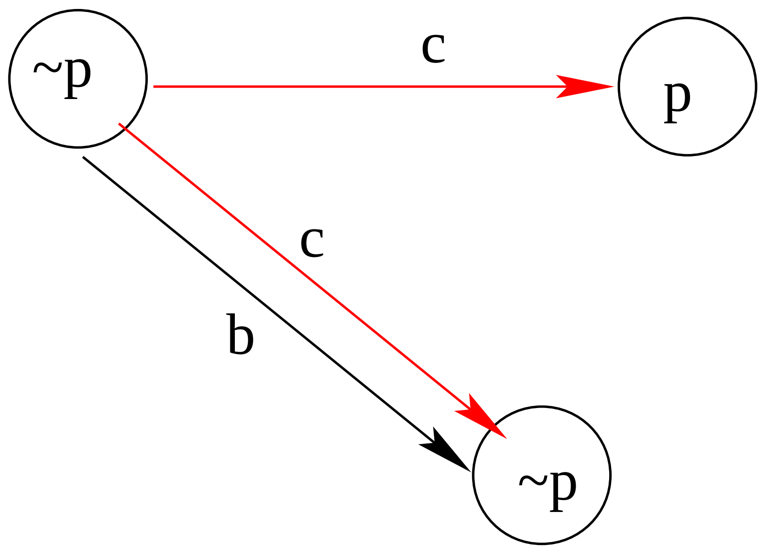
Formulating "Trying Ones Best"

Chitta Baral, Vladimir Lifschitz, and Jicheng Zhao

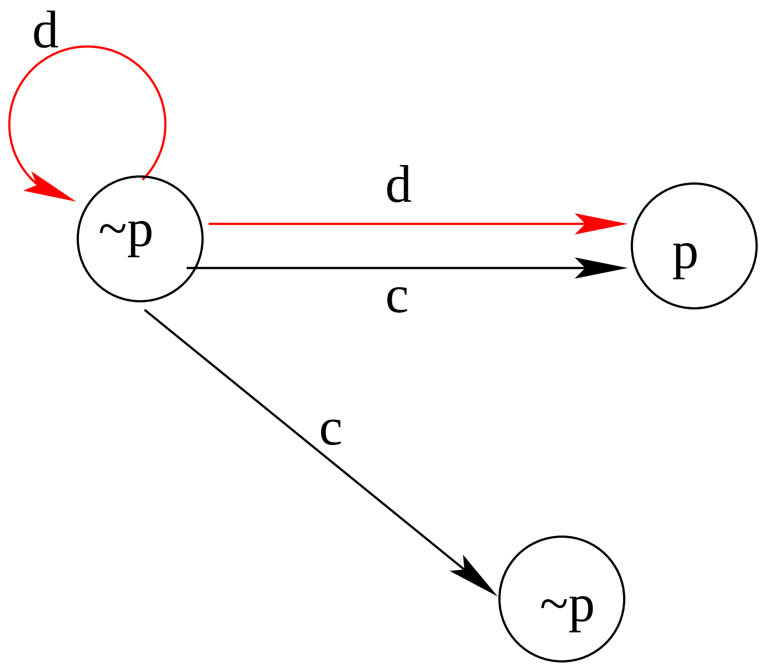
February 10, 2005

Introduction: The best course of action to reach p?



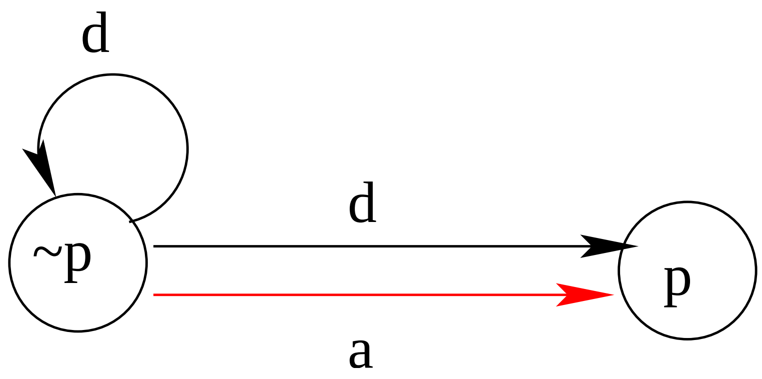
action c is better than b.

Introduction: The best course of action to reach p?



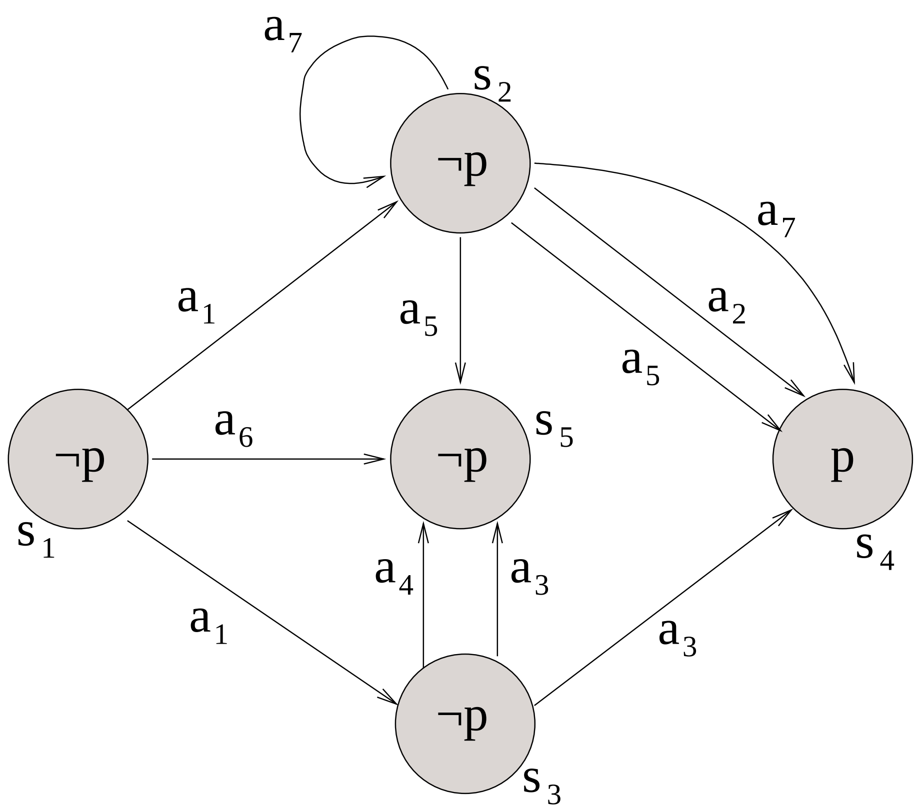
action d is better than c.

Introduction: The best course of action to reach p?



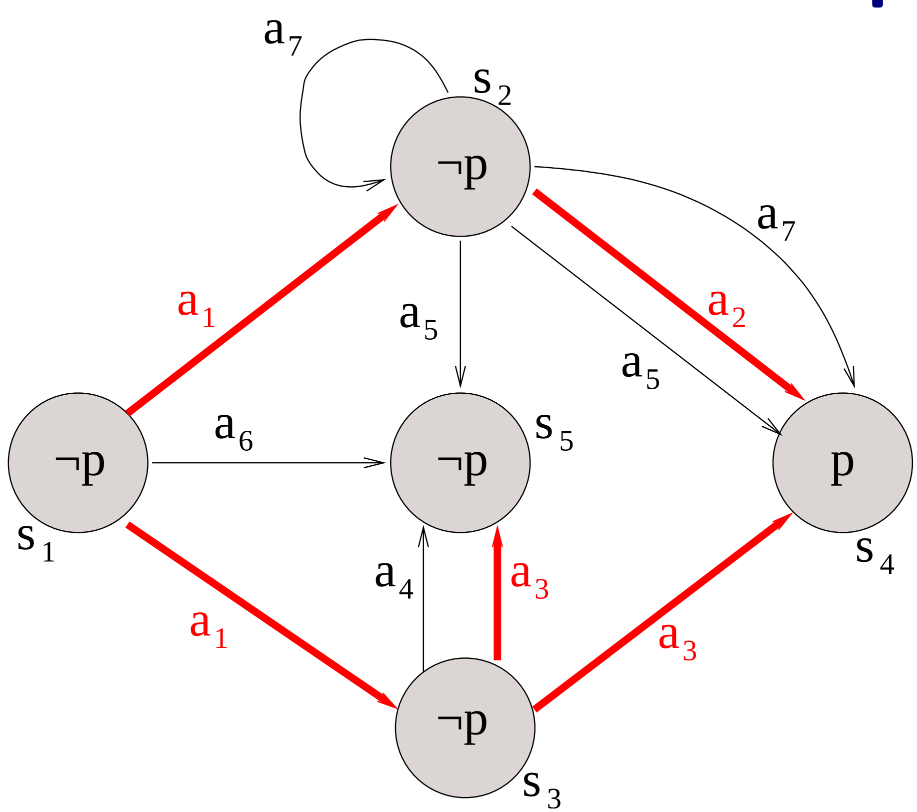
action a is better than d.

An example



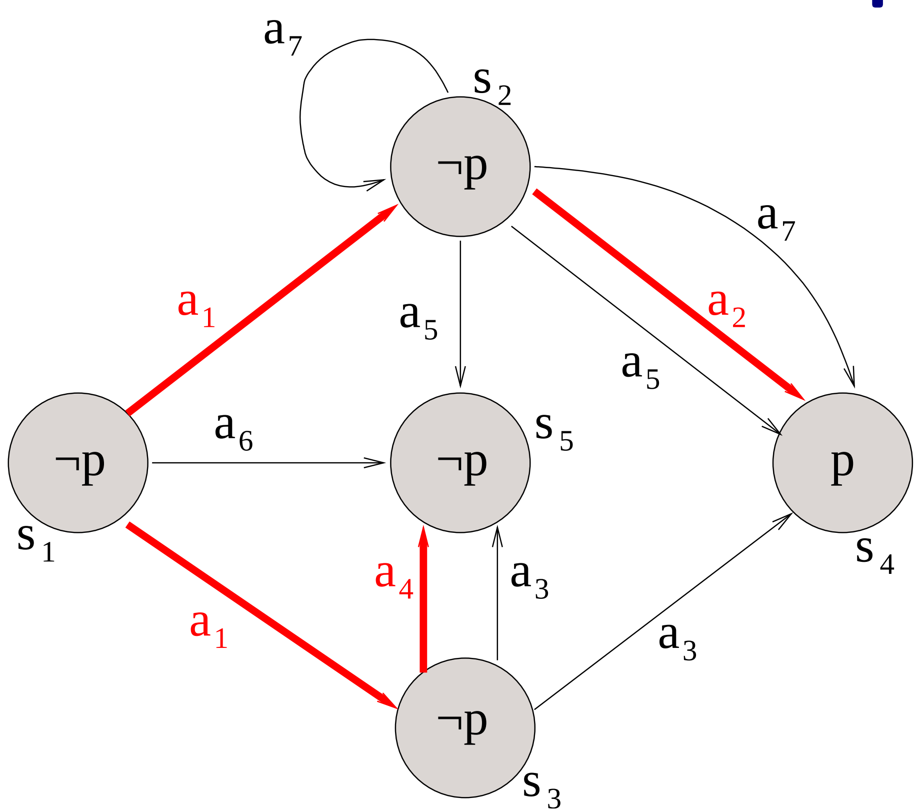
"Try your best to reach p"

An example



Policy π_1

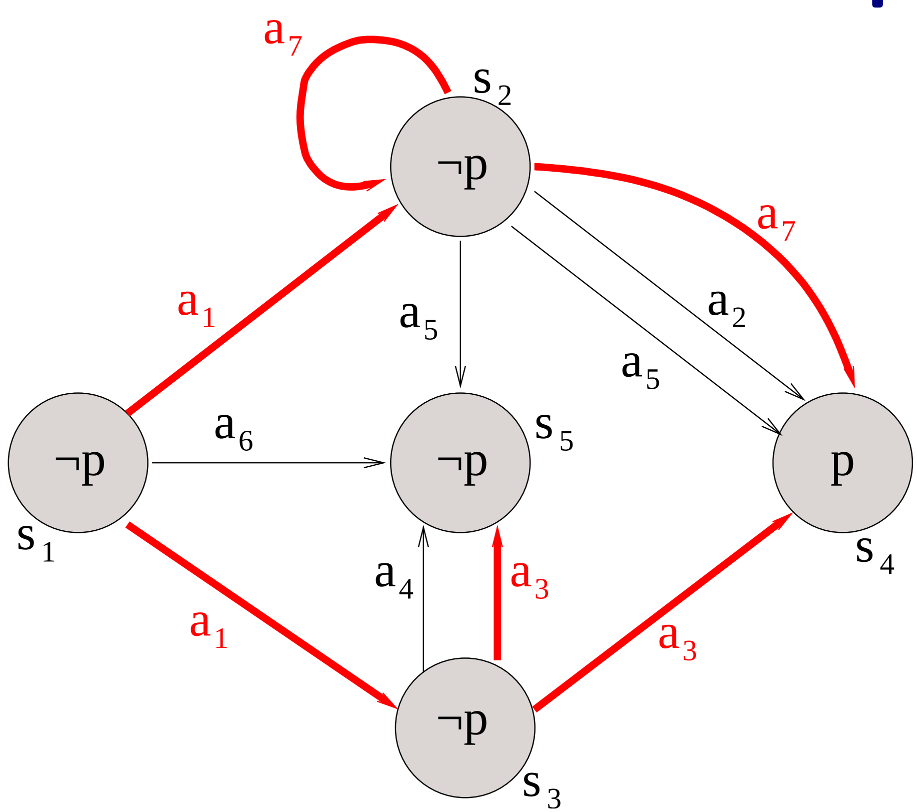
An example



Policy π_2

clearly *worse than* π_1!

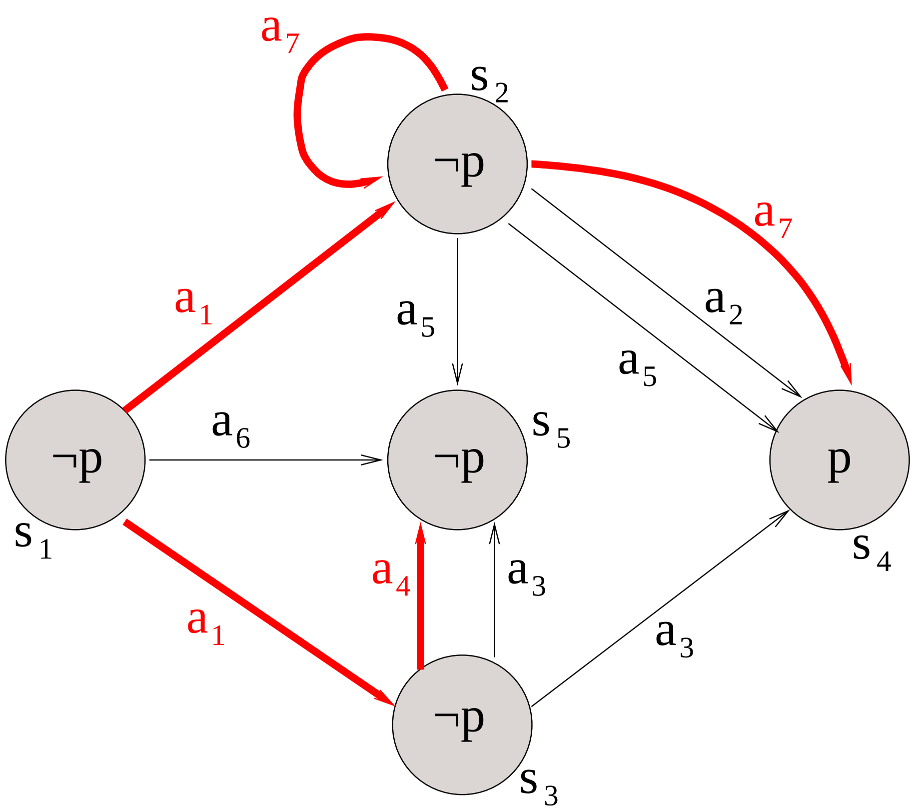
An example



Policy π_3

worse than π_1! but π_2?

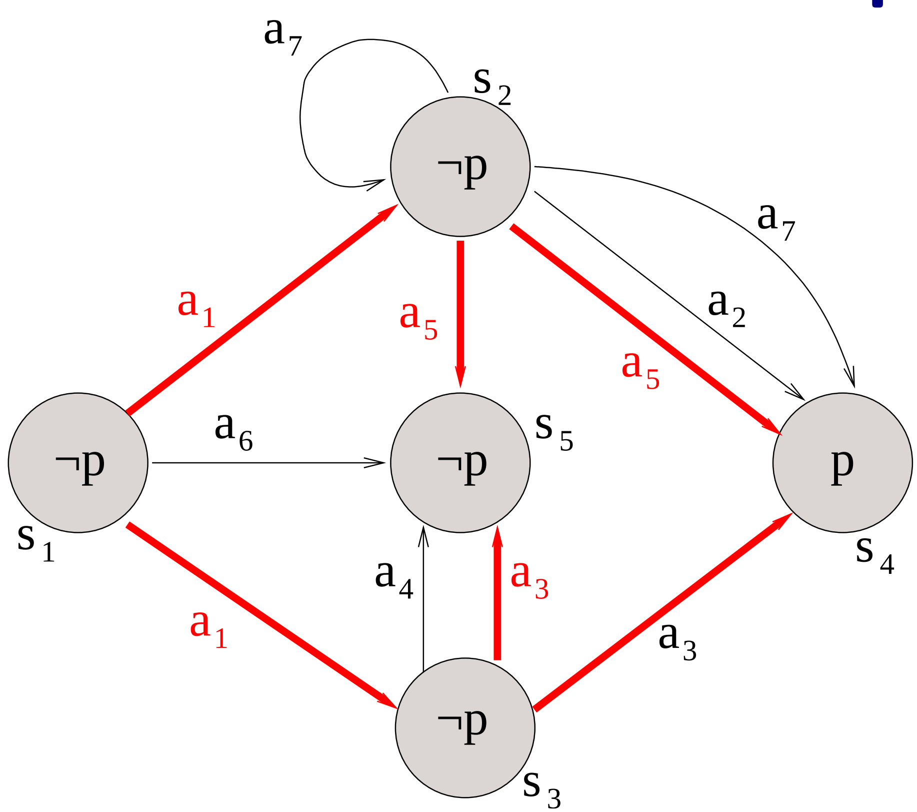
An example



Policy π_4

worse than π_2 and π_3

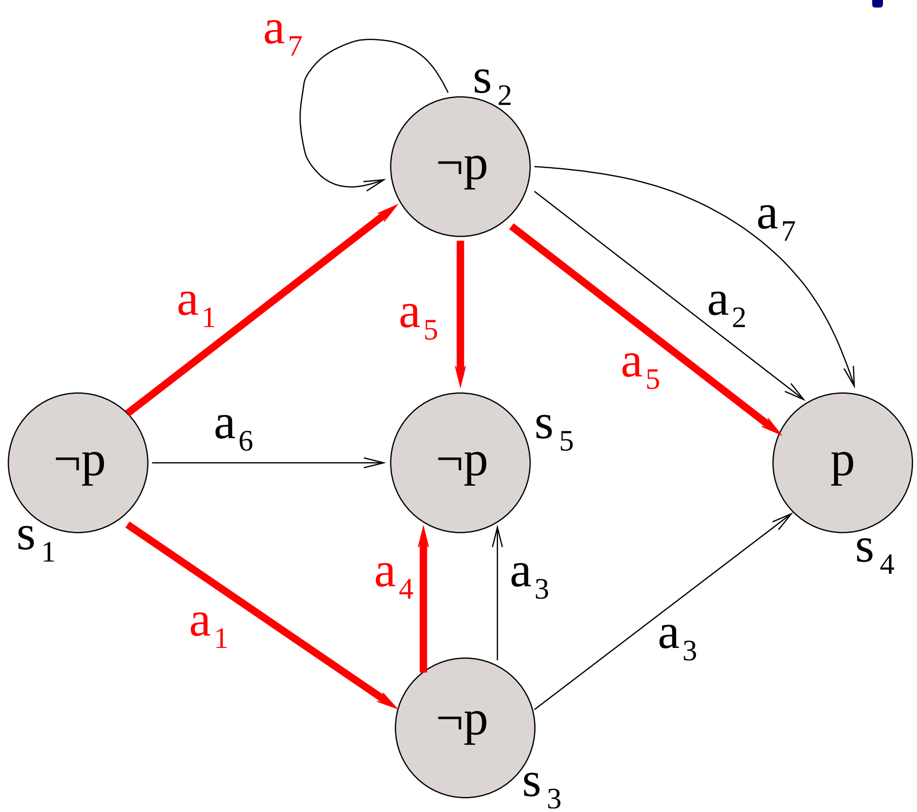
An example



Policy π_5

worse than π_3

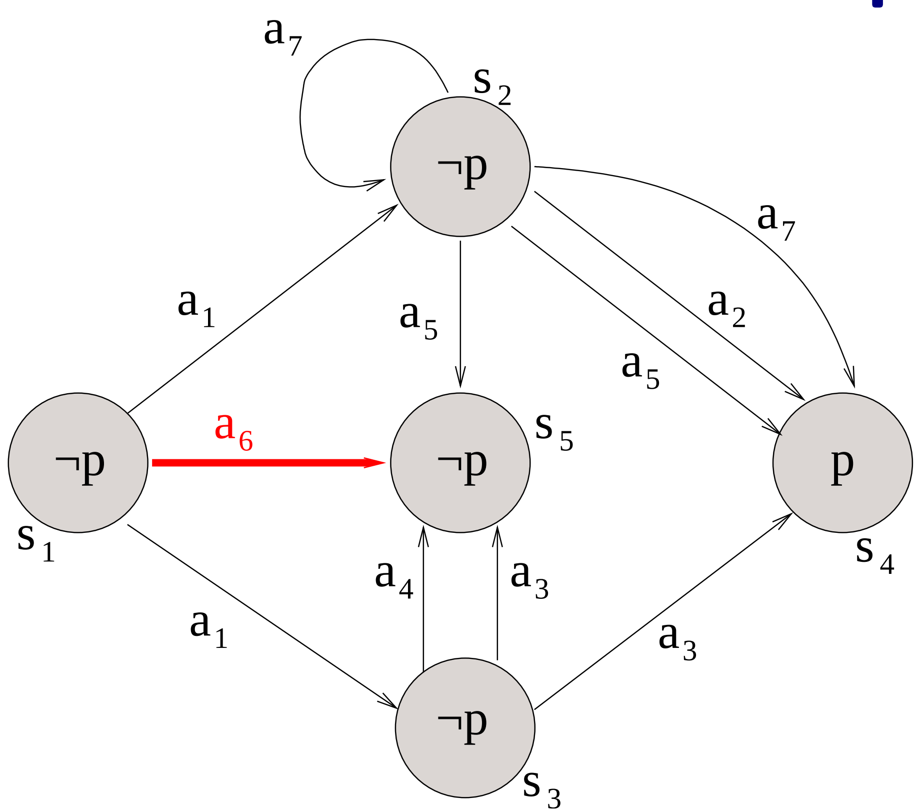
An example



Policy π_6

worse than π_4, π_5

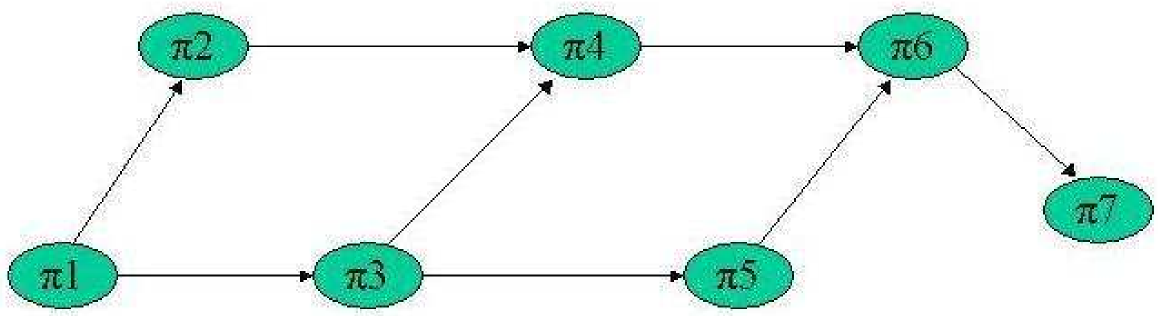
An example



Policy π_7

Really bad!

An example



Which is the best Policy?
How do we express "best policy"?

Existing Logic

- LTL: The property of a sequence of states besides the final state (if it exists);

Existing Logic

- LTL: The property of a sequence of states besides the final state (if it exists);

- CTL*: LTL $+$ properties of all pathes from each state;

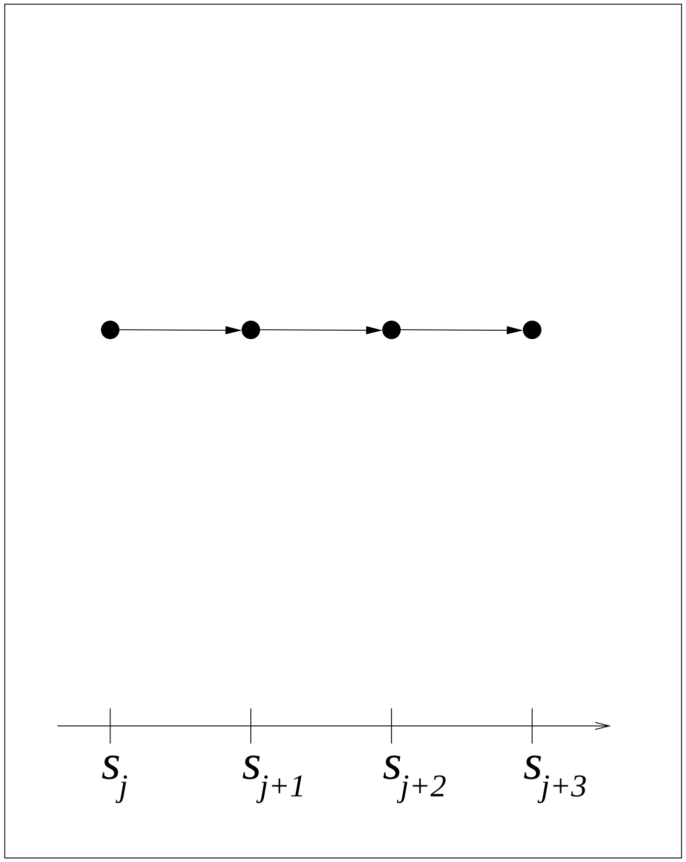
Existing Logic

- LTL: The property of a sequence of states besides the final state (if it exists);

- CTL*: LTL $+$ properties of all pathes from each state;

- π-CTL*: CTL* $+$ properties of all pathes in the policy from a state.

Linear Temporal Logic LTL



- Linear time: sequence of states
- Operators:

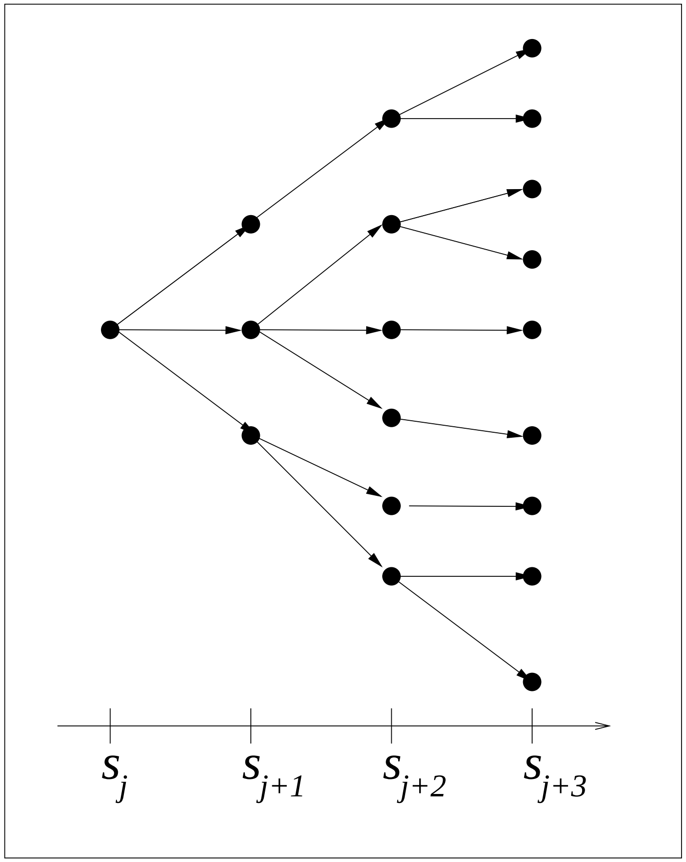
  $\Box p$ = always p

  $\Diamond p$ = eventually p

  $\bigcirc p$ = next p

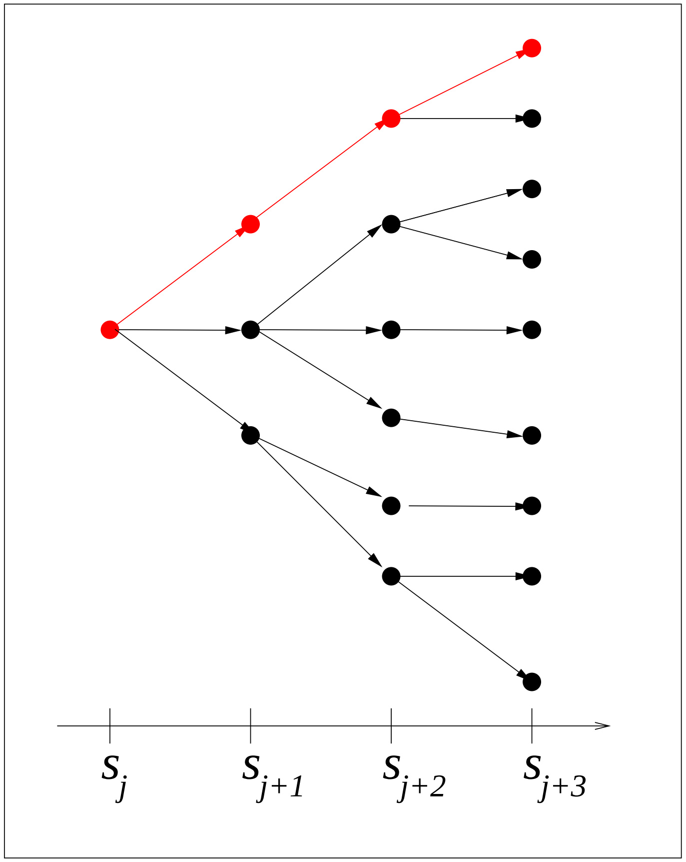
  $p \,\mathsf{U}\, q$ = p true until q

Branching Temporal logic CTL*



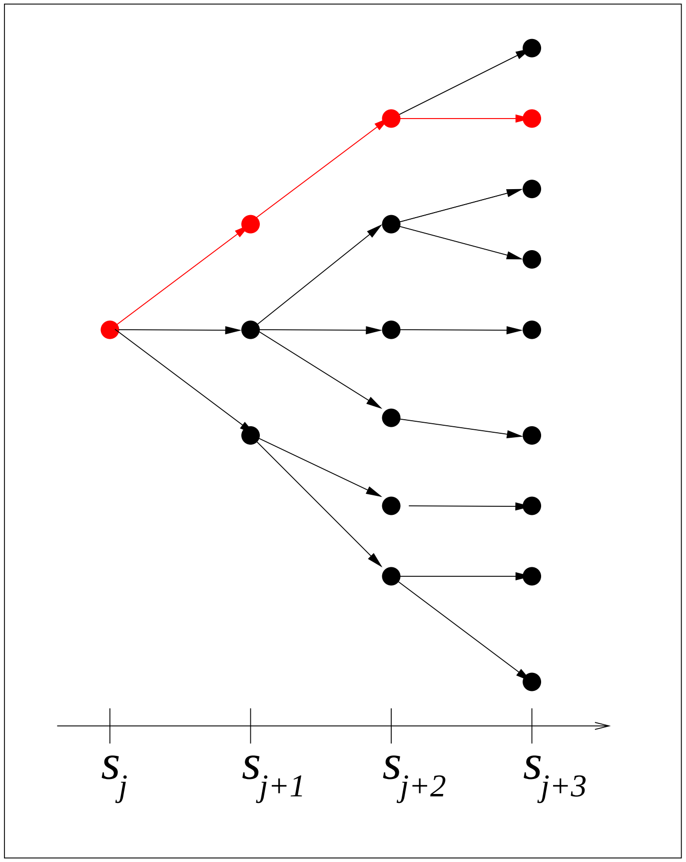
- Branching time
- New operators for paths

Branching Temporal logic CTL*



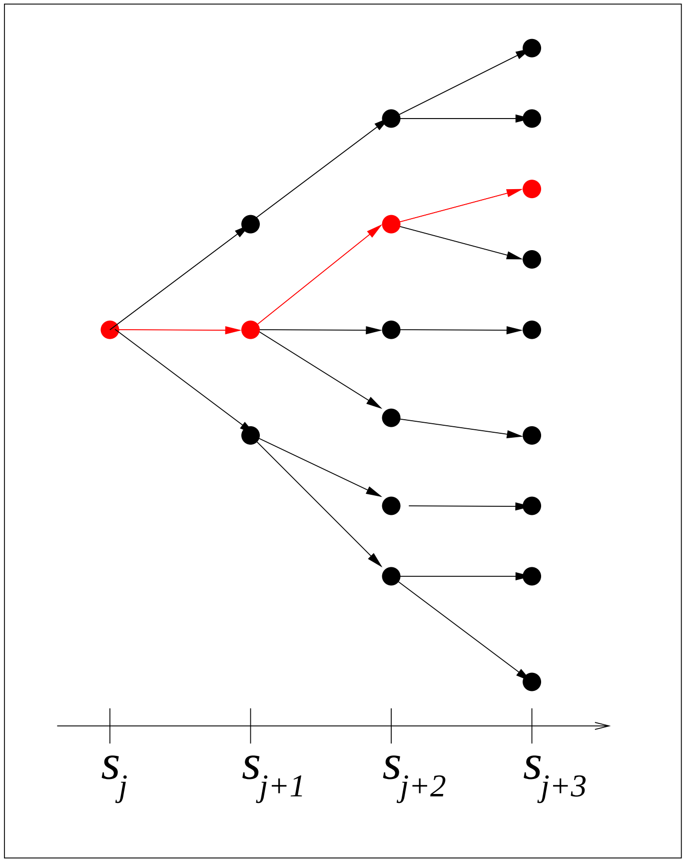
- Branching time
- New operators for paths

Branching Temporal logic CTL*



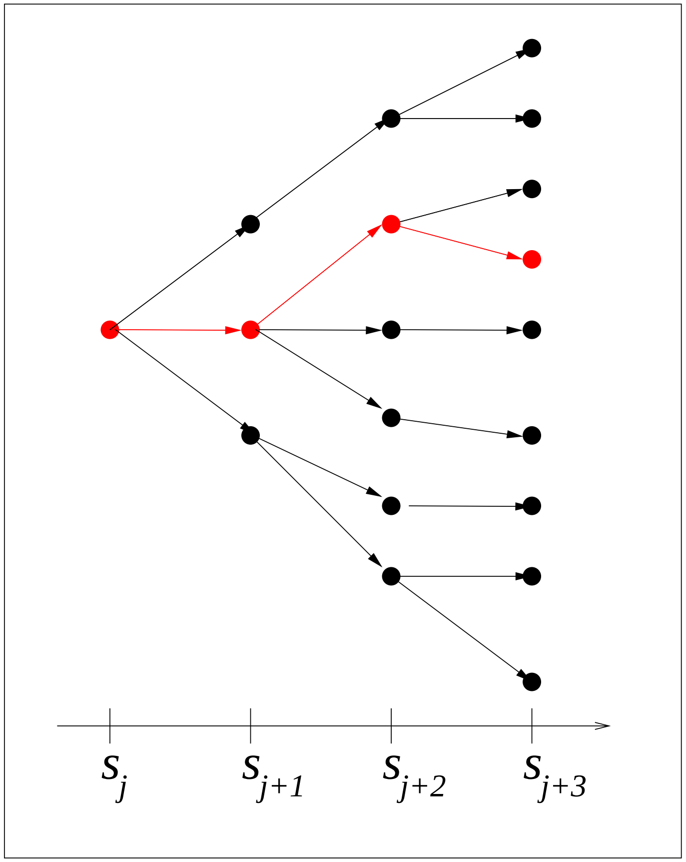
- Branching time
- New operators for paths

Branching Temporal logic CTL*



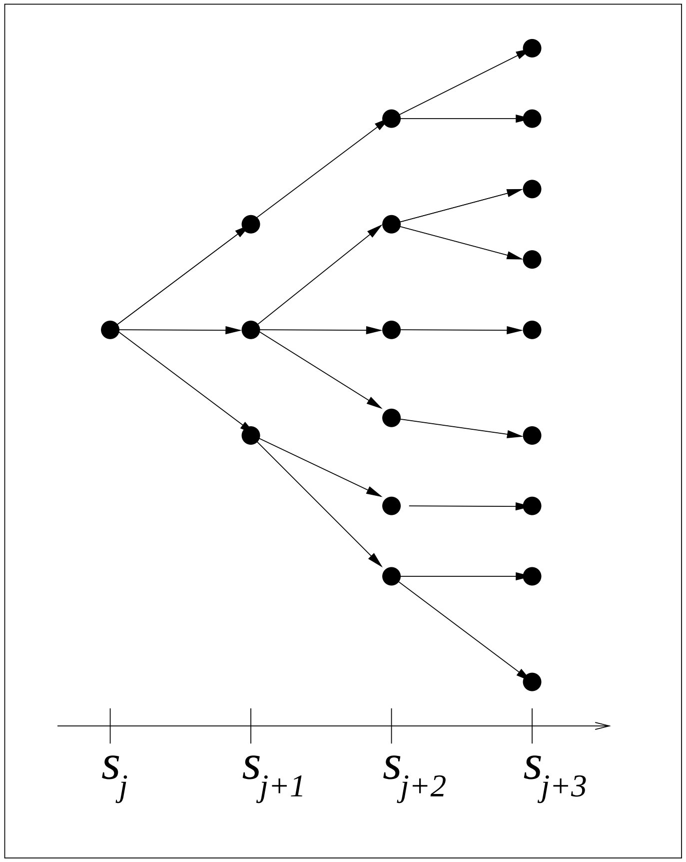
- Branching time
- New operators for paths

Branching Temporal logic CTL*



- Branching time
- New operators for paths

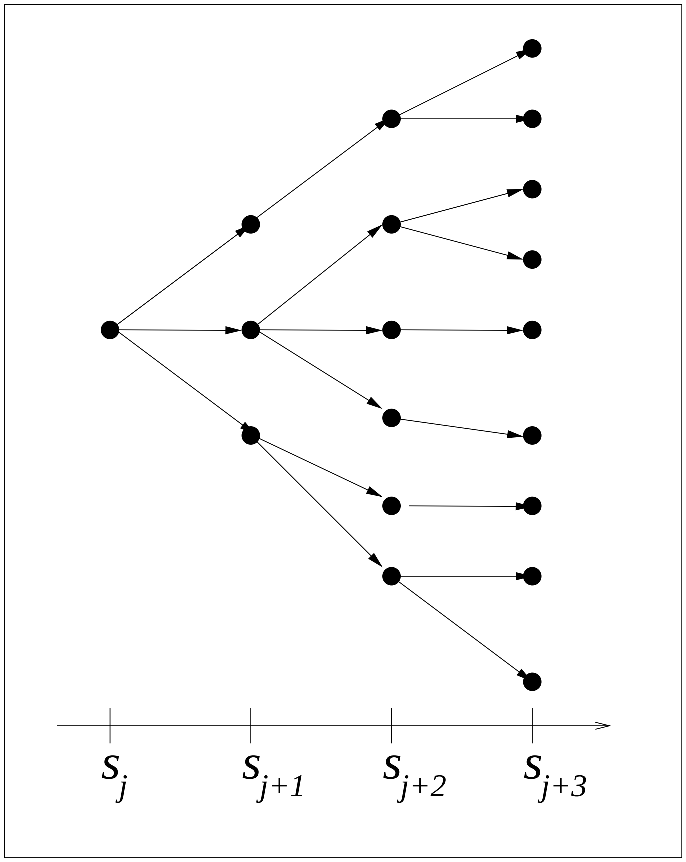
Branching Temporal logic CTL*



$\mathsf{A}\phi$ = for any path, ϕ holds

$\mathsf{E}\phi$ = for some path, ϕ holds

Branching Temporal logic CTL*



Examples:

$\mathsf{A}\Diamond p$ = all paths reach p

$\mathsf{E}\Box p$ = in some path, always p

Branching Temporal logic CTL*

Syntax:

$\langle p \rangle$ = propositional formula;

$\langle sf \rangle$ = "state" formula;

$\langle pf \rangle$ = "path" formula

Branching Temporal logic CTL*

Syntax:

$\langle p \rangle$ = propositional formula;

$\langle sf \rangle$ = "state" formula;

$\langle pf \rangle$ = "path" formula

$$\langle sf \rangle \quad ::= \quad \langle p \rangle \mid \langle sf \rangle \wedge \langle sf \rangle \mid \langle sf \rangle \vee \langle sf \rangle \mid \neg \langle sf \rangle \mid \mathsf{E} \langle pf \rangle \mid \mathsf{A} \langle pf \rangle$$

Branching Temporal logic CTL*

Syntax:

$\langle p \rangle$ = propositional formula;

$\langle sf \rangle$ = "state" formula;

$\langle pf \rangle$ = "path" formula

$$\langle sf \rangle \quad ::= \quad \langle p \rangle \mid \langle sf \rangle \wedge \langle sf \rangle \mid \langle sf \rangle \vee \langle sf \rangle \mid \neg \langle sf \rangle \mid \mathsf{E} \langle pf \rangle \mid \mathsf{A} \langle pf \rangle$$

$$\langle pf \rangle \quad ::= \quad \langle sf \rangle \mid \langle pf \rangle \vee \langle pf \rangle \mid \neg \langle pf \rangle \mid \langle pf \rangle \wedge \langle pf \rangle \mid$$
$$\langle pf \rangle \ \mathsf{U} \ \langle pf \rangle \mid \bigcirc \langle pf \rangle \mid \Diamond \langle pf \rangle \mid \Box \langle pf \rangle$$

The extension of CTL*: π-CTL*

Syntax:

$$\begin{array}{lcl} \langle sf \rangle & ::= & \langle p \rangle \mid \langle sf \rangle \wedge \langle sf \rangle \mid \langle sf \rangle \vee \langle sf \rangle \mid \neg \langle sf \rangle \mid \\ & & \mathsf{E} \langle pf \rangle \mid \mathsf{A} \langle pf \rangle \mid \underline{\mathsf{A}_\pi \langle pf \rangle \mid \mathsf{E}_\pi \langle pf \rangle} \\ \langle pf \rangle & ::= & \langle sf \rangle \mid \langle pf \rangle \vee \langle pf \rangle \mid \neg \langle pf \rangle \mid \langle pf \rangle \wedge \langle pf \rangle \mid \\ & & \langle pf \rangle \ \mathsf{U} \ \langle pf \rangle \mid \bigcirc \langle pf \rangle \mid \Diamond \langle pf \rangle \mid \Box \langle pf \rangle \end{array}$$

The extension of CTL*: π-CTL*

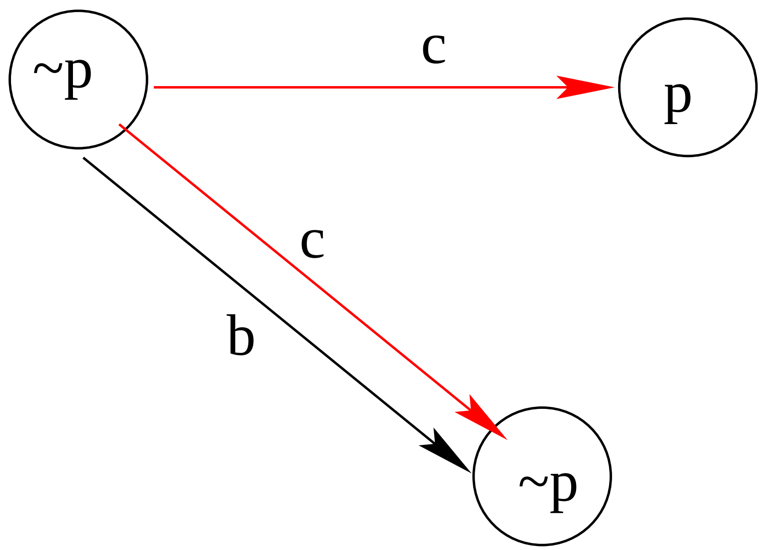
- group the set of paths from the initial state that all correspond to the same policy:

The extension of CTL*: π-CTL*

- group the set of paths from the initial state that all correspond to the same policy:
  - $\mathsf{A}_\pi\ pf$: 'for all paths that agree with the policy π, pf holds';
  - $\mathsf{E}_\pi\ pf$: 'there exists a path that agrees with the policy π for which pf holds'.
- By policy, we mean the mapping from states to actions.
- We now illustrate some goals in π-CTL*

Weak Plan

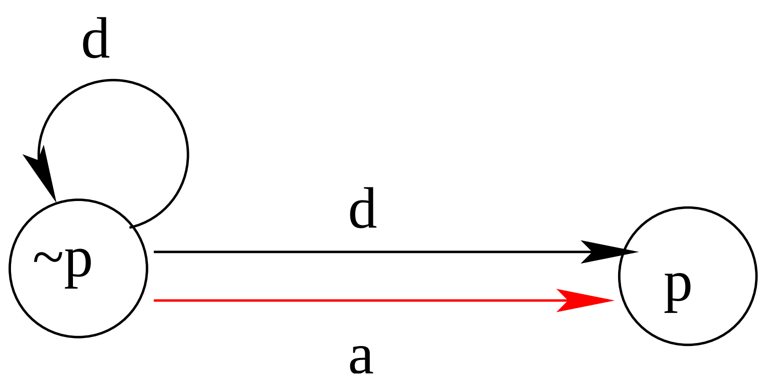
The weakest reachability goal "from the initial state there is a possibility that p can be reached" is expressed by $\mathsf{E}_\pi \Diamond p$.



(s_1, c) is a weak plan

strong plan

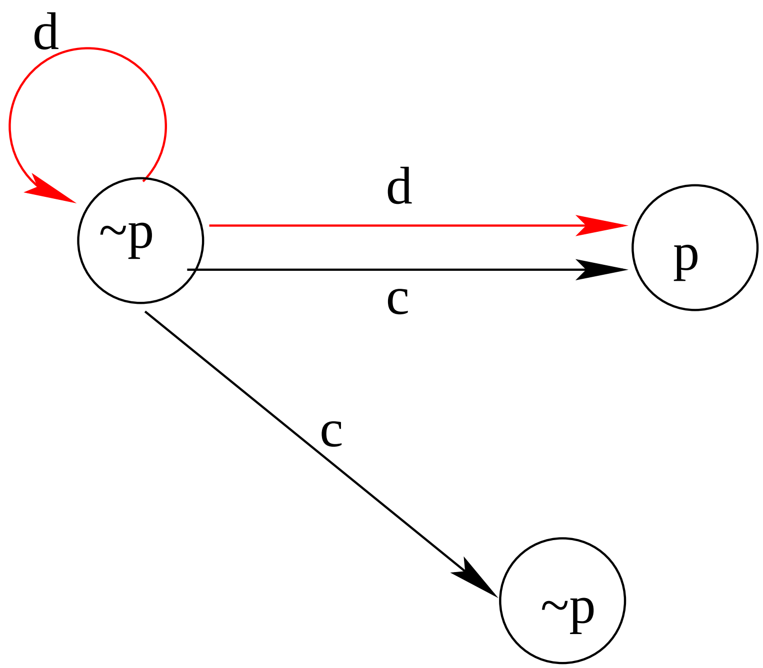
A stronger goal "from the initial state p must be reached" is expressed as $\mathsf{A}_\pi \Diamond p$.



(s_1, a) is a strong plan

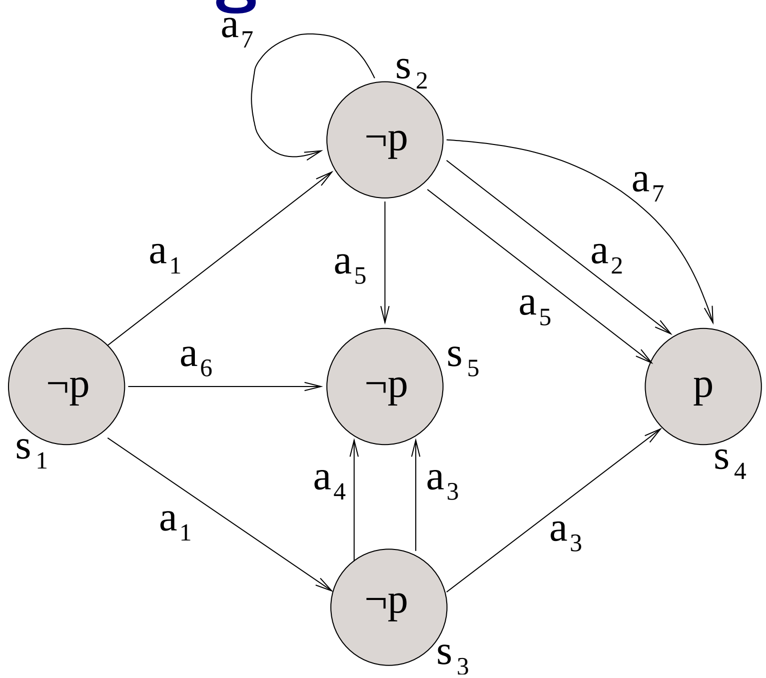
Strong cyclic plan

"All along the trajectory there is always a possible path to p by following the policy" is expressed as $\mathsf{A}_\pi \Box (\mathsf{E}_\pi \Diamond p)$.



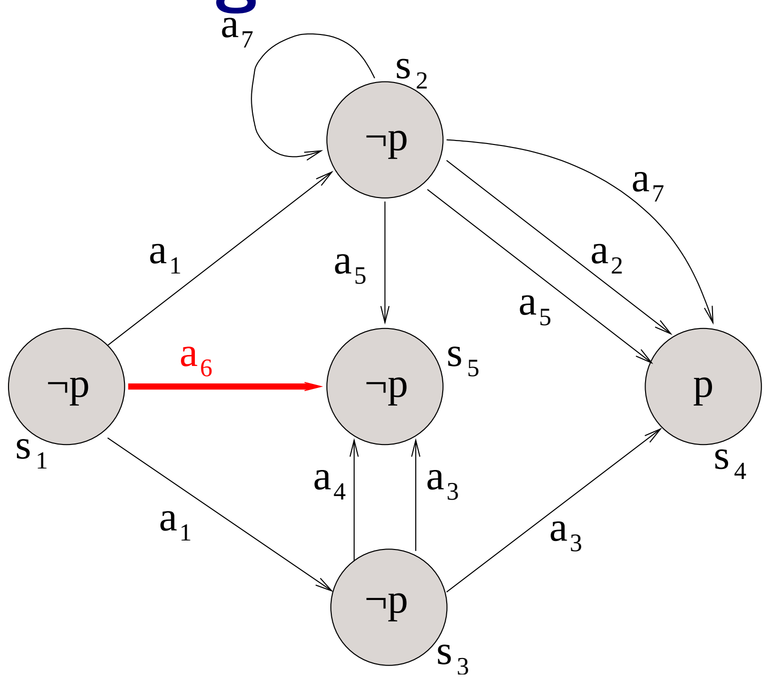
(s_1, d) is a strong cylcic plan.

Some goals in π-CTL*

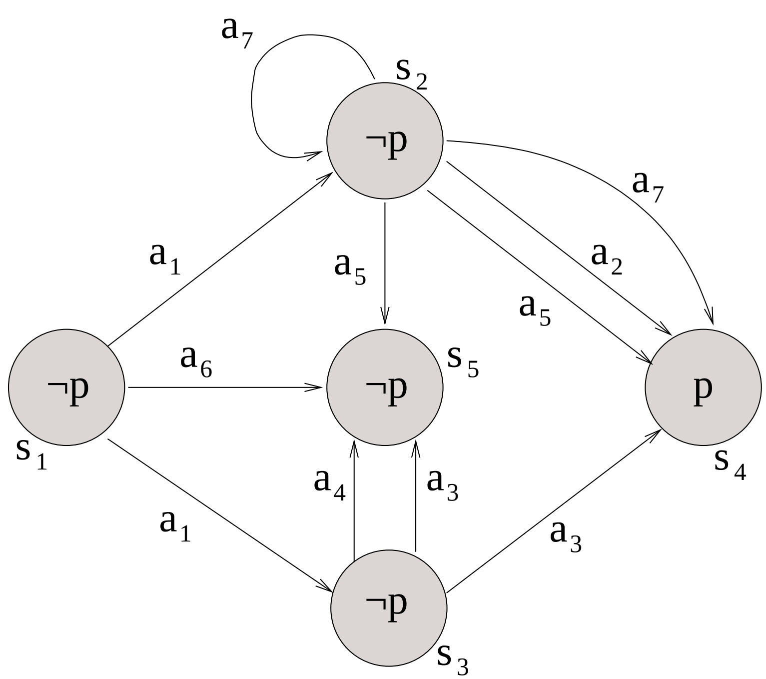


weak plan: The weakest reachability goal "from the initial state there is a possibility that p can be reached" is expressed by $\mathsf{E}_\pi \Diamond p$.

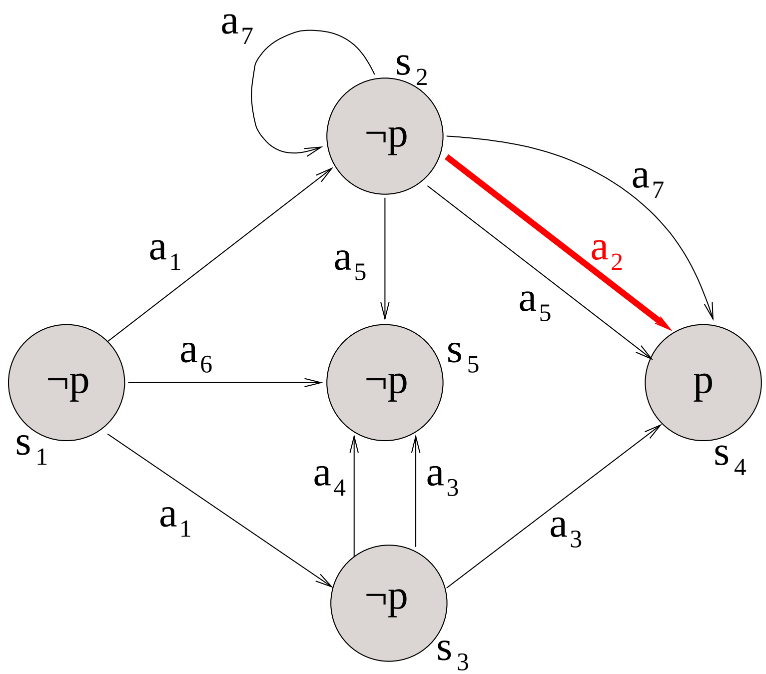
Some goals in π-CTL*



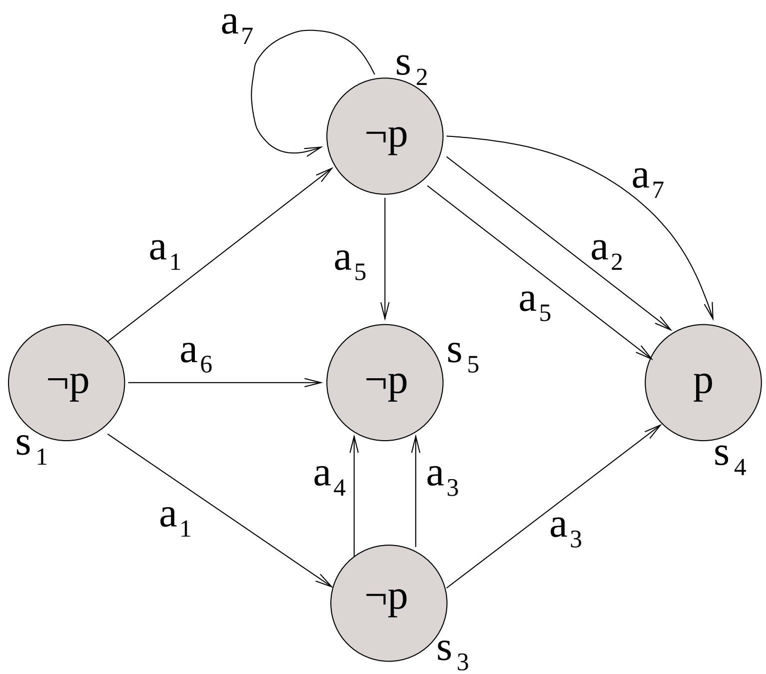
weak plan: The weakest reachability goal "from the initial state there is a possibility that p can be reached" is expressed by $\mathsf{E}_\pi \Diamond p$. From s_1, all policies but π_7 satisfy the goal.



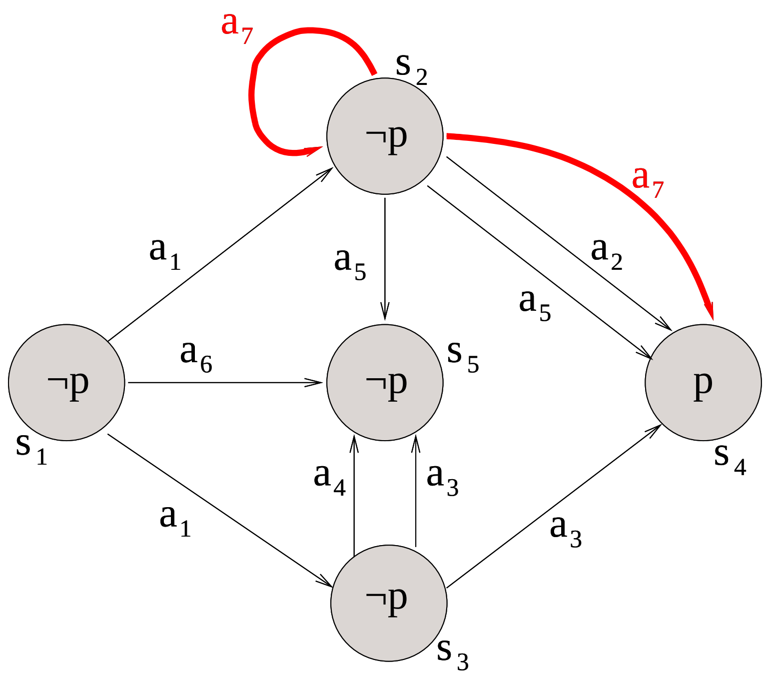
strong plan: A stronger goal "from the initial state p must be reached" is expressed as $\mathsf{A}_\pi \Diamond p$. For s_1, no policy makes it true.



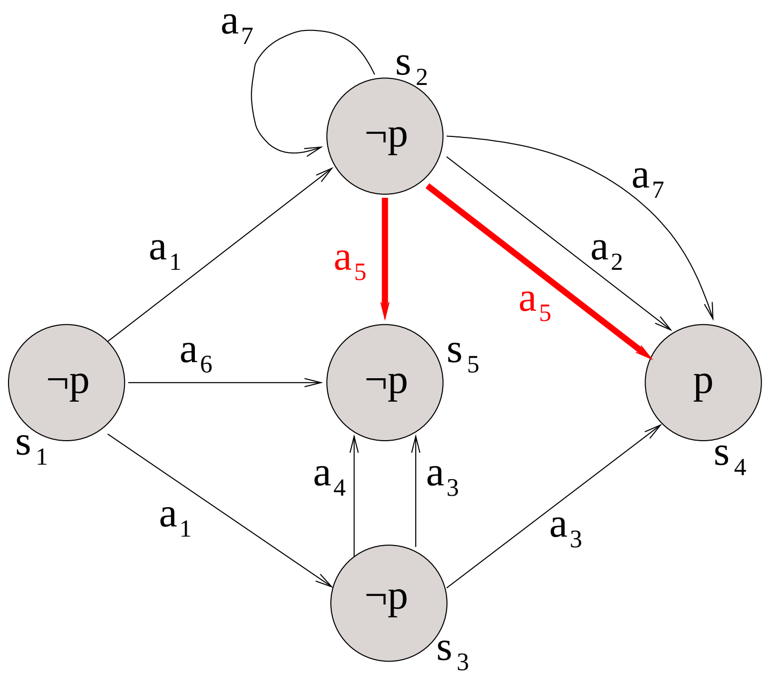
strong plan: A stronger goal "from the initial state p must be reached" is expressed as $\mathsf{A}_\pi \Diamond p$. For s_1, no policy makes it true. But, for instance, for s_2 the policy $\{(s_2, a_2)\}$ satisfies the goal.



"All along the trajectory there is always a possible path to p by following the policy" is expressed as $\mathsf{A}_\pi \Box (\mathsf{E}_\pi \Diamond p)$. For s_1, no policy.



"All along the trajectory there is always a possible path to p by following the policy" is expressed as $\mathsf{A}_\pi \Box (\mathsf{E}_\pi \Diamond p)$. For s_1, no policy. For s_2, policies $\{(s_2, a_2)\}$ and $\{(s_2, a_7)\}$ satisfy this goal.



However, policy $\{(s_2, a_5)\}$ does not, (we could go to s_5 from where p can not be reached).

More examples

- $\mathsf{A}_\pi(\mathsf{E}\Diamond p) =$ "All along the trajectory there is always a possible path to p, but this path is not necessary abide the policy the agent taken".

More examples

- $\mathsf{A}_\pi(\mathsf{E}\Diamond p)$ = "All along the trajectory there is always a possible path to p, but this path is not necessary abide the policy the agent taken".
- $\mathsf{A}(\mathsf{E}_\pi\Diamond p)$ = "For any state that is reachable from the initial state, there is always a path to p by following the policy."

More examples

- $\mathsf{A}_\pi(\mathsf{E}\Diamond p)$ = "All along the trajectory there is always a possible path to p, but this path is not necessary abide the policy the agent taken".
- $\mathsf{A}(\mathsf{E}_\pi\Diamond p)$ = "For any state that is reachable from the initial state, there is always a path to p by following the policy."
- $\mathsf{E}\Diamond p \rightarrow \mathsf{E}_\pi\Diamond p$ = "from the initial state, if it is possible to reach p, the agent should possibly reach p". Useful to allow the agent to pursue an alternative goal when it realizes that its initial goal is no longer achievable.

More examples

- $\mathsf{A}_\pi(\mathsf{E}\Diamond p)$ = "All along the trajectory there is always a possible path to p, but this path is not necessary abide the policy the agent taken".
- $\mathsf{A}(\mathsf{E}_\pi\Diamond p)$ = "For any state that is reachable from the initial state, there is always a path to p by following the policy."
- $\mathsf{E}\Diamond p \rightarrow \mathsf{E}_\pi\Diamond p$ = "from the initial state, if it is possible to reach p, the agent should possibly reach p". Useful to allow the agent to pursue an alternative goal when it realizes that its initial goal is no longer achievable.
- $\mathsf{A}_\pi\Box(\mathsf{E}\Diamond p \rightarrow \mathsf{E}_\pi\Diamond p)$ = idem, but now from *any* state in the trajectory (not only initial one).

P-CTL*

To find the best policy, the comparasion of policies is necessary. For example:

All along your trajectory
if from any state p can be achieved for sure,
then the policy being executed must achieve p,
else ......

P-CTL*

To find the best policy, the comparasion of policies is necessary. For example:

All along your trajectory
if from any state p can be achieved for sure,
then the policy being executed must achieve p,
else ......

- $\mathcal{AP}$: 'for all policies from the state, the property is hold';
- $\mathcal{EP}$: 'there exist a policy from the state such that the property is hold in the policy'.

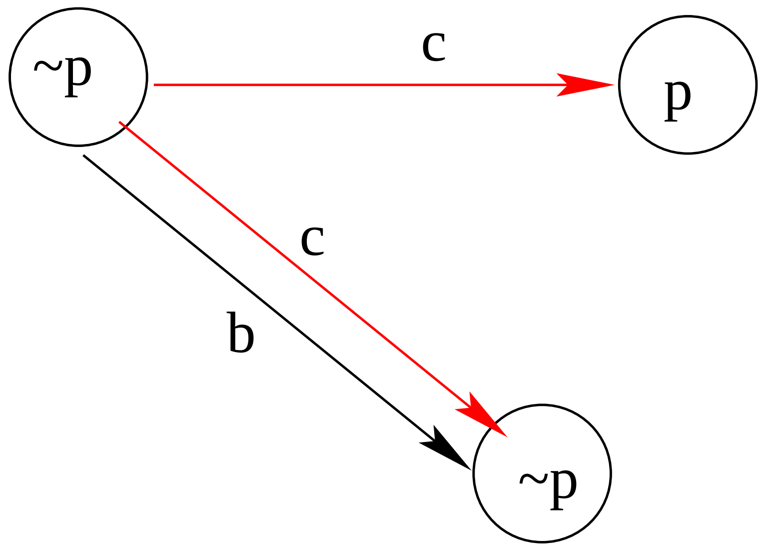
P-CTL*

Syntax:

$$
\begin{aligned}
\langle sf \rangle \quad ::= \quad & \langle p \rangle \mid \langle sf \rangle \wedge \langle sf \rangle \mid \langle sf \rangle \vee \langle sf \rangle \mid \neg \langle sf \rangle \mid \\
& \mathsf{E}\langle pf \rangle \mid \mathsf{A}\langle pf \rangle \mid \mathsf{A}_{\pi}\langle pf \rangle \mid \mathsf{E}_{\pi}\langle pf \rangle \mid \underline{\mathcal{AP}\langle sf \rangle \mid \mathcal{EP}\langle sf \rangle} \\
\langle pf \rangle \quad ::= \quad & \langle sf \rangle \mid \langle pf \rangle \;\vee\; \langle pf \rangle \mid \neg \langle pf \rangle \mid \langle pf \rangle \wedge \langle pf \rangle \mid \\
& \langle pf \rangle \;\mathsf{U}\; \langle pf \rangle \mid \bigcirc \langle pf \rangle \mid \Diamond \langle pf \rangle \mid \Box \langle pf \rangle
\end{aligned}
$$

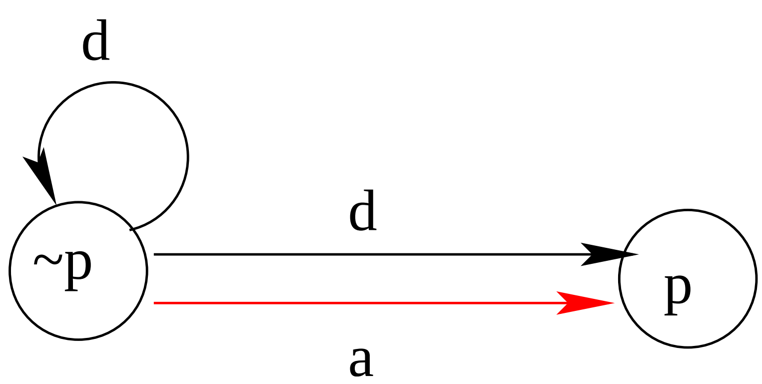
Goals in P-CTL*: Based on the weak plan

"from the initial state, if there is a policy such that p is possibilly reached, then in the policy chosen by the agent, p is possibilly reached" is expressed by $(\mathcal{EP}\mathsf{E}_\pi \Diamond p) \rightarrow (\mathsf{E}_\pi \Diamond p)$.

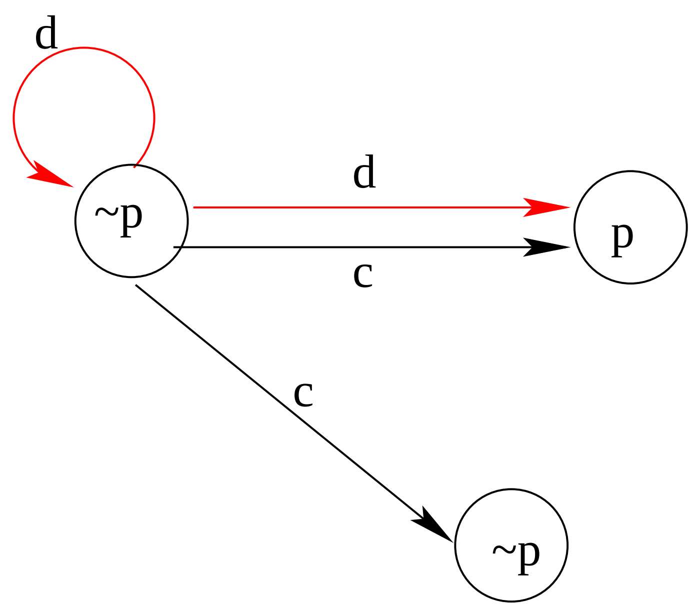


Based on the strong plan

"from the initial state, if there is a policy such that p must be reached, then in the policy chosen by the agent, p must be reached" is expressed by $(\mathcal{E}\mathcal{P}\mathsf{A}_\pi\Diamond p) \rightarrow (\mathsf{A}_\pi\Diamond p)$.



Based on the strong cyclic plan: "from the initial state, if there is a policy such that all along the trajectory there is always a possible path to p by following the policy, then in any state of the chosen policy, there is always a possible path to p" is expressed as $\mathcal{E}\mathcal{P}(\mathsf{A}_\pi\Box(\mathsf{E}_\pi\Diamond p)) \rightarrow (\mathsf{A}_\pi\Box(\mathsf{E}_\pi\Diamond p))$.



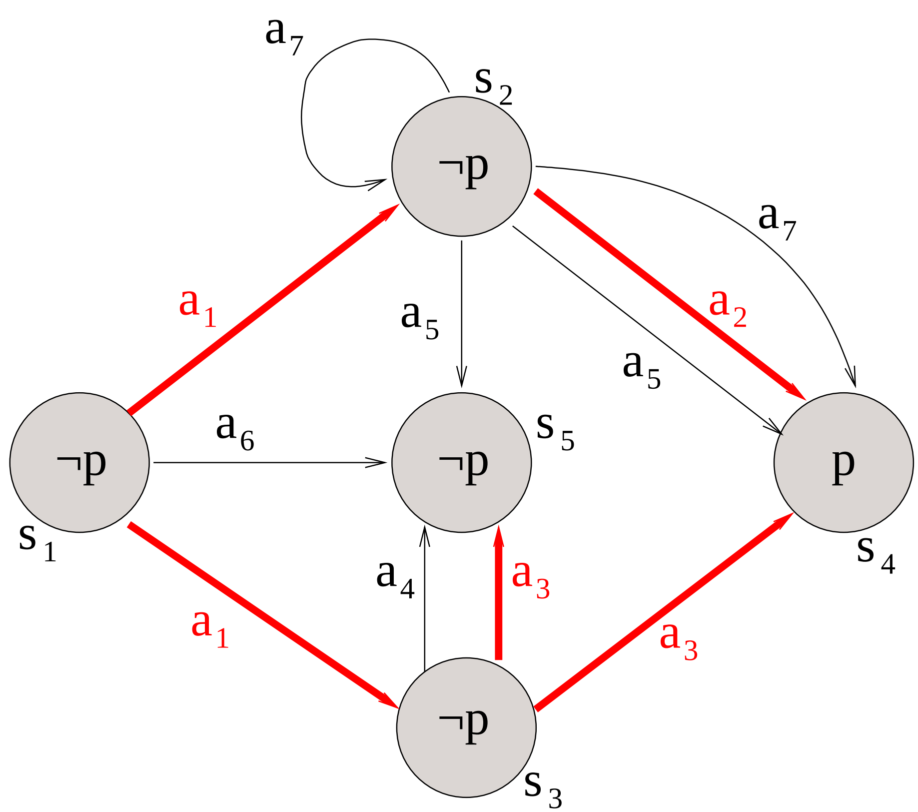
one version of Try your best to reach p

"In any state, if there is a policy that is possibly reach p, then the agnet should possibly reach p; if there is a policy that guarantees to reach p, then the agent should guarantee to reach p; if there is a policy such that in any state of the policy, there is a path to p, then in the policy chosen by the agent, there is always a path to p."

It is expressed as

$$\begin{aligned} &\mathsf{A}_\pi\Box((\mathcal{EP}\mathsf{E}_\pi\Diamond p) \rightarrow (\mathsf{E}_\pi\Diamond p)) \\ \wedge &\mathsf{A}_\pi\Box((\mathcal{EP}\mathsf{A}_\pi\Diamond p) \rightarrow (\mathsf{A}_\pi\Diamond p)) \\ \wedge &\mathsf{A}_\pi\Box(\mathcal{EP}(\mathsf{A}_\pi\Box(\mathsf{E}_\pi\Diamond p)) \rightarrow (\mathsf{A}_\pi\Box(\mathsf{E}_\pi\Diamond p))) \end{aligned}$$

Policy π_1 in the previous example is the "Best" policy



- In State s_1: There is a policy that has path to p, but no policy can guarantee to reach p
- In state s_2: There is an action (a_5) that has path to p , there is an action (a_7) that in any state of any path in the policy, there is alway a hope of reaching p, and there is an action (a_2) that guarantee to reach p.
- In state s_3: There is an action (a_3) that has path to p
- In state s_4: p is reached
- In state s_5: No policy has path to p, give up.

Goal presentation	Satisfiable policies
$\mathsf{E}_\pi \Diamond p$	$\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6$
$\mathsf{A}_\pi \Box(\mathcal{EP}\mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p) \rightarrow \mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p))$	$\pi_1, \pi_2, \pi_3, \pi_4, \pi_7$
$\mathsf{A}_\pi \Box(\mathcal{EP}\mathsf{E}_\pi \Diamond p \rightarrow \mathsf{E}_\pi \Diamond p)$	π_1, π_3, π_5
$\mathsf{E}_\pi \Diamond p \wedge \mathsf{A}_\pi \Box(\mathcal{EP}\mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p) \rightarrow \mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p))$	$\pi_1, \pi_2, \pi_3, \pi_4$
$\mathsf{A}_\pi \Box(\mathcal{EP}\mathsf{A}_\pi \Diamond p \rightarrow \mathsf{A}_\pi \Diamond p)$	π_1, π_2, π_7
$\mathsf{E}_\pi \Diamond p \wedge \mathsf{A}_\pi \Box(\mathcal{EP}\mathsf{A}_\pi \Diamond p \rightarrow \mathsf{A}_\pi \Diamond p)$	π_1, π_2
$\mathsf{A}_\pi \Box((\mathcal{EP}\mathsf{E}_\pi \Diamond p \rightarrow \mathsf{E}_\pi \Diamond p) \wedge (\mathcal{EP}\mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p) \rightarrow \mathsf{A}_\pi \Box(\mathsf{E}_\pi \Diamond p)))$	π_1, π_3
$\mathsf{A}_\pi \Box((\mathcal{EP}\mathsf{E}_\pi \Diamond p \rightarrow \mathsf{E}_\pi \Diamond p) \wedge (\mathcal{EP}\mathsf{A}_\pi \Diamond p \rightarrow \mathsf{A}_\pi \Diamond p))$	π_1
$\mathsf{A}_\pi \Diamond p$	$\emptyset$

More examples

- $\mathsf{A}_\pi \Box ((\mathcal{E}\mathcal{P}\mathsf{A}_\pi \Diamond p) \mathsf{U} q) =$ " reach q but want to make sure that all along the path if necessary it can make a new policy that can guarantee to reach q".

More examples

- $\mathsf{A}_\pi \Box((\mathcal{EP}\mathsf{A}_\pi \Diamond p)\mathsf{U}q)$ = " reach q but want to make sure that all along the path if necessary it can make a new policy that can guarantee to reach q".
- $\mathsf{A}_\pi \Box(\mathcal{APE}_\pi \neg \Box p \rightarrow \mathsf{A}_\pi(q\mathsf{U}p) \wedge \mathcal{EP}\mathsf{A}_\pi \Box p \rightarrow \mathsf{A}_\pi \Box p)$ = "Maintain p true and if that is not guaranteedly possible, then it must maintain q true until p becomes true."

Conclusions

- We extended π-CTL* to capable of comparing policies

- P-CTL* is a proper superest of mentioned existing languages

- P-CTL* is capable of capturing several degrees of "trying the best of reaching p"