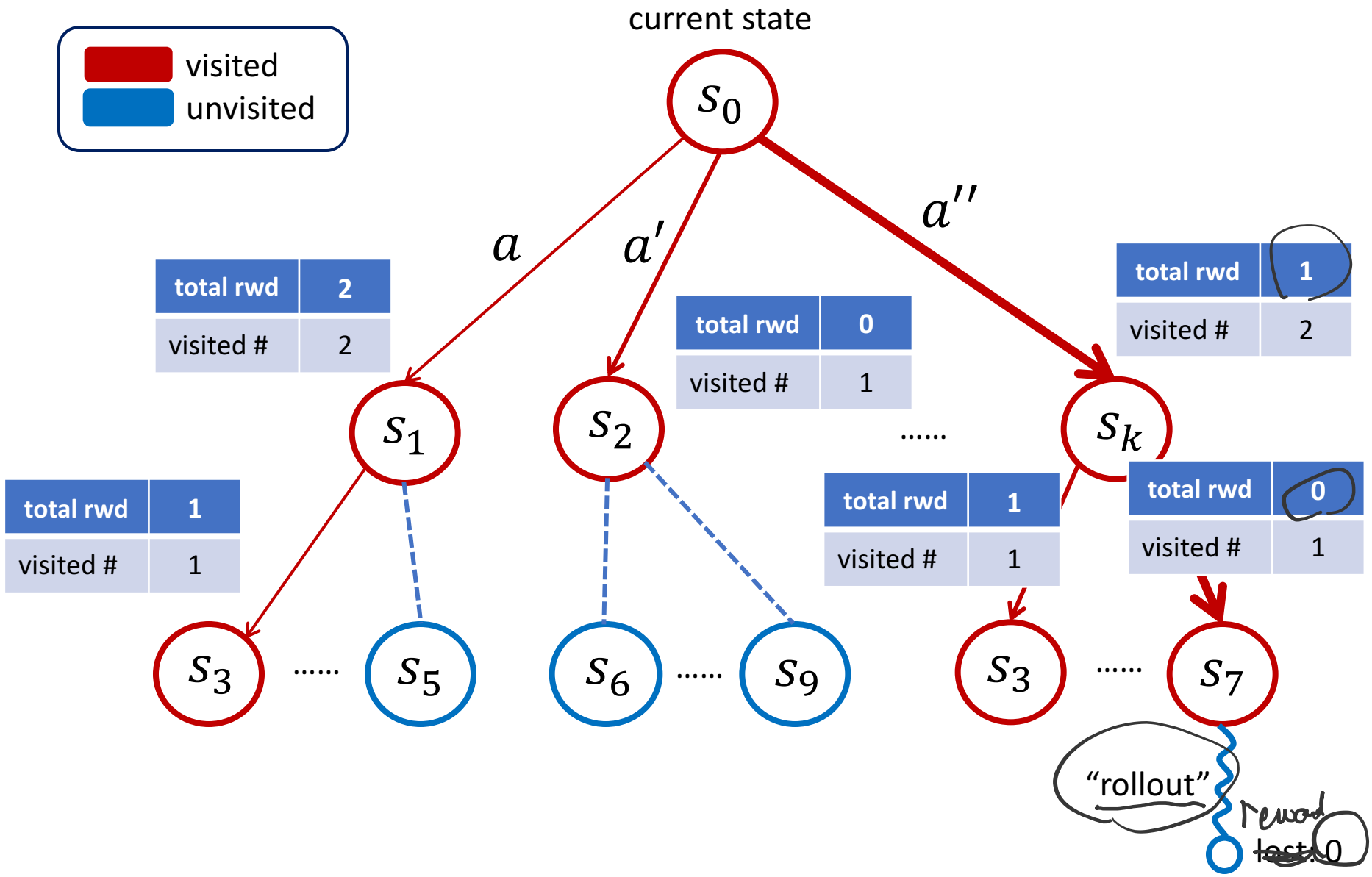
Stage 1: Traversal over Visited Part of the Tree



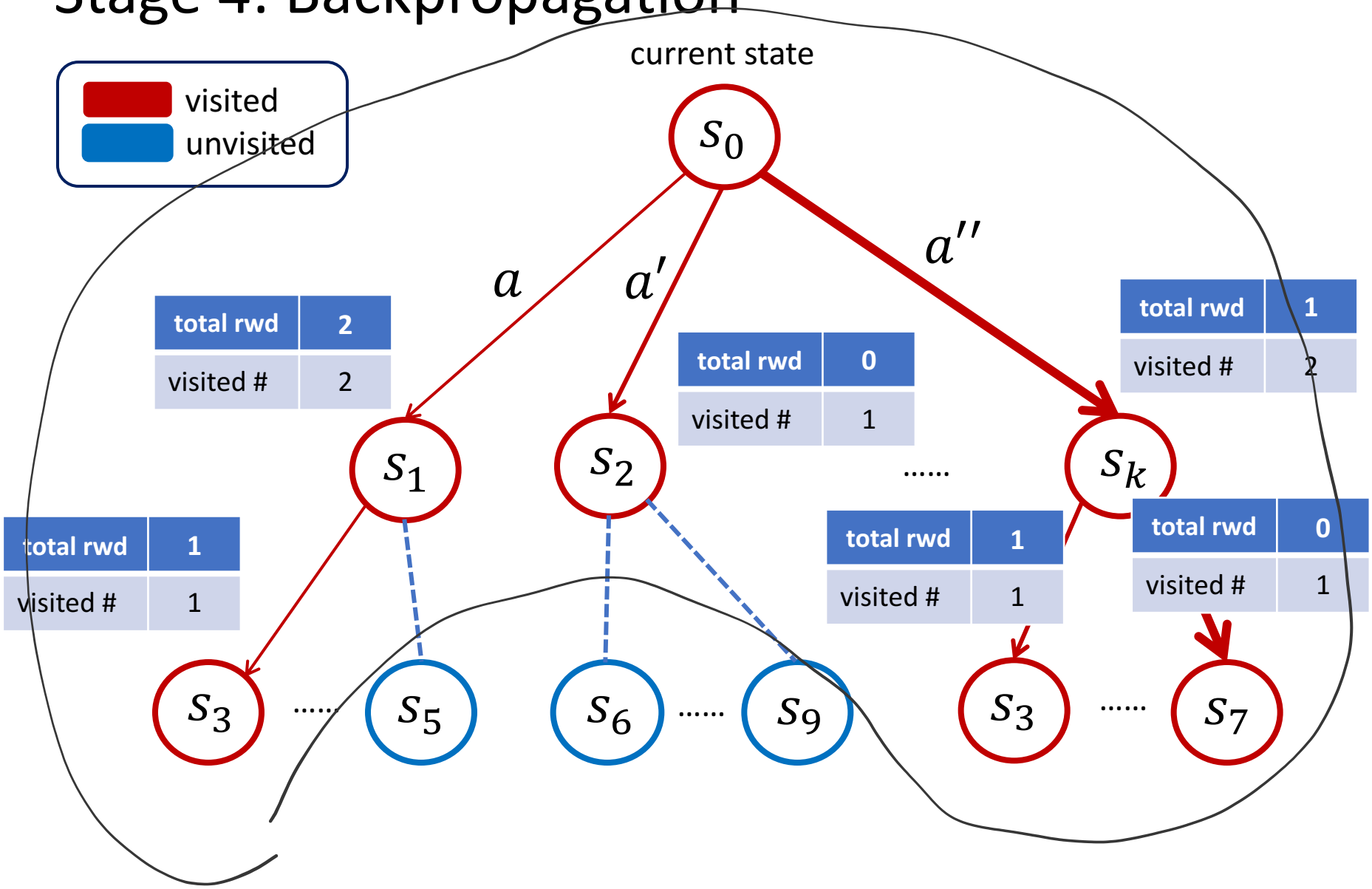
Stage 2: Expand Unvisited 'Child Node'



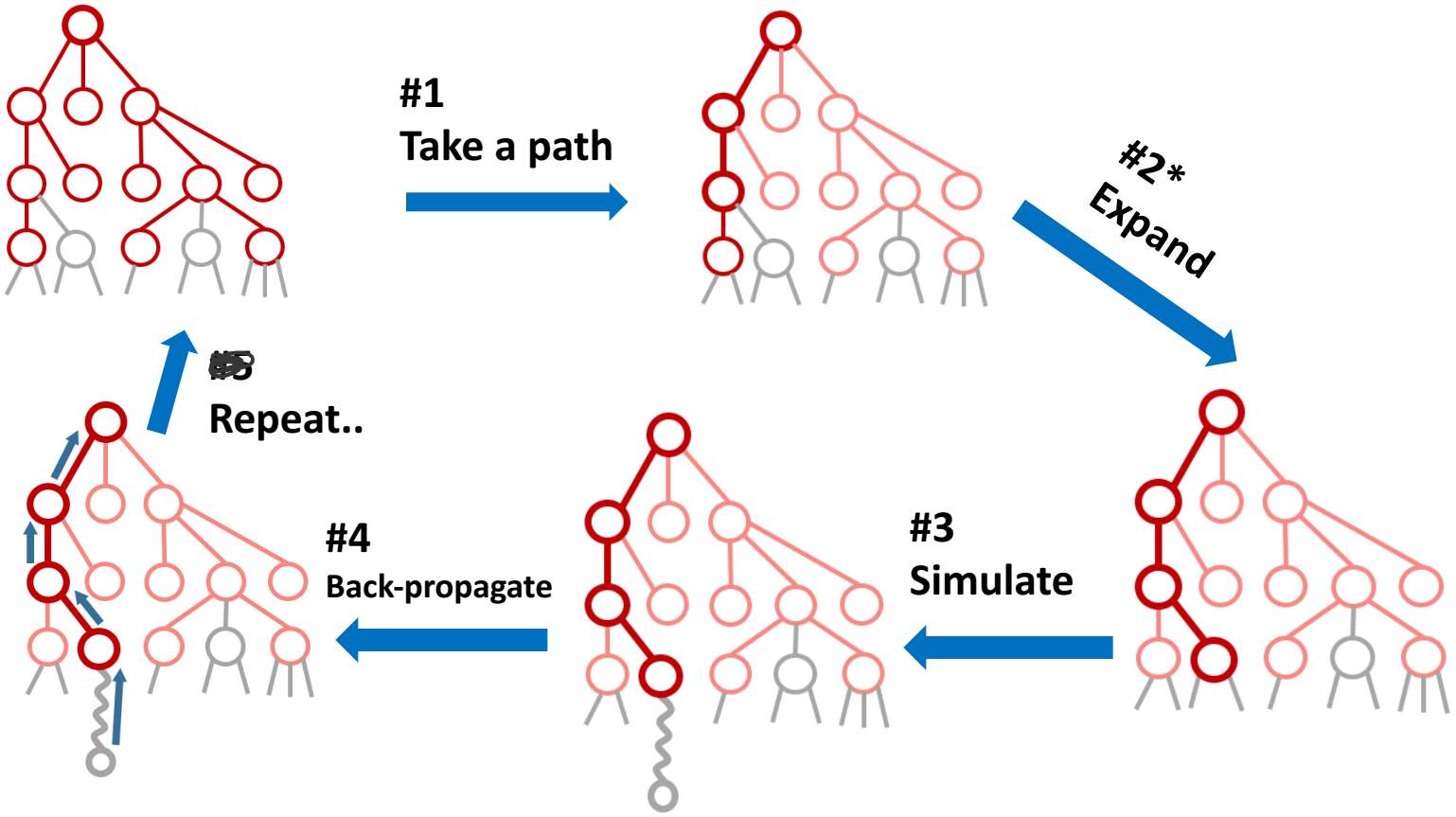
Stage 3: Simulation



Stage 4: Backpropagation

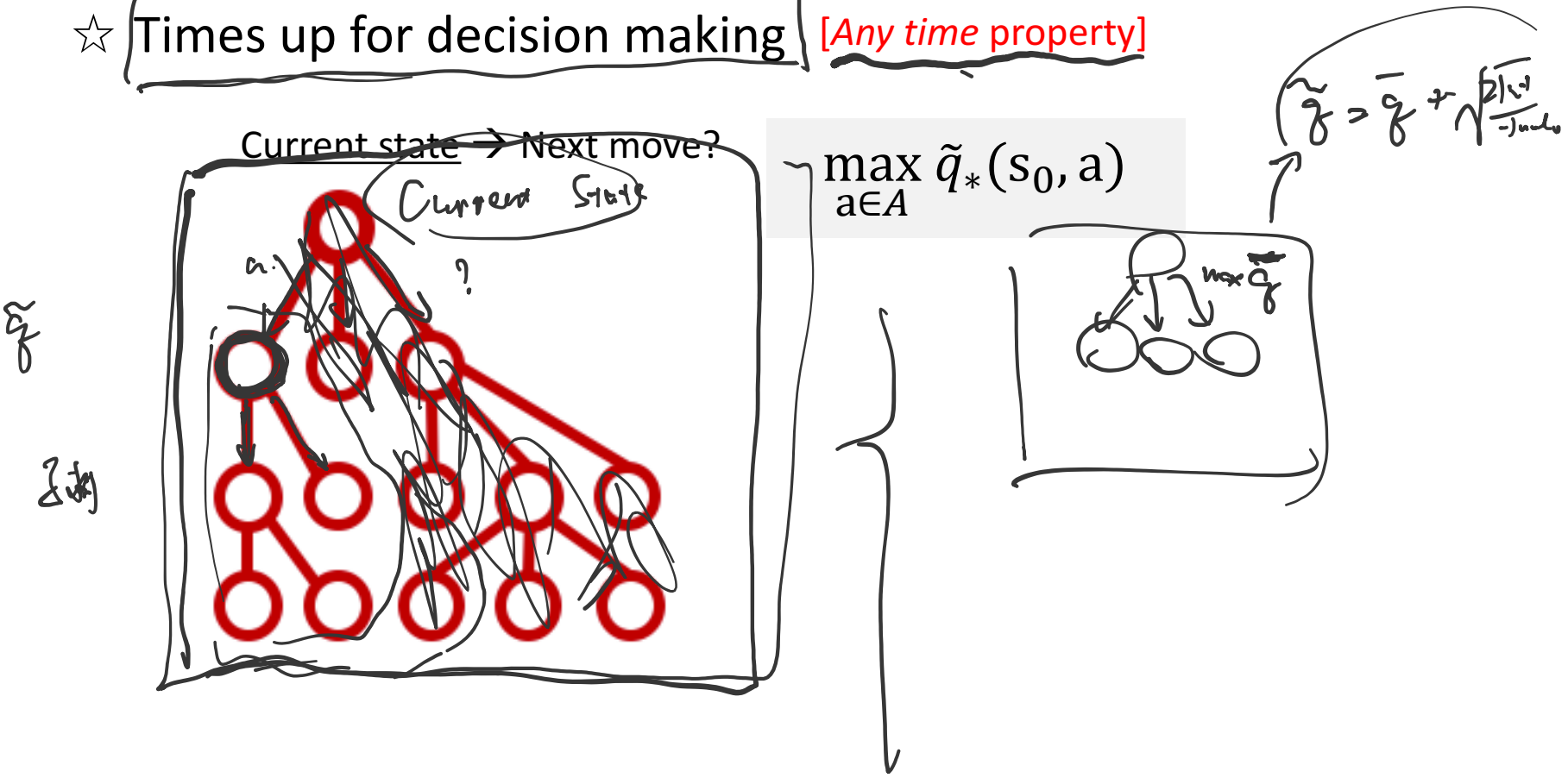


Another iteration is DONE!



Repeat iterations until …

- ☆ Reaching a pre-defined threshold for # of iterations
- ☆ Times up for decision making [*Any time* property]



Performance of MCTS

- ☆ For each state *s,* the estimated value-function is an unbiased *estimate of* the actual value-function as # of samples → ∞
- ☆ The bias converges to zero at rate of $O(\sum_{i=0}^{H-1} \ln(N_i)/N_i)$

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10

Advantages of MCTS

- ☆ Online w/ *any-time* property
- ☆ Requires no domain knowledge
- ☆ Structed, thus allows various degrees of freedom

  for improvements in each stage
- ☆ Requires no full knowledge of the env. dynamics

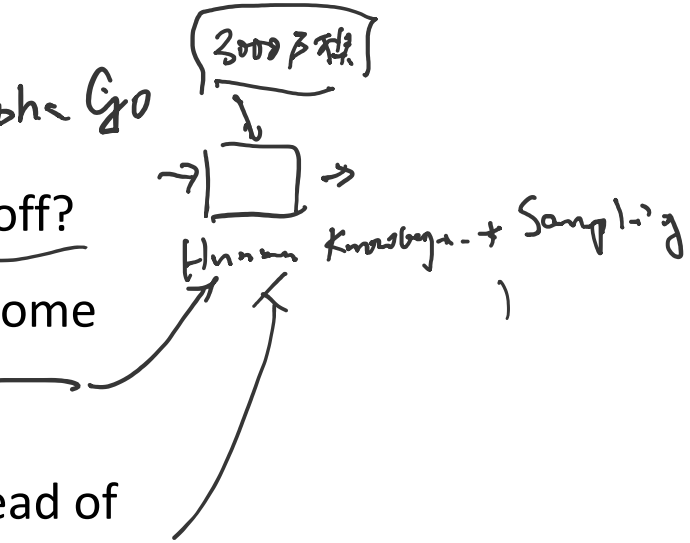
Limitations & Possible Improvements

☆ Traversal

☆ Does UCB achieve the best *e-e* tradeoff?

☆ What if samples are inadequate for some upcoming state?

☆ Can it be taught w/ experiences instead of sampling from scratch?



☆ Expansion

☆ When the branching factor is big..

☆ Simulation

☆ *Purely* random rollouts may not be a good idea

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10

Lecture Flow

Motivation

Brief History of Monte Carlo Tree Search (MCTS)

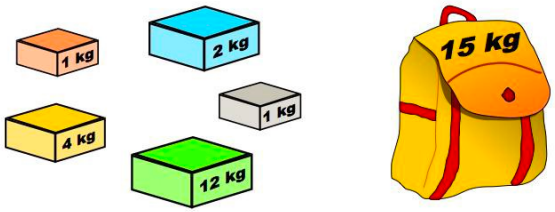
A Close Look at MCTS

Remarks

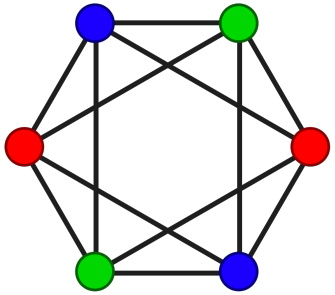
MCTS for Combinatorial Optimization Problems

MCMC HW Q10

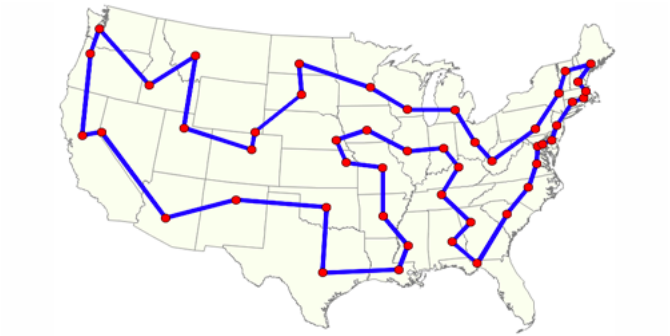
Combinatorial Optimization Problems



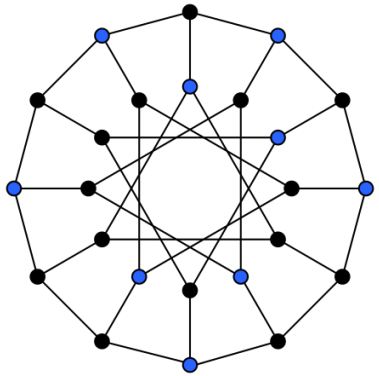
Knapsack Problem



Graph Coloring Problem

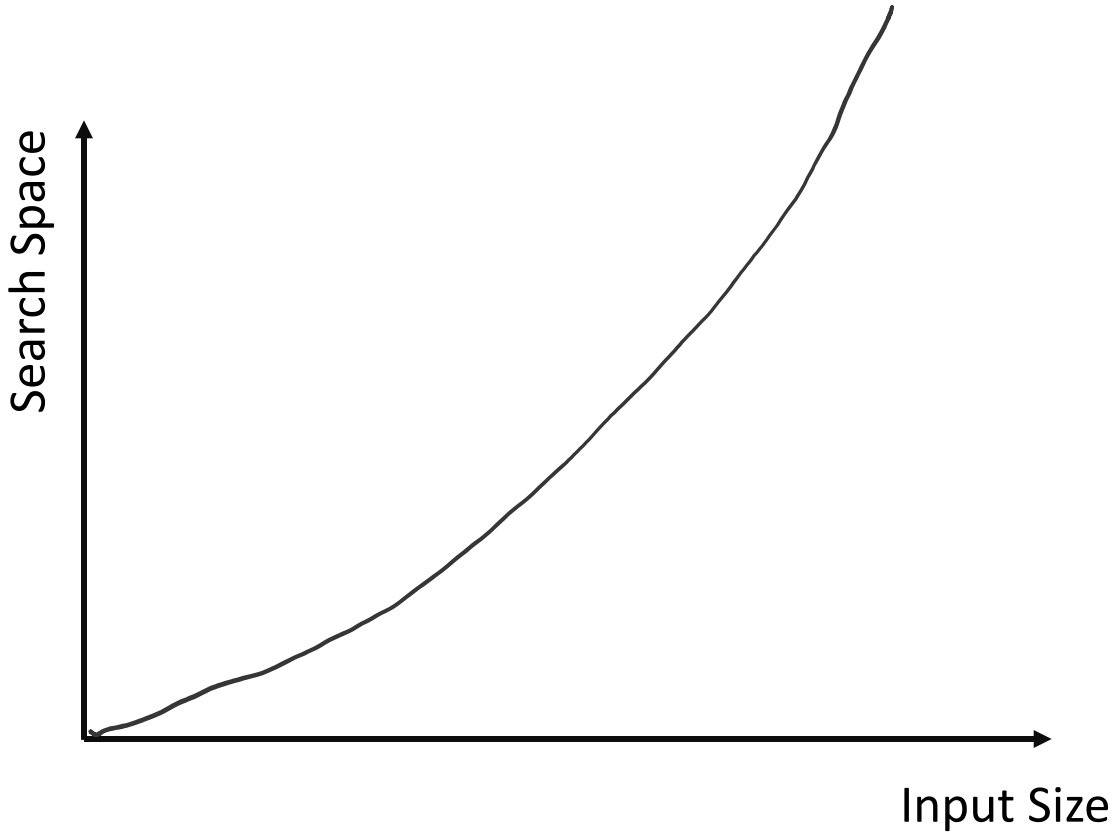


Traveling Salesman Problem



Maximum Independent Set Problem

Challenges in Problem Solving



Efficient & optimal solutions to only minority

Most are NP-hard

Curse-of-dimensionality

Zillions of problems in engineered networks

- CoFlow scheduling in DCN flow scheduling
- Maximum Weighted Independent Set in wireless networks
- Virtual Network Embedding in net. virtualization
- Function Placement in net. function virtualization
- Instance/Task Placement in stream processing systems
- …

Available Arsenal

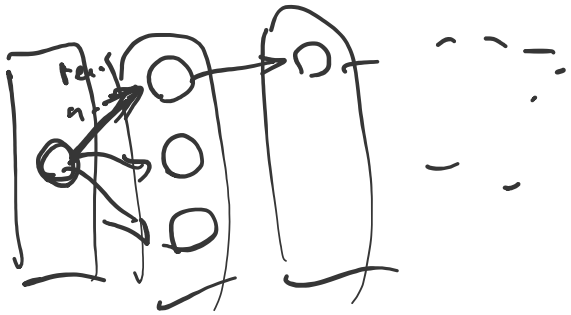
- Approximation algorithms
  - Greedy methods, DP, relaxation techniques, random sampling methods
- Exact algorithms
  - Branch-and-bound, branch-and-cut, *etc.*
- Heuristic
  - Most w/o performance guarantee but in pursuit of low complexity

etc.

MCTS for Combinatorial Optimization Problems

Convert to a seq. decision-making problem

Problem instance

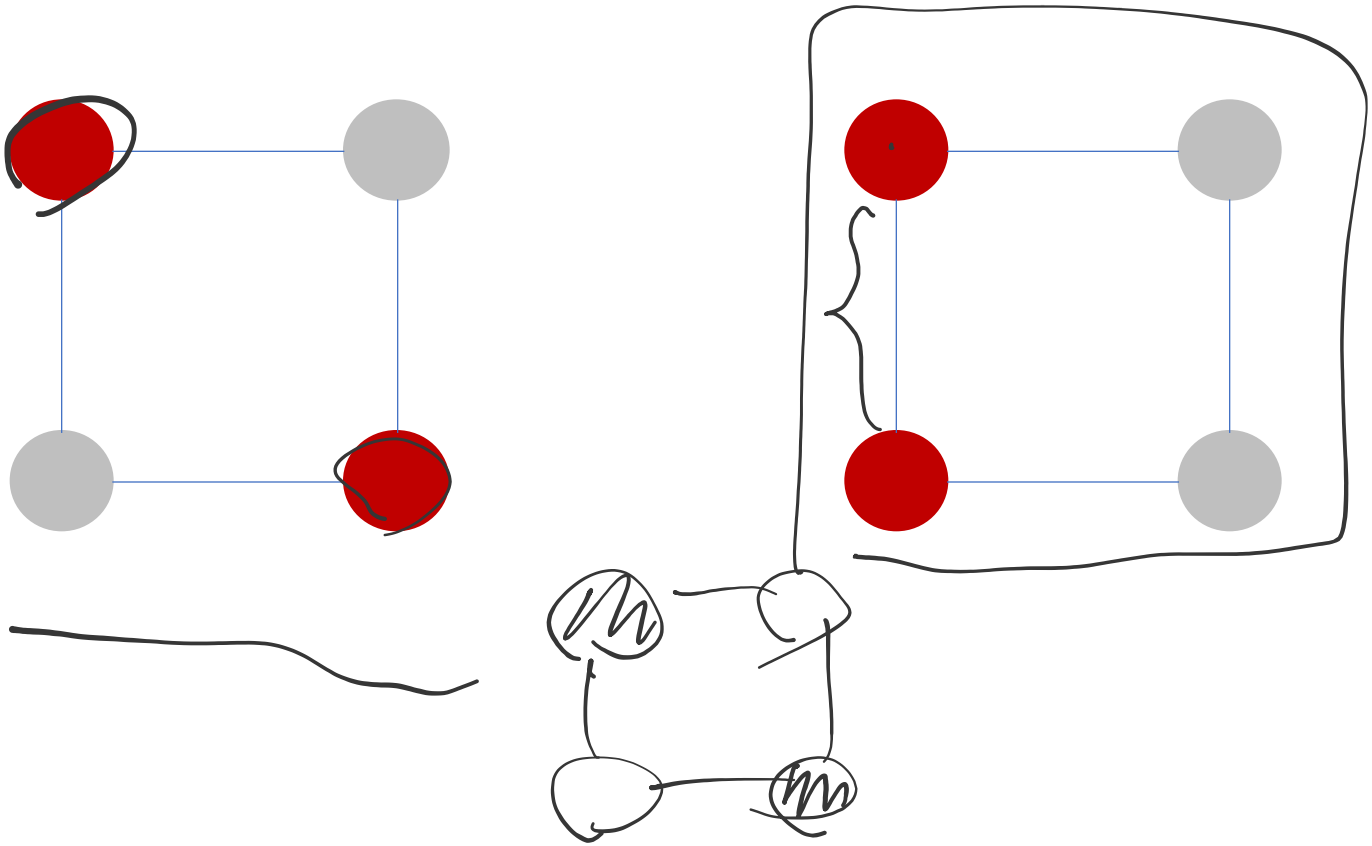


Define the “state”, ”action” & “value”

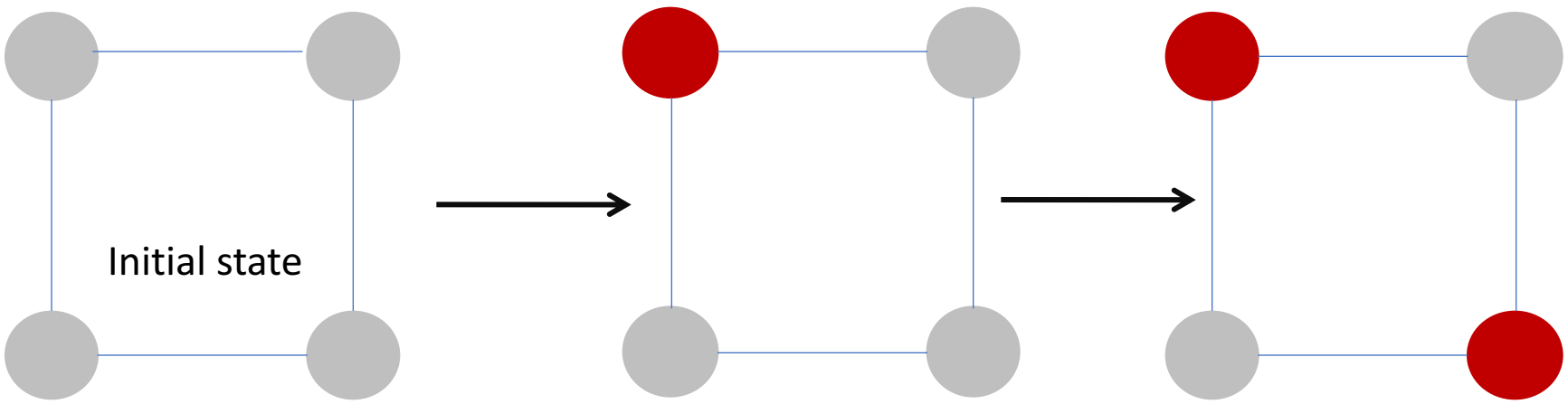
Conceive the search tree

MCTS comes into play

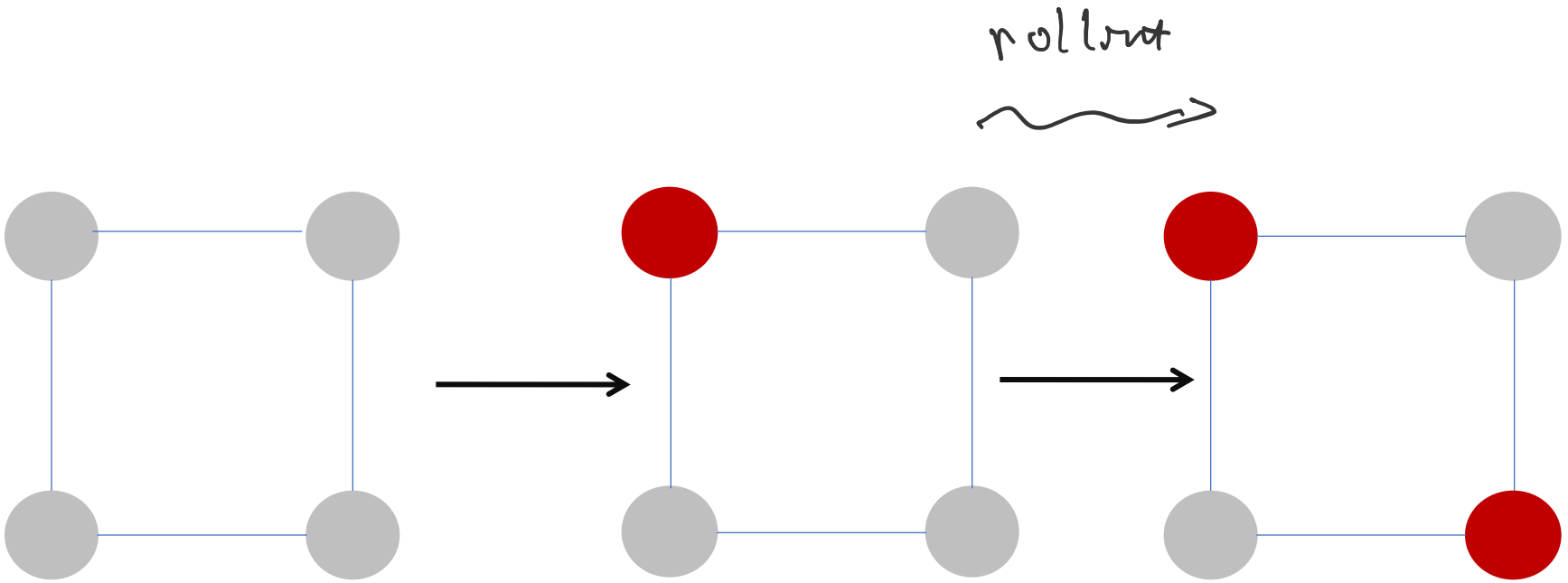
MCTS for ~~Maximum~~ Independent Set Problem



MIS as a Seq. Decision-making Problem



Search Tree for MIS



Lecture Flow

Motivation

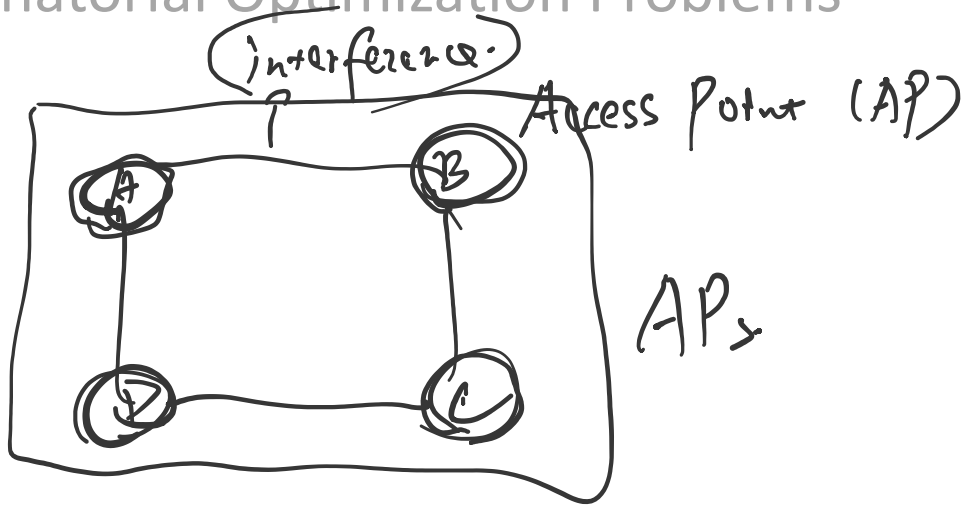
Brief History of Monte Carlo Tree Search (MCTS)

A Close Look at MCTS

Remarks

MCTS for Combinatorial Optimization Problems

MCMC HW Q10



MIS.

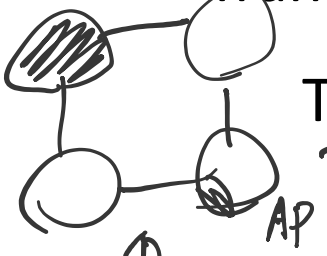
Design as a Discrete-Time Markov Chain

Desired PMF: $\pi_f^* = C^{-1}\exp(\beta|f|) \quad \forall f \in F$

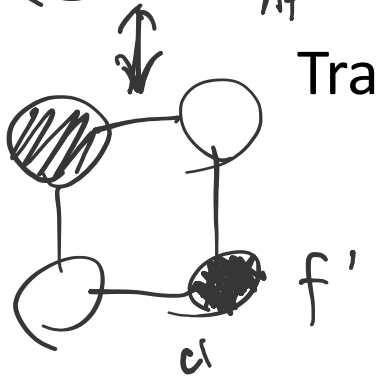
State space: Set of all feasible independent sets

Transition between states:

Topology of the chain:



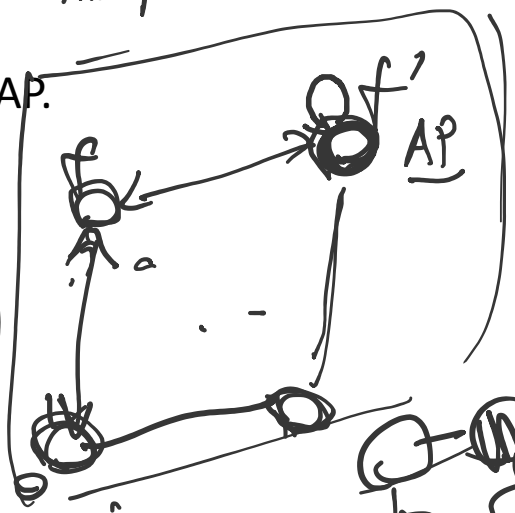
Any two adjacent states differ by one AP.



Transition probability:

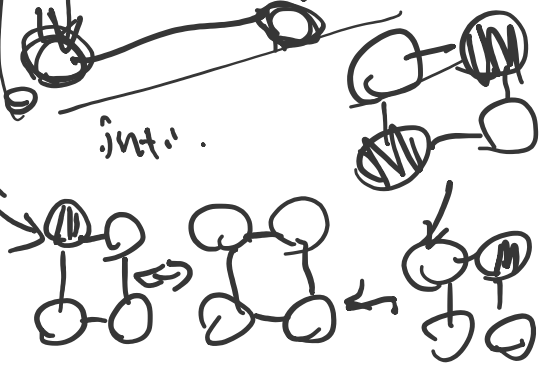
$$p_{f,f'} = \frac{\min\{1, \exp(\beta(|f'| - |f|))\}}{N}$$

$$p_{f,f} = 1 - \sum_{f' \in N(f)} p_{f,f'}$$

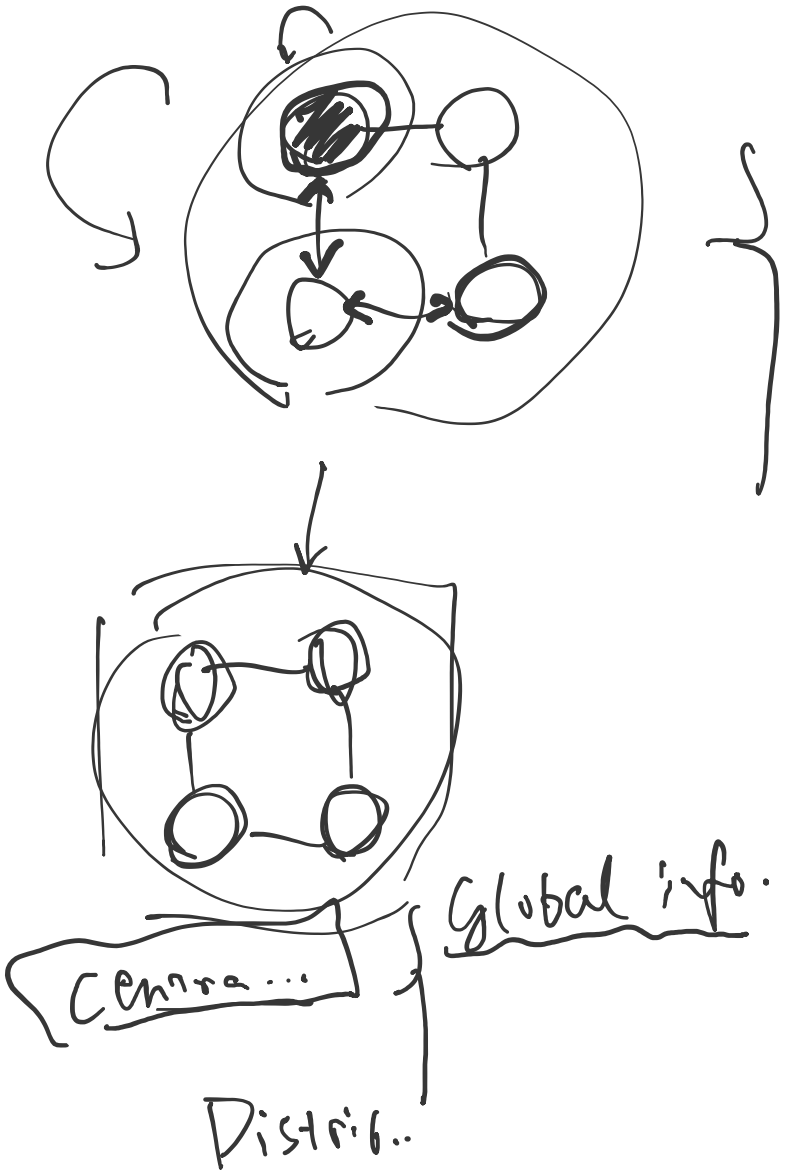


Chain Design?

$$\pi_f^* P_{f,f'} = \pi_{f'}^* P_{f',f}$$



Design as a Discrete-Time Markov Chain



Design as a Continuous-Time Markov Chain

Desired PMF: $\pi_f^* = C^{-1}\exp(\beta|f|), \forall\, f \in F$

State space: Set of all feasible independent sets

Transition between states:

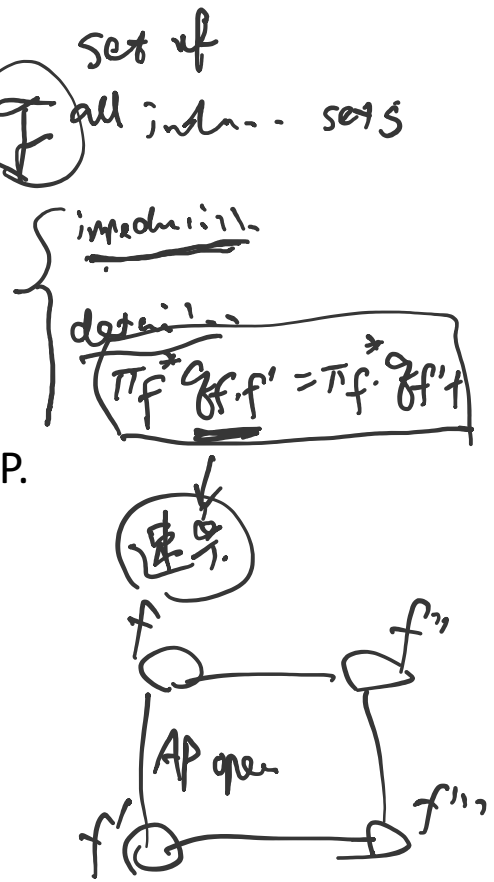
Topology of the chain:

Any two adjacent states differ by one AP.

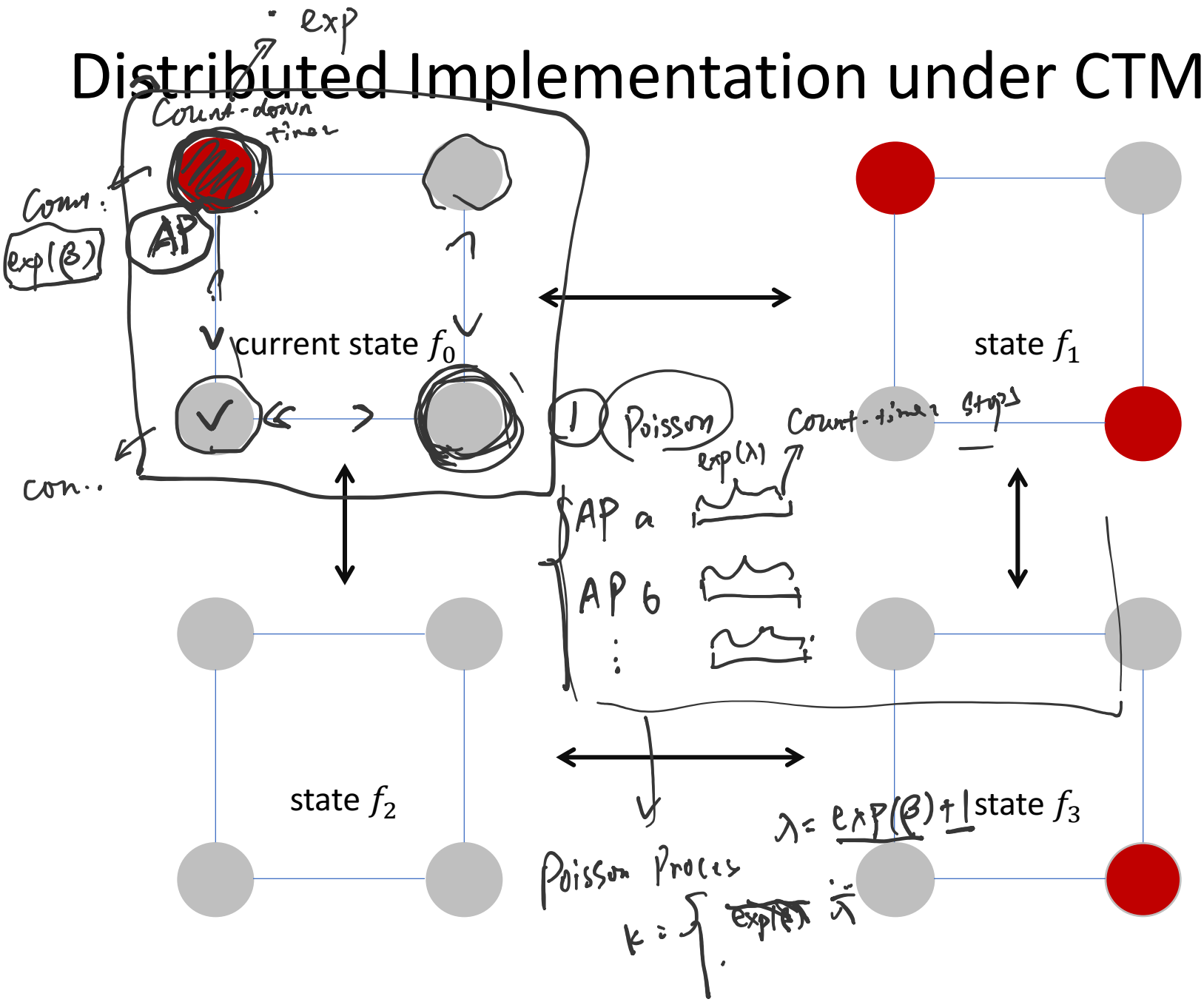
Transition rate:

$$q_{f,f'} = \min\{1, \exp\big(\beta(|f'| - |f|)\big)\}$$

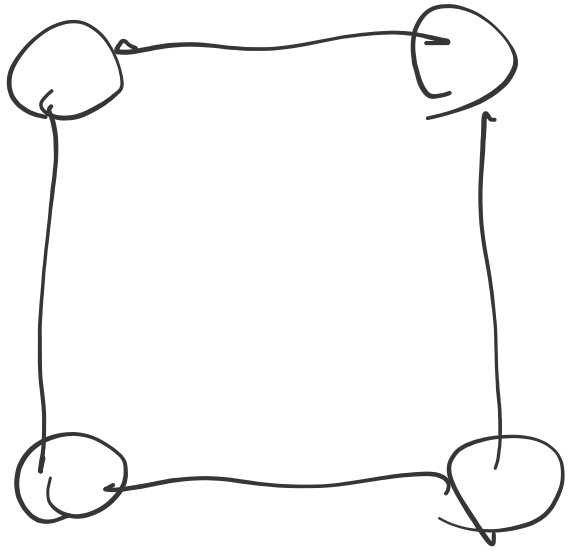
Chain Design:



Distributed Implementation under CTMC

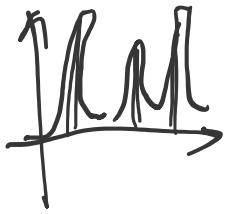


Design as a Continuous-Time Markov Chain

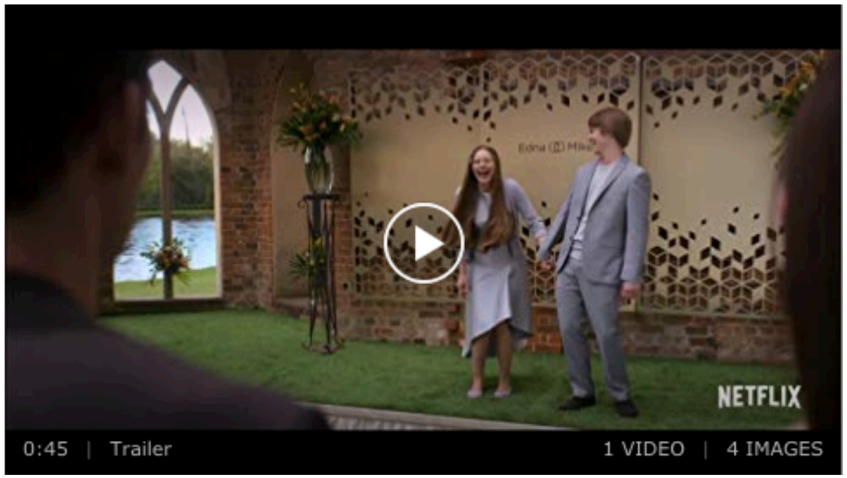


Comparison of MCMC & MCTS

	MCMC	MCTS
Performance Guarantee	✓	—
Degrees of Freedom		
Anytime Property	✗	✓



American Drama w/ Similar Idea to MCTS



Black Mirror Season 4, Ep. 4


Thank you!
Any questions?

This lecture only gives you a *taste* of

Reinforcement Learning.

Omnipresent Idea of *Approximation* in RL

Monte Carlo Methods

Temporal Difference (TD)

Bootstrapping

Learning-aided Planning