



2024 Report

<https://chc-comp.github.io/>

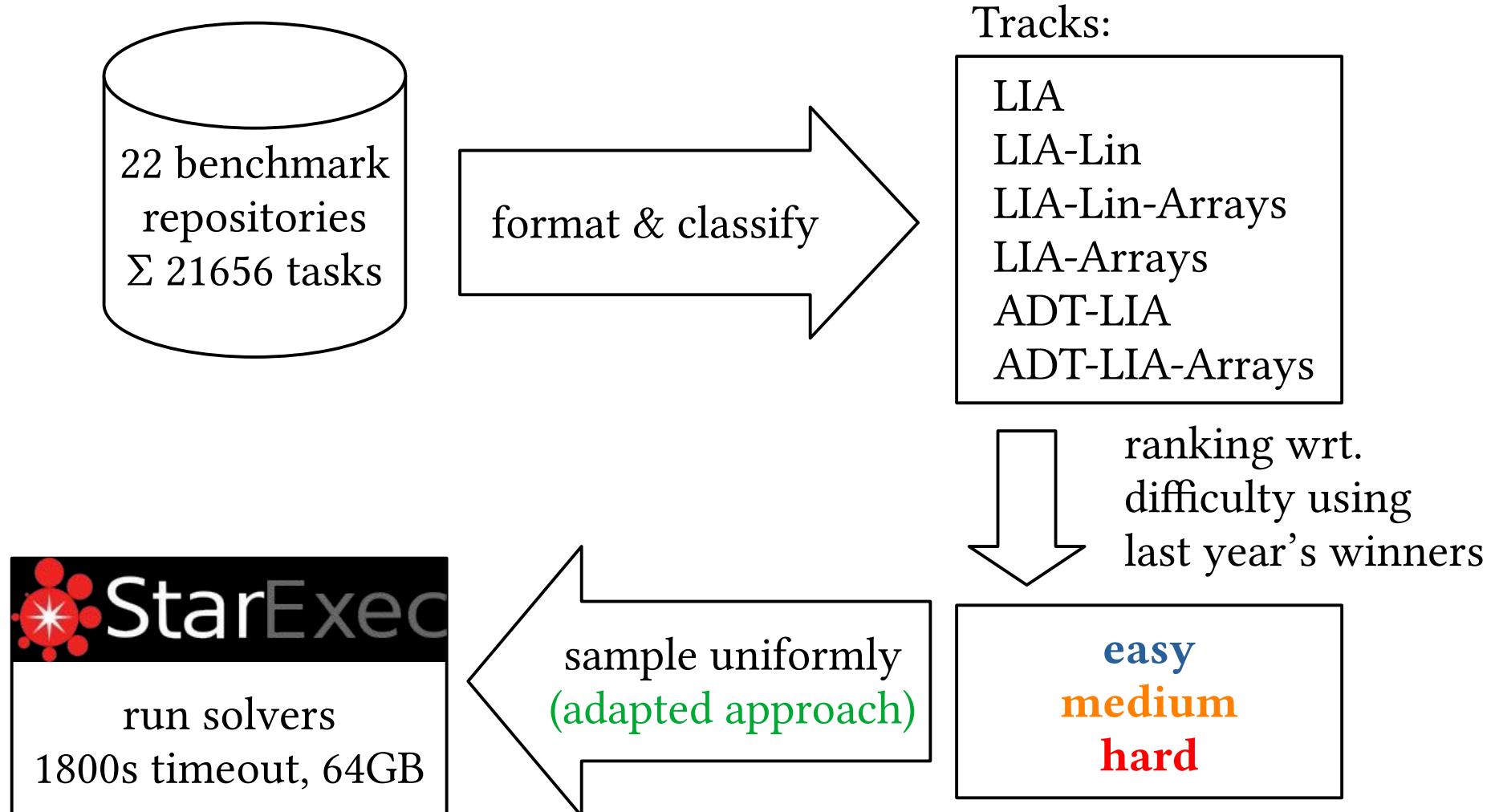
presented at
HCVS 2024, April 7, Luxembourg

Gidon Ernst, LMU Munich, Germany
Jose F. Morales, IMDEA Software Institute, Spain

Goals & Overview

- CHC-COMP: friendly but competitive evaluation of constrained Horn-clause solvers, since 2018
<https://chc-comp.github.io/>
- common task format (subset of SMT-LIB)
<https://chc-comp.github.io/format.html>
- public benchmark repository
<https://github.com/chc-comp>
- Timeline: Jan—April, results presented at HCVS

Benchmark Selection & Execution



Benchmarks

- Overall 22 repositories with 21656 tasks of varying characteristics, size, and difficulty
- Updated Benchmarks (thanks Zafer Esen)
<https://github.com/zafer-esen/tricera-adt-arr/>
- Please contribute!
- Formatting & processing toolchain at
<https://github.com/chc-comp/chc-tools>

6 Tracks

A track limits features that are admitted in the tasks

- Different SMT-LIB theories

LIA: linear integer arithmetic
Arrays (+ quantifiers)

ADT: algebraic data types

- LIA: linear clauses only

- omitted: LRA-TS (transition systems over reals)

LIA
LIA-Lin
LIA-Lin-Arrays
LIA-Arrays
ADT-LIA
ADT-LIA-Arrays

Benchmark Selection

Goal: have a good mix for each track

Approach 2024: weighted **random** sampling

- 20% easy, 40% medium, 40% hard tasks
- uniform across repositories
(normalize wrt. #tasks; avoid manual scaling)
- fix 300 tasks/track overall
(somewhat lower than 2023 due to short timeline)

Measuring difficulty

- Ranking: use two good but complementary solvers (typically last year's winners; small timeout 30s)
 - **easy** (A): both can solve a benchmark
 - **medium** (B/C): only one solves it, respectively
 - **hard** (D): neither solves it
- Solvers used
 - LIA-Lin: B: golem --engine spacer,lawi,split-tpa C: eld -portfolio
 - LIA: B: eld -portfolio C: golem --engine spacer
 - others: B: eld -portfolio C: z3

All benchmarks

repository	LIA				LIA-Lin				LIA-Lin-Arrays				LIA-Arrays				ADT-LIA				ADT-LIA-Arrays								
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D					
adt-purified-benchmarks																	5	1	30	31					67				
ADTRem																	52	0	38	161					251				
aeval-benchmarks	0	0	0	1	2	14	0	37																	54				
aeval-unsafe					0	26	0	28																	54				
eldarica-misc	24	2	12	31	19	107	0	23																	218				
extra-small-lia					0	17	0	38																	55				
hcai-bench	77	2	48	6	20	50	27	4	25	0	6	8	13	0	7	5									298				
hopv	59	3	5	1	36	13	0	0																	117				
jayhorn-benchmarks	3029	37	1192	3067	75	0	0	0																	7400				
kind2-chc-benchmarks	136	12	603	100																						851			
ldv-ant-med							0	0	9	2	0	0	69	273												353			
llreve-bench	24	0	29	6	16	40	0	10	9	34	2	12													182				
quic3							0	0	0	43															43				
ringen-adt-benchmarks																	15	3	20	400					438				
rust-horn	4	0	1	1	10	1	0	0									32	2	5	17					73				
solidity																	956	0	727	7					1252				
synthesis	33	0	46	40													284	0	2826	2261						5490			
tip-adt-lia																		10	0	32	278					320			
tricera-adt-arr																									7	0	46	103	156
tricera-benchmarks	4	0	0	0	5	39	0	361																			409		
ultimate	0	0	0	8													0	0	0	23						31			
vmt-chc-benchmarks					67	692	10	137																		906			
	3390	56	1936	3261	250	999	37	638	34	34	17	65	1253	0	3629	2569	114	6	125	887	1259	0	984	113	###				
	8643				1924			150					7451				1132			2356									

2024 Selection

selected	LIA				LIA-Lin				LIA-Lin-Arrays				LIA-Arrays				ADT-LIA				ADT-LIA-Arrays					
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D		
adt-purified-benchmarks																	5	1	21	25					52	
ADTRem																	20	0	19	44					83	
aeval-benchmarks	0	0	0	1	2	8	0	14																	25	
aeval-unsafe					0	12	0	14																	26	
eldarica-misc	14	2	5	19	8	9	0	14																	71	
extra-small-lia					0	12	0	18																	30	
hcai-bench	15	0	10	2	6	10	2	4	25	0	6	8	13	0	7	5									113	
hopv	16	3	5	1	9	10	0	0																	44	
jayhorn-benchmarks	13	11	19	31	14	0	0	0																	88	
kind2-chc-benchmarks	11	8	10	11																						40
ldv-ant-med							0	0	9	2	0	0	15	48											74	
llreve-bench	11	0	12	6	7	12	0	9	9	34	2	12													114	
quic3							0	0	0	43															43	
ringen-adt-benchmarks																	11	3	14	35					63	
rust-horn	4	0	1	0	9	1	0	0									18	2	5	14					54	
solidity																	36	0	27	7					78	
synthesis	14	0	11	22													27	0	32	64					170	
tip-adt-lia																	6	0	15	42					63	
tricera-adt-arr																									7	
tricera-benchmarks	4	0	0	0	5	7	0	40																	125	
ultimate	0	0	0	8													0	0	0	19					56	
vmt-chc-benchmarks					6	8	5	25																	27	
	102	24	73	101	66	89	7	138	34	34	17	65	76	0	81	143	60	6	74	160	85	0	122	93	1650	
	300				300				150				300				300				300					

Participants

new participant	LIA-Lin	LIA	LIA-Lin- Arrays	LIA- Arrays	ADT-LIA	ADT-LIA- Arrays
Cata					✓	
Eldarica	✓	✓	✓	✓	✓	✓
Golem	✓	✓				
LoAT	✓					
Theta	✓	✓	✓	✓		
U. Tree Automizer	✓	✓	✓	✓		
Ultimate Unihorn	✓	✓	✓	✓		

Remark: Z3/Spacer was not submitted

Big Thanks to



Aaron Stump for StarExec access and run queues

Hari Govind and Emanuele De Angelis for on-boarding & technical support

SoSyLab@LMU Munich for compute resources during preparations

LIA-Lin

solver	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Eldarica	300	230	154	76	70	70	0	143537	37761	4697	35
Golem	300	196	124	72	104	104	0	192781	64398	486	6
LoAT	300	156	86	70	144	144	0	259388	103696	1392	10
Theta	300	153	97	56	147	135	0	289323	265709	10212	0
UltimateUnihorn	300	136	82	54	164	155	0	291794	250001	11214	0
UltimateTreeAutomizer	300	91	48	43	209	181	0	329224	316679	10955	0

cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

LIA

tied 1st

solver	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Golem	300	246	147	99	54	54	0	104227	104237	148	20
Eldarica	300	246	153	93	54	54	0	111508	33500	4661	20
UltimateUnihorn	300	122	62	60	178	119	0	237190	197396	11202	0
UltimateTreeAutomizer	300	51	14	37	249	180	0	329963	303779	10966	0
Theta	300	42	14	28	258	250	0	469130	429984	10235	0

cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

LIA-Lin-Arrays

close!

solver	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Theta	150	88	62	26	62	23	0	102969	94021	5085	8
Eldarica	150	86	62	24	64	64	0	116090	32036	2335	5
UltimateUnihorn	150	53	39	14	97	24	0	52744	45018	5542	1
UltimateTreeAutomizer	150	38	25	13	112	17	0	31990	29950	5482	0

- cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

LIA-Arrays

solver	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Eldarica	300	157	80	77	143	143	0	263720	85060	4653	35
UltimateUnihorn	300	98	43	55	202	64	0	126538	106250	11068	2
Theta	300	63	30	33	237	149	0	294021	259199	10205	1
UltimateTreeAutomizer	300	62	5	57	238	119	0	218312	199681	10959	7

cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

ADT-LIA

newcomer :)

so	er	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Cata		300	214	14	200	86	3	0	8905	8110	761	117
Eldarica		300	139	51	88	161	161	0	295208	76486	4723	42

cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

ADT-LIA-Arrays

solver	cnt	ok	sat	uns	fld	to	mo	time	real	space	uniq
Eldarica	300	257	158	99	43	43	0	81564	26396	4633	257

cnt number of benchmarks
ok number of solved benchmarks (sat + unsat)
sat safety proofs found
uns counterexamples found
fld number of unsolved (failed)
to timeouts without result
mo memory exhausted without result

uniq number of unique benchmarks solved

Winners

LIA-Lin

LIA

LIA-Lin-
Arrays

LIA-
Arrays

ADT-LIA
ADT-LIA-
Arrays

Eldarica

Eldarica&Golem

Theta

Eldarica

Cata

Eldarica

Golem

U.Unihorn

Eldarica

U.Unihorn

Eldarica

Loat

U.Tree.Automizer

U.Unihorn

Theta



Discussion & Outlook

- publish scripts + report
- Wishlist from last year
 - General LRA track, more ADT benchmarks
 - Parallel/portfolio track
 - Model & CEX validation
- Proposal: non-goal directed, i.e., “best-effort” track
- Organizers of the next edition?