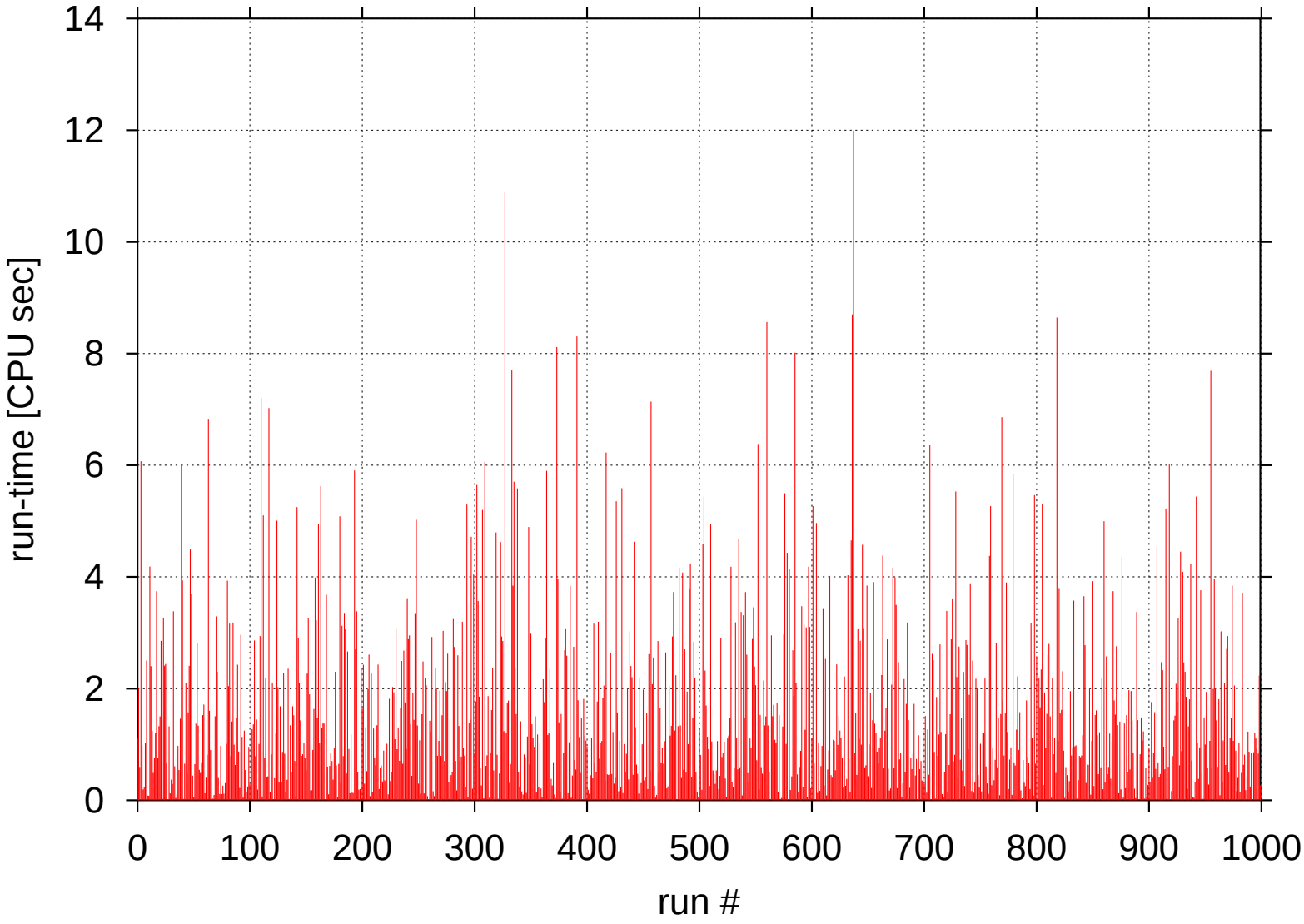
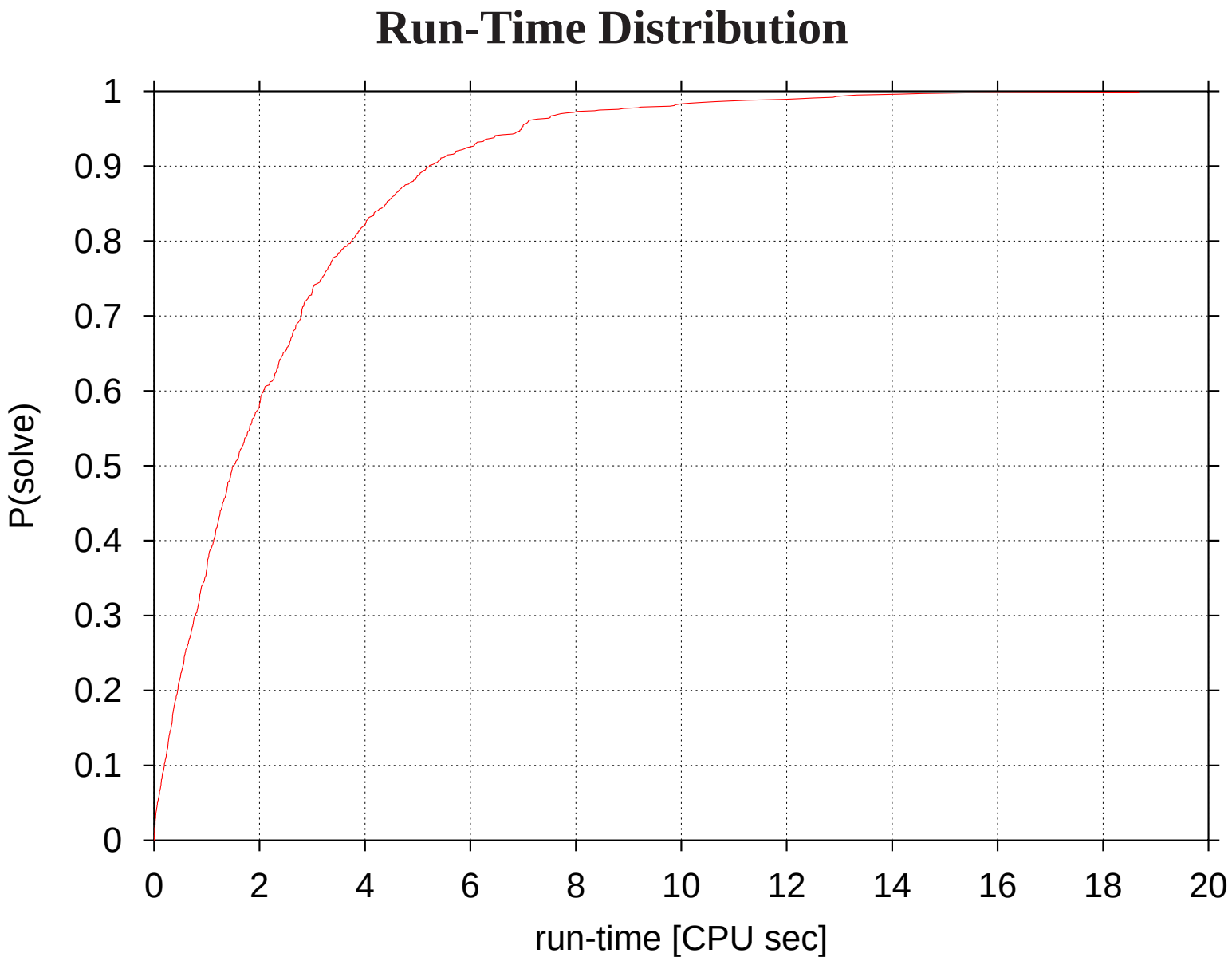
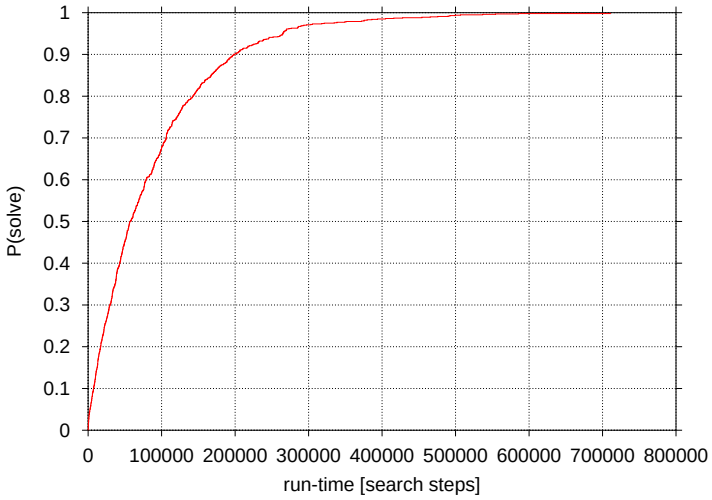
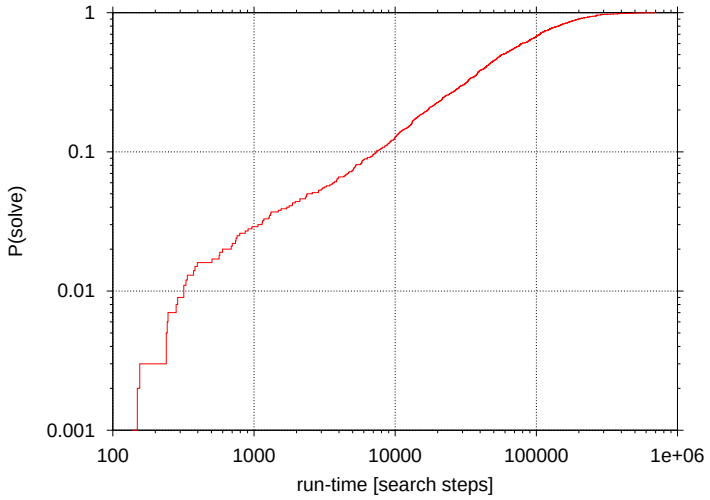
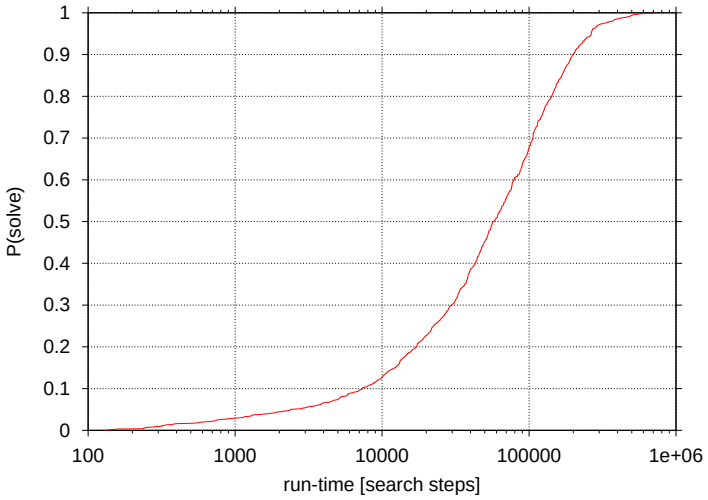

Raw run-time data (each spike one run)
run-time [CPU sec]
0
2
4
6
8
10
12
14
0
100
200
300
400
500
600
700
800
900
1000
run #

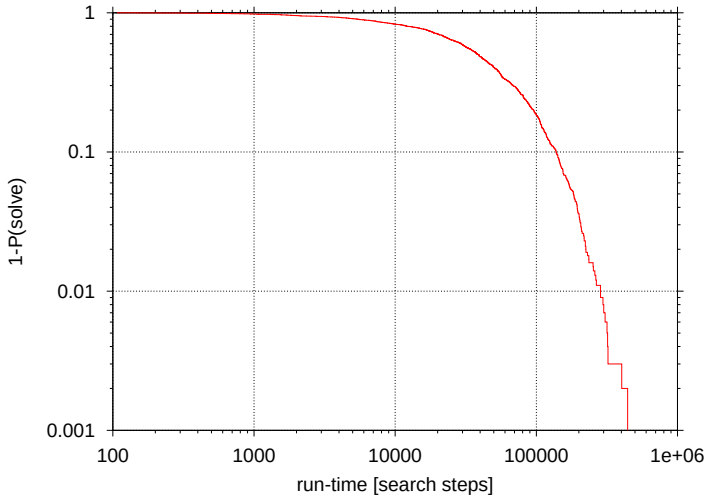

Run-Time Distribution
P(solve)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
run-time [CPU sec]

RTD Graphs


P(solve)
run-time [search steps]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100000
200000
300000
400000
500000
600000
700000
800000


P(solve)
run-time [search steps]
0.001
0.01
0.1
1
100
1000
10000
100000
1e+06

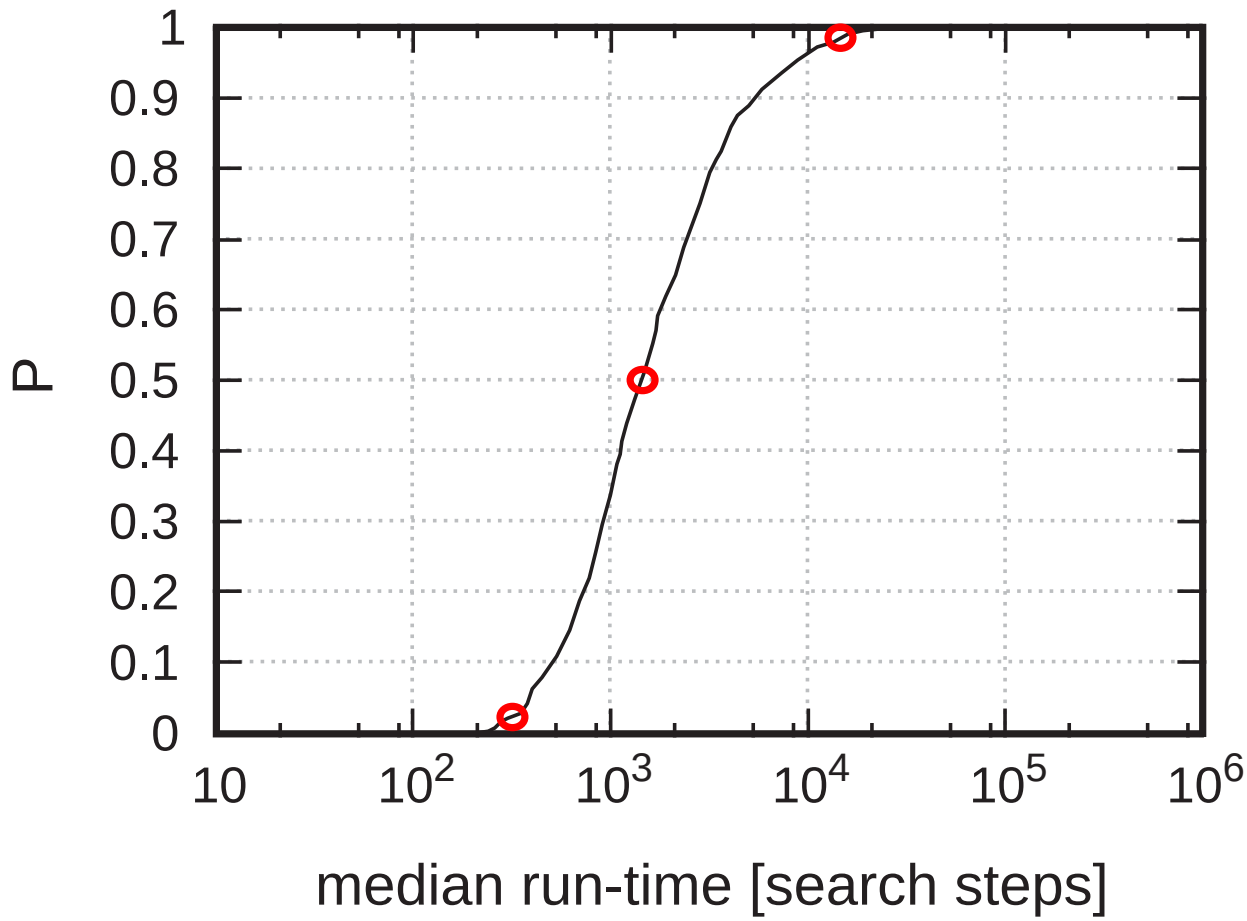

P(solve)
run-time [search steps]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100
1000
10000
100000
1e+06


1-P(solve)
run-time [search steps]
0.001
0.01
0.1
1
100
1000
10000
100000
1e+06

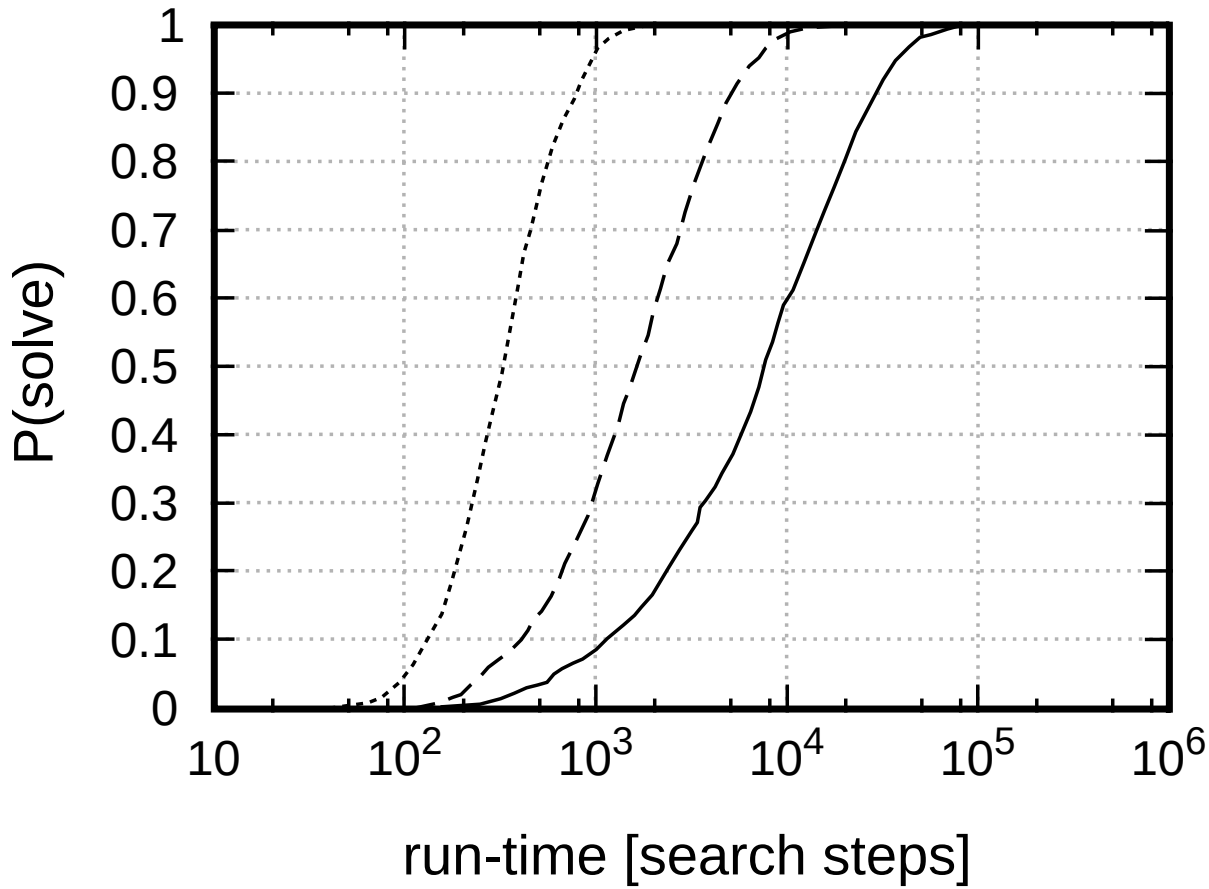
Protocol for obtaining the empirical RTD for an LVA A applied to a given instance π of a decision problem:

- Perform k independent runs of A on π with cutoff time t'. (For most purposes, k should be at least 50–100, and t' should be high enough to obtain at least a large fraction of successful runs.)
- Record number k' of successful runs, and for each run, record its run-time in a list L.
- Sort L according to increasing run-time; let $rt(j)$ denote the run-time from entry j of the sorted list ($j = 1, \ldots, k'$).
- Plot the graph $(rt(j), j/k)$, *i.e.*, the cumulative empirical RTD of A on π.

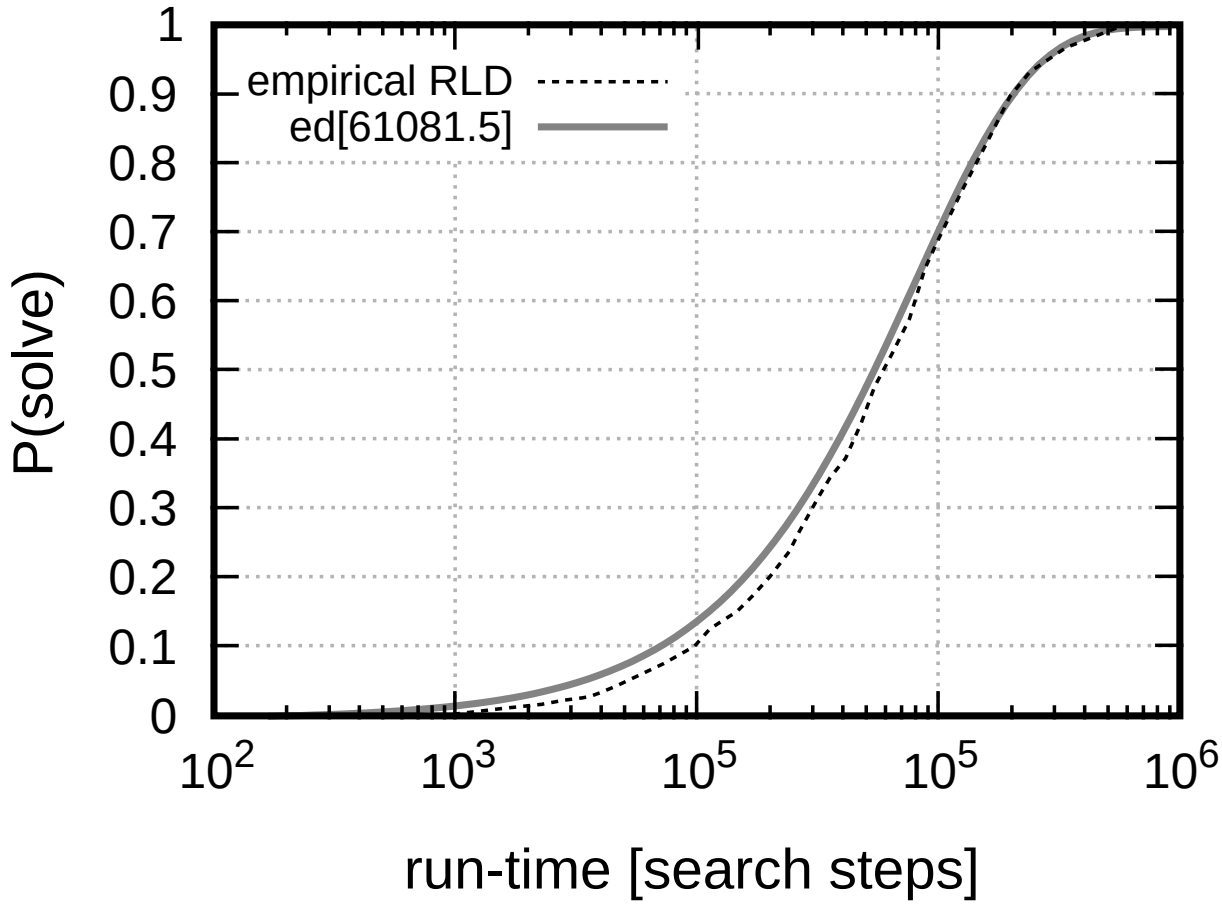
Distribution of median search cost for WalkSAT/SKC over set of 1000 randomly generated, hard 3-SAT instances:



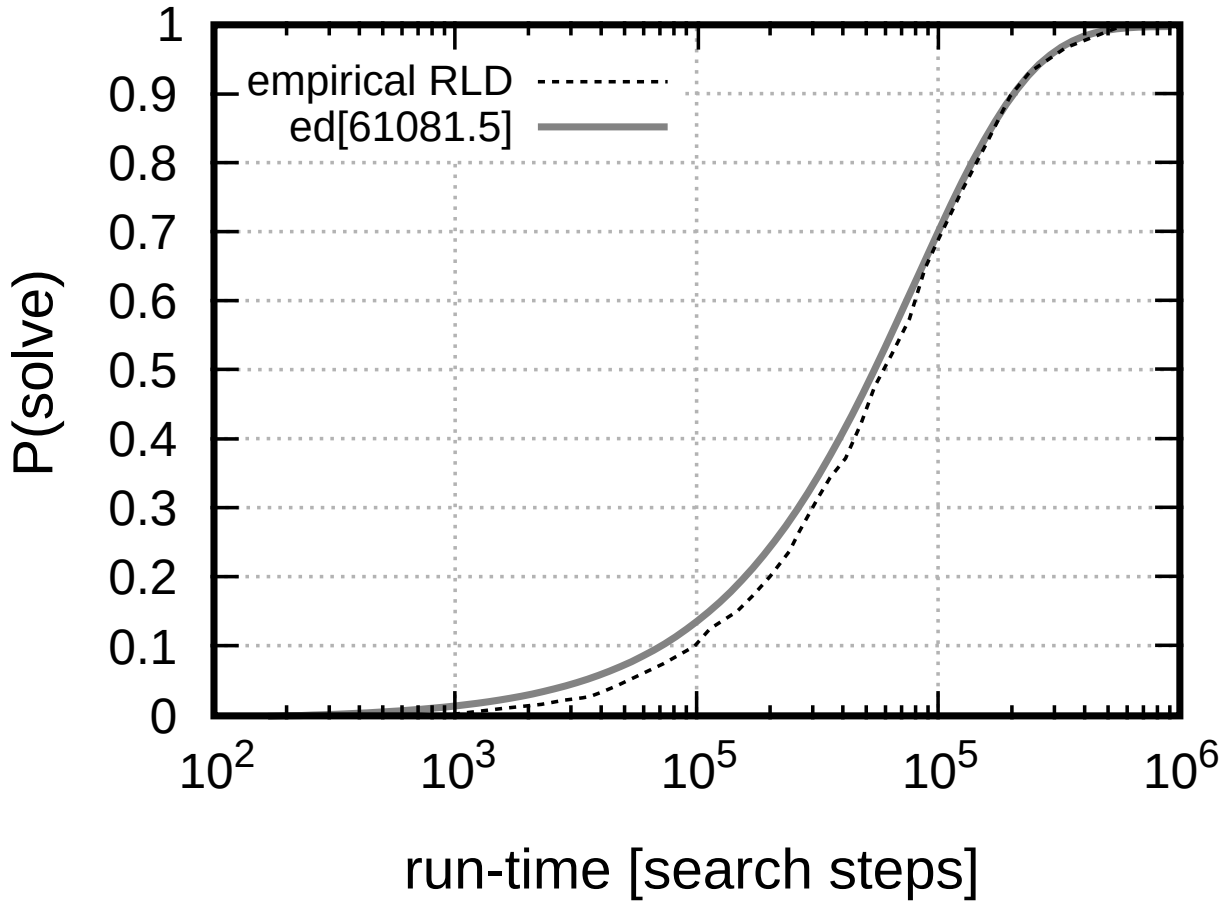
RTDs for WalkSAT/SKC, a prominent SLS algorithm for SAT, on three hard 3-SAT instances:



Approximation of an empirical RTD with an exponential distribution $\mathrm{ed}[m](x) := 1 - 2^{-x/m}$:

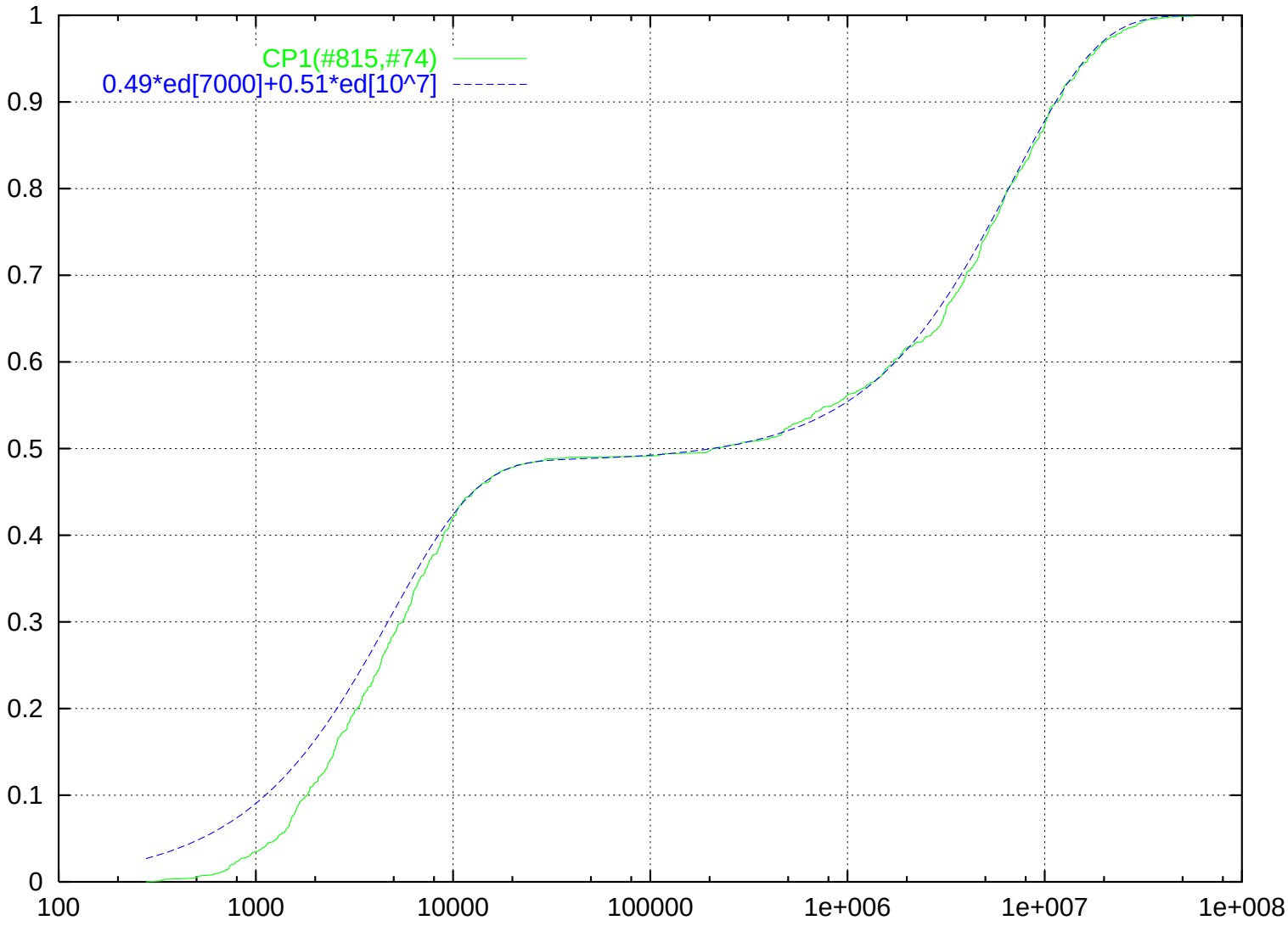


Approximation of an empirical RTD with an exponential distribution $\mathrm{ed}[m](x) := 1 - 2^{-x/m}$:

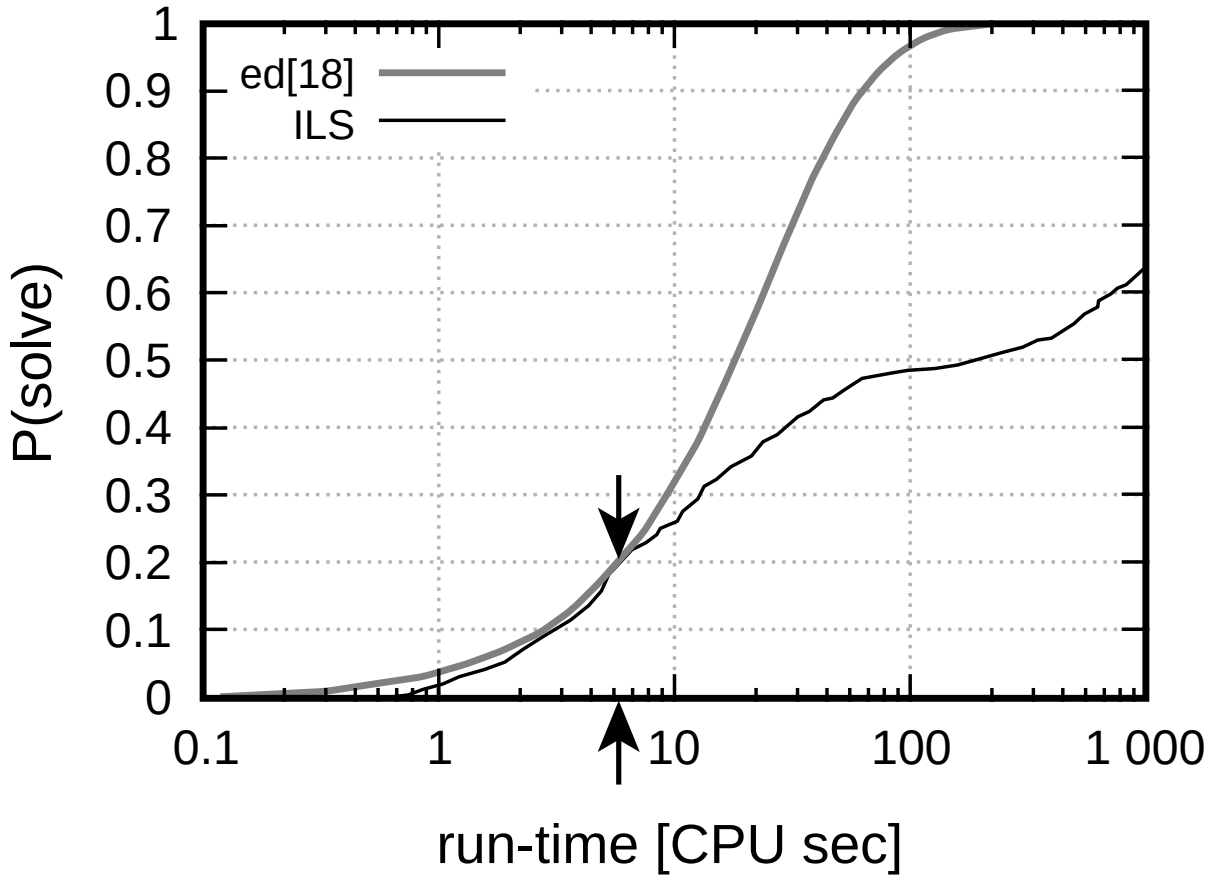


The optimal fit exponential distribution obtained from the Marquardt-Levenberg algorithm passes the χ^2 goodness-of-fit test at $\alpha = 0.05$.

RTD Approximation with Mixture of Exponential Distributions



Example of an empirical RTD of an SLS algorithm on a problem instance for which static restarting is effective:



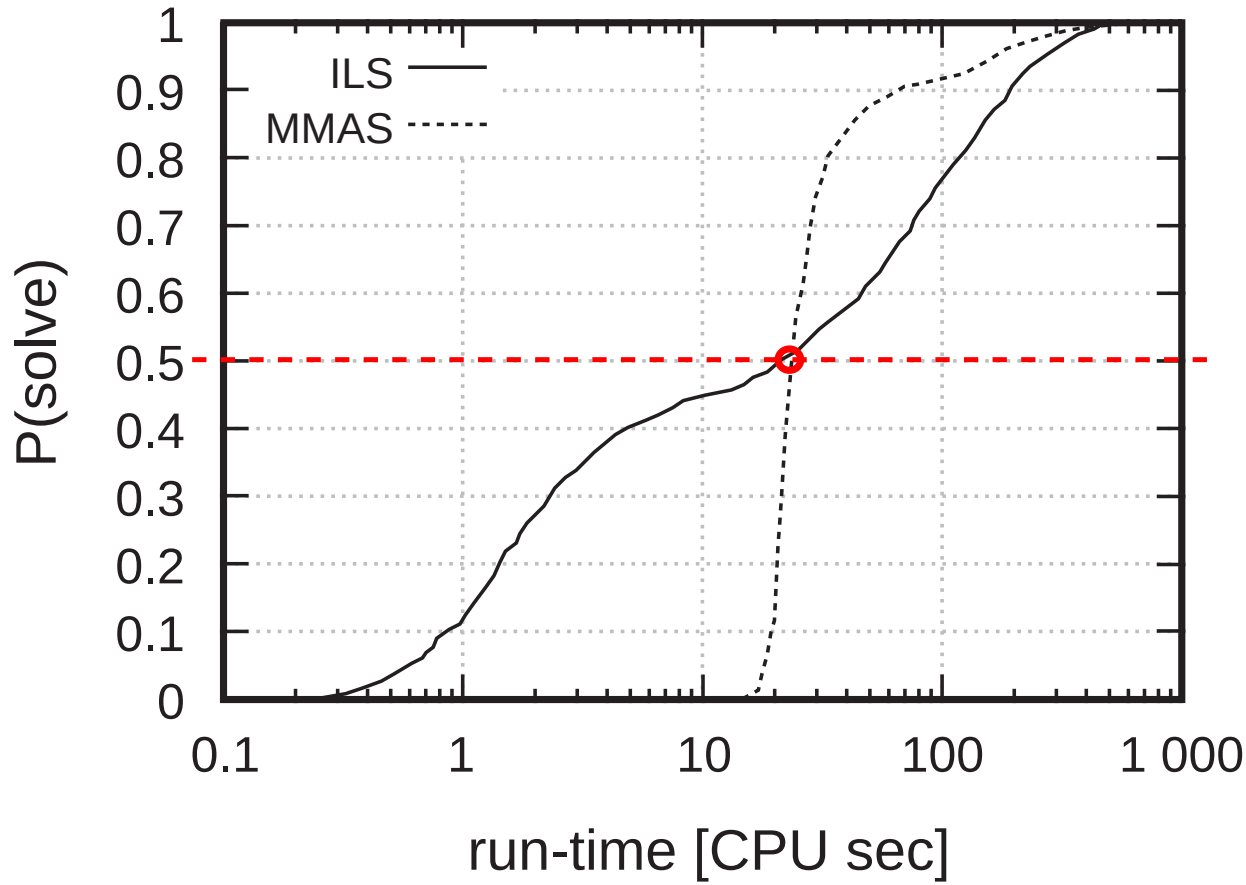
'ed[18]' is the CDF of an exponential distribution with median 18; the arrows mark the optimal cutoff-time for static restarting.

Performance differences detectable by the Mann-Whitney U-test for various sample sizes (sign. level 0.05, power 0.95):

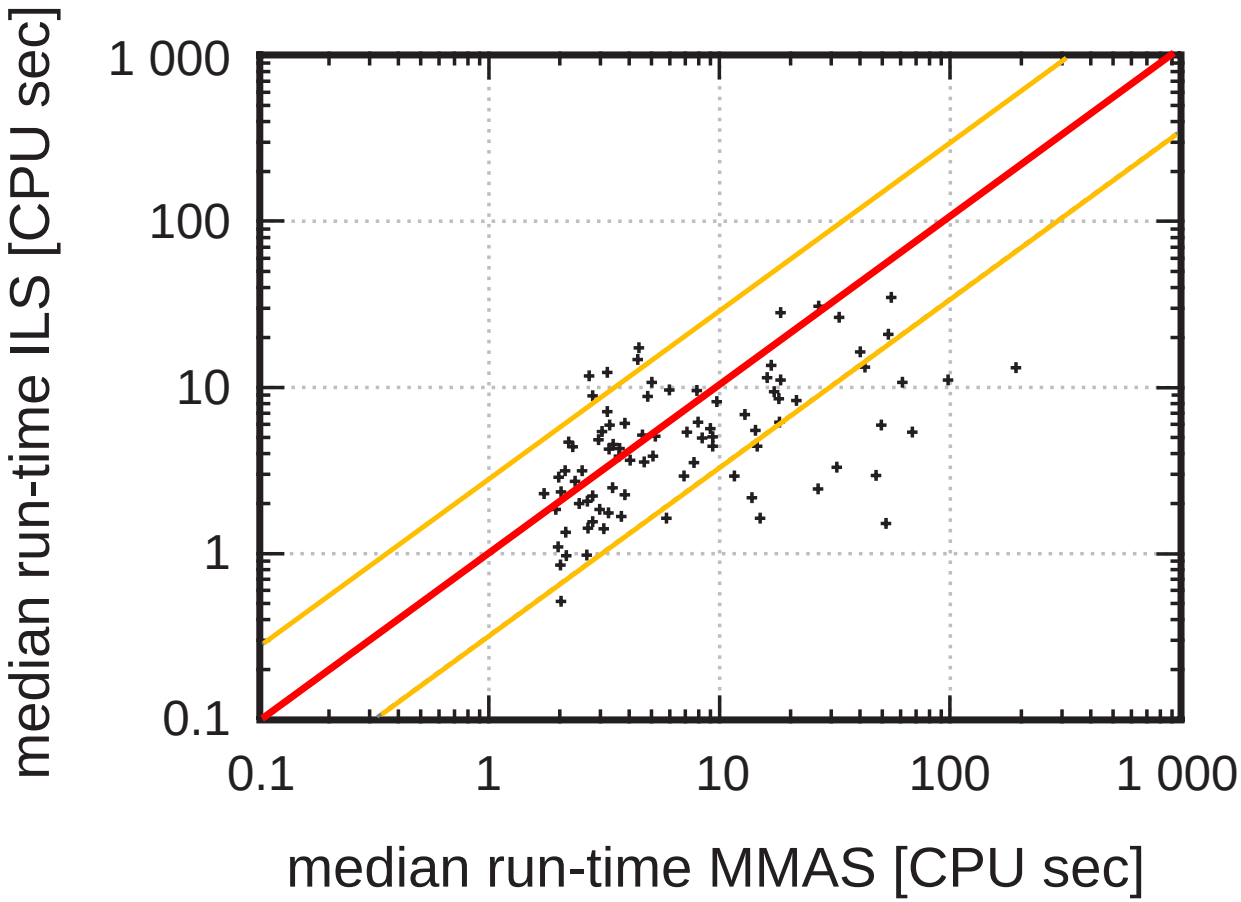
sample size	m_1/m_2
3 010	1.1
1 000	1.18
122	1.5
100	1.6
32	2
10	3

m_1/m_2 is the ratio between the medians of the two empirical distributions.

Example of crossing RTDs for two SLS algorithms for the TSP applied to a standard benchmark instance (1000 runs/RTD):



Correlation between median run-time for two SLS algorithms for the TSP over a set of 100 randomly generated instances:



10 runs per instance.