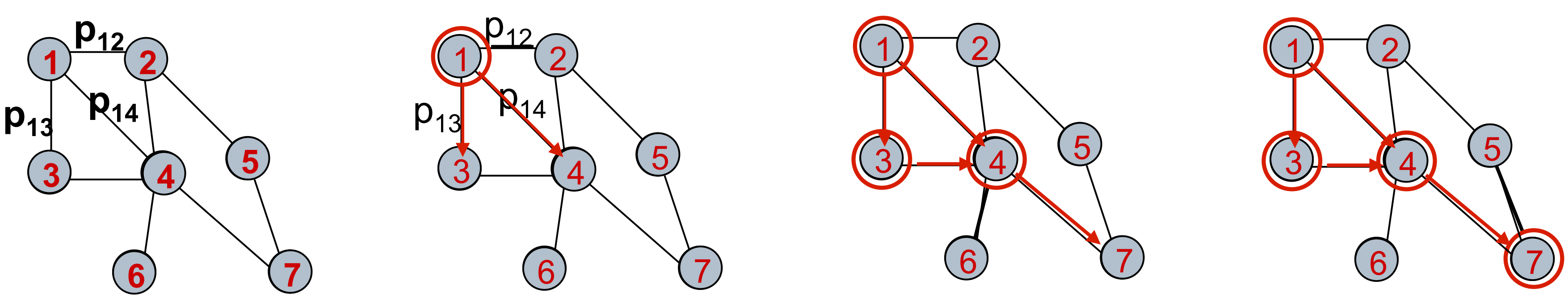
An Empirical Study of Optimization for Maximizing Diffusion in Networks

Kiyan Ahmadizadeh
Bistra Dilkina, Carla P. Gomes, Ashish Sabharwal
Cornell University
Institute for Computational Sustainability

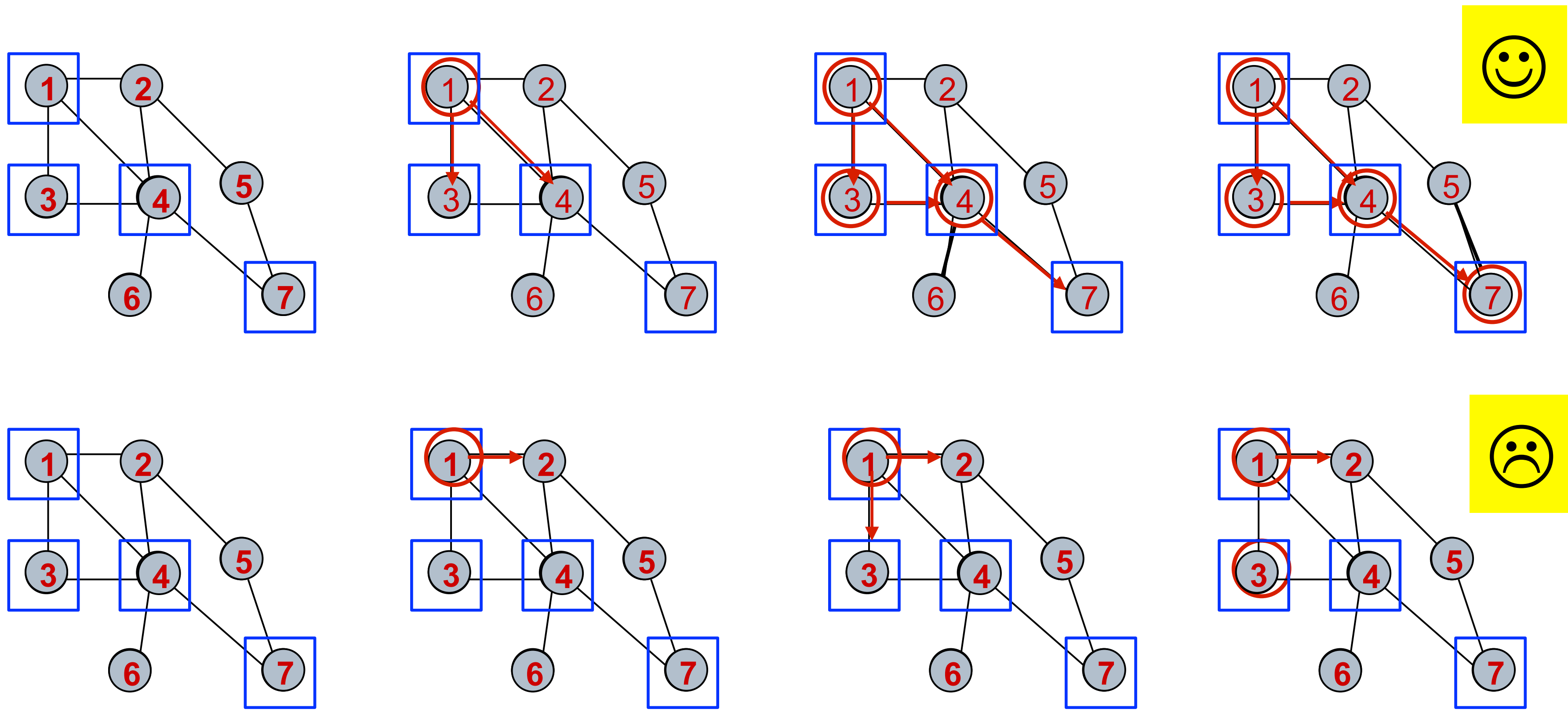
Diffusion in Networks: Cascades

- Our diffusion model: *cascades*
- A network: $G=(V,E)$
- Initial set of active nodes $S \subseteq V$
- Diffusion process as local stochastic activation rules of spread from active nodes to their neighbors
  - Independent cascade: probability of spread across each edge: $p_{vw}\ \forall (v,w) \in E$ (independent of cascade history)



Influencing Cascades

- Assume cascades can only spread to nodes acquired by some *action.*



Maximizing Node Activity in Cascades

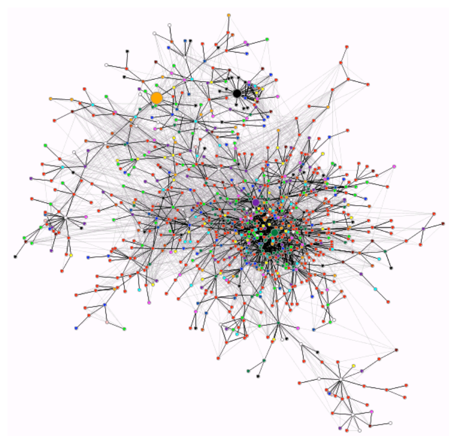
- A set of actions $A = \{a_1..a_L\}$, $a \subseteq A$
- a_i : cost $c(a_i)$, buys nodes $V_i \subseteq V$. Total budget B.
- Time horizon H (discrete).
  - Typically many years.
- $X_v(a,t)$: random variable indicating whether node v becomes activated in cascade under action set a at time t

$$\max_a \sum_{v \in V} E[X_v(a,H)] \text{ s.t. } \sum_{a_i \in a} c(a_i) \leq B$$

Influencing Cascades: Motivating Examples

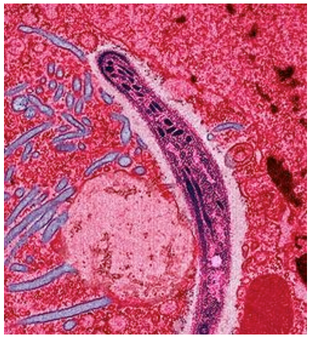
- Human Networks: Technology adoption among friends/ peers.
- Social Networks:
  - Spread of rumor/news/articles on Facebook, Twitter, or among blogs/websites.

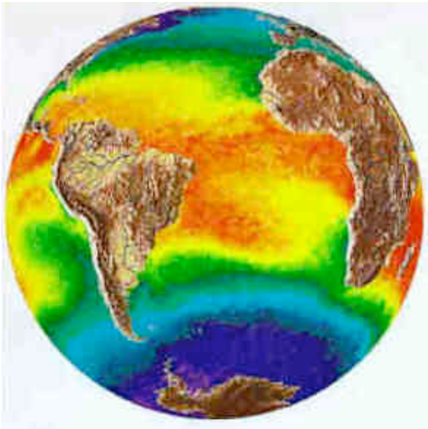
Targeted-actions (e.g. marketing campaigns) can be chosen to optimize the spread of these phenomena.



Influencing Cascades: Motivating Examples

- Epidemiology: Spread of disease is a cascade.
  - In human networks, or between networks of households, schools, major cities, etc.
  - In agriculture settings.





- Contamination: The spread of toxins / pollutants within water networks.

Mitigation strategies can be chosen to minimize the spread of such phenomena.

Our Application: Species Conservation

- Intuition: Buy land as future species habitat.
- Nodes: Land patches suitable as habitat (if conserved).
- Actions: Purchasing a real-estate parcel (containing a set of land patches).

Our Application: Species Conservation

- Given existing populations in some patches, a limited budget, and cascade model of species dispersion:
  - Which real-estate parcels should be purchased as conservation reserves to ***maximize*** the expected number of populated patches at the time horizon?
- Target species: the Red-Cockaded Woodpecker
  - Federally listed rare and endangered species [USA Fish and Wildlife Service, 2003].

RCW Cascade Model

- Recall spread probabilities: $p_{vw}\ \forall (v,w) \in E$
- Spread probability between pairs of land patches:
  - Distance.
  - Suitability score.
- Land patches remain active between time-steps based on a *survival probability*.
- Cascade model based on *meta-population* model [Walters et al., 2002]

Past Work

- [Kempe et al., 2003] – Initiating cascades.
  - Limited to choosing start nodes for cascade.
  - Problem is *sub-modular* (greedy methods apply).
  - Sub-modularity does not hold in more general settings.

- [Sheldon et al., 2010] – Single-stage node acquisition for cascades.
  - Unrealistic in many planning situations.
    - Large planning horizons => multiple rounds of purchases.

Talk Goals

- Study and compare three problem variants
  - (A) Single-stage up-front budget.
  - (B) Single-stage split budget.
  - (C) Two-stage split budget.

- Explore the computational difficulty of this problem.

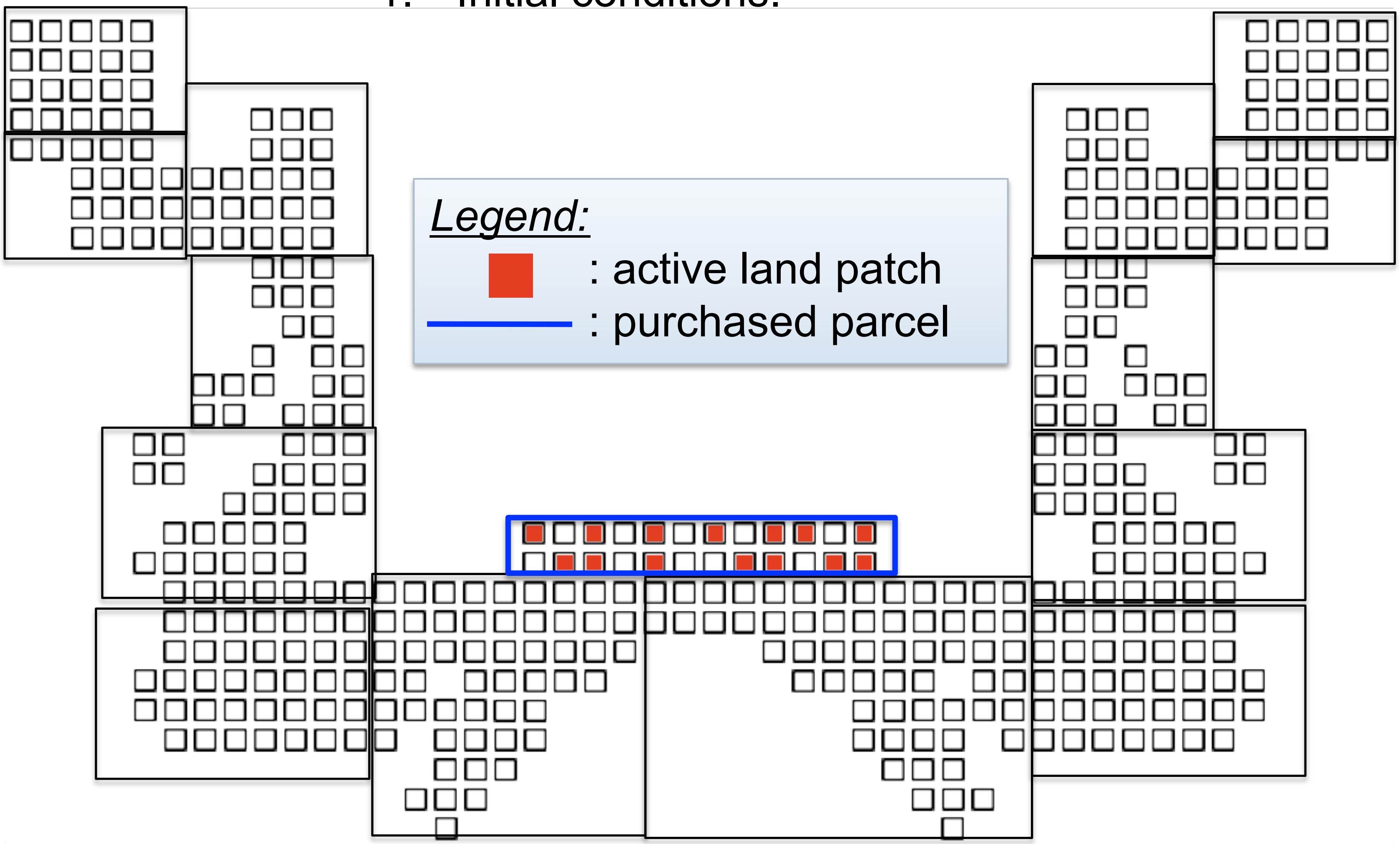
- Explore the tradeoffs in solution quality (expected number of active nodes) obtained from these three models.
  - Informs planners and planning policy makers.

Single-stage Decision Making

- Commit to all purchase decisions at t=0.
  - Decisions not informed by cascade progress *(closed loop)*.
- (A) Single-stage Up-front Budget:
  - Commit to purchases at t=0.
  - Make purchases at t=0.
  - Already computationally difficult.

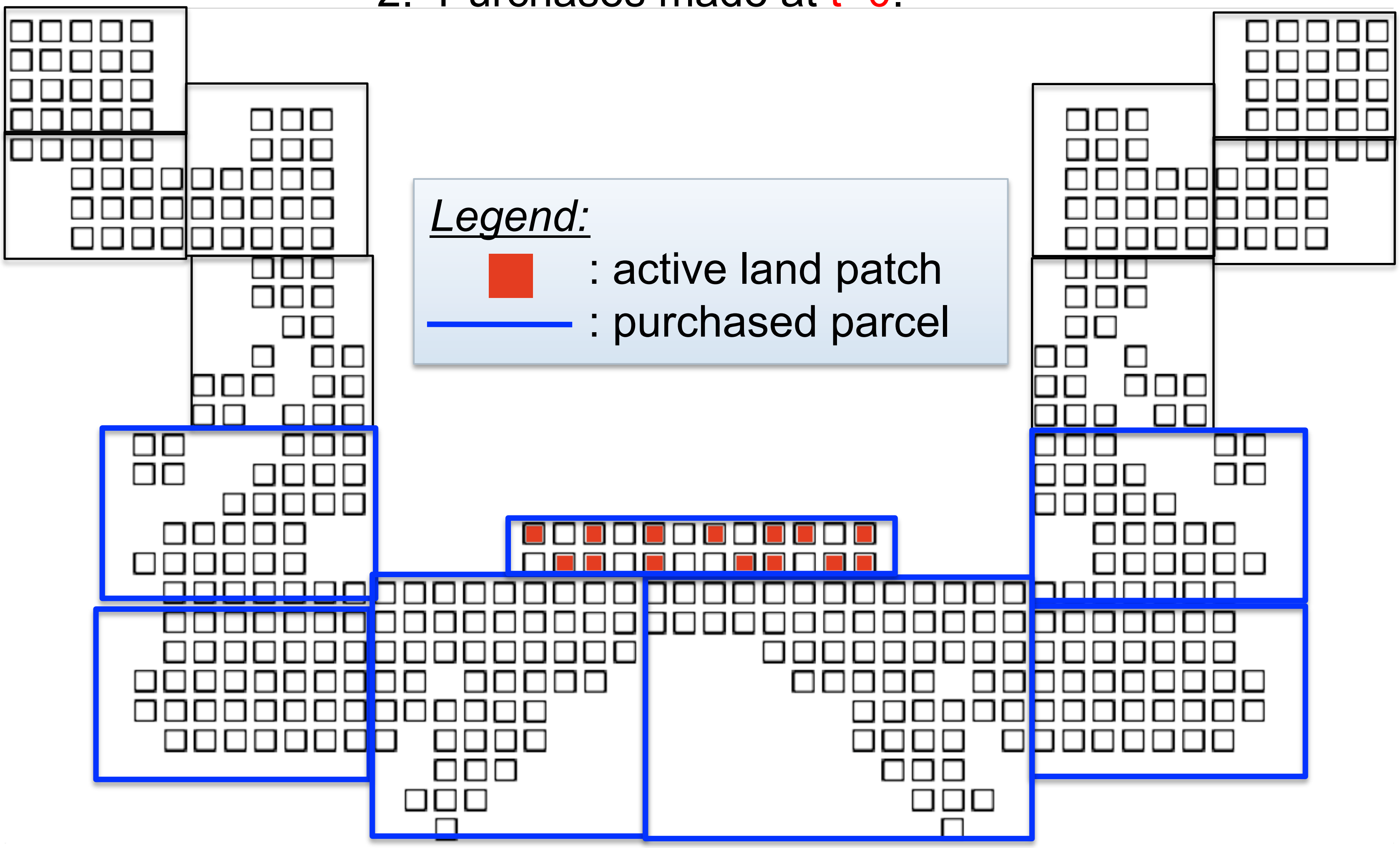
RCW Single-Stage Decision Making

1. Initial conditions.



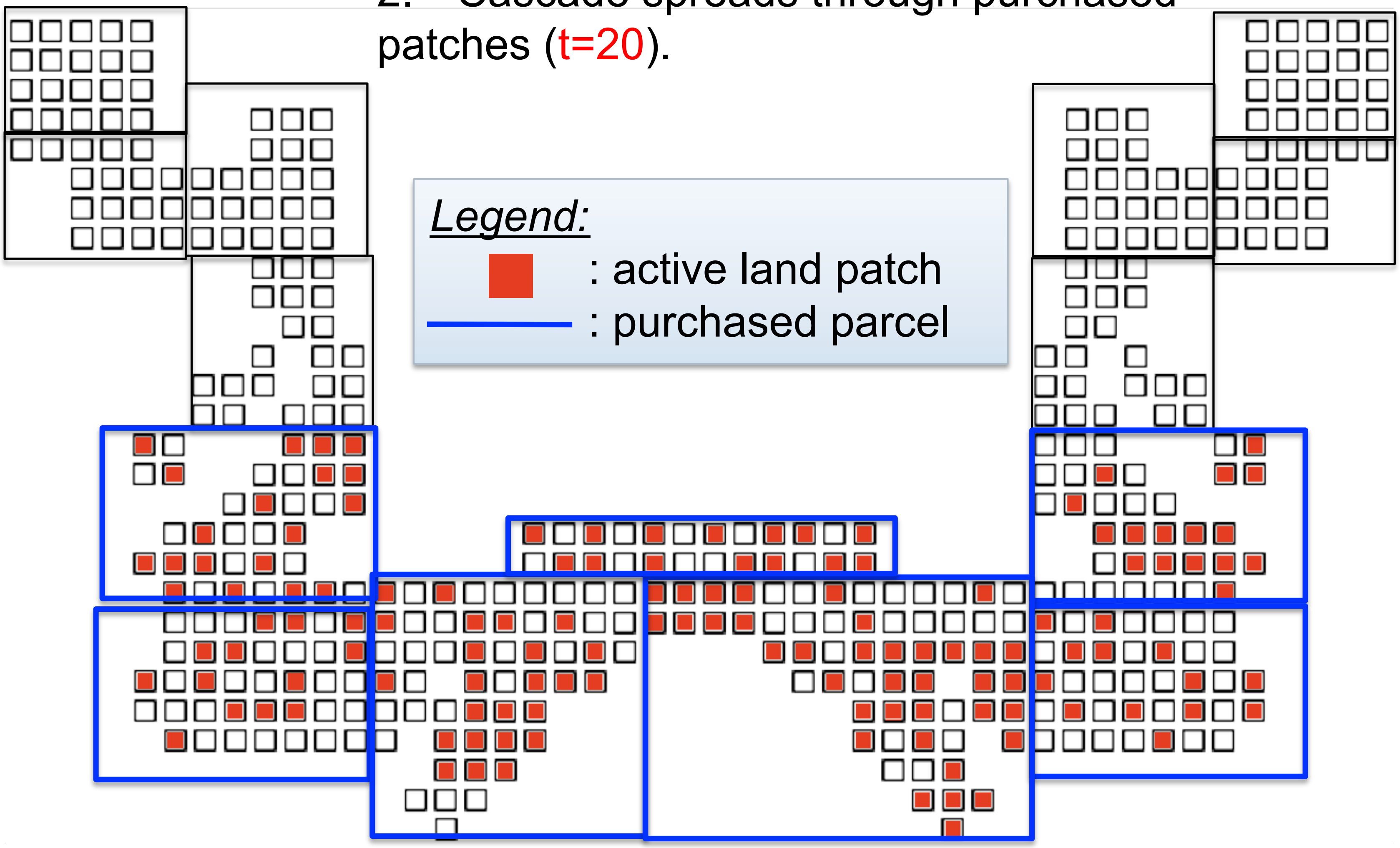
RCW Single-Stage Decision Making

2. Purchases made at t=0.



RCW Single-Stage Decision Making

2. Cascade spreads through purchased patches (t=20).



Legend:

: active land patch

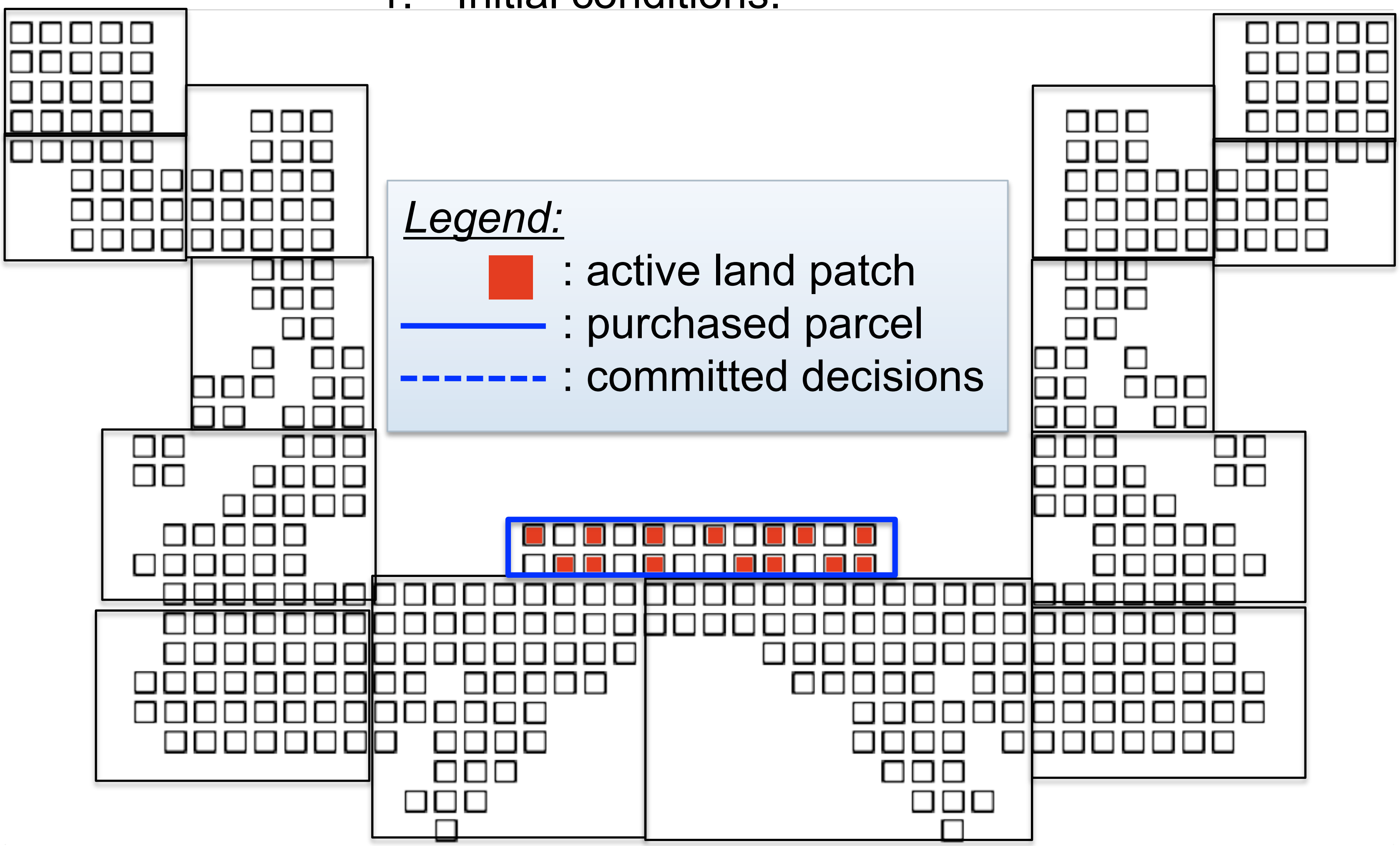
: purchased parcel

Single-stage Decision Making

- (B) Single-stage Split Budget:
  - Purchases in two time-steps with budget split.
  - Commit to purchase decisions in first time-step
    - No adjustment for observations on cascade progression.

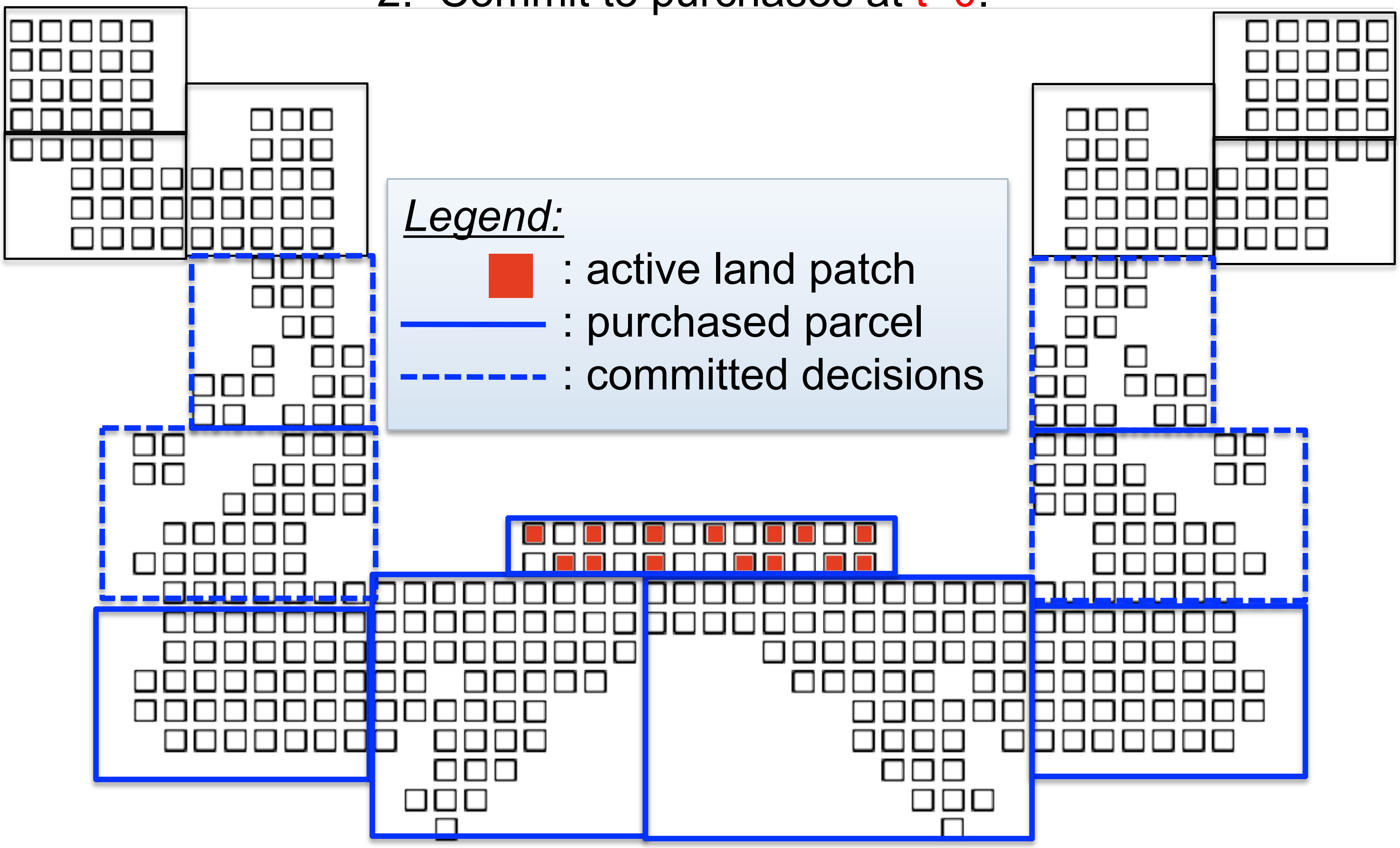
RCW Single-Stage Decision Making

1. Initial conditions.



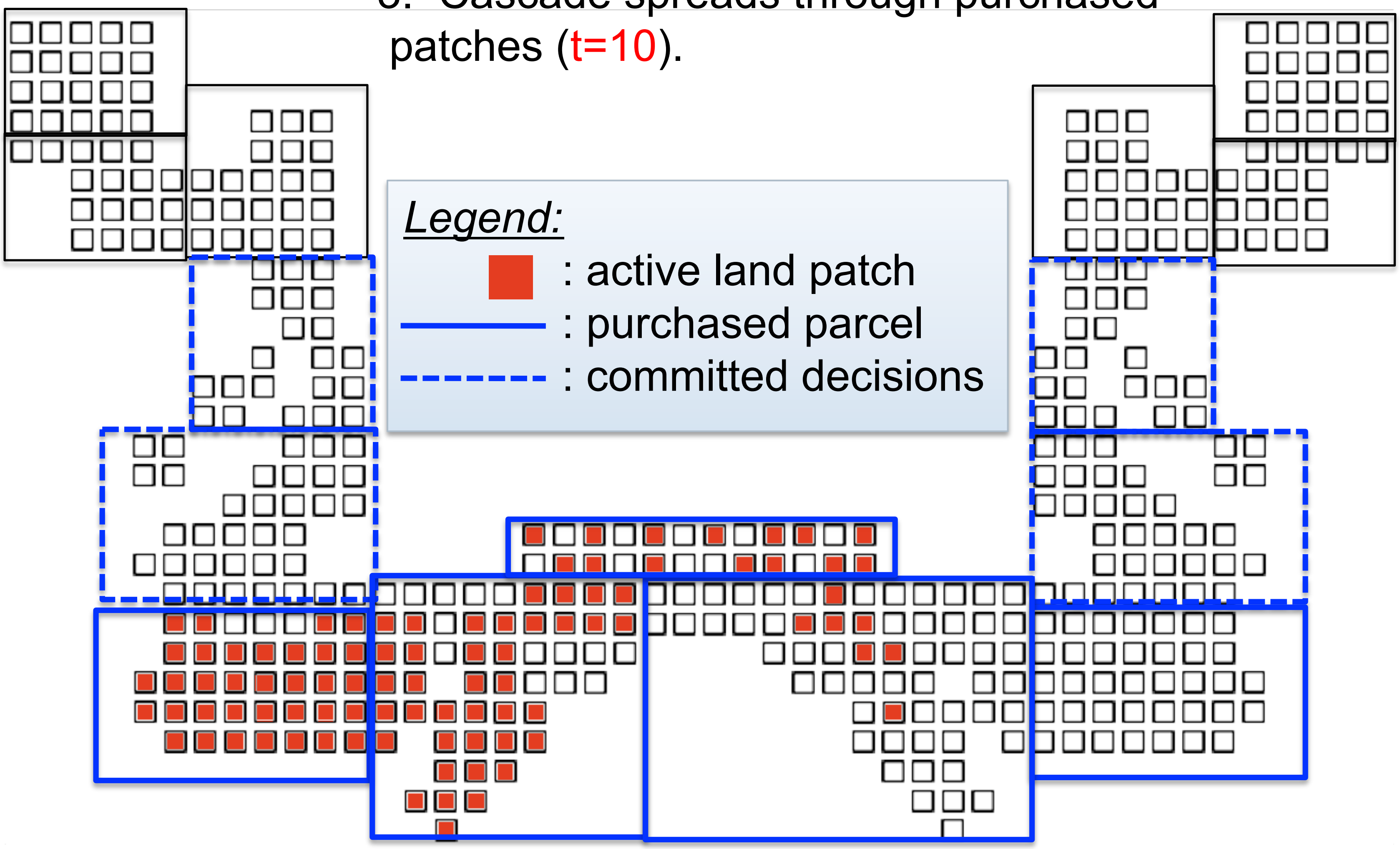
RCW Single-Stage Decision Making

2. Commit to purchases at t=0.



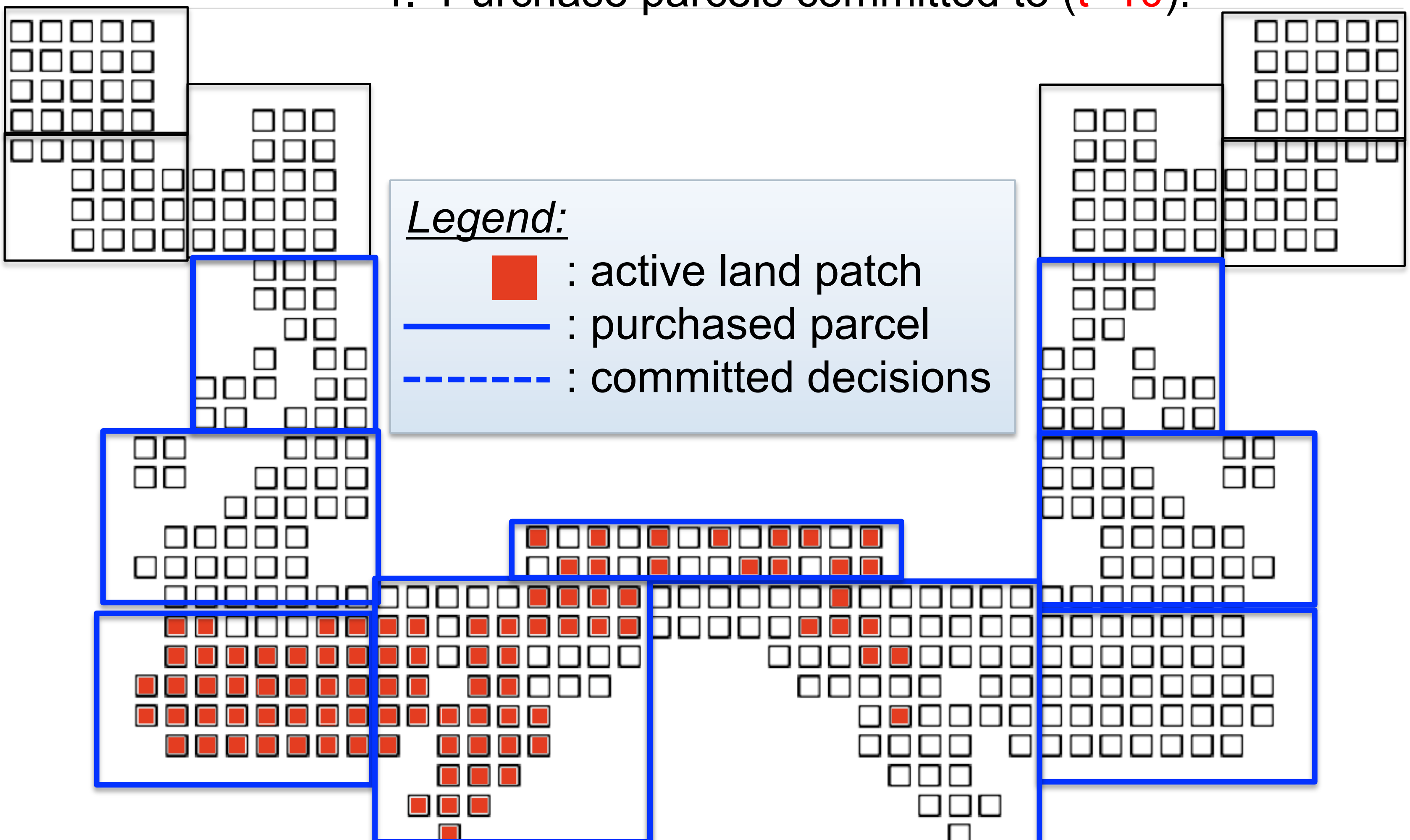
RCW Single-Stage Decision Making

3. Cascade spreads through purchased patches (t=10).



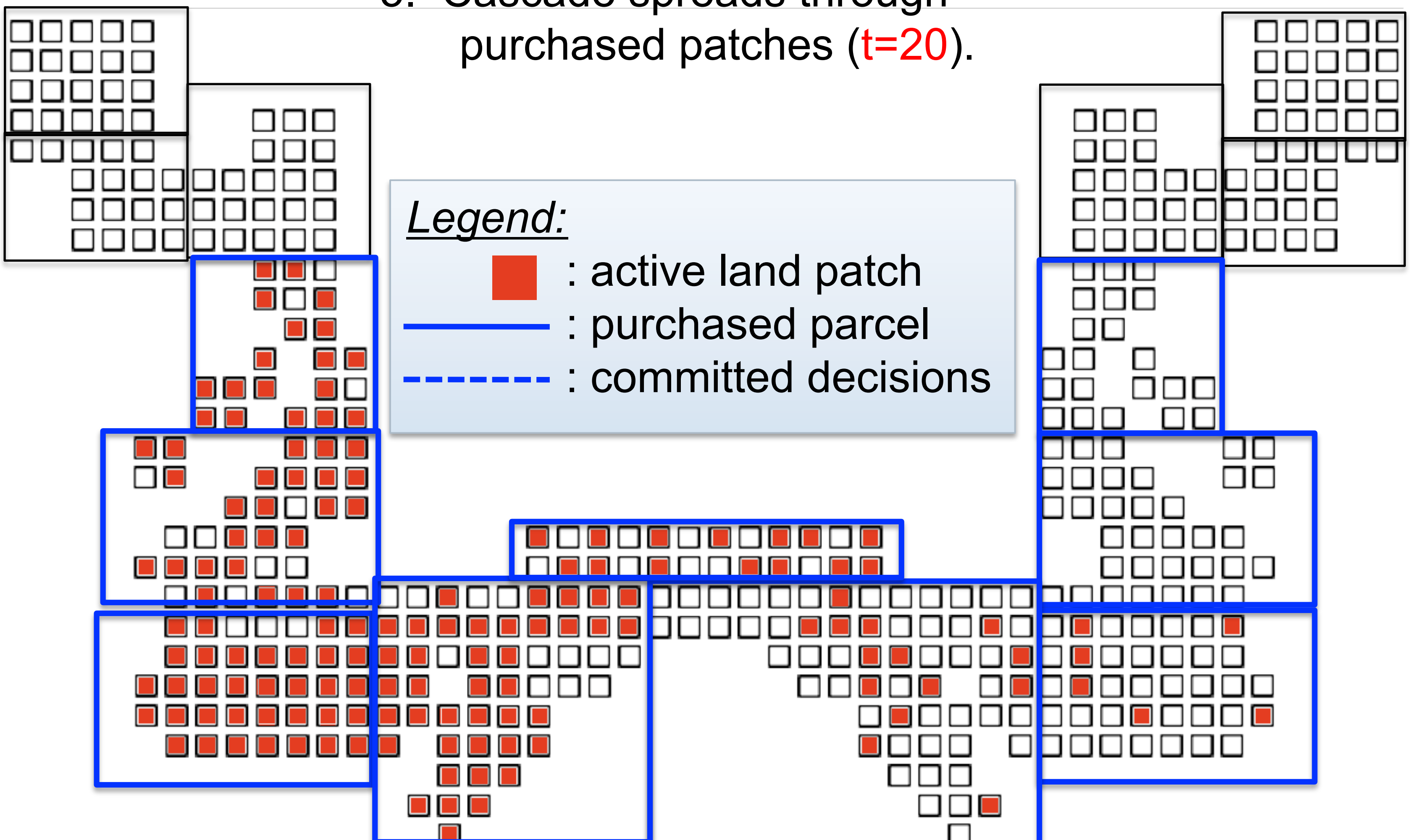
RCW Single-Stage Decision Making

4. Purchase parcels committed to (t=10).



RCW Single-Stage Decision Making

5. Cascade spreads through purchased patches (t=20).



Split-Budget in Applications

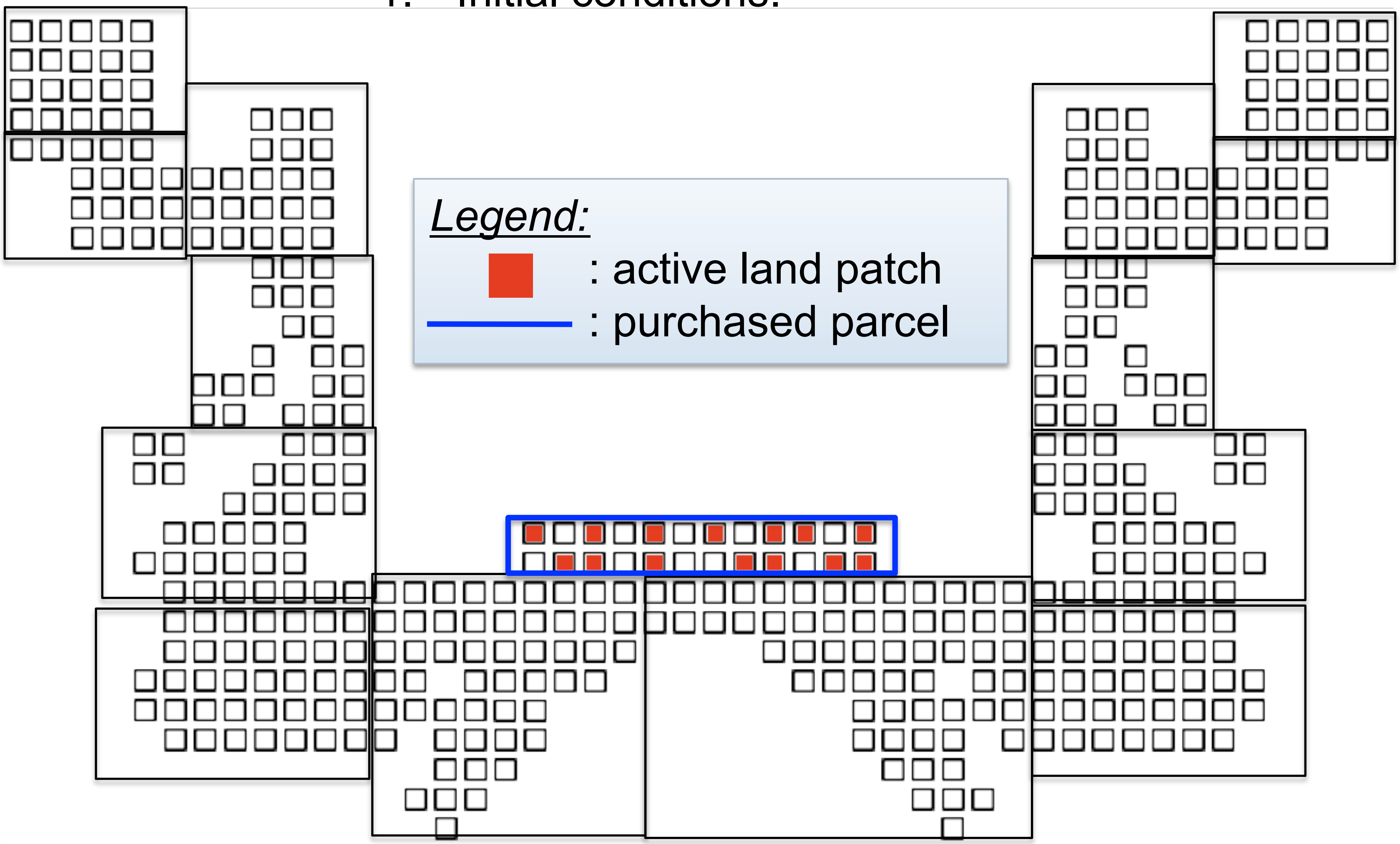
- Why not adjust based on observations?
  - Call for proposals, grants, government funding, etc. often require strict, projected budgets.
    - Requires making purchase decisions in a single-stage at t=0.
  - Little variation in stochastic behavior of cascade.
- First step toward true two-stage model.
  - Significantly more difficult than single-stage upfront budget.

Two-stage Decision Making

- Purchase decisions are made in two time-steps (stages).
  - (C) Second stage decisions can be informed by the outcome of the first stage (open loop).
- Complete solution specifies
  - first-stage decisions
  - second-stage decisions for *every possible scenario* from the first stage
    => a "policy tree"
- Goal: Compute first-stage decisions that maximize expected outcome of second-stage.

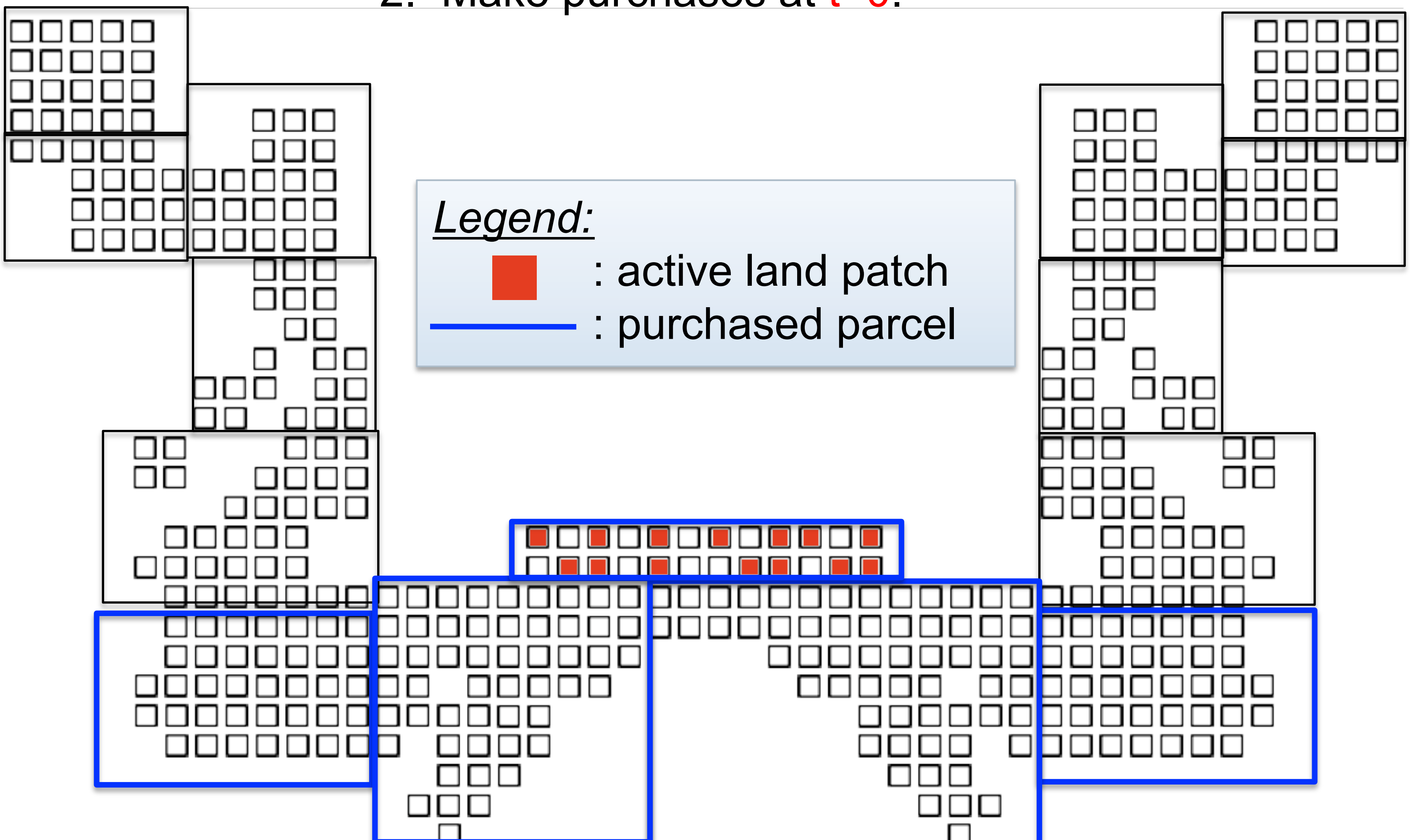
RCW Two-Stage Decision Making

1. Initial conditions.



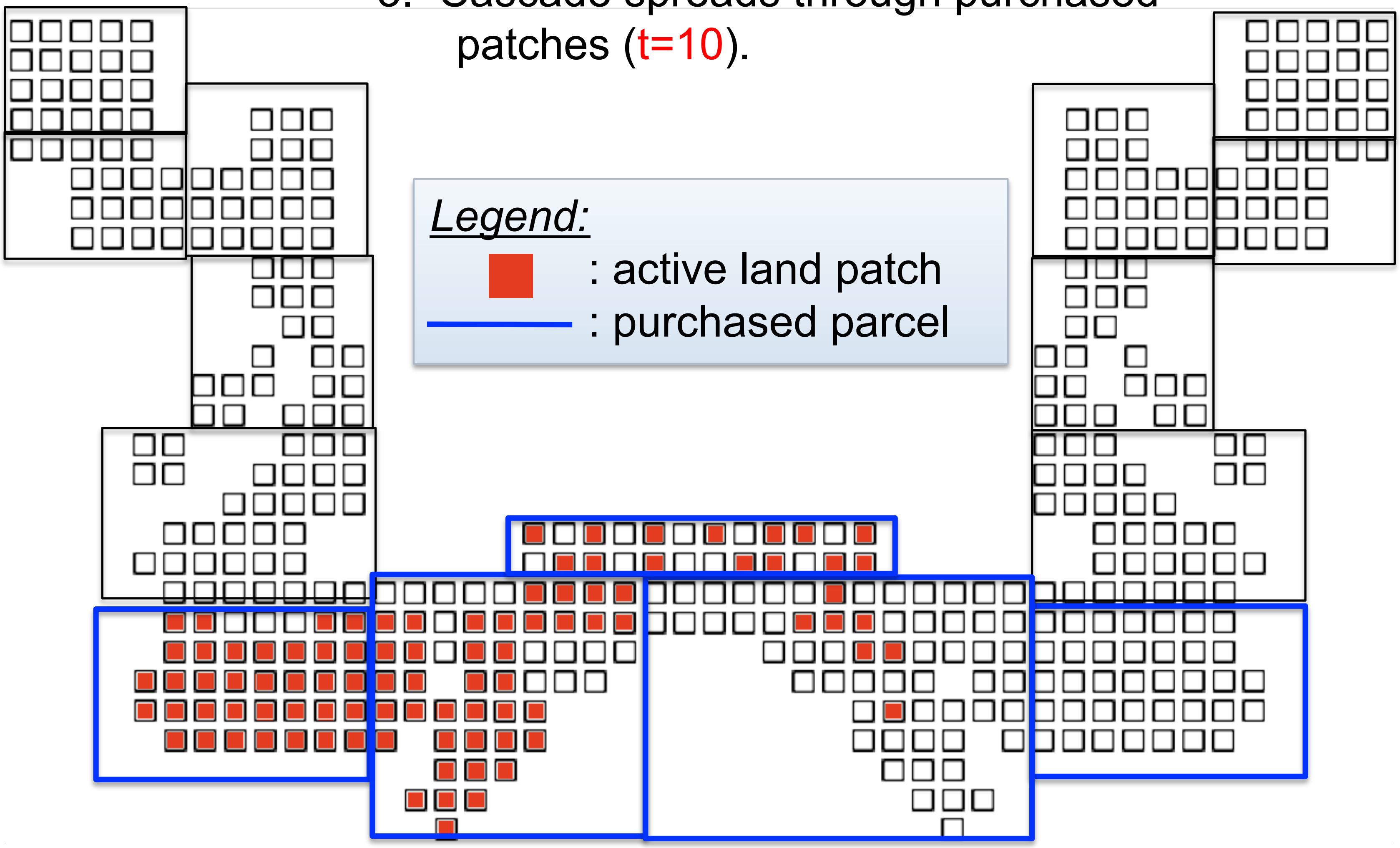
RCW Two-Stage Decision Making

2. Make purchases at t=0.



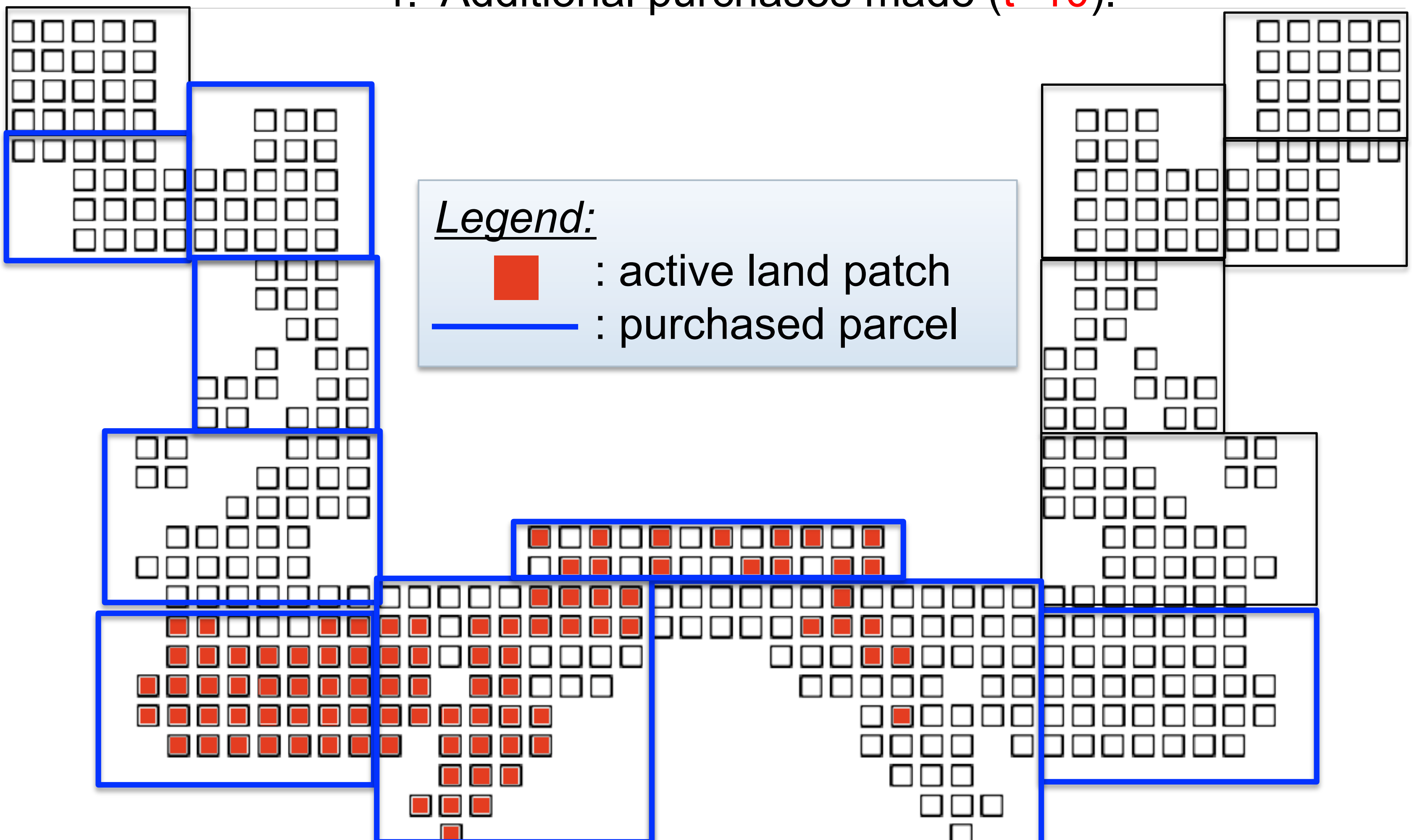
RCW Two-Stage Decision Making

3. Cascade spreads through purchased patches (t=10).



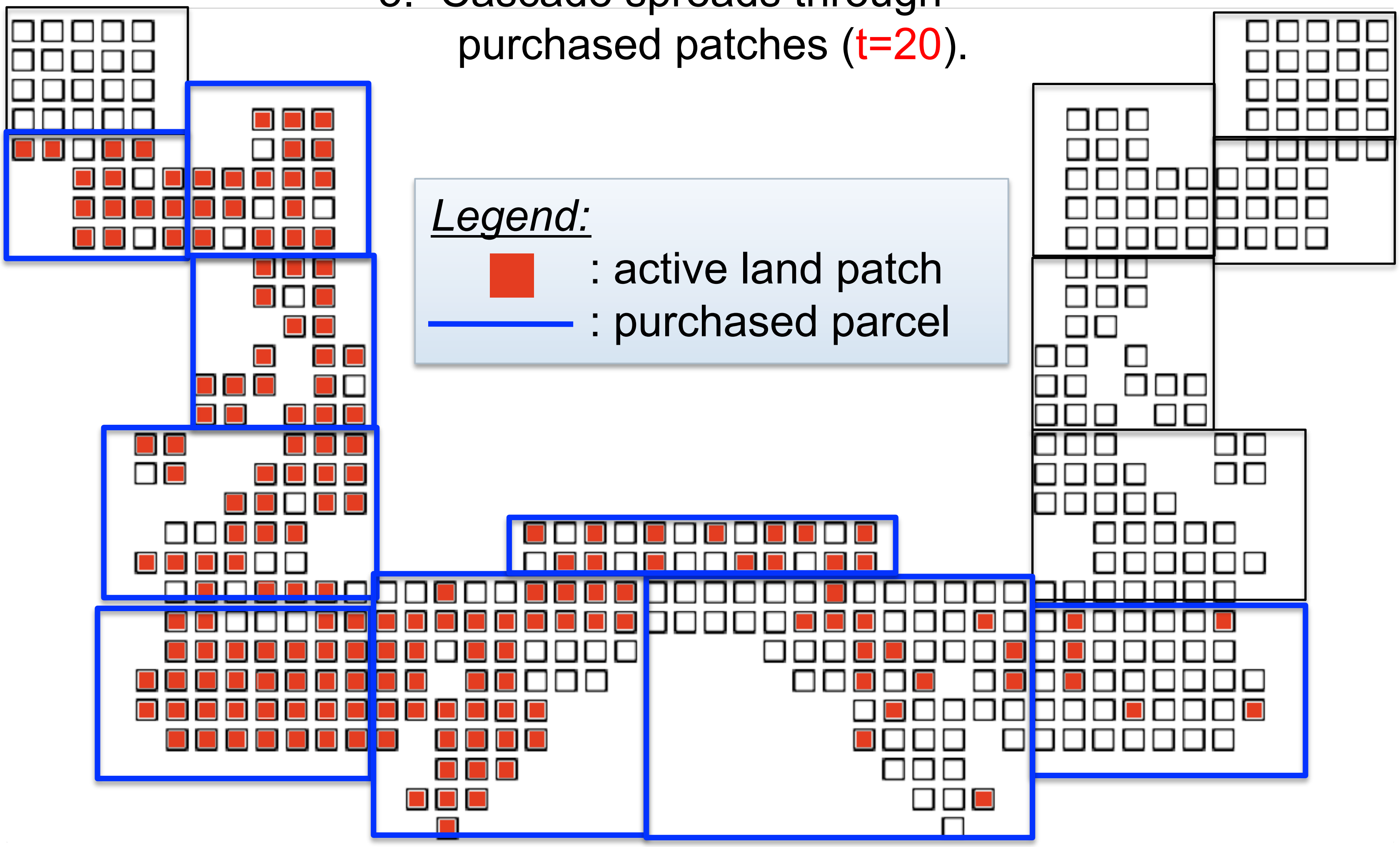
RCW Two-Stage Decision Making

4. Additional purchases made (t=10).



RCW Two-Stage Decision Making

5. Cascade spreads through purchased patches (t=20).

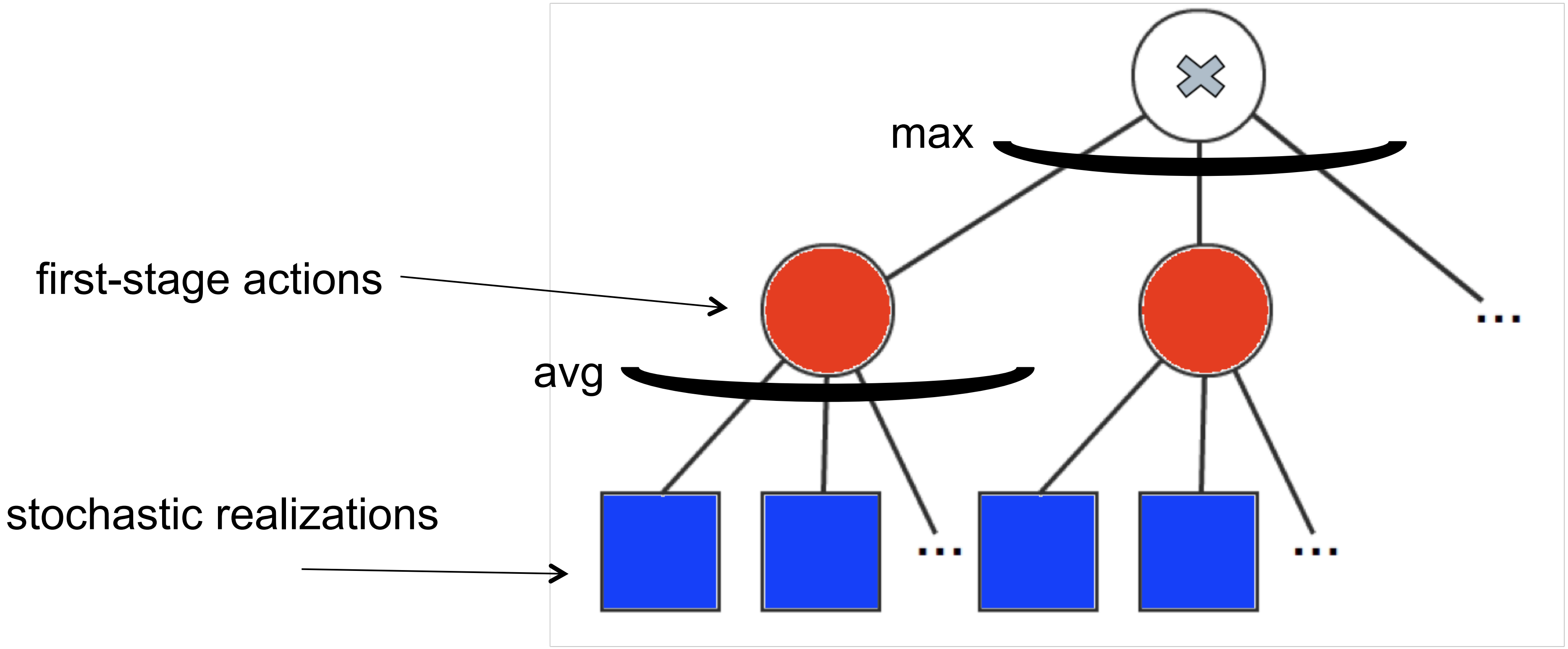


Search Space Complexity

- Complexity of stochastic optimization illustrated by *scenario tree.*
  - Goal: Choose the actions that maximize the expected outcome of stochastic behavior.

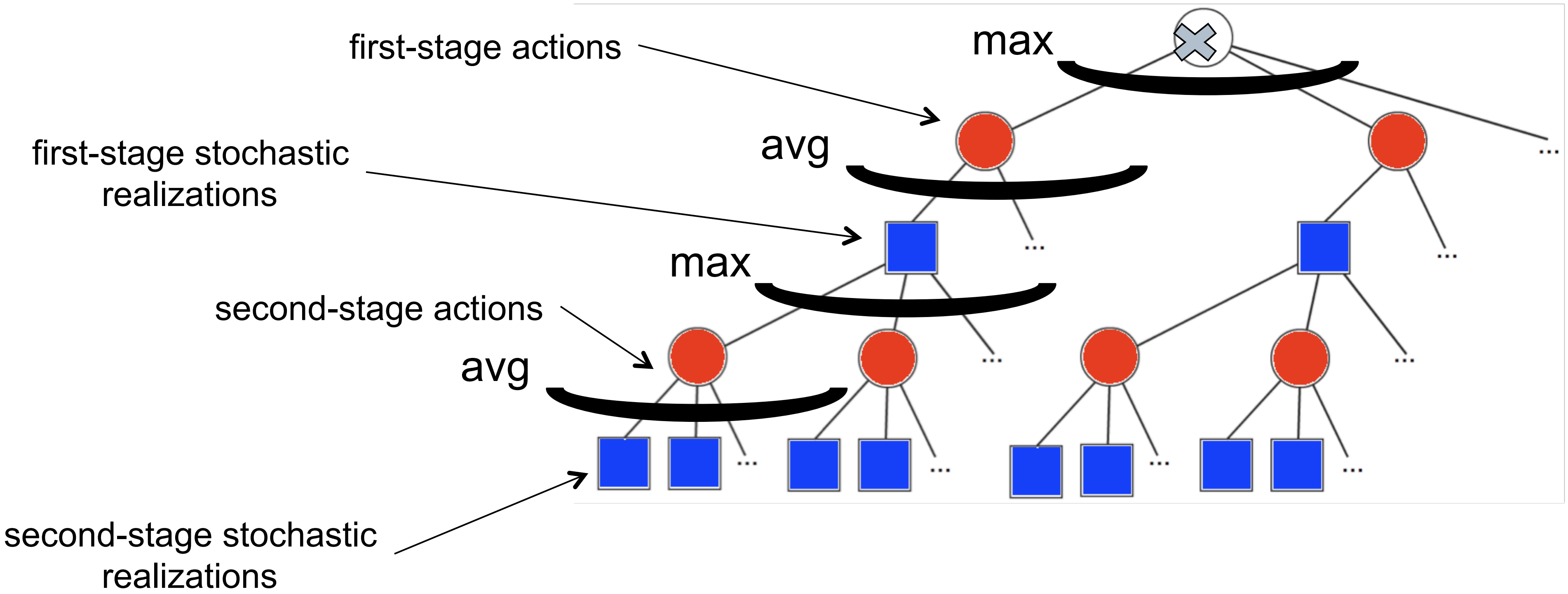
Search Space (Single-stage)

- Single-stage problems: *scenario tree* with fan-out linear in scenario space.



Search Space (Two-stage)

Two-stage problem: scenario tree with quadratic fan-out in scenario space. Largely intractable.



Solution Methods

Stochastic MIP Formulation

$$\text{maximize } \mathbb{E}\left[\sum_{r=0}^{R} \mathrm{x}(r,H)\right] \text{ such that}$$

$$\mathrm{y}(p,0) = 1 \qquad \forall p \in \text{initial (free) parcels} \qquad (1)$$

$$\sum_{p=1}^{P} \mathrm{C}(p) \times \mathrm{y}(p,t) \leq \mathrm{B}(t) \qquad \forall t \in \{0..\mathrm{H}-1\} \qquad (2)$$

$$\sum_{t=0}^{\mathrm{H}-1} \mathrm{y}(p,t) \leq 1 \qquad \forall p \in \{1..P\} \qquad (3)$$

$$\mathrm{x}(r,t) \leq \sum_{t'=0}^{t} \mathrm{y}(P(r),t') \qquad \forall r \in \{1..R\}, \forall t \in \{1..\mathrm{H}\} \qquad (4)$$

$$\mathrm{x}(r,t) \leq \sum_{r'=1}^{R} \xi_{r',r}^{t-1} \mathrm{x}(r',t-1) \qquad \forall r \in \{1..R\}, \forall t \in \{1..\mathrm{H}\} \qquad (5)$$

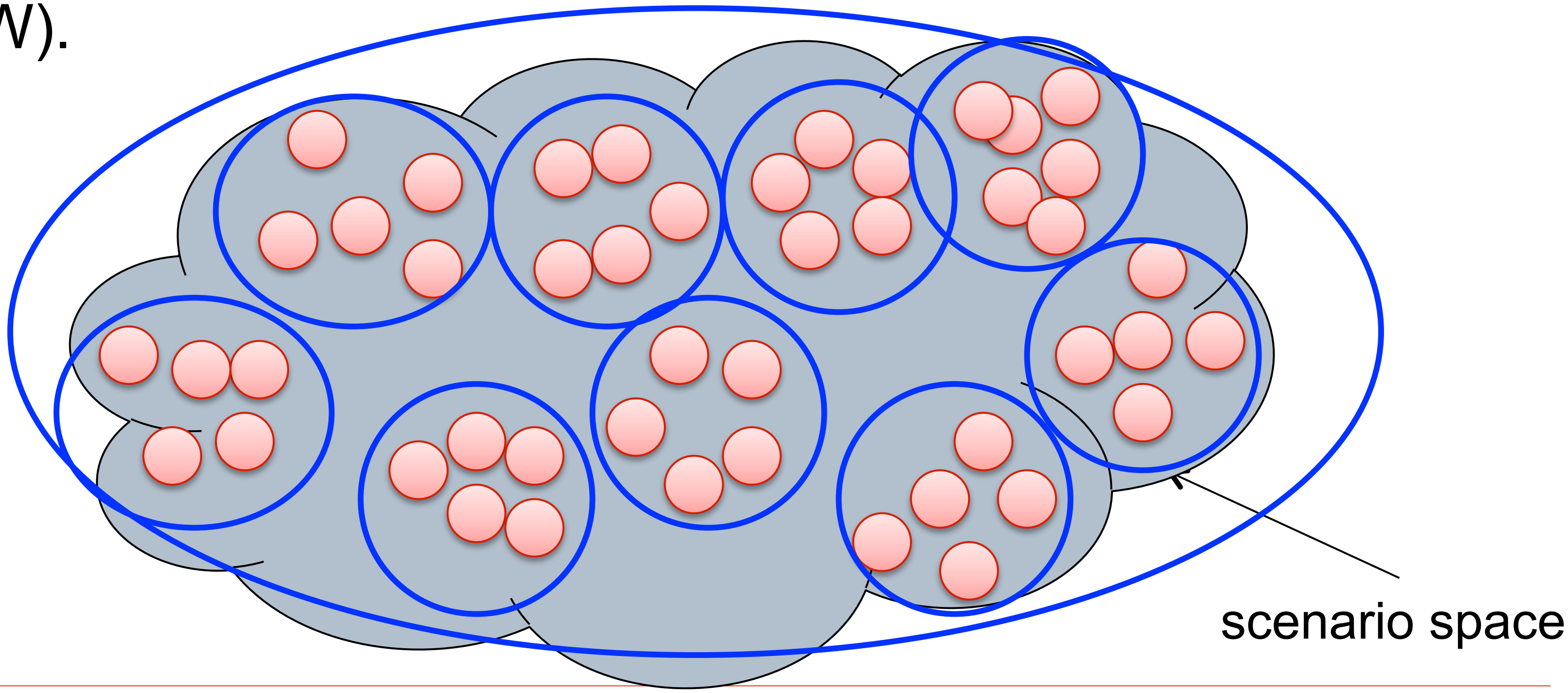
$$\mathrm{x}(r,0) = I(r) \qquad \forall r \in \{1..R\} \qquad (6)$$

r : habitat patches
p : real-estate parcels
$P(r)$: parcel of patch r
$I(r)$: true if r is initially occupied.
t : time-steps
H : planning-horizon
P : number of parcels
R : number of territories
$x(r,t)$: true if territory r active at time t
$y(r,t)$: true if parcel p purchased at time t
$\xi_{r',r}^{t}$: stochastic coefficients following dispersion model

- Maximizes expected active land patches at time horizon.
- Applies to single-stage problems (A) upfront budget and (B) split budget
- Deterministic analogue (finite *scenario* set) => building block for solution procedures.
  - *scenario :* cascade realization

Sample Average Approximation

- Stochastic optimization by solving series of deterministic analogues [Shapiro, 2003]
- Sample a set of N *scenarios*.
- Optimal solution for one sampled set *over-fits* to that set.
- Larger N increases MIP complexity (max 20 scenarios tractable for RCW).



Sample Average Approximation (Single-stage)

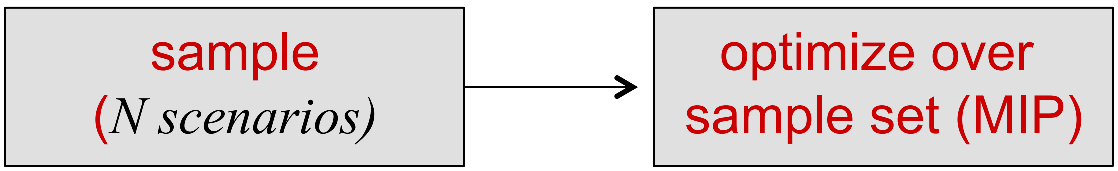
- Sample a finite set of N cascade scenarios.

sample
(N *scenarios*)

Sample Average Approximation (Single-stage)

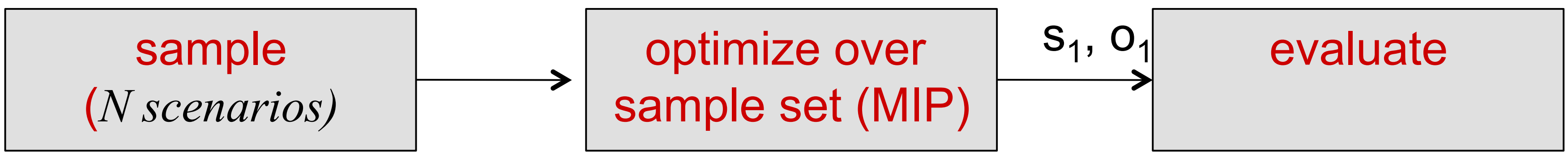
- Maximize the empirical average over this set.

$$\max_{a} \frac{1}{N} \sum_{k=1}^{N} \sum_{v \in V} X_v^k(a, H) \text{ s.t. } \sum_{a_i \in a} c(a_i) \leq B$$



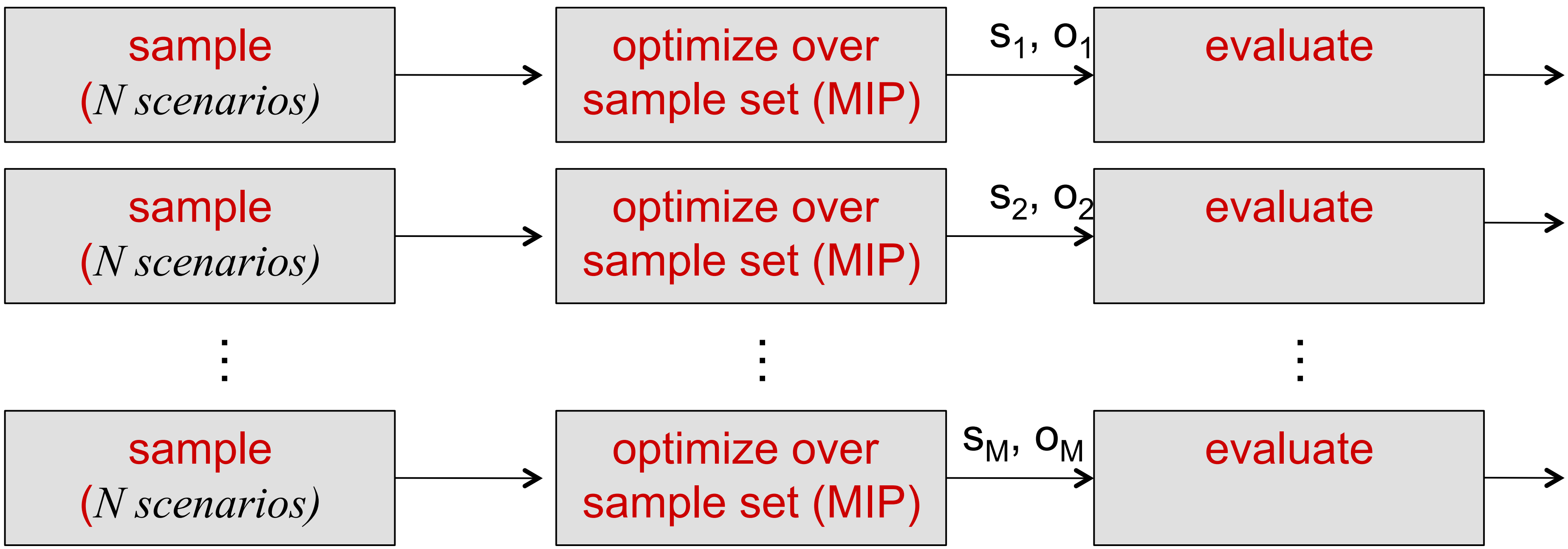
Sample Average Approximation (Single-stage)

- Evaluate obtained solution *s* on small set of independent scenarios.



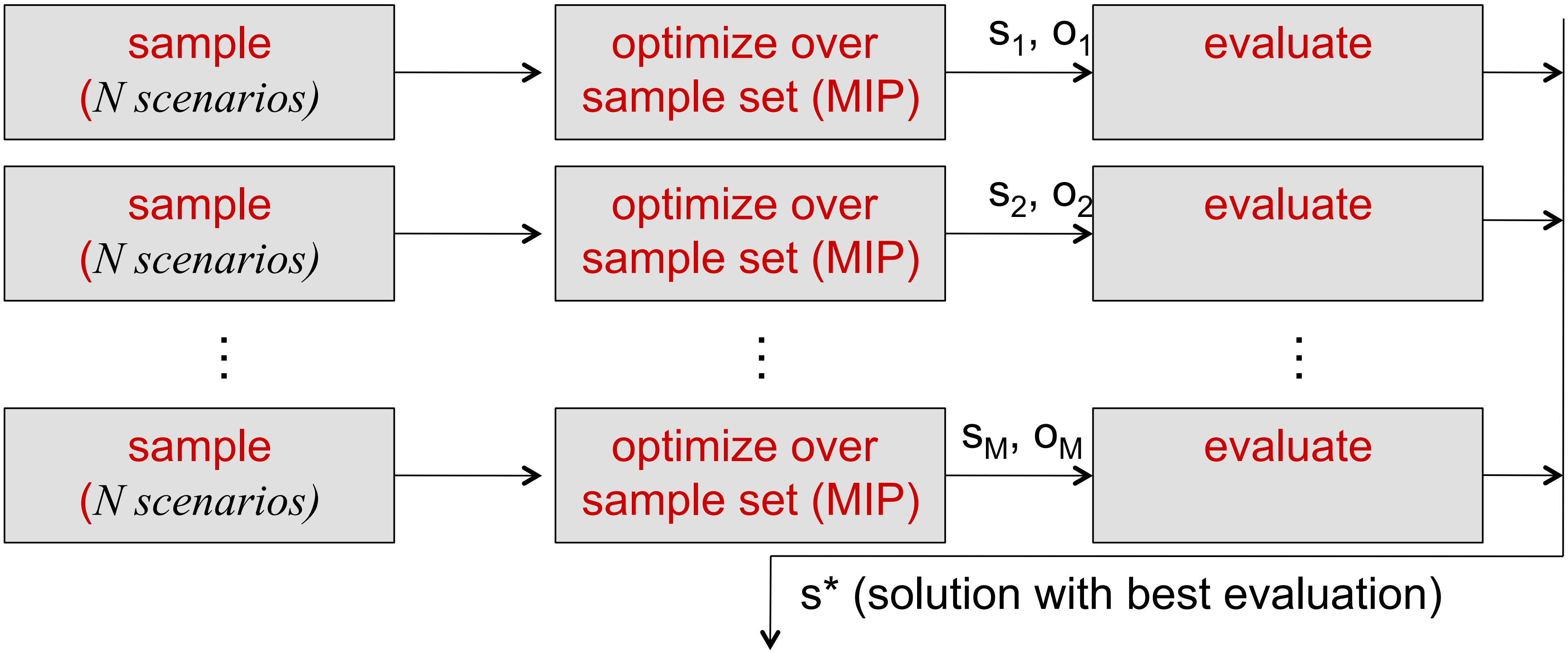
Sample Average Approximation (Single-stage)

- Repeat process *M* times to obtain *M* candidate solutions.



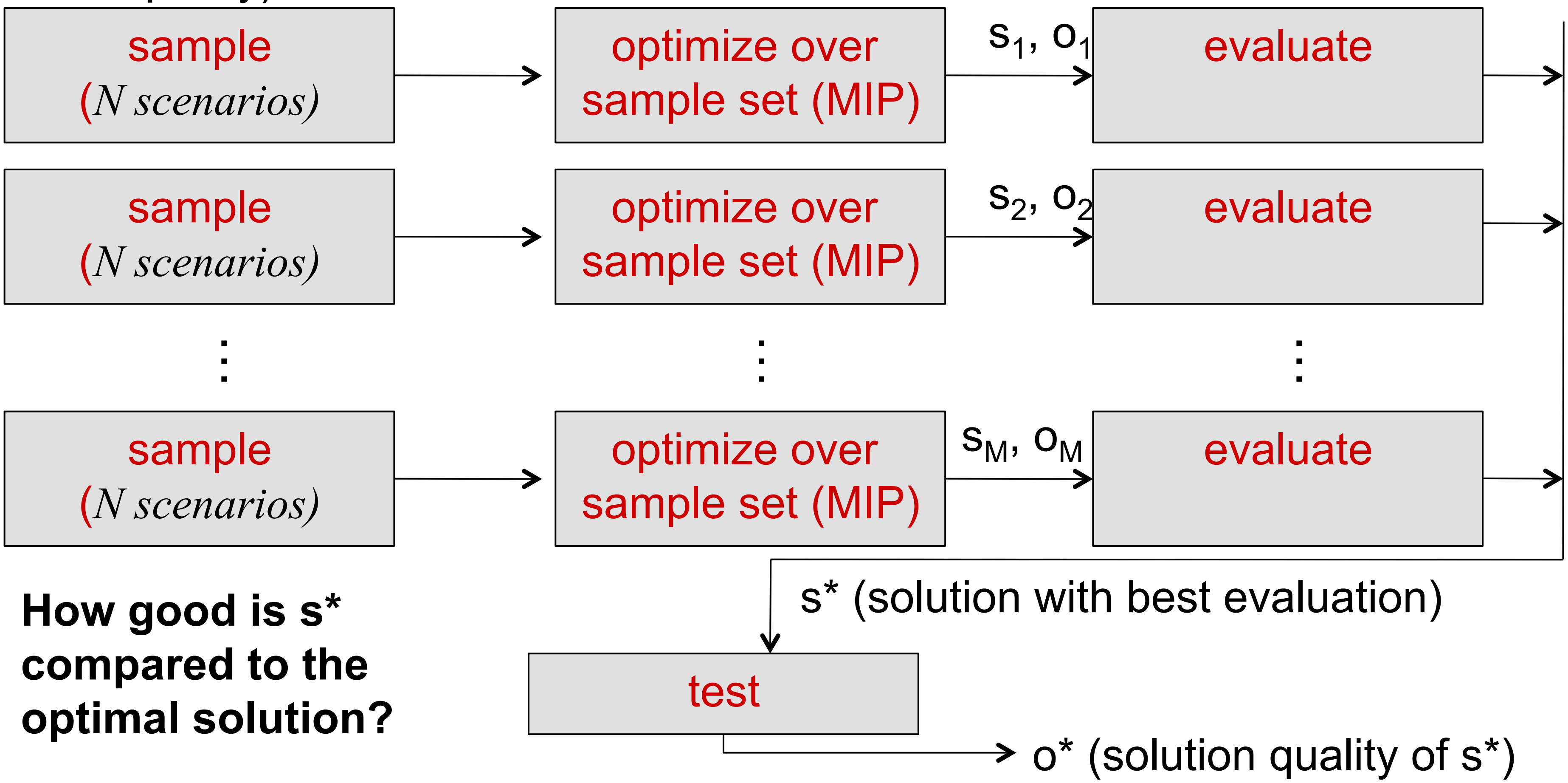
Sample Average Approximation (Single-stage)

- Take candidate solution s* with best evaluation as solution obtained by process.



Sample Average Approximation (Single-stage)

- Evaluate s* on large, independent test set of scenarios (final solution quality).



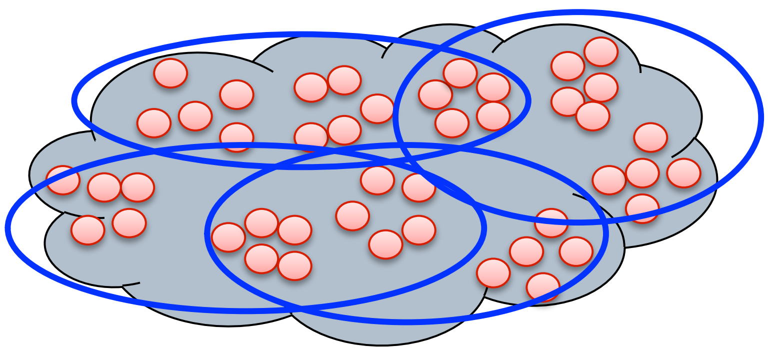
How good is s* compared to the optimal solution?

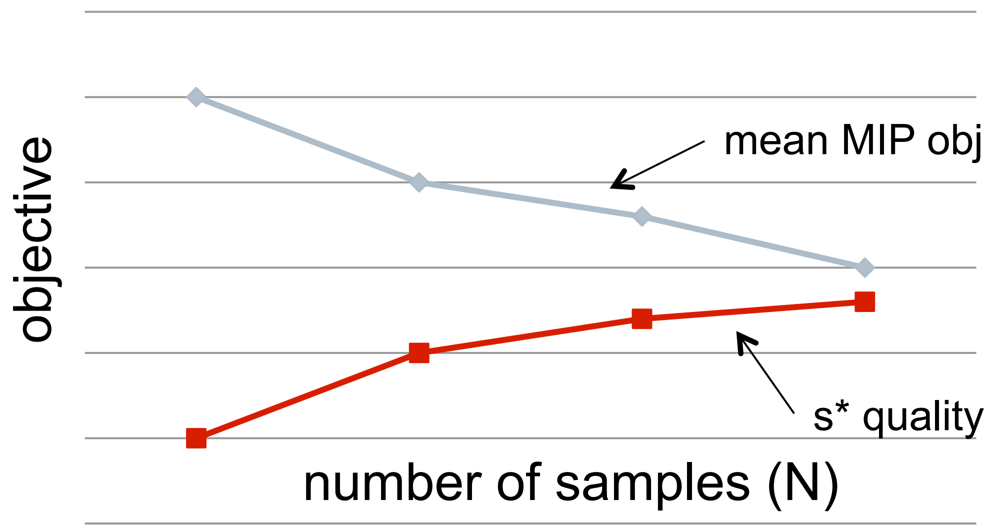
Stochastic optimality guarantees

- Expected utility of s* gives a *lower bound* on true optimum
  => o* gives a *stochastic lower bound* on true optimum

- E[o] gives an *upper bound* on the true optimum.
  => Sampled average of o gives *stochastic upper bound*

- Convergence of bounds guaranteed for increasing sample size *N*.





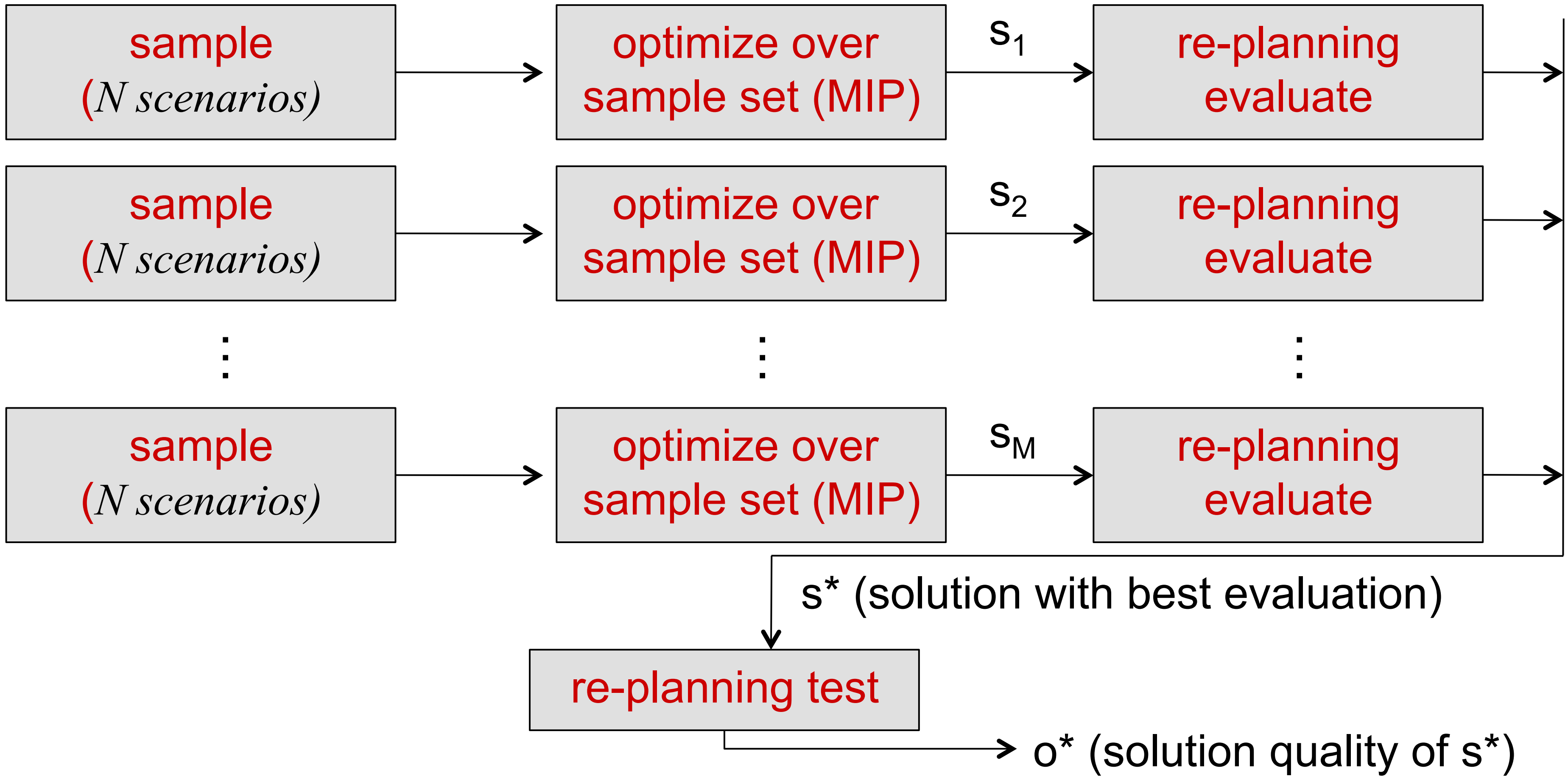
Two-stage Re-planning with SAA

- Purchase decisions made in time-steps *0* and T_1 over horizon *H*. Budgets b_1 and b_2.

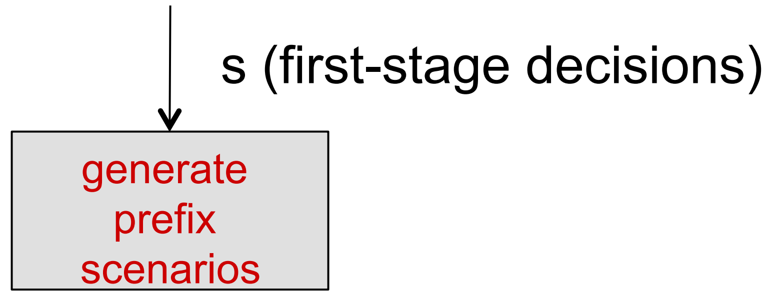
- Re-planning approximates solution to (C) Two-Stage Split Budget
  - Computes set of first-stage decisions.
  - *Nested* SAA procedure used to evaluate candidate first-stage decisions.

Two-Stage Re-planning

- Obtain M candidate *first-stage* decisions using SAA for (B) Single-stage split budget

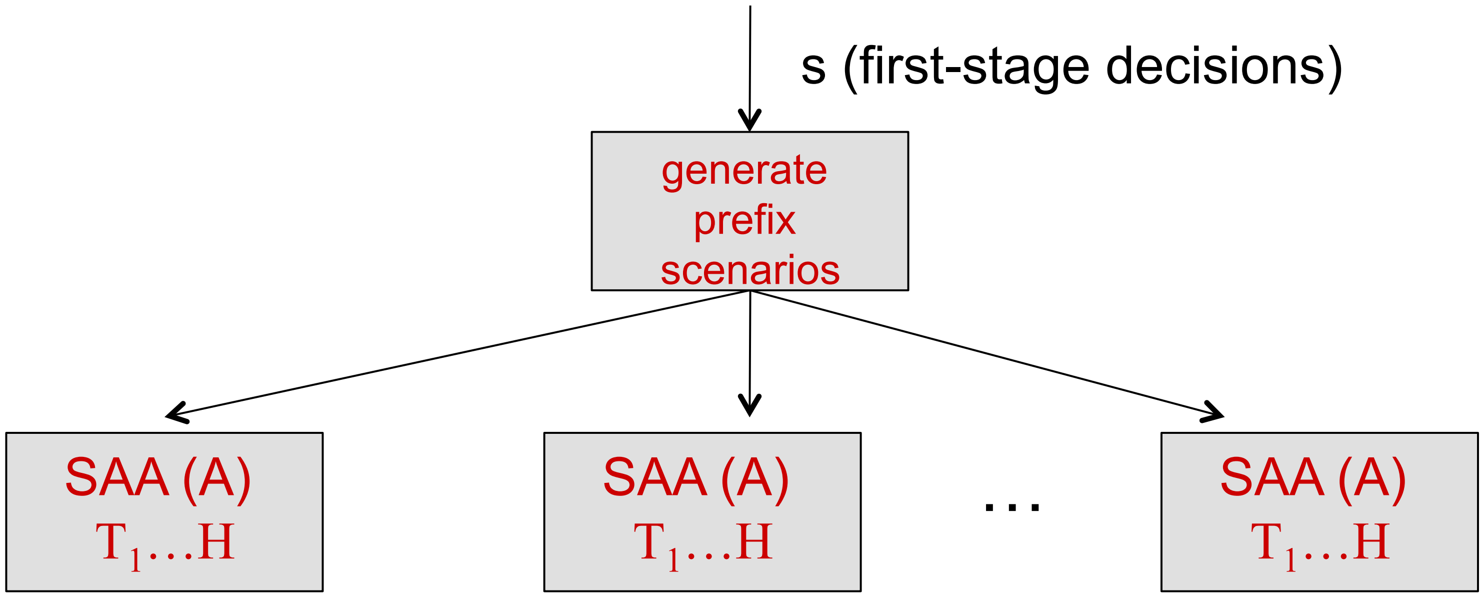


SAA Re-planning Evaluation



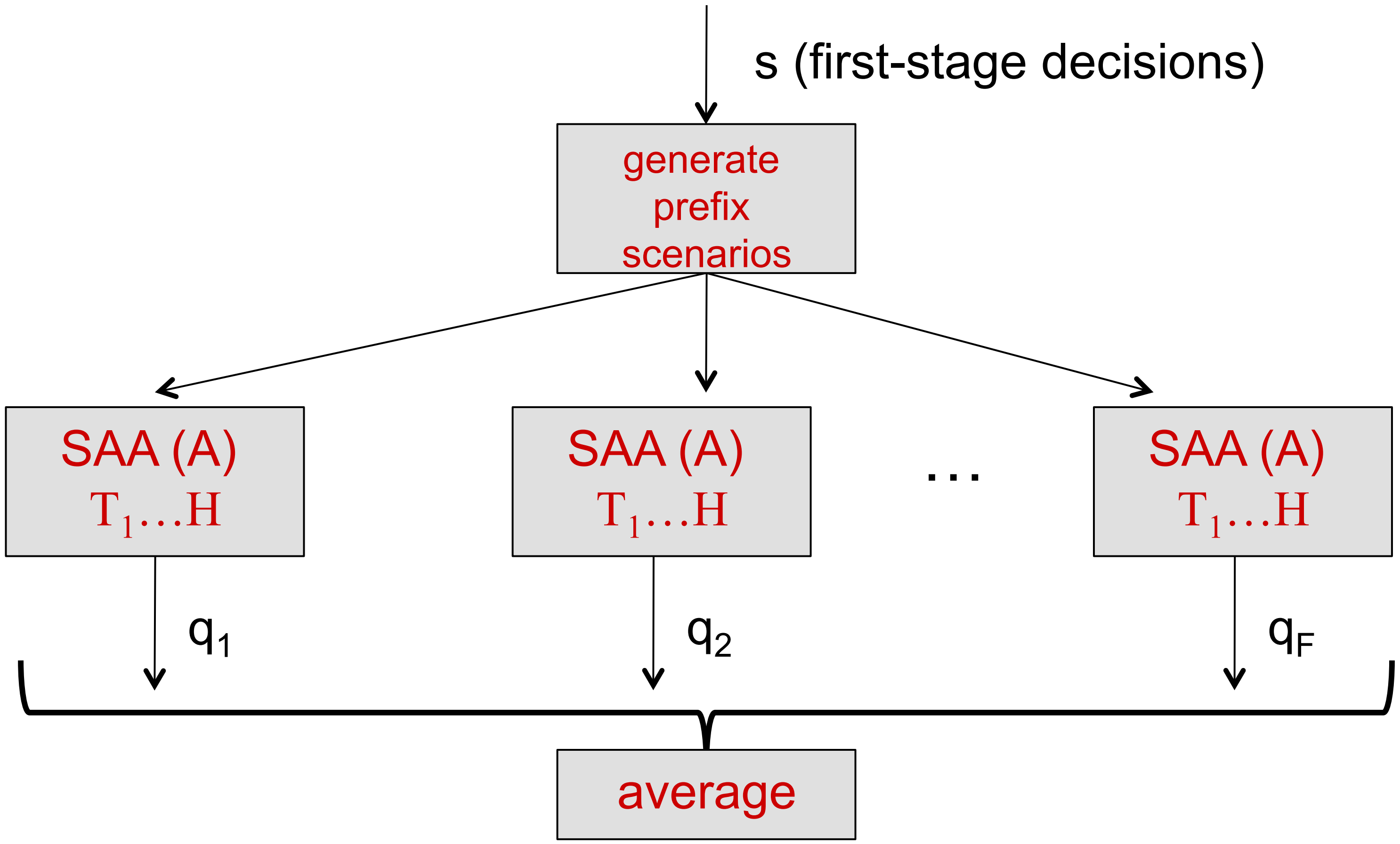
- Generate F *prefix scenarios*, realizing first stage under s

SAA Re-planning Evaluation



- For each prefix scenario: SAA single-stage upfront budget for years $T_1 \ldots. H$.
  - Occupied patches at end of prefix scenario are initial
  - First-stage purchases available for free

SAA Re-planning Evaluation

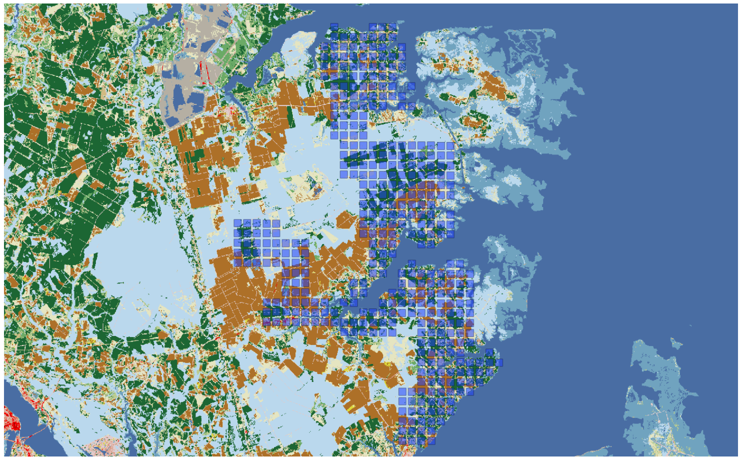


- Evaluation performance of s: average second-stage performance (q_i)

Experiments

Experimental Background

- Red-cockaded Woodpecker (RCW) Conservation in North Carolina.
  - Listed by U.S.A government as rare and endangered [USA Fish and Wildlife Service, 2003].
- The Conservation Fund: conserve RCW on North Carolina coast.
- **Nodes** = land patches large enough to be RCW habitat (411 patches).
- **20 initial territories**
- *H*=20 time horizon.
- **Actions** = parcels of land for purchase that contain potential territories (146 parcels).
- N = 10 (finite sample set size for forming MIPs)
- CPLEX used for MIP solving.

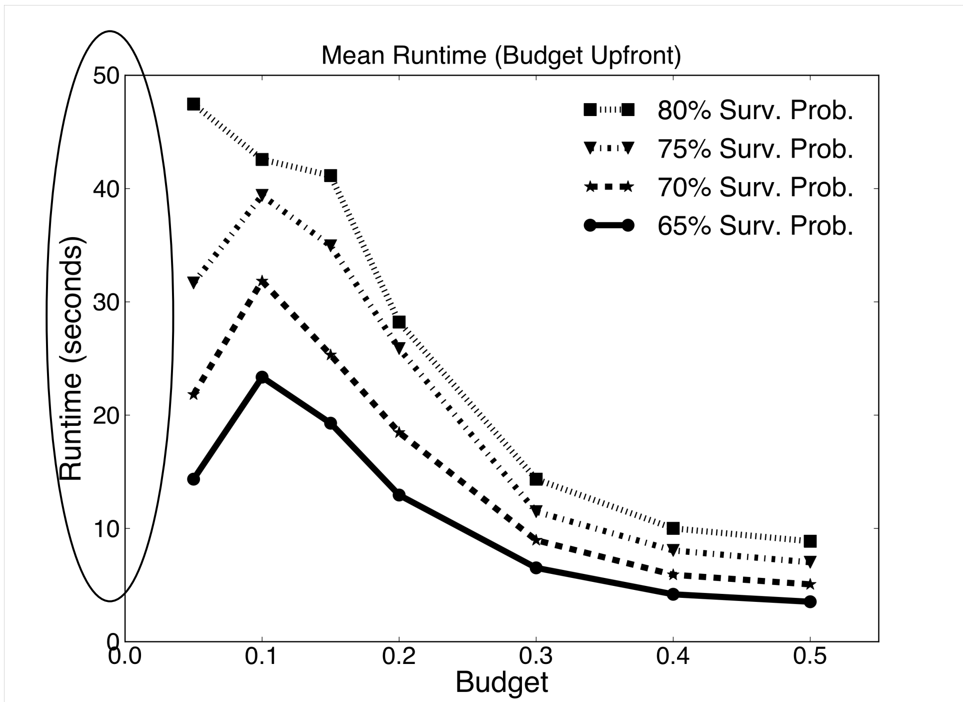


RCW Problem Generator

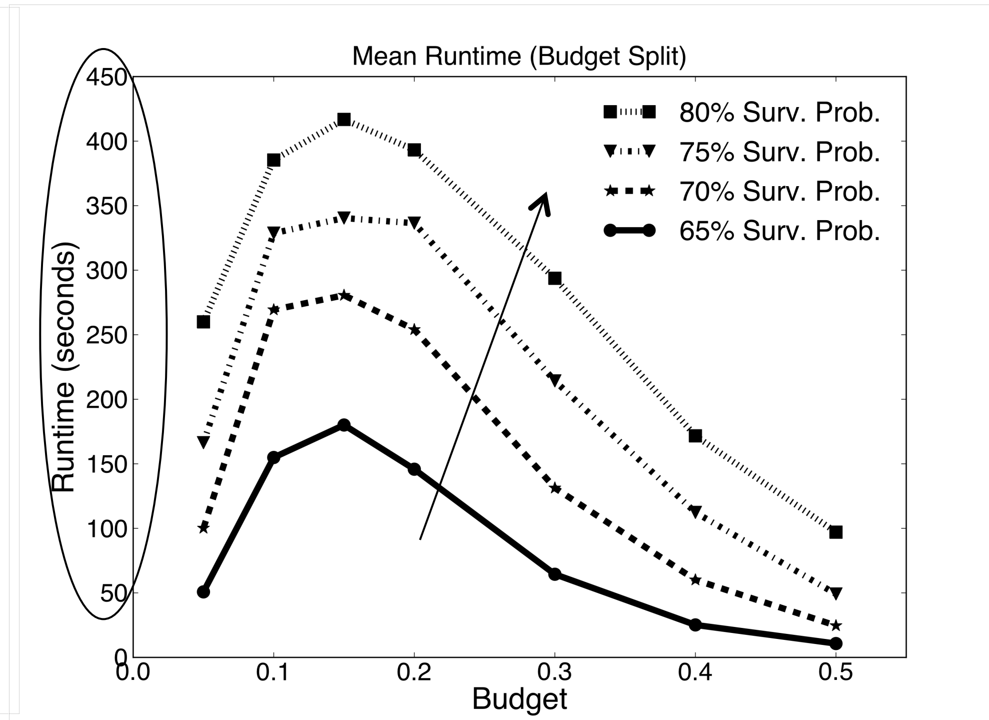
- Instances generated from base map by random perturbation.
  - Perturbs territory suitability scores around base values.
  - Randomly choose initial territories in high suitability parcels.
  - Assign parcel costs with inverse correlation to parcel suitability.
- Used to generate large set of maps which we use to study runtime distributions.
- Generator available online (C++ Implementation). www.cs.cornell.edu/~kiyan/rcw/generator.htm

Runtime Distributions

(A) Single-stage upfront budget



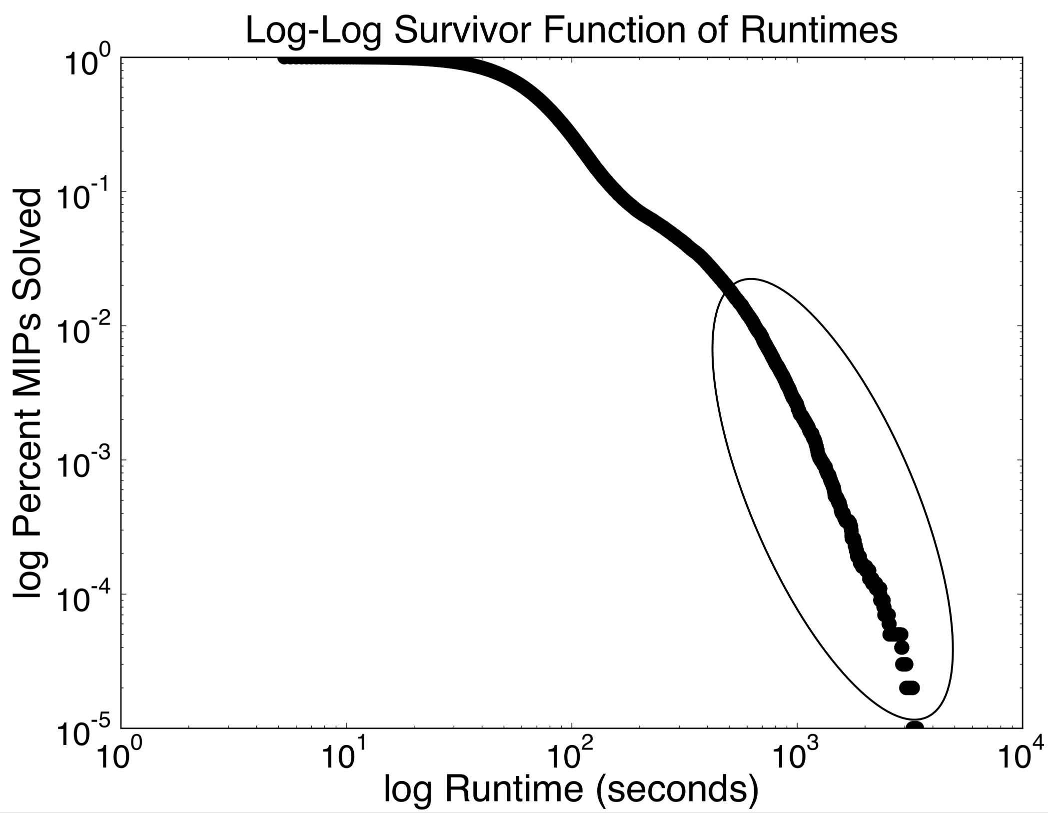
(B) Single-stage split budget



1. Easy-hard-easy pattern
2. Increasing difficulty with survival probability
3. Split budget 10x harder than upfront budget

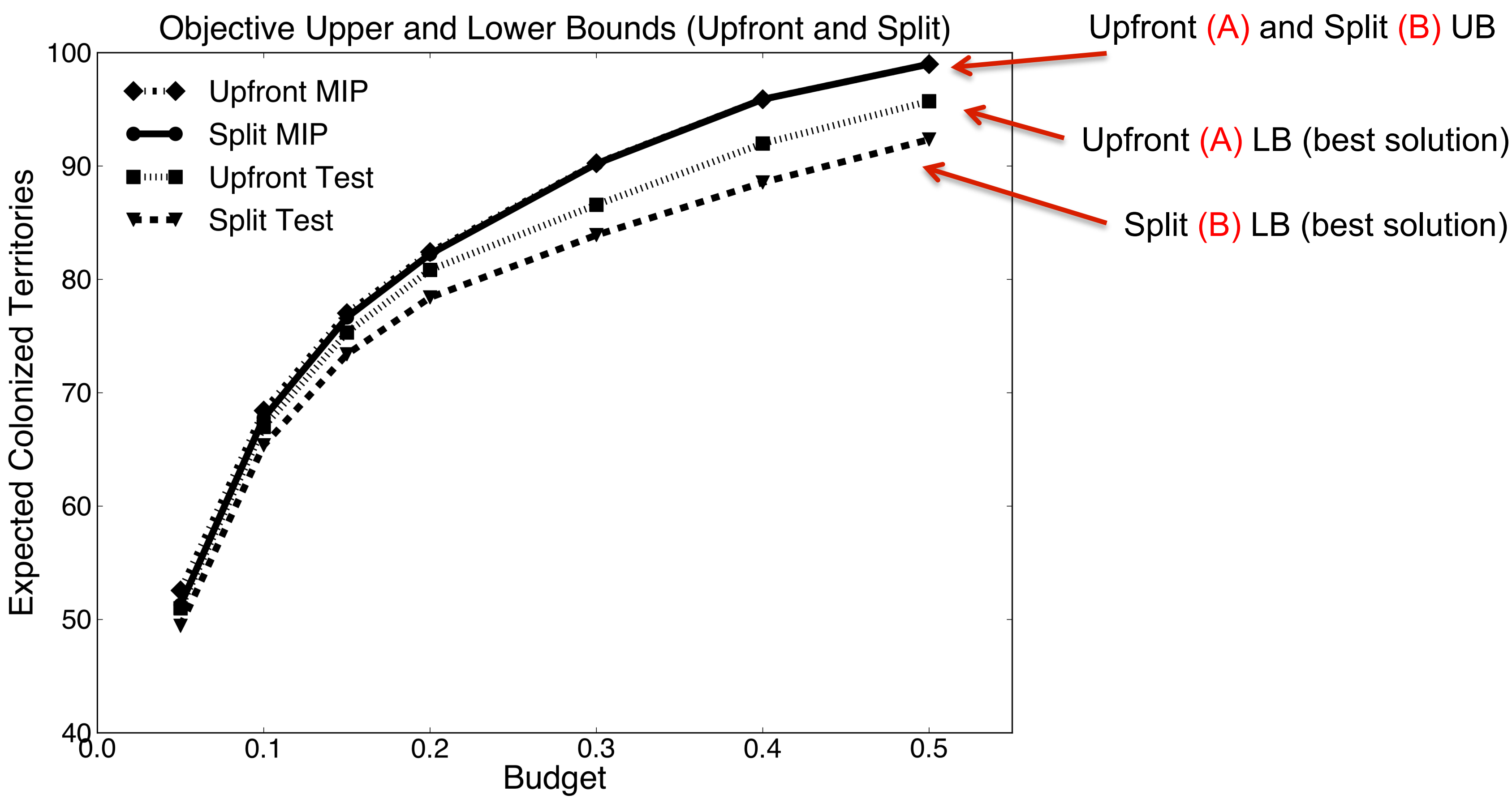
Single Instance Difficulty: Power-Law Decay

"survivor function":
Fraction of instances unsolved in time t



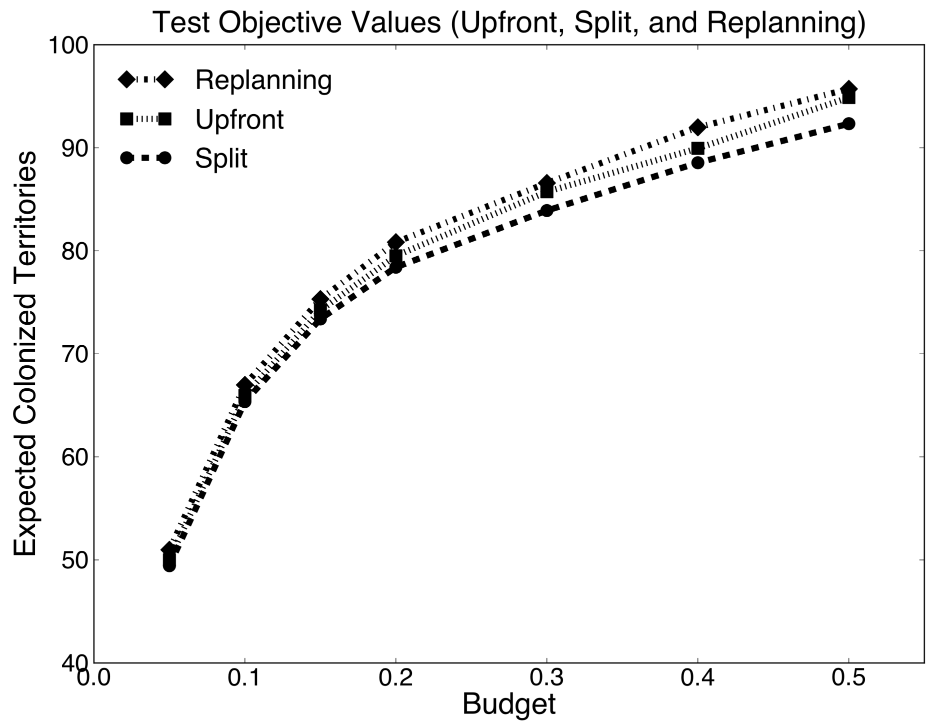
Data: 100,000 MIPs formed from 10 random scenarios on map-30714.

Single-stage: upfront vs. split budget



1. Close bounds indicate solution close to optimal.
2. Upfront variant obtains higher quality solutions than split.

Boosting Solution Quality with Re-planning



1. Re-planning with observations provides significant improvements over committing to all decisions upfront.
2. Re-planning can outperform single-stage upfront budget.

The Balance in Re-planning

- Re-planning sensitive to balance in budget split:
  - Benefits with 30-70% split budget with T_1=5
  - Does worse with, e.g., 50-50% split with T_1=10, and many other combinations
  - Spending too much upfront limits actions available to re-plan in second stage.
  - Spending too little upfront leaves little variation re-planning can take advantage of.
- Re-planning outperforms either two problem methods under the correct planning conditions.
- Decision on budget split could be encoded in optimization problem (future work).

Summary and Conclusions

- Presented cascade model for stochastic diffusion in many interesting networks (conservation, epidemiology, social networks).
- Extended SAA sampling methodology for stochastic optimization to a multistage setting for cascades.
- Significant complexity and variation when solving deterministic analogues of stochastic problems.
  - Easy-hard-easy patterns, power-law decay
- When decisions are made in multiple stages, re-planning based on stochastic outcomes can have significant benefit (when budget split is carefully chosen).