ADVANCES IN ALIGNMENTS

B.F. van Dongen

ALIGNMENTS: CHALLENGES

- Various model classes require/allow for model specific techniques
  - Petrinets, Reset/Inhibitor nets, Process trees, Colored Petri nets,
  - Petri nets with IDs?
  - Mixed paradigm models?

- Theoretical Bounds are unclear

- No proper heuristic solutions exist
  - Online approaches do not have a global view,
  - Decomposition approaches do not guarantee to return alignments.

THEORETICAL PROPERTIES

- Finding an alignment for a given model `M` and trace `t` for cost function `c`:
  - Identify the (a) cheapest firing sequence from m_i to m_f in a synchronous product model `S`, with a cost function c'

- For Reset/Inhibitor nets this problem is *undecidable*, (if $c'=0$ for all transitions then this is the same as asking: is m_f reachable from m_i)
- For Petri nets this problem is EXPSPACE hard,
- For 1-safe nets, this problem is PSPACE hard,
- For Free-choice Petri nets this problem is NP hard,
- For marked graphs, this problem is polynomial…

PRACTICAL SOLUTIONS

- Assume m_f is reachable from m_i in the synchronous product model S, (lazy soundness)
- Use a shortest-path algorithm building the search space from m_i until m_f is reached, never visiting the same state twice
- Since we know m_f is reachable from m_i, we know that the search space is bounded (decidability is guaranteed), but worst case enormous
- Use an estimator to *underestimate* the remaining cost from m_c to m_f

ESTIMATING: THE MARKING EQUATION

- For any given Petri net N, with initial marking m_i, firing sequence s leading to final marking m_f, holds: $\mathbf{m_i} + \mathbb{A} \cdot \mathbf{x} = \mathbf{m_f}$ (marking equation)

  with $\mathbf{x}$ the parikh vector of s

- Minimizing $c(\mathbf{x})$ leads to an underestimate for c(s), and hence an underestimate for the cheapest path from m_i to m_f. (again an NP-hard problem)

- x does, in no way, correspond to a realizable firing sequence, but

- For *free-choice nets*: if $\mathbf{x}$ is an integer solution to the marking equation for $\mathbf{m_i}$ and $\mathbf{m_f}$, then there exists a firing sequence s from m_i to m_f.

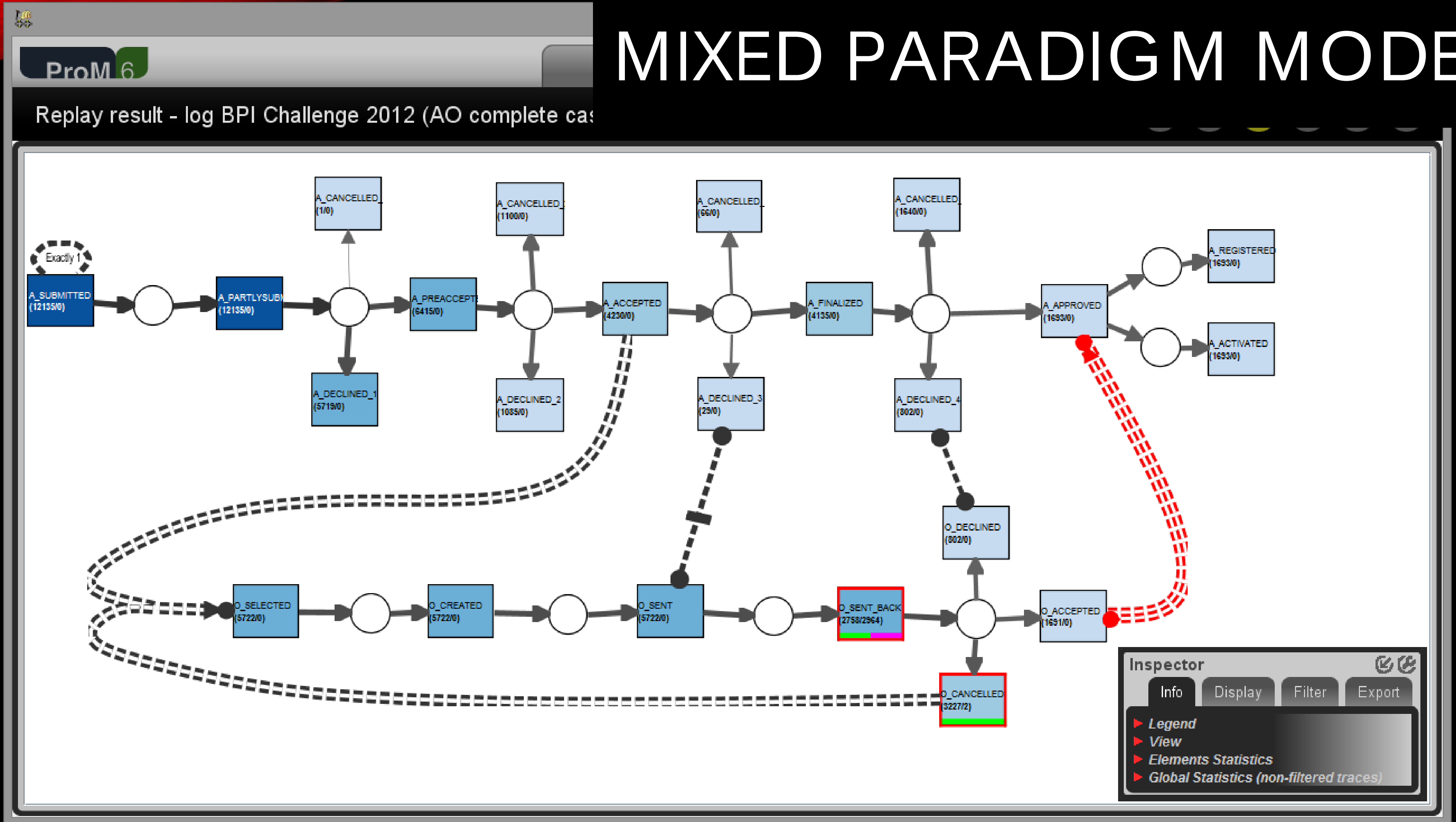
PROPERTIES OF MODELS

- Process trees are 1-safe, free-choice Petri nets with nice structural properties, so finding alignments on them is easier, but still NP hard.

- Restricting the behavior in a Petri net (with inhibitor arcs, data guards, declarative constraints, etc.) makes the marking equation a worse and worse estimator, but it remains an *underestimating* estimator

- For mixed-paradigm models, an implementation now also exists allowing to assign costs to violating declarative constraints (work with Claudio Di Ciccio)

MIXED PARADIGM MODELS



HEURISTICS FOR ALIGNMENTS

- Decomposition approaches typically give an underestimate of the costs, but no alignment (actual path) is returned,
- Online approaches provide alignments overestimating the costs, but do not guarantee termination in the model (and finishing the model is again an alignment problem),

- We propose a new technique that:
  - Keep a global view on both trace and model,
  - Provides proper alignments,
  - In predictable time (and memory),
  - Overestimating the optimal alignment costs.

AGAIN, THE MARKING EQUATION

- For any given Petri net N, with initial marking m_i, firing sequence s leading to final marking m_f, holds:

$$\mathbf{m_i} + A \cdot \mathbf{s} = \mathbf{m_f}$$

- The firing rule of Petri nets can also be translated in this form, namely, if t is enabled in marking m:

$$\mathbf{m} + A^{-} \cdot \mathbf{1}_t \geq \mathbf{0}$$

  Where A^- is the consumption matrix of the Petri net and a $\mathbf{1}_t$ vector with value 1 for transition t and 0 otherwise.

- This can be combined into a larger ILP problem for k of x prefix aligments.

K OF X PREFIX ALIGMENTS

- Consider a firing sequence: <t1,t2,t3,t4,t5,…..>

- We can construct an ILP, that guarantees that:
  - The first x transitions can be executed in that order (*using the firing rule equations*)
  - The remainder of the trace is a solution to the marking equation (*and hence for free choice nets, the final marking is reachable from the marking reached after firing the first x transitions*)
  - k out of the first x transitions correspond to an event (*log-move or sync-move*)
  - The cost function is minimal.

- Then iteratively, we keep the x transitions and iteratively solve the problem for the remaining elements of the trace.

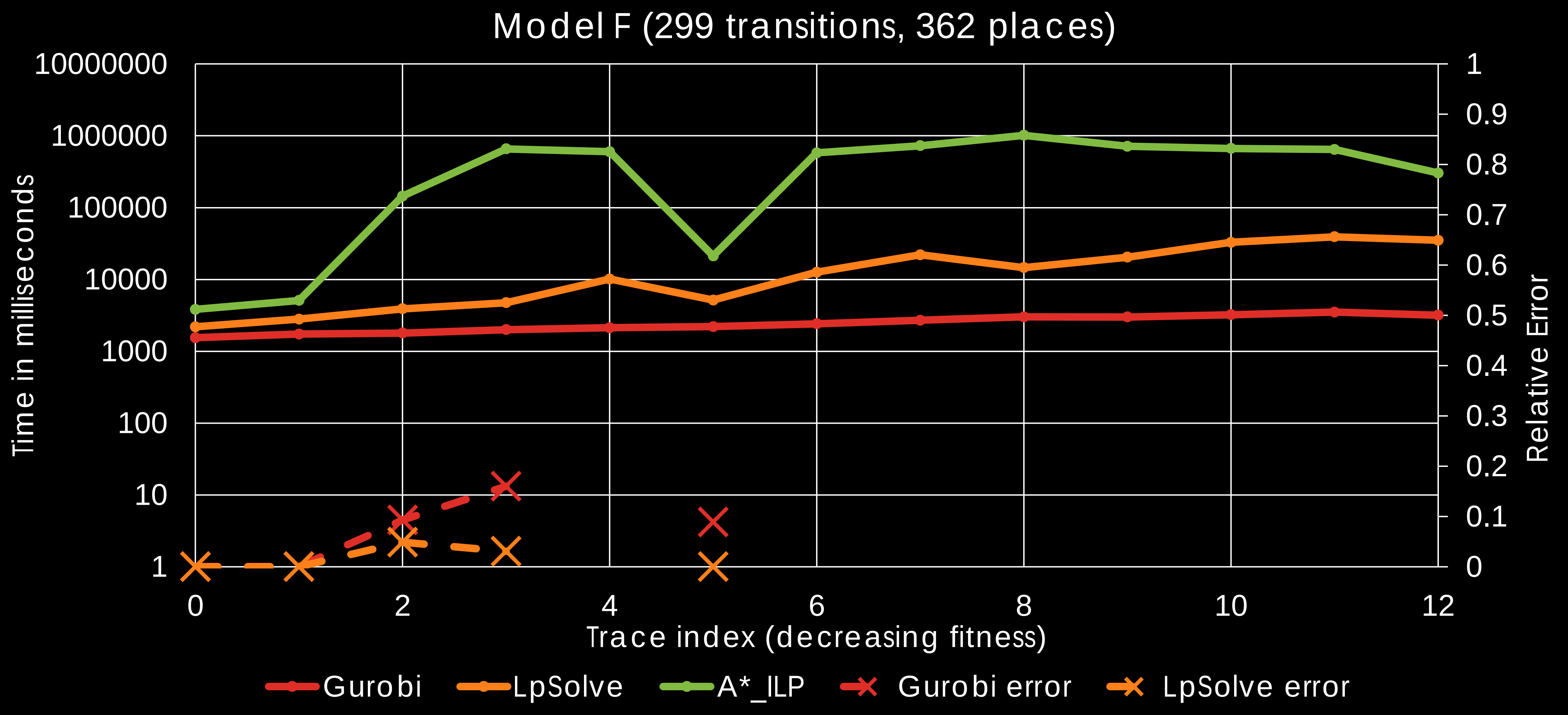
EXPERIMENTAL RESULTS

- We used A* as a baseline, and two solvers, LpSolve and Gurobi,
- We implemented backtracking in case deviation from expected costs becomes too high, and
- We used various values for x (and k).

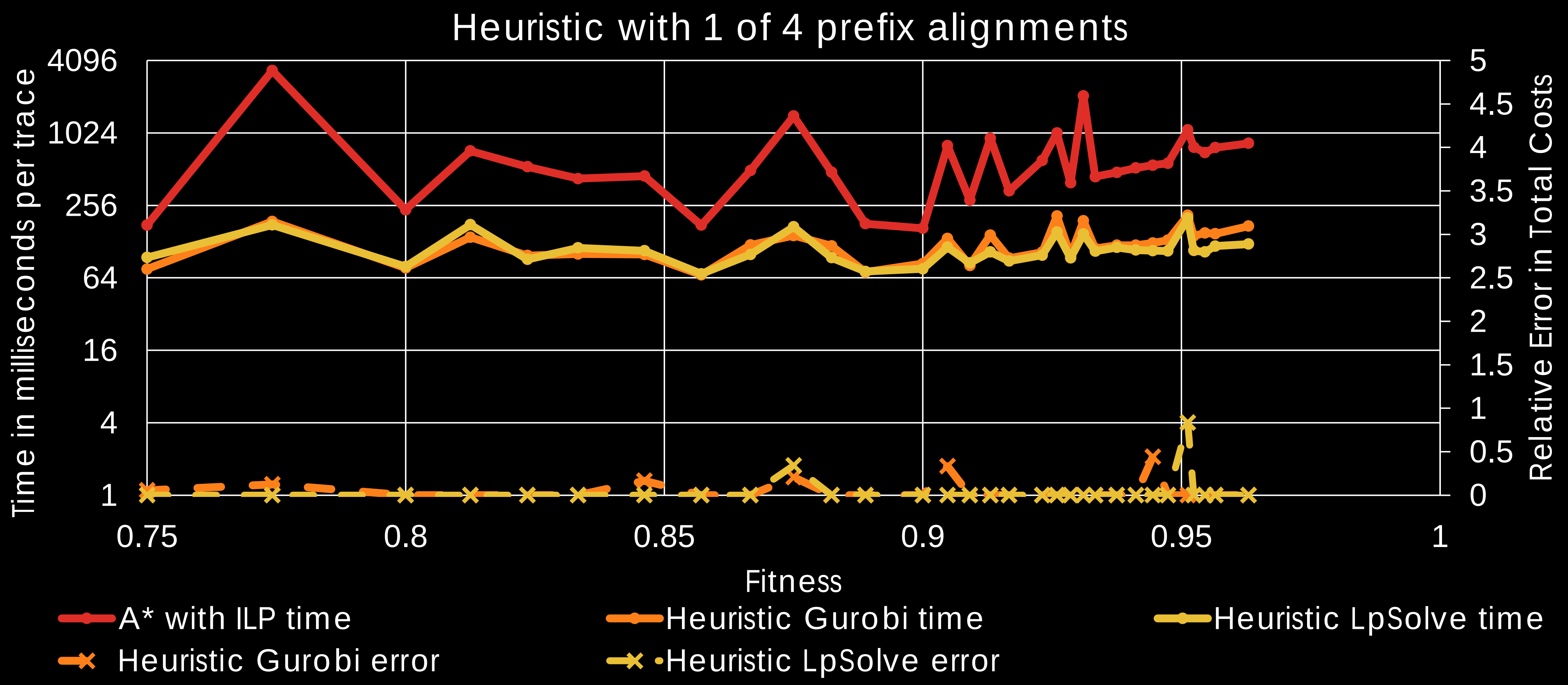
- We used both real-life logs and models and artificial ones,
- With traces on the entire fitness spectrum [0..1],

- Traces are processed one by one (to eliminate caching effects),
- All multi-threading is removed (to make the comparison fair).

ARTIFICIAL MODEL



REAL MODEL (FELIX' SEPSIS MODEL)



CONCLUSIONS

Computing Alignments is hard, but

- Current implementations are fairly efficient, but
- They occasionally fail on specific trace/model combinations, and
- Their performance is hard to predict.

So, we provide a heuristic algorithm, which

- Provides sub-optimal alignments,
- In predictable time and memory,

CHALLENGES

- Handling models with lots of tau transitions,
  - Tau transitions are hard to predict!

- In many structured loops
  - Structured loops imply transition invariants which are very bad for the matrix equations

- Handling partially ordered traces like in the regular alignments
  - Purely an implementation issue