

# MaxSAT Evaluation 2024

Jeremias Berg

University of Helsinki

Matti Järvisalo

University of Helsinki

Ruben Martins  
Carnegie Mellon  
University

Andreas Niskanen  
University of Helsinki

Tobias Paxian  
University of Freiburg

<https://maxsat-evaluations.github.io/>

SAT 2024, August, 2024

# What is Maximum Satisfiability?

- ▶ Maximum Satisfiability (MaxSAT):
  - ▶ Set of (hard) clauses.
  - ▶ Linear objective with positive coefficients
    - ▶ Alternatively, a set of (weighted) soft clauses
- ▶ **Goal:** Minimize the objective w.r.t. the clauses.
  - ▶ Equivalent to minimizing the sum of the weights of unsatisfied soft clauses

# Setup

Same structure as the one used in MaxSAT Evaluations 2017-2023:

- ▶ Source disclosure requirement:
  - ▶ Increase the dissemination of solver development
- ▶ Solver description using IEEE Proceedings style:
  - ▶ Better understanding of the techniques used by each solver
- ▶ Benchmark description using IEEE Proceedings style
  - ▶ Better understanding of the nature of each benchmark
- ▶ **Descriptions collected in the proceeding, published at MSE website.**

# News for this year

- ▶ Focus on the robustness of solvers
  - ▶ Mandatory regression suite
    - ▶ <https://github.com/tobipaxe/MaxSATRegressionSuite/>
  - ▶ Open-source fuzzer.
    - ▶ <https://github.com/tobipaxe/MaxSAT-Fuzzer>
  - ▶ External binaries of **closed-source** third-party libraries not allowed anymore.

# Evaluation tracks

## Evaluation tracks:

- ▶ Exact:
  - ▶ Weighted
  - ▶ Unweighted
- ▶ Anytime (60s and 300s):
  - ▶ Weighted
  - ▶ Unweighted
- ▶ Incremental:
  - ▶ No new solvers
  - ▶ Showcase solvers from 2022 on a new benchmark

# Execution environment

The exact track of MSE 2024 was run on the **new** nodes at StarExec:

- ▶ <https://www.starexec.org/>
- ▶ Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz
- ▶ 18432 KB Cache, 264 GB Memory
- ▶ Two solvers per node

Anytime and incremental track run on a computing cluster at the University of Helsinki, part of the Finnish Computing Competence Infrastructure (FCCI).

- ▶ <https://www2.helsinki.fi/en/infrastructures/fcci>
- ▶ 2 x Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz
- ▶ 27.5 MB Cache, 381 GB Memory

# Execution environment

Execution environment:

- ▶ Exact track:
  - ▶ Time limit: 3600 seconds
  - ▶ Memory limit: 32 GB
- ▶ Anytime track:
  - ▶ Two time limits: 60 seconds and 300 seconds
  - ▶ Memory limit: 32 GB
- ▶ Incremental showcase:
  - ▶ Time limit: 7200 seconds
  - ▶ Memory limit: 32 GB

# Benchmark Selection

- ▶ Exact track:
  - ▶ Benchmarks selected **randomly** with a limit on the number of instances per family
- ▶ Anytime track:
  - ▶ Hard instances of the exact track.
    - ▶ Instances that **cannot be solved optimally in 60 seconds** by any participants of the exact tracks.
- ▶ Incremental showcase:
  - ▶ Use **all instances** submitted with the new benchmark.



# New benchmarks

Unweighted:

- ▶ Privilege Escalation (127 instances)
  - ▶ Wenxi Wang, Yang Hu
- ▶ Nonmonotonic c-Inference (100 instances)
  - ▶ Martin von Berg, Arthur Sanin, Aron Spang, Christoph Beierle
- ▶ Minimizing Pentagons in the Plane (65 instances)
  - ▶ Bernardo Subercaseaux, John Mackey, Marijn J. H. Heule, Ruben Martins

Thank you to everyone who submitted benchmarks!

# New benchmarks

Weighted:

- ▶ Learning Balanced Rules (314 instances)
  - ▶ Antonio Carlos Souza Ferreira Junior, Thiago Alves Rocha
- ▶ Synthesizing Pareto-Optimal Interpretations for Black-Box Models (60 instances)
  - ▶ Aniruddha Joshi, Hazem Torfah, Shetal Shah, Supratik Chakraborty, S. Akshay, Sanjit A. Seshia

Thank you to everyone who submitted benchmarks!

# New benchmarks

Incremental:

- ▶ Incremental MaxSAT benchmarks from dynamic discretizations of train scheduling problems (72 instances)
  - ▶ Bjørnar Luteberget

Thank you to everyone who submitted benchmarks!

# MSE 2024 benchmarks

Exact track:

- ▶ Unweighted (553 instances)
- ▶ Weighted (571 instances)

Anytime track:

- ▶ Unweighted (216 instances)
- ▶ Weighted (229 instances)

# Exact Track

# Solvers

MaxSAT approaches in MSE 2024:

Solver	HS	Unsat-based	SIS	B & B	Other
UWrMaxSat		✓	✓		ILP, Pre
CASHWMaxSAT		✓	✓		ILP
CGSS2		✓			ILP
EvalMaxSAT		✓			ILP, SLS, Pre
Exact		✓			PB
MaxCDCL				✓	
WMaxCDCL				✓	
Pacose24			✓		Pre

- ▶ ILP is widely adopted by MaxSAT solvers as a preprocessing step
- ▶ Each solver has multiple versions with different features
  - ▶ 15 submissions in total

# Results

Only one version of each solver is shown in top-3

**Unweighted:** 553 instances

Solver	#Solved	Time (Avg)

**Weighted:** 571 instances

Solver	#Solved	Time (Avg)

# Results

Only one version of each solver is shown in top-3

**Unweighted:** 553 instances

Solver	#Solved	Time (Avg)
WMaxCDCL-openwbo1200	421	369.13
MaxCDCL-openwbo300	418	218.03
UWrMaxSat-SCIP-MaxPre	415	223.63
EvalMaxSAT	412	422.74
MaxCDCL	411	207.90

**Weighted:** 571 instances

Solver	#Solved	Time (Avg)

Erroneus behaviour on at least one instance.



# Results

Only one version of each solver is shown in top-3

**Unweighted:** 553 instances

Solver	#Solved	Time (Avg)
WMaxCDCL-openwbo1200	421	369.13
MaxCDCL-openwbo300	418	218.03
UWrMaxSat-SCIP-MaxPre	415	223.63
EvalMaxSAT	412	422.74
MaxCDCL	411	207.90

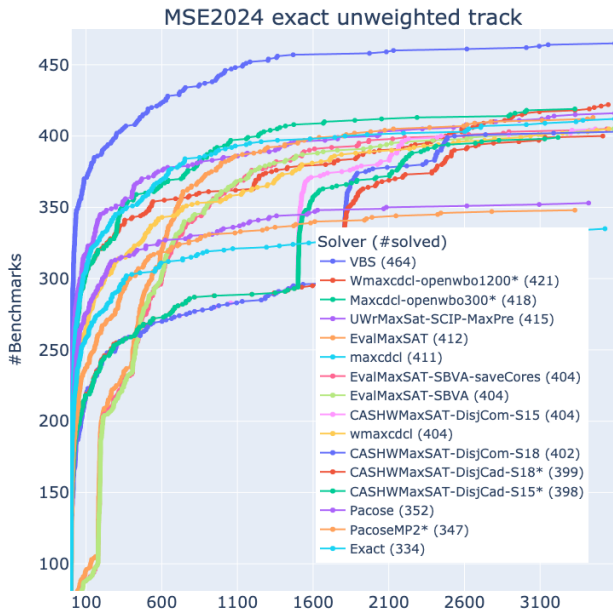
**Weighted:** 571 instances

Solver	#Solved	Time (Avg)
CASHWMaxSAT-DisjCom-S6	448	478.06
UWrMaxSat-SCIP	442	245.14
EvalMaxSAT	435	128.04

Erroneous behaviour on at least one instance.

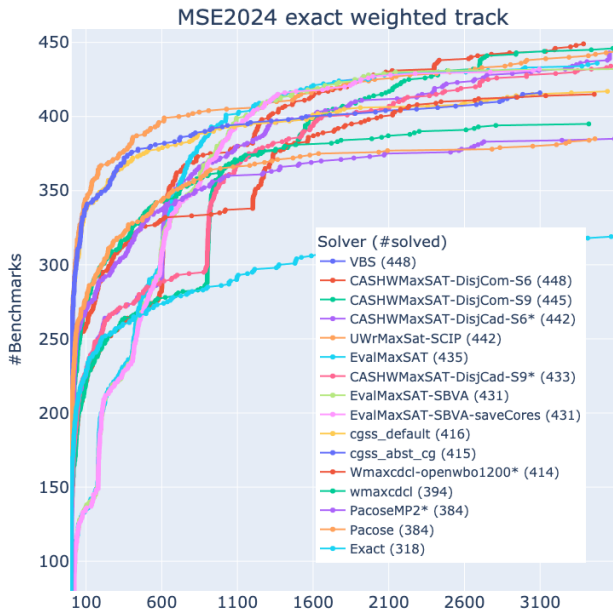
# Detailed

\*erroneous behaviour on at least one instance



# Detailed

\*erroneous behaviour on at least one instance



# Anytime Track

# Ranking for anytime tracks

- ▶ Anytime score: ratio between the best solution found by a given solver and the best-known solution:
  - ▶  $\text{SCORE}(\text{solver}, i) = \frac{(\text{cost of best known solution for } i + 1)}{(\text{cost of solution for } i \text{ found by solver} + 1)}$
  - ▶ Score 0 if no solution found
- ▶ Ranking based on average of all scores.

# Solvers

- ▶ noSAT-MaxSAT (2 versions)
  - ▶ Ole Lübke
  - ▶ SLS search that can initialize a solution with CaDiCaL.
- ▶ NuWLS-c-IBR
  - ▶ Menghua Jiang and Yin Chen
  - ▶ SLS combined with TT-Open-WBO-Inc
- ▶ SPB-MaxSAT-c (three variants)
  - ▶ Mingming Jin, Kun He, Jiongzhi Zheng, Jinghui Xue, Zhuo Chen
  - ▶ SLS combined with TT-Open-WBO-Inc
- ▶ Loandra
  - ▶ Jeremias Berg Christoph Jabs, Hannes Ihalaenen and Matti Järvisalo
  - ▶ Preprocessing, core-guided, SLS, and solution improving search.
- ▶ TT-Open-WBO-Inc (two variants)
  - ▶ Alexander Nadel,
  - ▶ SAT-solver-based local search with intelsat or glucose.
- ▶ Exact
  - ▶ Jo Devriendt
  - ▶ Conflict Driven PB solving

# Anytime track: Unweighted

Only best version of each solver shown

# Anytime track: Unweighted

Only best version of each solver shown

60s	Score (avg)
SPB-MaxSAT-c-Band	0.827
NuWLS-c-IBR	0.826
tt-open-wbo-inc-glucose	0.817
Loandra	0.751
noSAT-MaxSAT-c	0.720
Exact	0.478



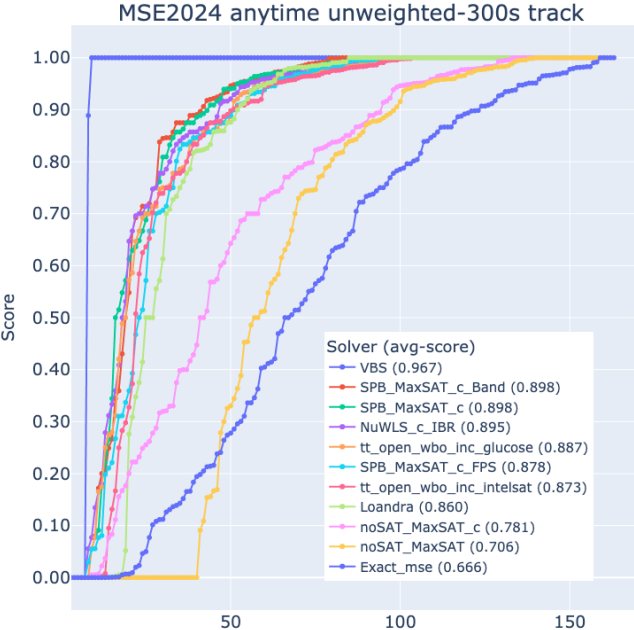
# Anytime track: Unweighted

Only best version of each solver shown

60s	Score (avg)
SPB-MaxSAT-c-Band	0.827
NuWLS-c-IBR	0.826
tt-open-wbo-inc-glucose	0.817
Loandra	0.751
noSAT-MaxSAT-c	0.720
Exact	0.478

300s	Score (avg)
SPB-MaxSAT-c.Band	0.898
NuWLS-c-IBR	0.895
tt-open-wbo-inc-glucose	0.887
Loandra	0.860
noSAT-MaxSAT-c	0.781
Exact	0.666

# Detailed



# Anytime track: Weighted

Only best version of each solver shown

# Anytime track: Weighted

Only best version of each solver shown

60s	Score (avg)
SPB-MaxSAT-c-FPS	0.876
tt-open-wbo-inc-glucose	0.866
NuWLS-c-IBR	0.861
Loandra	0.822
noSAT-MaxSAT-c	0.734
Exact	0.623

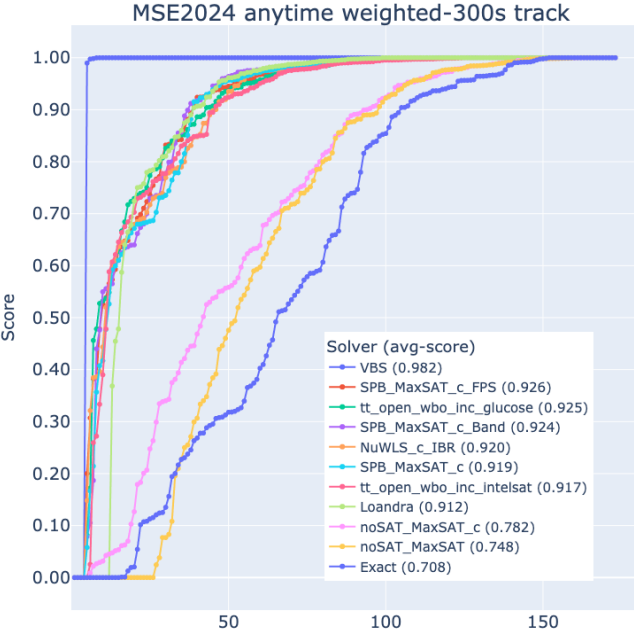
# Anytime track: Weighted

Only best version of each solver shown

60s	Score (avg)
SPB-MaxSAT-c-FPS	0.876
tt-open-wbo-inc-glucose	0.866
NuWLS-c-IBR	0.861
Loandra	0.822
noSAT-MaxSAT-c	0.734
Exact	0.623

300s	Score (avg)
SPB-MaxSAT-c-FPS	0.926
tt-open-wbo-inc-glucose	0.925
NuWLS-c-IBR	0.920
Loandra	0.912
noSAT-MaxSAT-c	0.782
Exact	0.708

# Detailed



# Incremental Showcase

# Organization and Results

- ▶ No new solvers submitted, one new benchmark with 72 instances
- ▶ Showcase incremental solvers from MaxSAT Evaluation 2022
  - ▶ EvalMaxSAT, UWMaxSat
- ▶ Results:
  - ▶ Both solvers solved 68/72 instances.
  - ▶ Only 2/72 instances took more than 20 seconds to solve: EvalMaxSAT clearly outperforms UWMaxSat
  - ▶ PAR-2 scores: EvalMaxSAT 810.11, UWMaxSat 886.45



# Webpages

## MaxSAT Evaluation 2024 webpage

<https://maxsat-evaluations.github.io/2024/>

- ▶ Detailed results for each instance
- ▶ Description of the solvers and benchmarks
- ▶ Source code of the solvers
- ▶ Benchmarks and log files.

# The future of the MSE

- ▶ Organized since 2006.
- ▶ Most of current organizers since 2020.

We are looking for new organizers to:

# The future of the MSE

- ▶ Organized since 2006.
- ▶ Most of current organizers since 2020.

**We are looking for new organizers to:**

# The future of the MSE

- ▶ Organized since 2006.
- ▶ Most of current organizers since 2020.

## We are looking for new organizers to:

- ▶ Provide new ideas
- ▶ Build benchmark sets.
- ▶ Run the tracks
- ▶ Summarize results for conference and website.

# The future of the MSE

- ▶ Organized since 2006.
- ▶ Most of current organizers since 2020.

## We are looking for new organizers to:

- ▶ Provide new ideas
- ▶ Build benchmark sets.
- ▶ Run the tracks
- ▶ Summarize results for conference and website.

If interested, get in touch:

[maxsatevaluation@gmail.com](mailto:maxsatevaluation@gmail.com)

(or any current organizer)

# Thanks

Thanks to everyone who contributed solvers and benchmarks!  
Without you, this evaluation would not have been possible!

Thanks to StarExec for allowing us to use their cluster for the exact track:

<https://www.starexec.org/>



Thanks to FCCI for supporting the anytime and incremental track with computational and data storage resources:

<https://www2.helsinki.fi/en/infrastructures/fcci>

