An Overview of the International Planning Competition Part 2: Classical Tracks

Amanda Coles[1] Andrew Coles[1]
Álvaro Torralba[2] Florian Pommerening[3]

[1]King's College London

[2]Saarland University

[3]University of Basel, Switzerland

January 27, 2019

Classical Tracks

Classical Planning:

- Deterministic and Fully-observable environment
- Find a sequence of actions that leads to the goal

Several Tracks:

- Optimal Track: find a plan of minimum cost
- Satisficing Track: find a plan as good as possible (but not necessarily optimal)
- Agile Track: find a plan as quickly as possible
- Cost-Bounded Track: find a plan whose cost is below a bound

Benchmarks

How to evaluate a general solver?

- The goal in planning is to develop a decision-making tool that can work in any situation

How to evaluate a general solver?

- The goal in planning is to develop a decision-making tool that can work in any situation
- But we evaluate it in specific situations!

How to evaluate a general solver?

- The goal in planning is to develop a decision-making tool that can work in any situation
- But we evaluate it in specific situations!
- Different planners may do best on different situations so a "good" benchmark selection is essential for the competition.
- Ideally benchmarks should:
  - be diverse: so that planners are evaluated in different scenarios avoiding "overfitting" to a particular class of planning problems
  - be inspired in real-world problems: so that the evaluation targets cases that are relevant for real-world applications
  - be challenging: so that research can be conducted on how to extend the planners to be effective in more scenarios

IPC Benchmarks

Benchmarks published in each IPC:

- IPC 1998: assembly, gripper, logistics, movie, mprime, mystery
- IPC 2000: blocks, elevators, freecell, logistics, schedule
- IPC 2002: depot, driverlog, freecell, rovers, satellite, zenotravel
- IPC 2004: airport, optical-telegraphs, philosophers, pipesworld, psr-large, psr-middle, psr-small
- IPC 2006: openstacks, pathways, pipesworld, rovers, storage, tpp, trucks
- IPC 2008: cybersec, elevators, openstacks, parcprinter, pegsol, scanalyzer, sokoban, transport, woodworking
- IPC 2011: barman, elevators, nomystery, openstacks, parcprinter, parking, pegsol, scanalyzer, sokoban, tidybot, transport, visitall, woodworking
- IPC 2014: barman, cavediving, childsnack, citycar, floortile, ged, hiking, maintenance, openstacks, parking, tetris, thoughtful, tidybot, transport, visitall

All of them publicly available to evaluate new planning algorithms!
`http://planning.domains`

New domains in 2018

PDDL features

- no new PDDL feature this time but . . .
- . . . stronger focus on conditional effects and grounding

$\rightsquigarrow$ In 2 domains, we used two different formulations using the tool by Bustos et al., (2014)

Grounding a Planning Task

Lifted Task

```
(:action move
  :parameters
   (?t - truck
   ?from ?to - location)
  :precondition
   (and (CONNECTED ?from ?to)
        (truck-at ?t ?from))
  :effect
   (and (not (truck-at ?t ?from))
        (truck-at ?t ?to))
(:objects
   t1 t2 - truck
   l1 l2 l3 - location
   )
(:init
  (CONNECTED l1 l2)
  (CONNECTED l2 l3)
  (truck-at t1 l1)
  (truck-at t2 l3)
)
```

Grounded Task

```
(:action move-t1-l1-l2
  :precondition (truck-at t1 l1))
  :effect
    (and (not (truck-at t1 l1))
         (truck-at t1 l2))

(:action move-t2-l1-l2
  :precondition (truck-at t2 l1))
  :effect
    (and (not (truck-at t2 l1))
         (truck-at t2 l2))

(:action move-t1-l2-l3
  :precondition (truck-at t1 l2))
  :effect
    (and (not (truck-at t1 l2))
         (truck-at t1 l3))
        . . .
```

Grounding a Planning Task

Lifted Task

```
(:action move
  :parameters
   (?t - truck
   ?from ?to - location)
  :precondition
   (and (CONNECTED ?from ?to)
        (truck-at ?t ?from))
  :effect
   (and (not (truck-at ?t ?from))
        (truck-at ?t ?to))
(:objects
   t1 t2 - truck
   l1 l2 l3 - location
   )
(:init
  (CONNECTED l1 l2)
  (CONNECTED l2 l3)
  (truck-at t1 l1)
  (truck-at t2 l3)
)
```

Grounded Task

```
(:action move-t1-l1-l2
  :precondition (truck-at t1 l1))
  :effect
    (and (not (truck-at t1 l1))
         (truck-at t1 l2))

(:action move-t2-l1-l2
  :precondition (truck-at t2 l1))
  :effect
    (and (not (truck-at t2 l1))
         (truck-at t2 l2))

(:action move-t1-l2-l3
  :precondition (truck-at t1 l2))
  :effect
    (and (not (truck-at t1 l2))
         (truck-at t1 l3))
        . . .
```

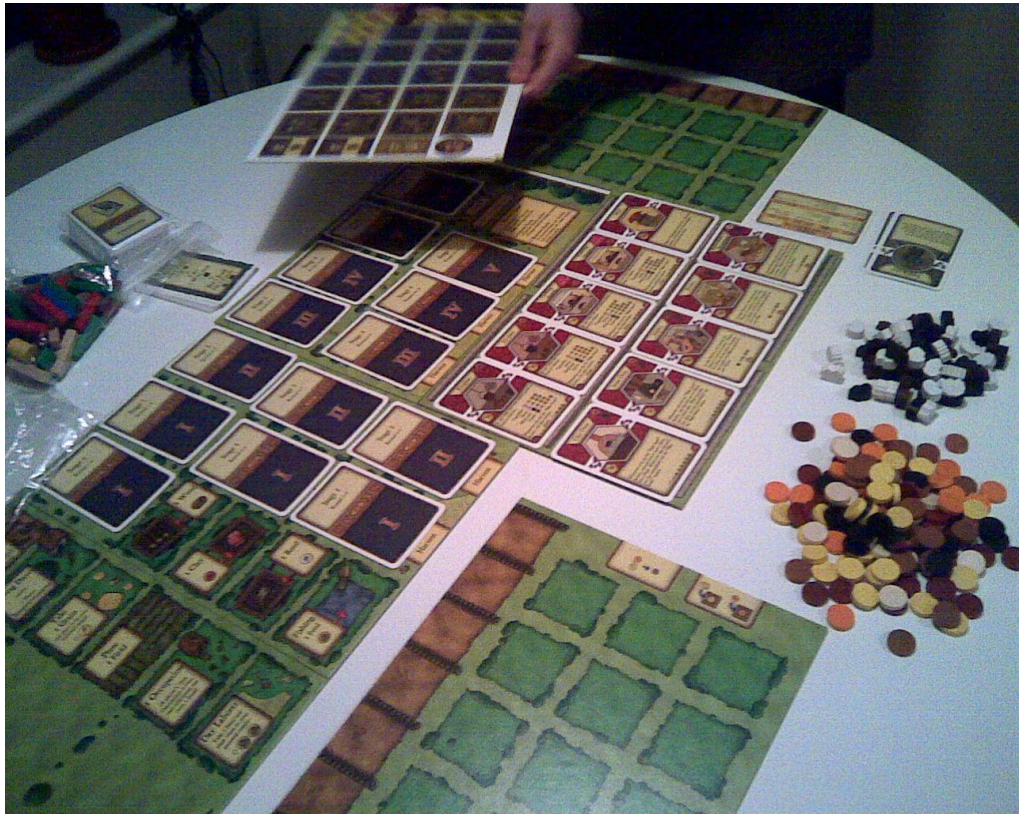
Conditional Effects

```
(:action toggle
  :parameters ?b - lightbulb
  :precondition ()
  :effect (and
     (when (on ?b) (not (on ?b)))
     (when (not (on ?b)) (on ?b))
     )
)
```

$\rightsquigarrow$ We do not know if the lightbulb will be on or off after applying toggle

Agricola

Submitted by: Tomás de la Rosa, Universidad Carlos III de Madrid

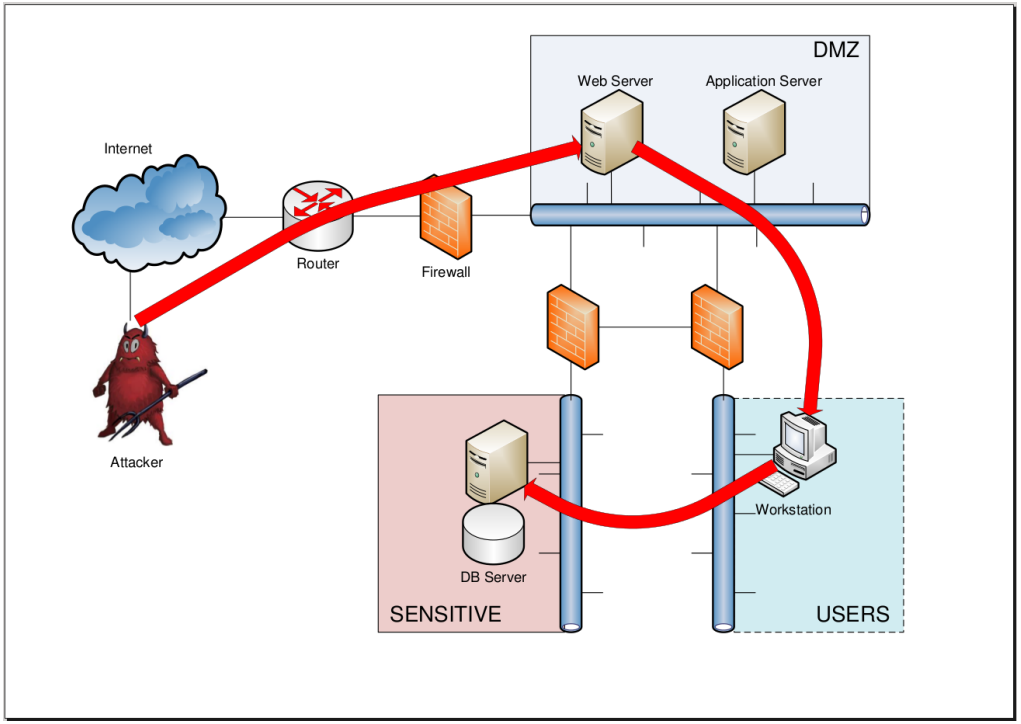


Loosely based on the board game “Agricola”.

- ⇝ dead-ends

Caldera

Submitted by: Andy Applebaum, Doug Miller, and Blake Strom, MITRE.

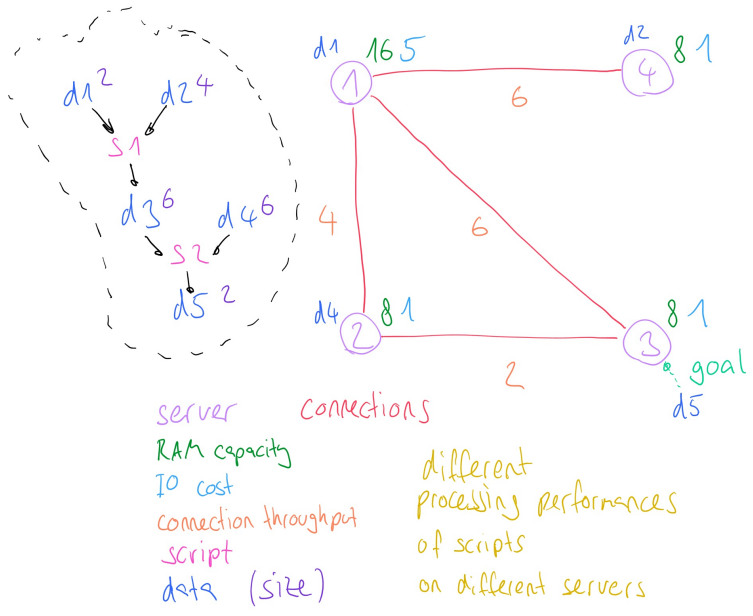


Cybersecurity domain based on a real-world application.

- ⇝ Delete-free domain
- ⇝ Quantified Conditional Effects
- ⇝ Hard to ground

Data Network

Submitted by: Submitted by: Manuel Heusner, Basel University



Process and send data accross a computer network.

$\rightsquigarrow$ Our Logistics variant ☺

Flash Fill

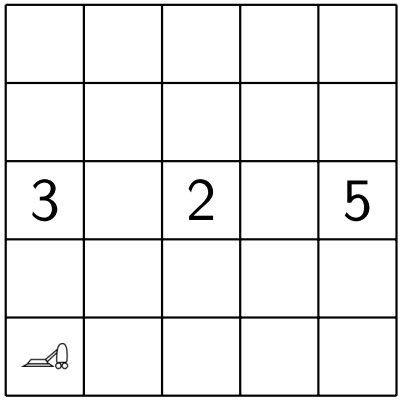
Submitted by: Javier Segovia, Universitat Pompeu Fabra

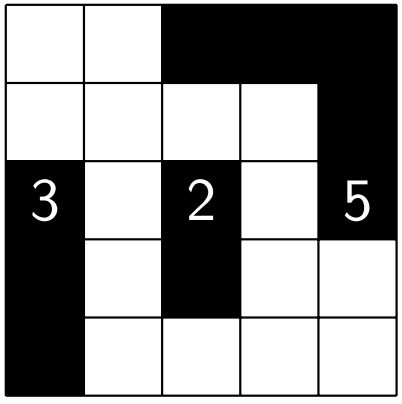
	A	B
1	Florian Pommerening	F P
2	Alvaro Torralba	A T
3	Foo Bar	F B
4	Bar Foo	
5		
6		
7		

Excel Flashfill feature modelled as a classical planning problem by using the planning programs compilation by Segovia et al.

- $\rightsquigarrow$ Quantified conditional effects (hard to handle)

Nurikabe



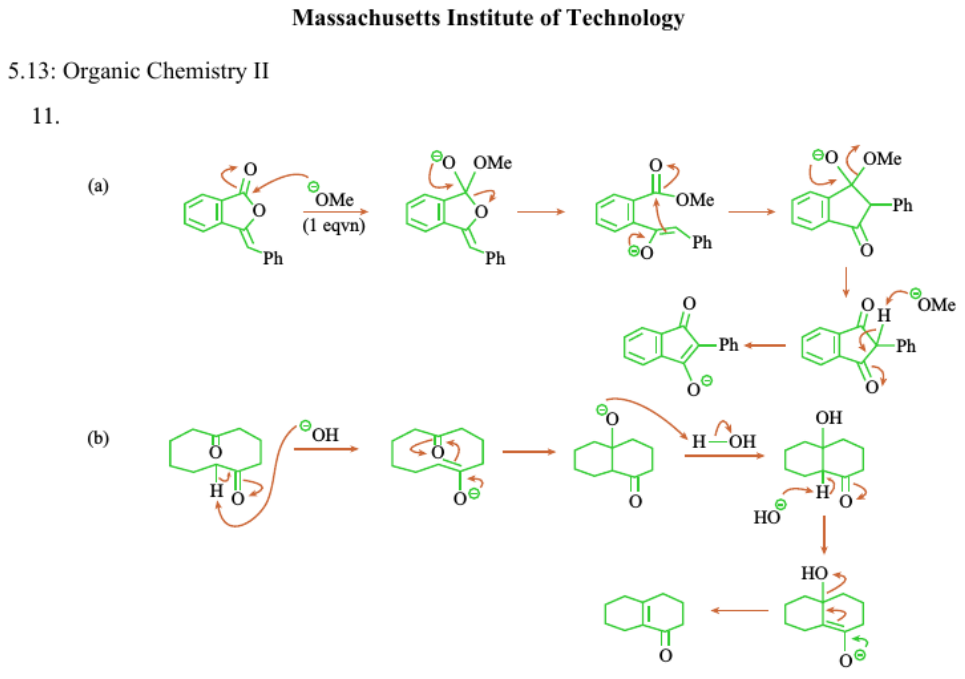


Version of Floortile where the robot must decide the painting pattern

$\rightsquigarrow$ Quantified conditional effects (easy to handle)

Organic Synthesis

Submitted by: Hadi Qovaizi, Arman Masoumi, Anne Johnson, Russell Viirre, Andrew McWilliams, and Mikhail Soutchanski, Ryerson University



Find a sequence of reactions that produces the target molecule from given initial molecules. The instances are based on real exam questions.

- ⇝ Hard to ground

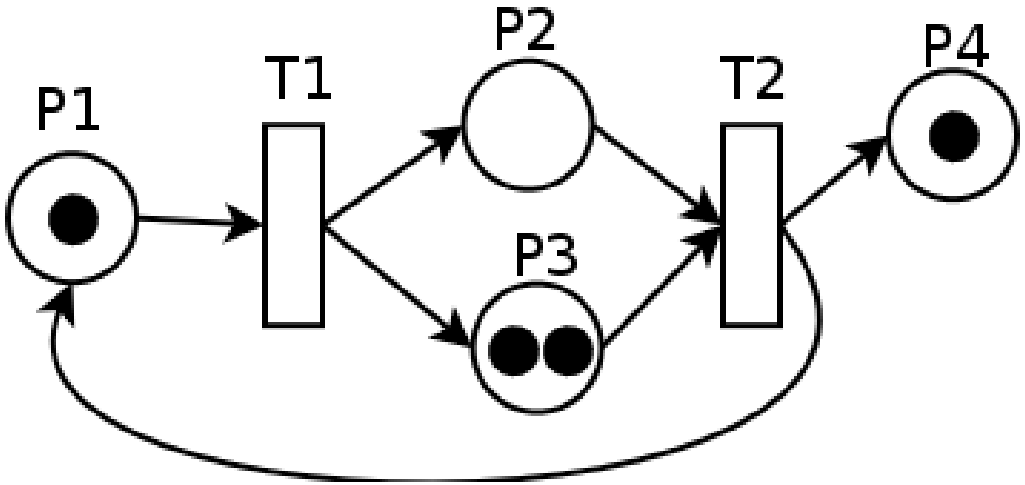
Organic Synthesis

(**:action** aldolcondensation
:parameters (?o4 ?o7 ?o10 ?o15 – oxygen ?na14 – sodium
?h8 ?h9 ?h11 ?h12 ?h13 – hydrogen
?c1 ?c5 ?c6 – carbon ?r12 ?r13 – rgroup)
:precondition ....
:effect .....
)

(**:objects**
c1 c2 c3 c4 – carbon
cl1 cl3 cl4 cl5 – chlorine
h1 h17 h18 h19 h2 h3 h4 h5 h56 h6 h7 – hydrogen
na1 – sodium
o1 o2 o3 o4 – oxygen
rg1 rg2 – rgroup
)

Petri Net Alignment

Submitted by: Massimiliano de Leoni and Andrea Marrella, Eindhoven University of Technology



Align the execution of a petri net to a sequence of events

$\rightsquigarrow$ 0-cost actions

Settlers

Submitted by: Marcel Steinmetz, Saarland University

Resource-constrained version of the numeric domain Settlers

$\rightsquigarrow$ Quantified conditional effects (hard to compile away)

Conditional Effects in Settlers

Action unload-ore (?v - vehicle ?p - place) has effects:

- (available-ore ?v) -= 1
- (available-ore ?p) += 1

Effect (available-ore ?p) -= 1 encoded as:

```
( forall (?rpold ?rpnew - ore_level)
  (when (and (available-ore ?p ?rpold)
             (DIFF-ORE ?rpnew ol1 ?rpold))
        (and (not (available-ore ?p ?rpold))
             (available-ore ?p ?rpnew))))
```

After grounding:

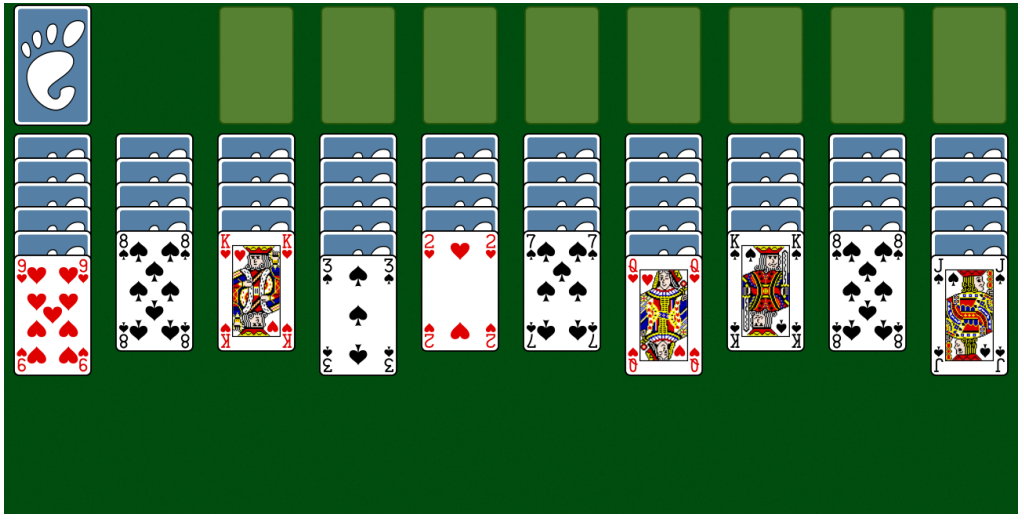
- (when (available-ore ?p) == 1 → (available-ore ?p) := 0
- (when (available-ore ?p) == 2 → (available-ore ?p) := 1
- (when (available-ore ?p) == 3 → (available-ore ?p) := 2
- (when (available-ore ?p) == 4 → (available-ore ?p) := 3

Snake

Version of the Snake game where the location where apples will spawn is known in advance.

- ⇝ Many facts (snake representation)

Spider



Variant of the Spider card game where all cards are faced up from the beginning.

- ⇝ Conditional effects
- ⇝ 0-cost actions

Termes

Submitted by: Sven Koenig and Satish Kumar

Single agent variant of Harvard TERMES robots, based on termites.

- $\rightsquigarrow$ Long plans

Overview

	Agricola	Caldera	Data Network	Flashfill	Nurikabe	Organic Synthesis	Petri Net Alignment	Settlers	Snake	Spider	Termes	SUM
C. Eff		✓!		✓!!	✓			✓!!		✓		5
0 cost							✓			✓		2
costs	✓	?	✓	✓		?		✓				4-6

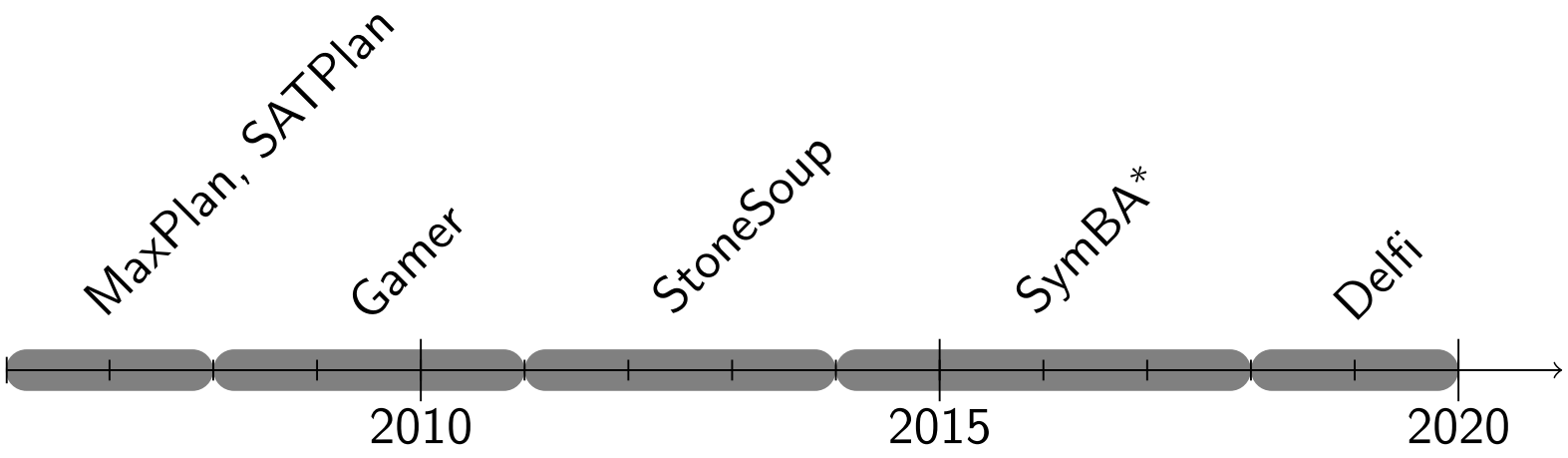
- 4 domains from "applications" (not developed for the IPC): Caldera, Flashfill, Organic Synthesis, Petri-net Alignment

Optimal Track

Rules of the Optimal Track

- Goal: Find an optimal plan
- Metric: number of plans solved

Trends and Breakthroughs: Optimal Planning



- SAT planners (MaxPlan, SATPlan)
- Symbolic Search planners (Gamer, SymBA*)
- Heuristic search planners
- Portfolios (StoneSoup, Delfi)

Techniques used in 2018

- abstraction heuristics
  - many and most sucessful submissions
- landmark heuristics
- critical path heuristics
- decoupled search
- symbolic search
  - hard-to-beat baseline: blind symbolic bi-directional search

Coverage	agricola	caldera	data-net.	flashfill	nurikabe	org.-syn.	settlers	snake	spider	termes	SUM
Delfi1	12	13	13	12	13	**20**	9	11	11	12	126
Complementary2	6	12	12	12	13	18	9	**14**	12	16	124
Complementary1	10	11	**14**	**13**	13	17	8	11	11	16	124
Planning-PDBs	6	12	**14**	11	13	18	8	13	11	16	122
symb. Bi-dir.	**15**	10	13	11	13	19	8	4	7	**18**	118
Scorpion	2	12	**14**	**13**	13	0	**10**	**14**	**17**	14	109
Delfi2	11	11	13	11	13	9	8	7	7	15	105
FDMS2	14	12	9	12	13	2	8	11	11	12	104
FDMS1	9	12	10	12	13	2	9	11	11	12	101
DecStar	0	8	**14**	11	**14**	8	8	11	13	12	99
Metis1	0	13	12	12	**14**	9	9	7	11	6	93
MSP	7	8	13	8	12	10	0	11	6	16	91
Metis2	0	**15**	12	12	**14**	9	0	7	12	6	87
Blind	0	8	7	11	10	7	8	12	11	10	84
Symple-2	1	8	9	7	13	2	0	0	5	13	58
Symple-1	0	8	9	8	13	2	0	0	4	13	57
maplan-2	2	2	9	0	6	0	0	**14**	1	12	46
maplan-1	0	2	12	0	6	0	0	7	10	6	43

	A^*	Symbolic	PDB	M&S	CEGAR	LM-cut	h^m	Symmetry	POR	# best	Score
Delfi1	✓	✓	✓	✓		✓		✓	✓	1	126
Complementary2	✓		✓							1	124
Complementary1	✓		✓							2	124
Planning-PDBs	✓		✓							1	122
symb. Bi-dir.		✓								2	118
Scorpion	✓				✓					**5**	109
Delfi2	✓	✓	✓	✓		✓		✓	✓	0	105
FDMS2	✓			✓						0	104
FDMS1	✓			✓						0	101
DecStar	✓		✓			✓		✓	✓	2	99
Metis1	✓					✓		✓	✓	1	93
MSP	✓	✓				✓				0	91
Metis2	✓					✓		✓	✓	2	87
Blind	✓									0	84
Symple-2		✓								0	58
Symple-1		✓								0	57
maplan-2	✓						✓			1	46
maplan-1	✓						✓			0	43

	A^*	Symbolic	PDB	M&S	CEGAR	LM-cut	h^m	Symmetry	POR	# best	Score
Delfi1	✓	✓	✓	✓		✓		✓	✓	1	126
Complementary2	✓		✓							1	124
Complementary1	✓		✓							2	124
Planning-PDBs	✓		✓							1	122
symb. Bi-dir.		✓								2	118
Scorpion	✓				✓					**5**	109
Delfi2	✓	✓	✓	✓		✓		✓	✓	0	105
FDMS2	✓			✓						0	104
FDMS1	✓			✓						0	101
DecStar	✓		✓			✓		✓	✓	2	99
Metis1	✓					✓		✓	✓	1	93
MSP	✓	✓				✓				0	91
Metis2	✓					✓		✓	✓	2	87
Blind	✓									0	84
Symple-2		✓								0	58
Symple-1		✓								0	57
maplan-2	✓						✓			1	46
maplan-1	✓						✓			0	43

Coverage	caldera	flashfill	nurikabe	settlers	spider	Ceff	Rest	SUM
Delfi1	13	12	13	9	11	58	68	**126**
Complementary2	12	12	13	9	12	58	66	124
Complementary1	11	**13**	13	8	11	56	68	124
Planning-PDBs	12	11	13	8	11	55	67	122
symb. Bi-dir.	10	11	13	8	7	49	**69**	118
Scorpion	12	**13**	13	**10**	**17**	**65**	44	109
Delfi2	11	11	13	8	7	50	55	105
FDMS2	12	12	13	8	11	56	48	104
FDMS1	12	12	13	9	11	57	44	101
DecStar	8	11	**14**	8	13	54	45	99
Metis1	13	12	**14**	9	11	59	34	93
MSP	8	8	12	0	6	34	57	91
Metis2	**15**	12	**14**	0	12	53	34	87
Blind	8	11	10	8	11	48	36	84
Symple-2	8	7	13	0	5	33	25	58
Symple-1	8	8	13	0	4	33	24	57
maplan-2	2	0	6	0	1	9	37	46
maplan-1	2	0	6	0	10	18	25	43

In 2008...

In 2008... Blind baseline solved 1 more instance than the winner of the competition

In 2008... Blind baseline solved 1 more instance than the winner of the competition

Coverage	**agricola**	**caldera**	**data-net.**	**flashfill**	**nurikabe**	**org.-syn.**	**settlers**	**snake**	**spider**	**termes**	**SUM**
Delfi1	12	13	13	12	13	20	9	**11**	11	12	126
Complementary2	6	12	12	12	13	18	9	14	12	16	124
Complementary1	10	11	14	13	13	17	8	**11**	11	16	124
Planning-PDBs	6	12	14	11	13	18	8	13	11	16	122
symb. Bi-dir.	15	10	13	11	13	19	8	**4**	**7**	18	118
Scorpion	2	12	14	13	13	**0**	10	14	17	14	109
Delfi2	11	11	13	11	13	9	8	**7**	**7**	15	105
FDMS2	14	12	9	12	13	**2**	8	**11**	11	12	104
FDMS1	9	12	10	12	13	**2**	9	**11**	11	12	101
DecStar	0	8	14	11	14	8	8	**11**	13	12	99
Metis1	0	13	12	12	14	9	9	**7**	11	**6**	93
MSP	7	8	13	**8**	12	10	**0**	**11**	**6**	16	91
Metis2	0	15	12	12	14	9	**0**	**7**	12	**6**	87
Blind	0	8	7	11	10	7	8	12	11	10	84

Conclusions: Optimal Track

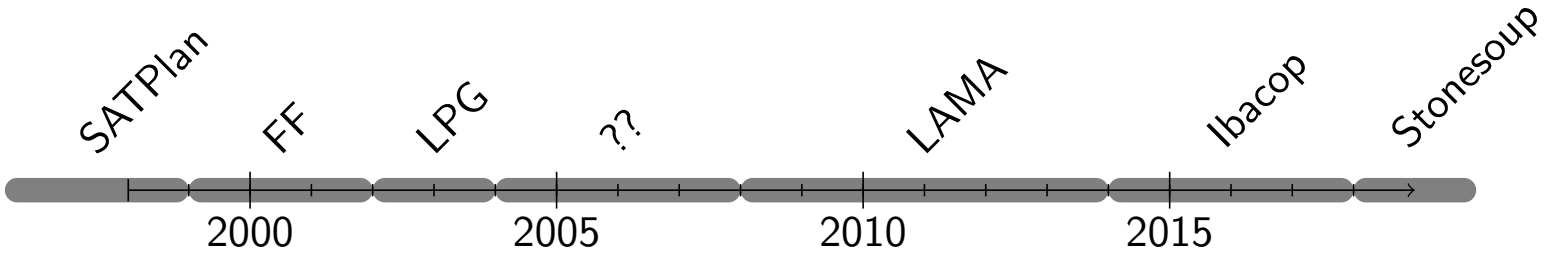
- Lots of research done in abstraction heuristics has paid off: PDBs, CEGAR, M&S
- A portfolio won the track but non-portfolio planners are still very competitive
- Symbolic search and A* are two competitive approaches for optimal planning

Satisficing Track

Rules of the Satisficing Track

- Goal: Find a plan with high quality
- Metric: $C/C*$
  - same as in 2008 but different from 2011, 2014
  - reference plans by many different means

Trends and Breakthroughs: Satisficing Planning



- SAT-based planners (SAT-plan, Madagascar)
- Heuristic search planners (FF, LPG, Fast Downward, LAMA)
- Portfolios (Ibacop, Stonesoup)

Techniques used in 2018

- Delete-relaxation heuristics
  - many variants of partial delete relaxation: Red-black, hCFF
- Landmarks
- Novelty: in many different forms
- Search algorithms
  - Greedy Best First Search
  - Decoupled search
  - Enforced Hill Climbing

Sat score	agricola	caldera	data-net.	flashfill	nurikabe	org.-syn.	settlers	snake	spider	termes	SUM
Stone Soup	13	14	10	13	18	9	**16**	7	10	8	123
Remix	13	14	10	12	**18**	9	16	7	10	6	120
DUAL-BFWS	12	**17**	11	**16**	14	11	6	9	12	5	119
Saarplan	**14**	11	12	13	16	11	9	8	10	7	116
DecStar	12	13	10	13	12	9	15	4	12	6	111
Cerberus	10	10	11	9	15	**12**	9	5	13	7	108
LAMA 2011	9	13	7	13	10	**12**	15	3	13	7	107
BFWS-Pref.	11	15	8	11	12	7	8	**15**	10	5	106
Cerberus-gl	10	10	11	9	15	**12**	9	5	**14**	6	106
OLCFF	13	11	12	0	17	9	0	7	11	7	92
POLY-BFWS	13	17	11	8	10	5	9	2	8	2	90
IBaCoP	10	5	**14**	0	6	8	0	8	8	**10**	73
IBaCoP2	11	6	11	0	7	8	0	7	7	7	66
MERWIN	10	0	10	0	5	**12**	0	4	11	7	62
mercury	12	0	8	0	5	11	0	3	12	7	61
DFS+	10	12	6	1	5	5	7	4	7	0	60
fs-sim	11	6	5	0	10	4	0	7	4	3	53
fs-blind	3	6	5	0	12	4	0	7	5	7	50
freelunch-dr	8	1	0	0	0	6	0	5	0	0	22
freelunch-ma	0	2	2	0	3	8	0	1	0	0	16
Symple-2	1	2	0	0	2	5	0	0	0	1	11
Symple-1	1	2	0	0	2	5	0	0	0	1	11
alien	4	0	1	0	0	4	0	0	0	0	9

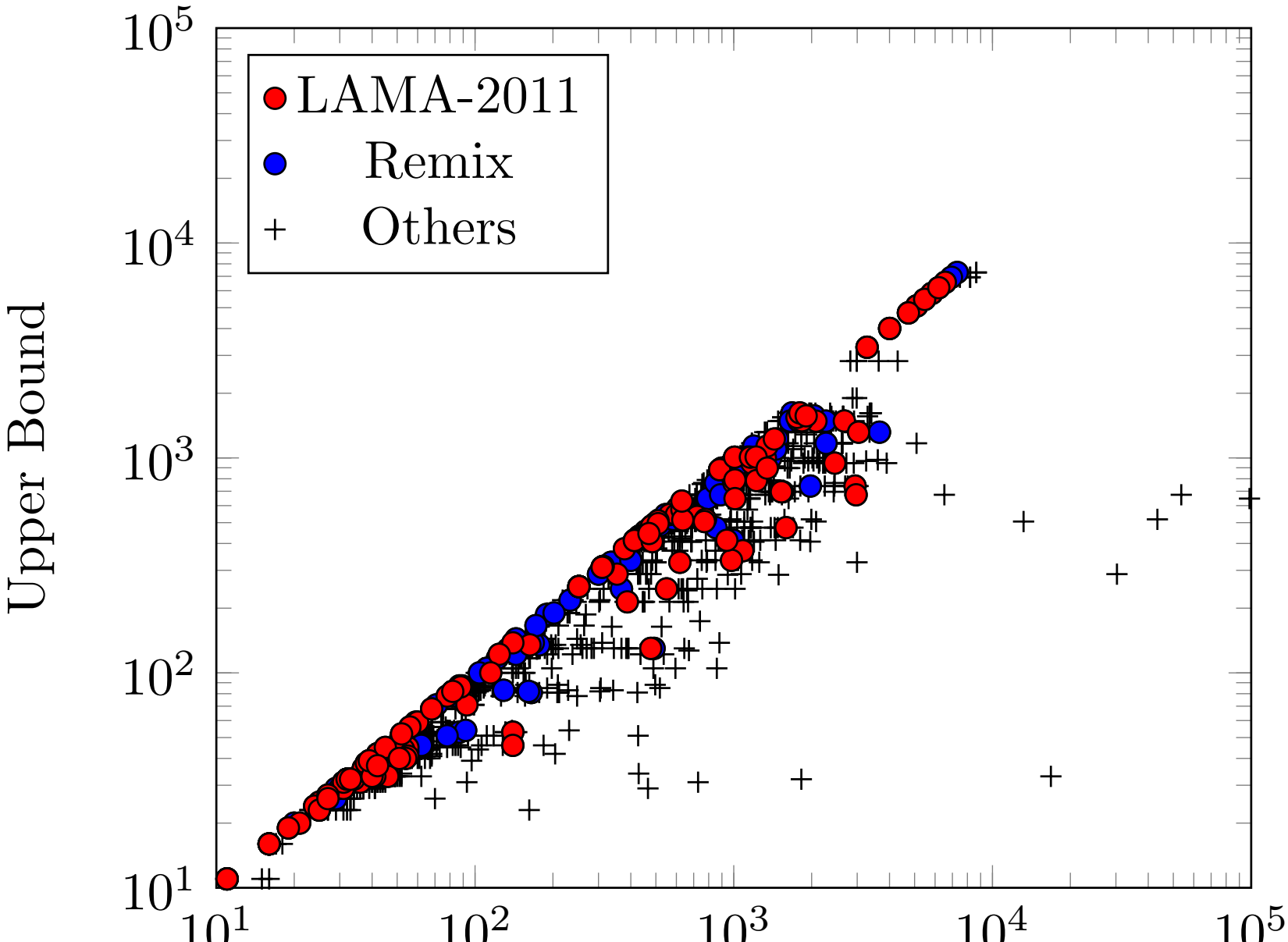
Sat score	GBFS	EHC	SAT	hFF	hRB	hCFF	Lm	Nov	# best	Score
FF	✓	✓		✓						
LPG				✓						
LAMA 2011	✓			✓			✓		1	107
IbaCop 2014	✓		✓	✓			✓			
Stone Soup	✓			✓			✓		1	123
Remix	✓			✓			✓		1	120
DUAL-BFWS	✓			✓			✓	✓	2	119
Saarplan	✓	✓		✓	✓	✓	✓		1	116
DecStar	✓			✓			✓		0	111
Cerberus | -gl	✓				✓		✓	✓	1 | 2	108 | 106
BFWS-Pref.	✓			✓			✓	✓	1	106
OLCFF	✓	✓		✓		✓	✓	✓	0	92
POLY-BFWS	✓							✓	0	90
IBaCoP 1–2	✓		✓	✓	✓		✓		2	73 | 66
MERWIN	✓				✓		✓	✓	1	62
mercury	✓				✓		✓		0	61
DFS+	✓							✓	0	60
fs-sim	✓							✓	0	53
fs-blind	✓							✓	0	50
freelunch-dr | -ma			✓						0	22
Symple-1 | 2									0	11 | 11
alien	✓			✓					0	9

Coverage	caldera	flashfill	nurik.	settlers	spider	Ceff	Rest	SUM
Stone Soup	14	13	**18**	**16**	10	**71**	52	**123**
Remix	14	12	**18**	**16**	10	**70**	50	**120**
DUAL-BFWS	**17**	**16**	14	6	12	65	54	119
Saarplan	11	13	16	9	10	59	**57**	116
DecStar	13	13	12	15	12	65	46	111
Cerberus	10	9	15	9	13	56	52	108
LAMA 2011	13	13	10	15	13	64	43	107
BFWS-Pref.	15	11	12	8	10	56	50	106
Cerberus-gl	10	9	15	9	**14**	57	49	106
OLCFF	11	0	17	0	11	39	53	92
POLY-BFWS	**17**	8	10	9	8	52	38	90
IBaCoP	5	0	6	0	8	19	54	73
IBaCoP2	6	0	7	0	7	20	46	66
MERWIN	0	0	5	0	11	16	46	62
mercury	0	0	5	0	12	17	44	61
DFS+	12	1	5	7	7	32	28	60
fs-sim	6	0	10	0	4	20	33	53
fs-blind	6	0	12	0	5	23	27	50
freelunch-dr	1	0	0	0	0	1	21	22
freelunch-ma	2	0	3	0	0	5	11	16
Symple-2	2	0	2	0	0	4	7	11
Symple-1	2	0	2	0	0	4	7	11
alien	0	0	0	0	0	0	9	9

Coverage	caldera	flashfill	nurik.	settlers	spider	Ceff	Rest	SUM
Stone Soup	14	13	**18**	**16**	10	**71**	52	**123**
Remix	14	12	**18**	**16**	10	**70**	50	**120**
DUAL-BFWS	**17**	**16**	14	6	12	65	54	119
Saarplan	11	13	16	9	10	59	**57**	116
DecStar	13	13	12	15	12	65	46	111
Cerberus	10	9	15	9	13	56	52	108
LAMA 2011	13	13	10	15	13	64	43	107
BFWS-Pref.	15	11	12	8	10	56	50	106
Cerberus-gl	10	9	15	9	**14**	57	49	106
OLCFF	11	0	17	0	11	39	53	92
POLY-BFWS	**17**	8	10	9	8	52	38	90
IBaCoP	5	0	6	0	8	19	54	73
IBaCoP2	6	0	7	0	7	20	46	66
MERWIN	0	0	5	0	11	16	46	62
mercury	0	0	5	0	12	17	44	61
DFS+	12	1	5	7	7	32	28	60
fs-sim	6	0	10	0	4	20	33	53
fs-blind	6	0	12	0	5	23	27	50
freelunch-dr	1	0	0	0	0	1	21	22
freelunch-ma	2	0	3	0	0	5	11	16
Symple-2	2	0	2	0	0	4	7	11
Symple-1	2	0	2	0	0	4	7	11
alien	0	0	0	0	0	0	9	9

Coverage	caldera	flashfill	nurik.	settlers	spider	Ceff	Rest	SUM
Stone Soup	14	13	**18**	**16**	10	**71**	52	**123**
Remix	14	12	**18**	**16**	10	**70**	50	**120**
DUAL-BFWS	**17**	**16**	14	6	12	65	54	119
Saarplan	11	13	16	9	10	59	**57**	116
DecStar	13	13	12	15	12	65	**46**	111
Cerberus	10	9	15	9	13	56	**52**	108
LAMA 2011	13	13	10	15	13	**64**	43	107
BFWS-Pref.	15	11	12	8	10	56	**50**	106
Cerberus-gl	10	9	15	9	**14**	57	**49**	106
OLCFF	11	0	17	0	11	39	**53**	92
POLY-BFWS	**17**	8	10	9	8	52	38	90
IBaCoP	5	0	6	0	8	19	54	73
IBaCoP2	6	0	7	0	7	20	**46**	66
MERWIN	0	0	5	0	11	16	**46**	62
mercury	0	0	5	0	12	17	**44**	61
DFS+	12	1	5	7	7	32	28	60
fs-sim	6	0	10	0	4	20	33	53
fs-blind	6	0	12	0	5	23	27	50
freelunch-dr	1	0	0	0	0	1	21	22
freelunch-ma	2	0	3	0	0	5	11	16
Symple-2	2	0	2	0	0	4	7	11
Symple-1	2	0	2	0	0	4	7	11
alien	0	0	0	0	0	0	9	9

Plan Quality Analysis



Conclusions: Satisficing Track

- hFF still very relevant today: top 8 planners use it or a variant thereof (red-black or CFF)
- Many planners using different variants of novelty
- best-first width search
  - best single-planner performance
  - very agile
- SAT planning
  - entries not competitive on the 2018 domains
- Good support for Conditional Effects has been very relevant

Agile Track

Rules of the Agile Track

- Goal: Find a plan quickly
- Metric: $1 - \log(t)/\log(300)$, or 1 if solved in first second
  - different from 2014
  - independent of reference time
  - stronger emphasis on solving in short time
- Instance selection:
  - Same instances as in satisficing track

Agile Track

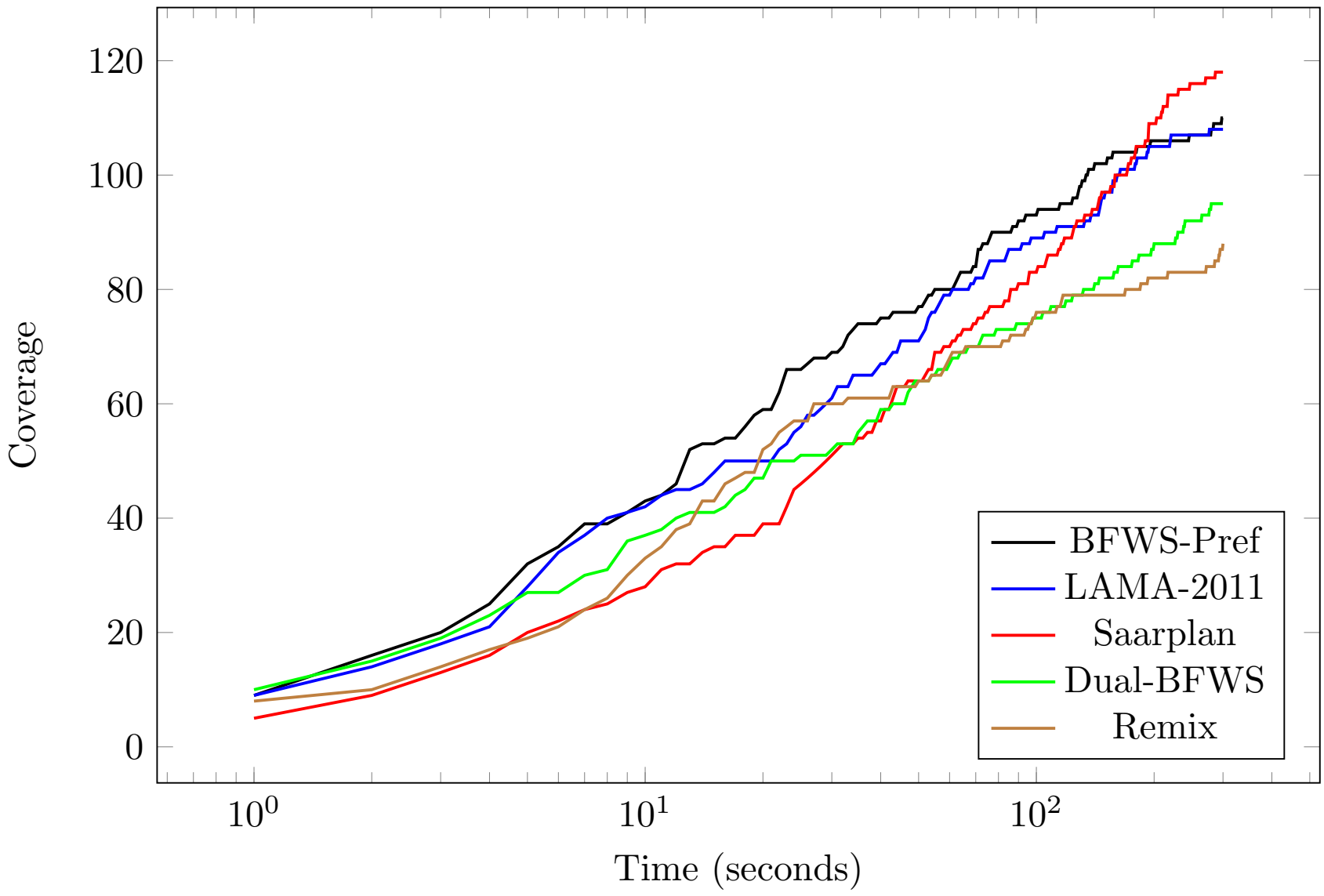
- Recently introduced in 2014
- Techniques similar to those for satisficing planning
  - Less room for preprocessing
  - Portfolios less effective

Agl score	agricola	caldera	data-net.	flashfill	nurikabe	org.-syn.	settlers	snake	spider	termes	SUM
BFWS-Pref.	2.3	6.9	6.1	**8.8**	**7.5**	3.9	2.4	8.9	**5.6**	3.8	56.1
LAMA 2011	0.8	6.6	7.6	7.4	6.3	2.9	**7.1**	2.0	4.6	**7.6**	52.7
Saarplan	1.4	6.6	**9.5**	3.4	7.5	1.9	3.8	3.8	2.0	6.3	46.3
DUAL-BFWS	1.6	**7.6**	4.4	8.0	7.1	**4.8**	1.7	3.8	4.2	3.1	46.2
Remix	1.2	6.1	6.6	6.0	7.1	3.3	5.6	1.6	1.5	5.4	44.3
POLY-BFWS	2.2	7.5	5.4	6.7	7.4	2.8	1.8	2.5	4.8	1.9	43.0
DecStar	1.4	5.8	5.3	3.9	6.4	2.3	5.6	1.8	2.6	6.3	41.4
OLCFF	1.3	6.6	9.1	0.4	7.4	1.7	0.0	3.8	1.7	6.0	38.1
Cerberus	0.5	5.9	4.8	2.4	7.4	1.5	1.7	2.7	0.7	6.8	34.4
Cerberus-gl	0.4	5.8	4.8	2.4	7.5	1.6	1.7	2.6	0.7	3.6	31.0
LAPKT-DFS+	2.4	6.6	1.9	0.3	4.1	2.0	2.6	0.7	3.5	0.0	24.1
mercury2014	1.1	0.0	7.1	0.0	1.1	2.5	0.0	1.9	3.4	6.4	23.5
fs-blind	0.5	3.4	2.4	0.0	7.4	0.2	0.0	4.7	1.5	3.4	23.5
fs-sim	**2.5**	3.3	3.2	0.0	6.5	0.4	0.0	2.7	1.1	3.0	22.8
MERWIN	0.9	0.0	7.0	0.0	1.1	2.5	0.0	1.8	2.8	5.7	21.7
freelunch-dr	1.1	0.9	3.4	0.0	0.0	1.2	0.0	**10.8**	1.8	0.0	19.2
IBaCoP	0.3	0.4	1.5	0.0	0.1	1.0	0.0	0.4	0.0	0.8	4.5
freelunch-ma	0.0	1.4	0.8	0.0	0.6	1.0	0.0	0.0	0.0	0.0	3.9
alien	0.6	0.0	1.1	0.0	0.0	1.8	0.0	0.0	0.0	0.0	3.5
IBaCoP2	0.3	0.1	1.1	0.0	0.0	0.5	0.0	0.4	0.0	0.6	3.0
Symple-2	0.0	0.1	0.0	0.0	0.4	1.7	0.0	0.0	0.0	0.0	2.1
Symple-1	0.0	0.1	0.0	0.0	0.5	1.4	0.0	0.0	0.0	0.0	2.1

Agl score	GBFS	EHC	SAT	hFF	hRB	hCFF	Lm	Nov	# best	Score
BFWS-Pref.	✓			✓			✓	✓	3	56.1
LAMA 2011	✓			✓			✓		2	52.7
Saarplan	✓	✓		✓	✓	✓	✓		1	46.3
DUAL-BFWS	✓			✓			✓	✓	2	46.2
Remix	✓			✓			✓			44.3
POLY-BFWS	✓							✓		43.0
DecStar	✓			✓			✓			41.4
OLCFF	✓	✓		✓		✓	✓	✓		38.1
Cerberus | -gl	✓				✓		✓	✓		34.4 | 31.0
LAPKT-DFS+	✓							✓		24.1
mercury2014	✓				✓		✓			23.5
fs-blind	✓							✓		23.5
fs-sim	✓							✓	1	22.8
MERWIN	✓				✓		✓	✓		21.7
freelunch-dr | -ma			✓						1	19.2 | 3.9
IBaCoP 1–2	✓		✓	✓	✓		✓			4.5 | 3.0
alien	✓			✓						3.5
Symple-1 | 2										2.1

Coverage	caldera	flashfill	nurik.	settlers	spider	Ceff	Rest	SUM
BFWS-Pref.	6.9	**8.8**	**7.5**	2.4	**5.6**	31.2	**24.9**	**56.1**
LAMA 2011	6.6	7.4	6.3	**7.1**	4.6	**32**	20.7	52.7
Saarplan	6.6	3.4	**7.5**	3.8	2	23.3	**23**	46.3
DUAL-BFWS	**7.6**	8	7.1	1.7	4.2	28.6	17.6	46.2
Remix	6.1	6	7.1	5.6	1.5	26.3	18	44.3
POLY-BFWS	7.5	6.7	7.4	1.8	4.8	28.2	14.8	43
DecStar	5.8	3.9	6.4	5.6	2.6	24.3	17.1	41.4
OLCFF	6.6	0.4	7.4	0	1.7	16.1	22	38.1
Cerberus	5.9	2.4	7.4	1.7	0.7	18.1	16.3	34.4
Cerberus-gl	5.8	2.4	7.5	1.7	0.7	18.1	12.9	31
LAPKT-DFS+	6.6	0.3	4.1	2.6	3.5	17.1	7	24.1
mercury2014	0	0	1.1	0	3.4	4.5	19	23.5
fs-blind	3.4	0	7.4	0	1.5	12.3	11.2	23.5
fs-sim	3.3	0	6.5	0	1.1	10.9	11.9	22.8
MERWIN	0	0	1.1	0	2.8	3.9	17.8	21.7
freelunch-dr	0.9	0	0	0	1.8	2.7	16.5	19.2
IBaCoP	0.4	0	0.1	0	0	0.5	4	4.5
freelunch-ma	1.4	0	0.6	0	0	2	1.9	3.9
alien	0	0	0	0	0	0	3.5	3.5
IBaCoP2	0.1	0	0	0	0	0.1	2.9	3
Symple-2	0.1	0	0.4	0	0	0.5	1.6	2.1
Symple-1	0.1	0	0.5	0	0	0.6	1.5	2.1

Coverage	caldera	flashfill	nurik.	settlers	spider	Ceff	Rest	SUM
BFWS-Pref.	6.9	**8.8**	**7.5**	2.4	**5.6**	31.2	**24.9**	**56.1**
LAMA 2011	6.6	7.4	6.3	**7.1**	4.6	**32**	20.7	52.7
Saarplan	6.6	3.4	**7.5**	3.8	2	23.3	**23**	46.3
DUAL-BFWS	**7.6**	8	7.1	1.7	4.2	28.6	17.6	46.2
Remix	6.1	6	7.1	5.6	1.5	26.3	18	44.3
POLY-BFWS	7.5	6.7	7.4	1.8	4.8	28.2	14.8	43
DecStar	5.8	3.9	6.4	5.6	2.6	24.3	17.1	41.4
OLCFF	6.6	0.4	7.4	0	1.7	16.1	22	38.1
Cerberus	5.9	2.4	7.4	1.7	0.7	18.1	16.3	34.4
Cerberus-gl	5.8	2.4	7.5	1.7	0.7	18.1	12.9	31
LAPKT-DFS+	6.6	0.3	4.1	2.6	3.5	17.1	7	24.1
mercury2014	0	0	1.1	0	3.4	4.5	19	23.5
fs-blind	3.4	0	7.4	0	1.5	12.3	11.2	23.5
fs-sim	3.3	0	6.5	0	1.1	10.9	11.9	22.8
MERWIN	0	0	1.1	0	2.8	3.9	17.8	21.7
freelunch-dr	0.9	0	0	0	1.8	2.7	16.5	19.2
IBaCoP	0.4	0	0.1	0	0	0.5	4	4.5
freelunch-ma	1.4	0	0.6	0	0	2	1.9	3.9
alien	0	0	0	0	0	0	3.5	3.5
IBaCoP2	0.1	0	0	0	0	0.1	2.9	3
Symple-2	0.1	0	0.4	0	0	0.5	1.6	2.1
Symple-1	0.1	0	0.5	0	0	0.6	1.5	2.1


120
100
80
60
40
20
0
Coverage
10^0
10^1
10^2
Time (seconds)
BFWS-Pref
LAMA-2011
Saarplan
Dual-BFWS
Remix

Conclusions: Agile Track

- Portfolios not dominant when a solution needs to be found quickly
- LAMA is still a very strong competitor
  - very stable on domains with conditional effects
- Best-first width search was a very dominant approach

Cost-Bounded Track

Cost-bounded Track

- Goal: Find a plan with costs below given bound
- Metric: number of plans solved
- Instance selection:
  - mix of instances from satisficing and optimal track
- Bound selection:
  - Very Tight: find an optimal solution (similar to the optimal track but there is no need to prove that it is optimal)
  - Very Loose: find any solution (similar to the agile track)
- To keep things interesting we used two tight bounds per instance

Techniques in the Cost-Bounded Track

- Most planners are configurations from either the optimal or the agile track adapted to return only solutions with a valid cost.

Coverage	agricola	caldera	data-net.	nurikabe	settlers	snake	spider	termes	SUM
Stone Soup	8	16	8	**15**	**14**	9	9	**8**	87
Remix	8	13	**10**	14	**14**	**10**	**10**	7	86
Saarplan	**9**	12	7	11	10	9	7	**8**	73
DUAL-BFWS	6	**17**	9	6	7	**10**	6	3	64
LAMA 2011	8	9	9	7	10	5	6	**8**	62
Complementary2	8	10	3	10	5	**10**	9	6	61
Cerberus-gl	3	8	7	11	9	8	7	5	58
Cerberus	2	8	7	11	9	8	7	5	57
Planning-PDBs	5	7	3	9	5	7	**10**	6	52
DecStar	5	8	2	10	8	5	8	5	51
Complementary1	6	8	2	12	4	5	8	6	51
OLCFF	5	12	4	12	0	**10**	5	1	49
MERWIN	8	0	9	4	0	6	4	3	34
Symple-2	0	2	0	2	0	0	2	4	10
Symple-1	0	2	0	2	0	0	2	4	10

Coverage	agricola	caldera	data-net.	nurikabe	settlers	snake	spider	termes	SUM
Stone Soup	8	16	8	**15**	**14**	9	9	**8**	87
Remix	8	13	**10**	14	**14**	**10**	**10**	7	86
Saarplan	**9**	12	7	11	10	9	7	**8**	73
DUAL-BFWS	6	**17**	9	6	7	**10**	6	3	64
LAMA 2011	8	9	9	7	10	5	6	**8**	62
Complementary2	8	10	3	10	5	**10**	9	6	61
Cerberus-gl	3	8	7	11	9	8	7	5	58
Cerberus	2	8	7	11	9	8	7	5	57
Planning-PDBs	5	7	3	9	5	7	**10**	6	52
DecStar	5	8	2	10	8	5	8	5	51
Complementary1	6	8	2	12	4	5	8	6	51
OLCFF	5	12	4	12	0	**10**	5	1	49
MERWIN	8	0	9	4	0	6	4	3	34
Symple-2	0	2	0	2	0	0	2	4	10
Symple-1	0	2	0	2	0	0	2	4	10

Conclusions: Cost-bounded Track

- Portfolios clearly dominate non-portfolio approaches
- Satisficing planning techniques are generally stronger than optimal planning techniques
  - $\rightarrow$ Even if the bound is the optimal solution cost!
- Great margin of improvement on designing specific algorithms for cost-bounded planning.

Summary

What the IPC 2018 brought us

- New domains with interesting challenges:
  - Hard to ground benchmarks
  - Domains with heavy use of conditional effects
- New planning algorithms
  - Stronger abstraction heuristics: PDBs, CEGAR, M&S, ...
  - Novelty
  - Decoupled Search
  - Comeback of Enforced Hill Climbing

Portfolios

- winners of 3/4 tracks
- recent trend, also in other competitions
- avoid weaknesses of single planners
- well suited for exponential scaling of benchmarks

Controversy

- complaints about attribution and interpretability
- move to separate track?
  - hard to clearly define (e.g., LAMA)

  $\rightsquigarrow$ Sparkle Planning Challenge 2019