Noname manuscript No.
(will be inserted by the editor)

Improving Swarming Using Genetic Algorithms

Etienne Renault

Abstract The verification of temporal properties against a given system may require the exploration of its full state space. In explicit model-checking this exploration uses a Depth-First-Search (DFS) and can be achieved with multiple randomized threads to increase performance.

Nonetheless the topology of the state-space and the exploration order can cap the speedup up to a certain number of threads. This paper proposes a new technique that aims to tackle this limitation by generating artificial initial states, using genetic algorithms. Threads are then launched from these states and thus explore different parts of the state space.

Our prototype implementation is 10% faster than state-of-the-art algorithms on a general benchmark and 40% on a specialized benchmark. Even if we expected a decrease of an order of magitude, these results are still encouraging since they suggest a new way to handle existing limitations. Empirically, our technique seems well suited for "linear" topology, i.e. the one we can obtain when combining model-cheking algorithms with partial order reductions techniques.

1 Introduction and Related Work

Model checking [10] aims to check whether a system satisfies a property. Given a model of the system and a property, it explores all the possible configurations of the system, i.e., the *state space*, to check the validity of the property. Typically two kinds of properties are distinguished, *safety* and *liveness* properties [25]. This paper mainly focuses on safety properties that are of special interest since they stipulate that some "bad thing" does not happen during execution. Nonetheless the adaptation of this work for checking liveness properties is detailed in Section 5.1.4.

LRDE, EPITA, Le Kremlin-Bicêtre, France

The state-space exploration techniques for debugging and proving correctness of concurrent reactive systems have proven their efficiency during the last decades [20, 27, 33, 5]. Nonetheless they suffer from the well known *state space explosion problem*, i.e., the state space can be far too large to be stored and thus explored in a reasonable time. This problem can be addressed using *symbolic* [9] or *explicit* [38] techniques even if we only consider the latter one in this paper.

Many improvements have been proposed for explicit techniques. *On-the-fly exploration* [11] computes the successors of a state only when required by the algorithm. As a consequence, if the property does not hold, only a subset of the state space is constructed. *Partial Order Reductions (POR)* [35, 30, 24] avoid the systematic exploration of the state space by exploiting the interleaving semantic of concurrent systems. *State Space Caching* [16] saves memory by "forgetting" states that have already been visited causing the exploration to possibly revisit a state several times. *Bit-state Hashing* [18] is a semi-decision procedure in which each state is associated to a hash value. When two states share the same hash value, one of this two states (and thus its successors) will be ignored.

These techniques focus on reducing the memory footprint during the state-space exploration. Combining these techniques with modern computer architectures, i.e., many-core CPUs and large RAM memories, tends to shift from a memory problem to an execution time problem which is: *How this exploration can be achieved in a reasonable time?*

To address this issue multi-threaded (as well as distributed) exploration algorithms (that can be combined with previous techniques) have been developed [19, 4, 14, 27]. Most of these techniques rely on the

swarming technique presented by Holzmann et al. [20]. In this approach, each thread runs an instance of a verification procedure but explores the state space with its own transition order.

Nowadays, best performance is obtained when combining swarming with *DFS*-based (Depth-First Search) verification[1] procedures [33, 5]. In these combinations, threads share information about states that have been *fully explored*, i.e. states where all successors have been visited by a thread. Such states are called *closed states*. These states are then avoided by other threads explorations since they can not participate in invalidating the property. These swarmed-DFS algorithms are linear but their scalability depends on two factors:

Topology problems. If the state space is linear (only one initial state, one successor per state), using more than one thread cannot achieve any speedup. This issue can be generalized to any state space that is deep but not wide.

Exploration order problems. States are tagged as *closed* following the DFS postorder of a thread. Thus, a state s can only be marked as *closed* after visiting at least N states, where N is the minimal distance between the initial state and s.

Table 1 highlights this scalability problem over the benchmark[2] used in this paper. It presents the cumulated exploration time (in a swarmed DFS fashion) for 38 models extracted from the literature. It can be observed that this algorithm achieves reasonable speedup up to 4 threads but is disappointing for 8 threads and 12 threads (the maximum we can test).

This paper proposes a novel technique that aims to keep improving the speedup as the number of threads increases and which is compatible with all memory reduction methods presented so far.

The basic idea is to use genetic algorithms to generate artificial initial states (Sections 2 and 3). Threads are then launched with their own verification procedure from these artificial states (Sections 4 and 5). We expect that these threads will explore parts of the state space that are relatively deep regarding to (many) DFS order(s). Thus, some states are tagged as *closed* without processing some path between the original initial state to these states.

Our prototype implementation (Section 6) has encouraging performances: the proposed approach runs 10% faster (with 12 threads) than state-of-the-art algorithms (with 12 threads). These results, even if they do not provide the expected gain of an order of magnitude, show that this novel approach is worth to be considered as a way to overcome existing limitations.

This paper is an extension of our work published at VECOS'18 [32] where we proposed new parallel exploration algorithms built upon the generation of artificial initial states using genetic programming. These artificial states were then used on-the-fly to generate new seeds for the various threads used during the exploration. In this approach half of the threads were spawned from artificial states while the others used the classical approach used in model-checking algorithms. To handle the verification of safety properties, this approach was adapted to avoid (1) early termination and (2) reporting false positive.

In addition to the above (common with our previous paper [32]), we investigate one variant: the number of threads using artificial states may have an impact on the performances of our algorithms (more details in Section 6). This section now integrates a full comparison with state-of-the-art algorithms as well as an evaluation for models with a linear topology. We also detail the full proof for Algorithm 6 in Section 5.1.3 (while only its sketch was given in the VECOS'18 version[32]). Section 5.1.2 provides a detailed example of our algorithm while Section 5.1.4 focuses on reporting counterexamples from our algorithms. Finally, the Section 5.2 details how our algorithms can be combined with classical algorithms for checking liveness properties.

Related Work. To our knowledge, the combination of parallel state space exploration algorithms with the generation of artificial initial states using genetic algorithms has never been done. The closest work is probably the one of Godefroid and Khurshid [15] that suggests to use genetic programming as an heuristic to help random walks to select the *best* successor to explore. The generation of other initial states has been proposed to maximize the coverage of random walks [34]: to achieve this, a bounded BFS is performed to obtain a pool of states that can be used as seed states. This approach does not help the scalability when the average number of successors is quite low (typically when mixing with POR).

In the literature there are some works that combine model checking with genetic programming but they are not related to the work presented here: Katz and Peled [22] use it to synthesize parametric programs, while all the other approaches are based on the work of Ammann et al. [1] and focus on the au-

[1] It should be noted that even if DFS-based algorithms are hard to parallelize [31] they scale better in practice than parallelized Breadth-First Search (BFS) algorithms.

[2] See Section 6 for more details about the benchmark.

	1 thread	2 threads	4 threads	8 threads	12 threads
Time in milliseconds	2 960 296	1 796 418	118 6344	981 222	978 711
Speedup	1	1.65	2.50	3.016	3.025

Table 1 Problem statement about swarmed DFS like approaches.

tomatic generation of *mutants* that can be seen as particular "tests cases".

2 Parallel State Space Exploration

Preliminaries. Concurrent reactive systems can be represented using *Transitions Systems* (TS). Such a system $T = \langle Q, \iota, \delta, V, \gamma \rangle$ is composed of a finite set of states Q, an initial state $\iota \in Q$, a transition relation $\delta \subseteq Q \times Q$, a finite set of integer variables V and $\gamma : Q \to \mathbb{N}^{|V|}$ a function that associates with each state a unique assignment of all variables in V. For a state $s \in Q$, we denote by $\texttt{post}(s) = \{d \in Q \mid (s,d) \in \delta\}$ the set of its direct successors. A *path* of length $n \geq 1$ between two states $q, q' \in Q$ is a finite sequence of transitions $\rho = (s_1, d_1) \dots (s_n, d_n)$ with $s_1 = q$, $d_n = q'$, and $\forall i \in \{1, \dots, n-1\}$, $d_i = s_{i+1}$. A state q is *reachable* if there exists a path from the initial state ι to q.

Swarming. Checking temporal properties involves the exploration of (all or some part of) the state space of the system. Nowadays, best performance is obtained by combining on-the-fly exploration with parallel DFS reachability algorithms [33, 5], such as Algorithm 1.

This algorithm is presented recursively for the sake of clarity. Lines 4 and 5 represent the main procedure: `ParDFS` takes two parameters, the transition system and the number n of threads to use for the exploration. Line 5 only launches n instances of the procedure `DFS`. This last procedure takes three parameters, s the state to process, *tid* the current thread number and a *color* used to tag new visited states. Procedure `DFS` represents the core of the exploration. This exploration relies on a shared hashmap *visited* (defined line 2)[3] that stores all states discovered so far by all threads and associates each state with a color (line 1):

- OPEN indicates that the state (or some of its successors) is currently processed by (at least) a thread,
- CLOSED indicates that the state and all its successors (direct or not) have been visited by some thread.

The `DFS` function starts (lines 7 to 8) by checking if the parameter s has already been inserted, by this thread or another one, in the *visited* map (line 7). If not, the state is inserted with the color OPEN (line 7). Otherwise, if s has already been inserted we have to check whether this state has been tagged CLOSED. In this case, s and all its successors have been visited: there is no need to revisit them. Line 10 grabs all the successors of the state s that are then shuffled to implement the swarming. Finally lines 11 to 15 perform the recursive DFS: for each successor s' of the current state, if s' has not been tagged CLOSED a recursive call is launched. When all successors have been visited, s can be marked CLOSED.

One can note that a shared Boolean *stop* is used in order to stop all threads as soon as a thread closes the initial state. This Boolean is useless for this algorithm since, when the first thread ends, all reachable states are tagged CLOSED and every thread is forced to backtrack. Nonetheless this Boolean will be useful later (see Section 4). Moreover the *visited* map is thread safe (and lock-free) so that it does not degrade performances of the algorithm (see [23, 3] for details about the implementation of parallel hash tables).

Problem statement. The previous algorithm (or some adaptations of it [33, 5]) obtains the best performance for explicit model checking. Nonetheless this swarmed algorithm suffers from a scalability problem. Figure 1 describes a case where augmenting the number of threads

Algorithm 1: Parallel DFS Exploration.

```
1  enum color = { OPEN, CLOSED }
2  visited: hashmap of (Q, color)        // Shared variable
3  stop ← ⊥                              // Shared variable
4  Procedure ParDFS(⟨Q, ι, δ, V, γ⟩ : TS, n : Integer)
5      DFS(ι, 1, OPEN) || ... || DFS(ι, n, OPEN)
6  Procedure DFS(s ∈ Q, tid : Integer, status : color)
7      if s ∉ visited then visited.add(s, status)
8      else if visited[s] = CLOSED then return
9      // Shuffle successors using tid as seed
10     todo ← shuffle(post(s), tid)
11     while (¬stop ∧ ¬todo.isempty()) do
12         s' ← todo.pick()
13         if s' is in the current recursive DFS stack then
               continue
14         if (s' ∉ visited) ∨ visited[s'] ≠ CLOSED) then
15             DFS(s', tid, status)
16     visited[s] ← CLOSED
17     if (s = ι) then stop ← ⊤
```

[3] Notice that we use the C++ convention where *visited.get*(*s*) is written *visited*[*s*].

will not bring any speedup[4]. This figure describes a transition system that is linear. The dotted transitions represent long paths of transitions. In this example, state x cannot be tagged CLOSED before state y and all the states between x and y have been tagged CLOSED. The problem here is that all threads start from state s. Since threads have similar throughput they will discover x and y approximately at the same time. Thus they cannot benefit from the information computed by the other threads. This example is pathological but can be generalized to any state space that is deep and narrow.

Suppose now that there are 2 threads and that the distance between s and x is the same as the distance between x and y. The only way to obtain the maximum speedup is to launch one thread with a DFS starting from s and launch the other thread from x. In this case, when the first thread reaches state x, x has just been tagged CLOSED: the first thread can backtrack and stop.

Fig. 1 Using more than one thread for the exploration is useless.

A similiar problem arises when performing on-the-fly model checking since (1) there is only one initial state and (2) all states are generated during the exploration. Thus a thread cannot be launched from a particular state. Moreover, the system's topology is only known after the exploration: we need a technique that works for any kind of topology.

The idea developed in this paper is the automatic generation of state x using genetic algorithms. The generation of *the perfect state* (the state x in the example) is a utopia. Nonetheless if we can generate a state relatively deep regarding to many DFS orders, we hope to avoid redundant work between threads, and thus maximize the information shared between threads. In practice we may generate states that do not belong to the state space, but Section 6 shows that more than 84% of generated states belongs to it.

3 Generation of Artificial Initial State

Genetic algorithms. For many applications the computation of an optimal solution is impossible since the set of all possible solutions is too large to be explored. To address this problem, Holland [17] proposed a new kind of algorithms (now called genetic algorithms) that are inspired by the process of natural selection. These algorithms are often considered as optimizers and used to generate high-quality solutions to search problems. Basically, genetic algorithms start by a population of candidate solutions and improve it using bio-inspired operators:

- *Crossover*: selects multiple elements in the population (the parents) and produces a child solution from them.
- *Mutation*: selects one element in the population and alters it slightly.

Applying and combining these operators produces a new generation that can be evaluated using a *fitness* function. This fitness function allows to select the best elements (w.r.t the considered problem) of this new population. These best elements constitute a new population on which mutation and crossover operations can be re-applied. This process is repeated until some satisfying solution is found (or until a maximal number of generations has been reached).

Genetic algorithms rely on a representation of solutions that is chromosome-like. In the definition of a transition system we observe that every state can be seen as a tuple of integer variables using the γ function. Each variable can be considered as a gene and the set of variables can be considered as a chromosome composed of 0 and 1. For instance, if a state is composed of two variables $a = 42$ and $b = 51$ the resulting chromosome (considering 8 bits integers) would be the one described in Figure 2.

a	b
00101010	00110011

Fig. 2 Chromosome representation.

Crossover. Concurrent reactive systems are generally composed of a set of N_p processes and a set of shared variables (or channels). Given a transition system $T = \langle Q, \iota, \delta, V, \gamma \rangle$ we can define $E : V \to [0, N_p[$, such that if v is a shared variable, $E(v)$ returns 0 and otherwise $E(v)$ returns the identifier of the process where the variable v is defined.

Algorithm 2 defines the crossover operation we use. This algorithm takes a parameter S which represents the population to use for generating a new state. Line 2 instantiates a new state s that will hold the result of the crossover operation. Lines 3 to 5 set up the values of the shared variables of s: for each shared variable v, an element of S is randomly selected to be the parent. Then, at line 5, one can observe that $\gamma(s)[v]$ (the value of v in s) is set according to $\gamma(parent)[v]$

[4] This particular case will certainly degrade performance due to contention over the shared hashmap.

	Process 1	
	a	b
$parent1$	00000000	00000000
$parent2$	11111111	11111111
`Crossover`(S)	00000000	11111111

Fig. 3 Possible Crossover.

(the value of v in the *parent*). Lines 6 to 9 perform a similar operation on all the remaining variables.

These variables are treated by batch, i.e., all the variables that belong to a same process are filled using only one parent (line 7). This choice implies that in our `Crossover` algorithm the local variable of a process cannot have two different parents: this particular processing helps to exploit the concurrency of underlying system. A possible result of this algorithm is represented in Figure 3 (with 8 bits integer variables, only one process, no shared variables, $S = \{parent1, parent2\}$ and *child* the state computed by `Crossover`(S)).

Algorithm 2: Crossover.

```
1  Procedure Crossover(S ⊆ Q)
2      s ← newState()
3      for v ∈ V s.t. E(v) = 0 do
4          parent ← pick random one of S
5          γ(s)[v] ← γ(parent)[v]
6      for i ∈ [0, N_p[ do
7          parent ← pick random one of S
8          for v ∈ V s.t. E(v) = i do
9              γ(s)[v] ← γ(parent)[v]
10     return s
```

Algorithm 3: Mutation.

```
1  Procedure Mutation(s ∈ Q)
2      for v ∈ V do
3          r ← random(0..1)
4          if r > THRESHOLD then
5              γ(s)[v] = random_flip_one_bit_in(γ(s)[v])
6              γ(s)[v] = bound_project(γ(s)[v])
```

Mutations. The other bio-inspired operator simulates alterations that could happen while genes are combined over multiple generations. In genetic algorithms, these mutations are performed by switching the value of a bit inside of a gene. Here, all the variables of the system are considered as genes.

Algorithm 3 describes this mutation. For each variable in the state s (line 2), a random number is generated. A mutation is then performed only if this number is above a fixed threshold (line 4): this restriction

	Process 1	
	a	b
s	00000100	00001000
`Mutation`(s)	00000101	00001000

Fig. 4 Possible Mutation.

limits the number of mutations that can occur in a chromosome. We can then select randomly a bit in the current variable v and flip it (line 5). Finally, line 6 exploits the information we may have about the system by restricting the mutated variable to its bounds.

Indeed, even if all variables are considered as integer variables there are many cases where the bounds are known a priori: for instance Boolean, enumeration types, characters, and so on are represented as integers but the set of values they can take is relatively small regarding the possible values of an integer. A possible result of this algorithm is represented in Figure 4 (with 8 bits integer variables and only two character variables, i.e., that have values between [0..255]).

Fitness. As mentioned earlier, every new population must be restricted to the only elements that help to obtain a better solution. Here we want to generate states that are (1) reachable and (2) deep with respect to many DFS orders. These criteria help the swarming technique by exploring parts of the state space before another thread (starting from the real initial state) reaches them.

We face here a problem that is: for a given state it is hard to decide whether it is a *good* candidate without exploring all reachable states. For checking *deadlocks* (i.e., states without successors) Godefroid and Khurshid [15] proposed a fitness function that will only retain states with few transitions enabled[5].

Since we have different objectives a new fitness function must be defined. In order to maximize the chances to generate a reachable state, we compute the average outgoing transitions (T_{avg}) of all the states that belong to the initial population. Then the fitness function uses this value as a threshold to detect *good* states. Many fitness function can be considered:

- **equality**: the number of successors of a *good* state is exactly equal to T_{avg}. The motivation for this fitness function is that if there are $N_p > 1$ independent processes that are deterministic then at every time, any process can progress. In this strategy, we consider that a good state has exactly N_p (equal to T_{avg}) outgoing transitions.

[5] Godefroid and Khurshid [15] do not generate states but finite paths and their fitness functions analyzes the whole paths to keep only those with few enabled transitions.

– **lessthan**: the number of successors of a *good* state is less than T_{avg}. The motivation for this fitness function is that if there are $N_p > 1$ independent deterministic processes that communicate then at any time each process can progress or two (or more) processes can be synchronized. This latter case will reduce the number of outgoing transitions.
– **greaterthan**: the number of successors of a *good* state is greater than T_{avg}. The motivation for this fitness function is that if there are $N_p > 1$ independent and non-deterministic processes then at any time each process can perform the same amount of actions or more.

Algorithm 4: The generation of new states.

```
1 Procedure Generate(S ⊆ Q)
2   for i ← 0 to NB_GENERATION do
3     S' ← ∅
4     for j ← 0 to POP_SIZE do
5       s ← Crossover(S)
6       Mutation(s)
7       if Fitness(s) then S' ← S' ∪ {s}
8     S ← S'
9   return S
```

Generation of artificial state. Algorithm 4 presents the genetic algorithm used to generate artificial initial states using the previously defined functions.

The only parameter of this algorithm is the initial population S we want to mutate: S is obtained by performing a swarmed bounded DFS and keeping trace of all encountered states. From the initial population S, a new generation can be generated (lines 4 to 8). At any time the next generation is stored in S' (lines 7 and 3). The algorithm stops after NB_GENERATION generations (line 2). Note that this algorithm can report an empty set according to the fitness function used.

4 State-Space Exploration with Genetic Algorithm

This section explains how Algorithm 1 can be adapted to exploit the generation of artificial initial states mentioned in the previous section. Algorithm 5 describes this parallel state-space exploration using genetic algorithm. The basic idea is to have a *collaborative portfolio* approach in which threads will share information about CLOSED states. In this strategy, half of the available threads runs the `DFS` algorithm presented in Section 2, while the other threads perform genetic exploration. This exploration is achieved by three steps:

1. Perform swarmed bounded depth-first search exploration that stores into a set $\mathcal{P}$ all encountered states (line 7). This exploration is *swarmed*, so that each thread has a different initial population $\mathcal{P}$. (Our *bounded*-DFS differs from the literature since it refers to DFS that stops after visiting N states.).
2. Apply Algorithm 4 on $\mathcal{P}$ to obtain a new population $\mathcal{P}'$ of artificial initial states (line 8).
3. Apply the `DFS` algorithm for each element of $\mathcal{P}'$ (lines 9 to 11). When the population $\mathcal{P}'$ is empty, just restart the thread with the initial state ι (see line 12).

One can note (line 1) that the *color* enumeration has been augmented with OPEN_GP. This new status may seem useless for now but allows to distinguish states that have been discovered by the genetic algorithm from those discovered by the traditional algorithm. In this algorithm OPEN_GP acts and means exactly the same as OPEN but: (1) this status is useful for the sketch of termination proof below and, (2) the next section shows how we can exploit similar information.

Algorithm 5: Parallel DFS Exploration using Genetic Algorithm.

```
1  enum color = { OPEN, OPEN_GP, CLOSED }
2  visited: hashmap of (Q, color)
3  stop ← ⊥
4  Procedure ParDFS_GP(⟨Q, ι, δ, V, γ⟩ : TS, n : Integer)
5    DFS(ι, 1, OPEN) || ... || DFS(ι, ⌊n/2⌋, OPEN) || DFS_GP(ι, ⌊n/2⌋+
       1) || ... || DFS_GP(ι, n)
6  Procedure DFS_GP(ι ∈ Q, tid : Integer)
7    P ← Bounded_DFS(ι, tid)     // Swarmed exploration
       using tid as a seed
8    P' ← Generate(P)            // Described Algorithm 4
9    while P' not empty ∧ ¬stop do
10     s ← pick one of P'
11     DFS(s, tid, OPEN_GP)
12   if ¬stop then DFS(ι, tid, OPEN)
```

Termination. Until now we have avoided mentioning one problem: there is no reason that a generated state is a reachable state. Nonetheless even if the state is not reachable, some of its successors (direct or not) may be reachable. Since the number of unreachable states is generally much larger than the number of reachable states, we have to ensure that Algorithm 5 terminates as soon as all reachable states have been explored.

First of all let us consider only threads running the `DFS` algorithm. Since this algorithm has already

been proven (see. [33] for more details), only the intuition is given here. When all the successors of an OPEN state have been visited, this state is tagged as CLOSED. Since all CLOSED states are ignored during the exploration, each thread will restrict parts of the reachable state space. At some point all the states will be CLOSED: even if a thread is still performing its `DFS` procedure, all the successors of its current state will be marked CLOSED. Thus the thread will be forced to backtrack and stop.

The problem we may have with using genetic algorithm is that all the threads performing the genetic algorithm may be running while all the other ones are idle since all the reachable states have already been visited. In this case, a running thread can see only unreachable states, i.e. OPEN_GP, or CLOSED ones. To handle this problem, a Boolean *stop* is shared among all threads (line 2). When this Boolean is set to $\top$ all threads stop regardless the exploration technique used (line 11, Algorithm 1). We observe line 9 that the use of other artificial states is also stopped, and no restart will be performed (line 12). This Boolean is set to $\top$ only when all the successors of the real initial state have been explored (line 17, Algorithm 1). Thus, one can note that even if a thread using the genetic algorithm visits first all reachable states it will stop all the other threads.

5 Checking Temporal Properties

5.1 Checking Safety Properties

5.1.1 The deadlock detection algorithm

Safety properties cover a wide range of properties: *deadlock freedom* (there is no state without successors), *mutual exclusion* (two processes execute some critical section at the same time), *partial correctness* (the execution satisfies a precondition but invalidates a postcondition), *etc.* One interesting characteristic of safety properties is that they can be checked using a reachability analysis (as described in Section 2). Nonetheless, our genetic reachability algorithm (Algorithm 5) cannot be directly used to check safety properties. Indeed, if a thread (using genetic programming) reports an error we do not know if this error actually belongs to the state space.

Algorithm 6 describes how to adapt Algorithm 5 to check safety properties. To simplify things we focus on checking deadlock freedom, but our approach can be generalized to any safety property. This algorithm[6] relies on both Algorithms 1 and 5. The basic idea is still to launch half of the threads from the initial state ι and the remaining ones from some artificial initial state (line 10).

- For a thread performing reachability with the genetic algorithm the differences are quite few. When a deadlock state is detected (line 36) we just tag this state as DEADLOCK_GP rather than CLOSED. This new status is used to mark all states leading to a deadlock state. Indeed since we do not know if the state is a reachable one we cannot report immediately that a deadlock has been found. Moreover we cannot mark this state CLOSED otherwise a counterexample could be lost. This new status helps to solve the problem: when such a state is detected to be reachable, a deadlock is immediately reported. The other modifications are lines 29 and 31: when backtracking, if a deadlock has been found no more states will be explored.
- For a thread performing reachability without genetic algorithm the differences are also quite few. Lines 24 to 27 only check if the next state to process has been marked DEADLOCK_GP. In this case this state is a reachable one and it leads to a deadlock state. We can then report that a deadlock has been found and stop all the other threads. A deadlock can also be reported directly (line 32), if the current state is a deadlock.

5.1.2 Detailed example.

Algorithm 6 is depicted step by step in Figure 5. For this example, we consider that lines 44 and 46 (computing the artificial initial states) have already been realized. Let us consider two threads: t_1 performing a *classical approach* and t_2 performing a *genetic programming* approach, i.e. $n = 2$, t_1 has been launched line 7, and t_2 line 9. Step (1) represents the state space (a, b, c, d, e, f) as well as some states that do not belong to it (g, h).

Step (2) represents t_1 starting from the initial state and exploring states a, b, c, and d. Step (3) represents t_2 starting from the artificial initial g state and exploring state h. In step (4), t_2 detects that all its successors are on its recursive stack. As a consequence, the state h will be marked CLOSED, line 39 of Algorithm 6. The thread t_2 can then backtrack state h and continue to explore the next successors of g, i.e. f.

Step (5) represents two concurrent actions. First, t_2 has explored states f and e. Second, t_1 has detected that all successors of d are on its recursive stack and then d has been marked as CLOSED. Then t_1 discovered that all the successors of c have been explored,

[6] Main differences have been highlighted to help the reader.

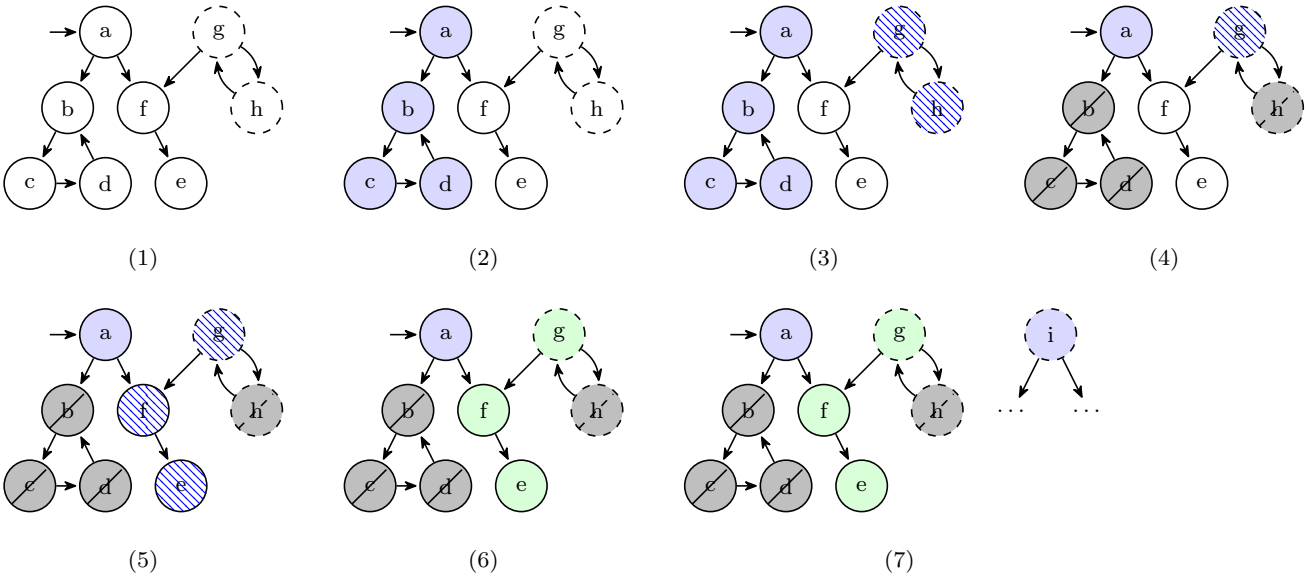

Fig. 5 Plain circles represents reachable states while dashed ones represents non reachable ones. Two threads explore this example, one from a and one from g (then from i). Blue states ○ represent OPEN states, and blue cross hatched states represent OPEN_GP ones. Forbidden signed states represent CLOSED states and green ones ○ the DEADLOCK_GP ones.

so this state can also be marked CLOSED line 39. The same operations are then applied to b which is then marked CLOSED. t_1 can then backtrack b and explores the remaining successors of a.

During step (6), thread t_2 discovers that state e is effectively a deadlock (line 37). State e is then tagged DEADLOCK_GP. When this state is backtracked (after the recursive call line 28), state f detects that its only successor can reach a deadlock state. As a consequence, state f will be immediately marked DEADLOCK_GP (line 30), and backtracked (line 31). For the same reasons, state g will also be tagged DEADLOCK_GP, and backtracked. Notice that lines 40–41 prevent stopping all the threads when backtracking state g.

From now, as soon as t_1 will discover state f it can report a counterexample, i.e. a deadlock has been detected. Indeed, lines 21–27 of Algorithm 6 detect such a situation. This situation can be described as follows: "a *classical thread* detects a state that can lead to a deadlock **but** was discovered by a *gp thread*". In this case, we can claim that there exists a path from the initial state to a deadlock. One should note that t_1 reports a deadlock without seeing f and e. When the deadlock is detected, all the other threads are stopped.

Finally step (7) depicts lines 46–48 of the algorithm. Thread t_2 finished its exploration from state g and picked another artificial state (here state i).

Discussion. The observer reader may notice three relevant informations:

1. Here, the *gp thread* starts two explorations, one from g and one from i. Both of these states have been generated and both of these states do not belong to the state space. There is no obligation for these states to be outside of the state space. If they belong to the state space, the algorithm works perfectly the same, without this information.
2. Suppose that in step (3), the algorithm choses state f rather than state h. In this case, a deadlock will be found, and all states are backtracked. Doing that will prevent the exploration of state h. We opted for this strategy in order to propagate as soon as possible the information about deadlock detection. Nonetheless, our algorithm is easily adaptable to force the exploration of remaining successors.
3. Our algorithm does not exploit the fact that two *gp threads* cooperate. Indeed, a first thread can detect a deadlock and backtracks. When the second *gp thread* discovers a DEADLOCK_GP state it can immediately backtrack while our current algorithm forces the exploration until the deadlock is re-discovered. Experiments (not presented in this paper) show that this optimization bring no benefits in practice.

5.1.3 Proof of the algorithm

This subsection details the proof that Algorithm 6 will report a deadlock if and only if there exists a reachable state that has no successors. To prove this algorithm, two theorems must be verified:

Algorithm 6: Parallel Deadlock Detection Using Genetic Algorithm.

```
1  enum color =
2        { OPEN, OPEN_GP, CLOSED , DEADLOCK_GP }
3  visited: hashmap of (Q, color)
4  stop ← ⊥
5  deadlock ← ⊥
6  Procedure
     ParDeadlockGP(⟨Q, ι, δ, V, γ⟩ : TS, n : Integer)
7        DeadlockDFS(ι, 1, OPEN) || ... ||
8        DeadlockDFS(ι, ⌊n/2⌋, OPEN) ||
9        DeadlockDFS_GP(ι, ⌊n/2⌋ + 1) || ... ||
10       DeadlockDFS_GP(ι, n)

11 Procedure
     DeadlockDFS(s ∈ Q, tid : Integer, status : color)
12   if s ∉ visited then
13     visited.add(s, status)
14   else if visited[s] = CLOSED then
15     return
16   todo ← shuffle(post(s), tid)
17   while (¬stop ∧ ¬todo.isempty()) do
18     s' ← todo.pick()
19     if s' is in the current recursive DFS stack then
20       continue
21     if (s' ∉ visited ∨ visited[s'] ≠ CLOSED) then
22       b ← (s' ∈ visited ∧ visited[s'] = DEADLOCK_GP
23             ∧ status = OPEN)
24       if b then
25         deadlock ← ⊤
26         stop ← ⊤
27         break
28       DeadlockDFS(s', tid, status)
29       if visited[s'] = DEADLOCK_GP ∧ status = OPEN_GP
           then
30         visited[s] ← DEADLOCK_GP
31         return
32   if post(s) = ∅ ∧ status = OPEN then
33     deadlock ← ⊤
34     stop ← ⊤
35     return
36   if post(s) = ∅ ∧ status = OPEN_GP then
37     visited[s] ← DEADLOCK_GP
38   else
39     v[s] ← CLOSED
40   if (s = ι) then
41     stop ← ⊤

42 Procedure DeadlockDFS_GP(ι ∈ Q, tid : Integer)
43   // Also check deadlock during this DFS
44   P ← Bounded_DFS(ι, tid)
45   P' ← Generate(P)
46   while P' not empty ∧ ¬stop do
47     s ← pick one of P'
48     DeadlockDFS(s, tid, OPEN_GP)
49   if ¬stop then
50     DeadlockDFS(ι, tid, OPEN)
```

Theorem 1. For all systems S, the algorithm terminates.

Theorem 2. A thread reports a deadlock iff $\exists s \in Q$, $post(s) = \emptyset$.

To simplify this proof, we denote by *classical thread* a thread that does not perform genetic algorithm while the other threads are called *gp threads*. The following invariants hold for all lines of Algorithm 6:

Invariant 1. If *stop* is $\top$ then no new state will be discovered.

Proof. *New states are computed line 16 but only discovered one-by-one line 18. Let us suppose that some thread sets the stop variable to $\top$ (lines 26, 34, or 40): this thread will quit the while loop line 17. Exiting the* DeadlockDFS *function will also exit the loop line 45 and exit this thread. For the other threads two situations may occur. First, the threads are backtracking from the call line 28: the next iteration will not be executed, and the threads will exit without discovering new states. If the threads are executing lines 18–27, then the call line 28 will be performed but the check line 17 will avoid new states to be discovered.*

Invariant 2. A deadlock state can only be OPEN, OPEN_GP or DEADLOCK_GP.

Proof. *From line 7–10, 47 and 49, the only statuses that can be used line 12–13 are* OPEN *and* OPEN_GP. *If a state has no successor, the condition in the loop (line 17) will not be satisfied and the thread jumps line 32. If the thread is using the classical approach, lines 32–34 are executed and a deadlock is reported (stopping the other threads). Otherwise, the state is only marked* DEADLOCK_GP *line 37.*

Invariant 3. No direct successor of a CLOSED state is a deadlock state.

Proof. *A state s is marked as* CLOSED *line 39 when all its successors have been visited lines 17 – 31. Lines 19 and 21 ensure that a recursive call is performed only on states that are (1)* OPEN *or not in the DFS stack. All the other direct successors are then explored and backtracked, and then marked* CLOSED *before s is marked* CLOSED.

Invariant 4. A state is marked CLOSED iff all its successors that are not on the thread's recursive stack are CLOSED.

Proof. *From Invariant 3, we know that all the direct successors of a state are either on the DFS stack or* CLOSED. *Since states are marked* CLOSED *in the DFS postorder line 39, all its successors that are not on the recursive stack are backtracked and then marked* CLOSED.

Invariant 5. Only *gp threads* can tag a state DEADLOCK_GP.

Proof. *Trivial. From the algorithm, the only places where the status is changed to* DEADLOCK_GP *are lines 30 and 37. For both, the previous line checks wheter the status is* OPEN_GP*. From line 9,10 and 48 only gp threads can have this status. One should note that line 50, the gp thread becomes a classical thread.*

Invariant 6. A state is DEADLOCK_GP iff it is a deadlock state or if one of its successors (direct or not) is a deadlock state.

Proof. *If a state is trivially a deadlock state, line 37 will mark it* DEADLOCK_GP *and return (line 35). Consequently, all its predecessors will be marked* DEADLOCK_GP *during the backtrack (line 30). Thus, a state can only be tagged* DEADLOCK_GP *iff one of its successors is a real deadlock.*

Invariant 7. Only *classical thread* can report that a deadlock has been found.

Proof. *A deadlock can be reported lines 25–26 or lines 33–34. Lines 33–34 can only be executed by a classical thread due to the condition line 32 (and the note of invariant 5). Lines 25–26 can also report that state s is a deadlock but the condition (status =* OPEN*) ensures that only a classical thread will report it. In this case a gp thread has discovered that some (possibly indirect) successor of s is a deadlock, without knowing that s is reachable from the initial state. The lines 25–26 detect that this state is reachable and can then report the deadlock.*

Invariant 8. If a state is reachable then all its direct successors are reachable.

Proof. *By construction, ι is the initial state then reachable. All threads starting from ι lines 7,8 and 50 will then start from a reachable state. Applying the transition relation line 16 will then only produce reachable states.*

Proof of Theorem 1. *From invariant 8, and since the system has a finite number of states, at least one thread will perform the exploration from the initial state ι. If no deadlock is found, the thread will backtrack and finally reach lines 40–41. The stop boolean will then be set to $\top$. From invariant 1. the algorithm stops. Reaching* CLOSED *states (line 14) will only prune the exploration and then have no impact on the termination for classical threads. Gp threads may only explore states that are not part of the state space. Nonetheless, invariant 1 ensures the termination for these threads when some thread will mark stop as $\top$. If a deadlock is discovered by a classical thread, invariant 1 and 4 will ensure the termination of all threads. If a gp thread detects it, invariant 7 combined to invariant 8 will force all threads to stop using lines 26 and 34.*

Proof of Theorem 2. *From invariant 2 we know that a deadlock state can never be marked* CLOSED *because (1) if discovered by a classical thread a deadlock is immediately reported, and (2) because otherwise this information must be propagated. Invariant 3, 5 and 6 ensure that the information is correctly propagated, while invariant 7 ensures that no gp thread can report that a deadlock has been found. The other direction of theorem 2 is quite evident. If a deadlock exists, then it is a reachable state. Since all reachable states are explored by classical thread, the report will be done (see invariant 7)*

5.1.4 Reporting a counterexample

In Algorithm 6, a classical thread can report the existence of a deadlock but cannot necessarily report the counterexample forming it. Indeed, the path from the initial state to the deadlock may be composed of several parts, computed by one *classical thread* and multiple *gp threads*.

If a *classical thread* detects a counterexample by itself, it can directly report the counterexample formed by its recursive stack. Otherwise, reporting the counterexample can be done as follows. First of all the recursive call stack forms the prefix, starting from the initial state ι to some state α. Note that we know that this prefix starts from ι since only *classical threads* can report deadlocks.

Then, the thread must compute the path from α to one deadlock state β. To do so, the thread will start a new DFS from α, by only considering states that are tagged DEADLOCK_GP. This restriction is then repeated at each step of the DFS. Indeed, after a *gp thread* has detected a deadlock, all the states on its DFS stack are tagged DEADLOCK_GP. Following a path of DEADLOCK_GP states will necessarily result in discovering a deadlock. Since the DFS will explore all paths, the state β will be rediscovered.

Combining the prefix and the path of DEADLOCK_GP states will build the whole counterexample.

5.2 Checking Liveness Properties

Until now, we only focused on the verification of safety properties. The verification of complex temporal properties involves the exploration of an automaton which

is the result of the synchronous product between the state space of the system and the property automaton [38].

In this settings, a state is composed of two parts: the system state and the property state. Thanks to the previous sections, we know how to build artificial states for the system part.

Generating artificial initial states for the property may be irrelevant. Indeed, these automata may have a huge impact (in term of state) on the synchronized product. Moreover, we know that only states synchronized with some real state of the property automaton will report a counterexample. The generation of artificial states for the property automaton may not be relevant.

Nonetheless, genetic algorithms presented so far can then be applied by considering that the property state is a variable just like the other system's variables. For a three states property automaton, we can consider that the actual state in the property automaton depends on a variable that can have three values: 1, 2 and 3 [7].

The adaptation of the artificial state generation is then straightforward: the system part is generated as described in Section 3 while the property part is generated as described earlier.

One should note that the generation of artificial initial states for the synchronized product is not sufficient for adapting Algorithm 6 for checking liveness properties. Indeed, checking liveness properties involves the use of an emptiness check in the automata theoretic approach for explicit LTL model-checking. An emptiness check is an algorithm looking for *accepting cycles* in the synchronized product, i.e. cycles that contain products states synchronized with some designated state of the property automaton.

Traditionally, two kinds of emptiness checks are used in explicit model-checking[8]:

- NDFS-based [13]: two nested dfs are used to detect accepting cycles. A first one looks for the accepting state, while the second one looks for a cycle around it.
- SCC-based [33]: one dfs is used to compute the *Strongly Connected Components* (SCC) of the synchronized product. As soon as an SCC containing an accepting state is discovered a counterexample can be reported.

[7] Notice that mutation can be done ensuring that this variable will not be less than 1 and not be greater than 3.

[8] Here, we only describe our approach on Büchi automata, but the adaptation for generalized Büchi automata is straightforward.

The adaptation of Algorithm 6 into these emptiness check can be done as follows.

- For NDFS-based algorithms, when a *gp thread* detects an accepting cycle, all the states forming it are tagged with an ACCEPTING_CYCLE status. When a *classical thread* detects such a state, a counterexample is raised.
- For SCC-based algorithms, when a *gp thread* detects an accepting SCC, all the states forming it are tagged with an ACCEPTING_SCC status. When a *classical thread* detects such an SCC, a counterexample is raised. One should note that tagging all the states of an SCC can be done in quasi-constant time using a union-find data structure (see Anderson and Woll [2] for a lock-free implementation of this structure and see Bloemen et al. [6] for its usage for computing SCC of a graph).

Notice that in both situations, as soon as a *classical thread* detects a counterexample, this latter one can be immediately reported and all the other threads stopped.

6 Evaluation

6.1 Overall evaluation

Benchmark Description. To evaluate the performance of our algorithms, we selected 38 models from the BEEM benchmark [28] that cover all types of models described by the classification of Pelánek [29]. All the models were selected such that Algorithm 1 with one thread would take at most 40 minutes on Intel(R) Xeon(R) @ 2.00GHz with 250GB of RAM. This six-core machine is also used for the following parallel experiments[9]. All the approaches proposed here have been implemented in Spot [12]. For a given model the corresponding system is generated on-the-fly using DiVinE 2.4 patched by the LTSmin team[10]. Notice, that, the swarming is "reproducible": for a given state and for a given threads, the exploration order will always be the same.

[9] For a description of our setup, including selected models, detailed results and code, see `http://www.lrde.epita.fr/~renault/benchs/ISSE-2020/results.html`. All experiments can be replayed using either the Dockerfile available at `https://github.com/etienne-renault/swarmedgp-docker` or directly the Docker available at `https://hub.docker.com/r/akaesus/swarmedgp`. Also note that the archive of all our experiments has been published on Zenodo with (DOI): `https://doi.org/10.5281/zenodo.3707234`

[10] See http://fmt.cs.utwente.nl/tools/ltsmin/#divine for more details. Also note that we added some patches (available in the web-page) to manage out-of-bound detection.

Reachability. To evaluate the performance of the Algorithm 5 presented in Section 4 we conducted 9158 experiments, each taking 30 seconds on the average. Table 2 reports selected results to show the impact of the fitness function and the mutation threshold over the performance of Algorithm 5 with 12 threads (the maximum we can test). For each variation, we provide *nb* the number of models computed within time and memory constraints, and *Time* the cumulated wall-time for this configuration (to run the whole benchmark). For a fair-comparison, we excluded from *Time* models that cannot be processed. Table 2 also reports state-of-the-art and **random** (used to evaluate the accuracy of genetic algorithms by generating random states as seed states). This latter technique is irrelevant since it is five time slower than state-of-the-art and only processes 32 models over 38.

If we now focus on genetic algorithms, we observe that the threshold highly impacts the results regardless the fitness function used: the more the threshold grows, the more models are processed within time and memory constraints.

The table also reports the best threshold[11] for all fitness function, i.e. 0.999. It appears that **greaterthan** only processed 37 models: this fitness function does not seem to be a good one since (1) it tends to explore useless parts of the state-space and (2) the variations of the threshold highly impact the performance of the algorithm. All the other fitness functions provide similar results for a threshold fixed at 0.999. Nonetheless we do not recommend **equality** since a simple variation of the threshold (0.7) could lead to extremely poor results. Our preference goes to **lessthan** and **lessstrict** since they seem to be less sensitive to threshold variation while achieving the benchmark 9% faster than state-of-the-art algorithm. Thus, while the speedup for 12 threads was 3.02 for state-of-the-art algorithm, our algorithm achieves a speedup of 3.31.

Note that the results reported in Table 2 include the computation of the artificial initial states. On the overall benchmark, this computation takes in average slightly less than 1 second per model (30 seconds for the whole benchmark). This computation has a negligible impact on the speedup of our algorithm.

We have also evaluated (not reported here, see web-page for more details) the impact of the size of the initial population and the size of each generation over the performance. It appears that augmenting (or decreasing) these two parameters deteriorate the performance. It is worth noting that the best values of all parameters are classical values regarding to state-of-the-art genetic algorithms. Finally, for each model (and **lessthan** as fitness), we compute a set of artificial initial states and run an exploration algorithm from each of these states. It appears that 84.6% of the 7 866 005 486 generated states are reachable states.

The chart presented in Figure 6 evaluates the percentage of reachable states from a population of artificial initial states (computed with the best parameter inferred from Table 2). The results are presented model per model. For each model and for a given artificial initial state we evaluated the percentage of visited states that are reachable from the real initial state. For a model, the boxplot displays this percentage from each artificial state in the population. Models are presented sorted according to their median value.

First of all, we can observe that all almost all models have at least one artificial state where all its successors are reachable from the initial state. Moreover one third of the models have more than a half of their artificial initial states with more than 50 percent of successors that are reachable. One can observe huge variations depending on models: for instance, almost all the states are reachable from any artificial state in *resistance.2* while there are few for *blocks.3*. A fine grained study of these models reveal that models with good results fall into two categories of the classification of Pelánek [29]: Mutex and Communication-Protocol. These models appear to be composed of large SCCs or long cycles. A fine grained analysis of these kind of model reveals that these state spaces are most of the time constructed using variables that are progressively incremented. We believe that this particular scheme is well suited for our techniques since it increase the probability of generating a valid state.

Finally, this chart suggests that the function used for the generation of artificial initial state may be crucial for our algorithms but may also be dependent of the kind of model targeted. We must admit that we hopped that a generic fitness function would have work for all the models: it appears empirically that it could be of interest to build one dedicated fitness function per category of model.

Safety properties. Now that we have detected empirically the best values for the parameters of the genetic algorithm we can evaluate the performance of our deadlock detection algorithm. In order to evaluate the performance of our algorithm we conduct 418 experiments. The benchmark contains 21 models with deadlocks and 17 models without. Table 3 compares the relative performance of state-of-the-art algorithm and Algorithm 6. For this latter algorithm, we only report the two fitness functions that give the best per-

[11] We evaluate other thresholds like 0.9999 or 0.99999 but it appears that augmenting the threshold does not increase performance, see the web-page for more details.

	Threshold							
	0.7		0.8		0.9		0.999	
	nb	*Time (ms)*	*nb*	*Time (ms)*	*nb*	*Time (ms)*	*nb*	*Time (ms)*
greaterthan	35	1 041 015	35	970 248	35	1 000 184	37	900 468
equality	35	3 217 183	35	965 259	35	934 947	38	907 148
lessthan	35	972 038	35	951 767	35	928 978	38	904 776
lessstrict	35	970 668	35	983 225	35	935 319	38	894 131
	No threshold							
random	(trivial comparator to evaluate genetic algorithms)						32	5 079 869
Algorithm 1	(state-of-the-art with 12 threads)						38	978 711

Table 2 Impact of the threshold and the fitness function on Algorithm 5 with 12 threads (NB_GENERATION=3, INIT=1000, POP_SIZE=50). The time is expressed in milliseconds and is the cumulated time taken to compute the whole benchmark (38 models); *nb* is the number of instances resolved with time and memory limits.

	Algorithm 1 (state-of-the-art)		Algorithm 6			
			lessthan		**lessstrict**	
	Time (ms)	States	Time (ms)	States	Time (ms)	States
Deadlocks	2 888	7.01E^6	3 713	5.87E^6	3 414	5.47E^6
No deadlocks	516 152	5.79E^8	462 881	6.73E^8	468 683	6.82E^8

Table 3 Comparison of algorithms for deadlock detection. Each runs with 12 threads, and we report the variation of two different fitness functions: **lessstrict** and **lessthan**. Results present the cumulated time and states visited for the whole benchmark.

formance for reachability. Indeed, since Algorithm 6 is based on Algorithm 5 we reuse the best parameters to obtain the best performance. Results for detecting deadlocks are quite disappointing since our algorithm is 15% to 30% slower. A closer look to these results shows that deadlocks are detected quickly and Algorithm 6 has degraded performance due to the computation of artificial initial states. Nonetheless (and unexpected behavior of our algorithms) the experiments tends to show that when a deadlock exists our deadlock detection algorithm require 15% less memory than state-of-the-art algorithm. The runtime for detecting deadlock is nonetheless too small to infer some particular property. A detailed evaluation of the length of counterexamples in models where finding deadlock is hard could be of interest to see if this trends is verified.

On the contrary we observe that our algorithm is 10% faster (regardless whether we use **lessthan** or **lesstrict**) than the classical algorithm when the system has no deadlock. One can note that this algorithm performs better than a simple reachability algorithm. Indeed, even if the system has no deadlock: the algorithm can find non-reachable deadlock. In this case, the algorithm backtracks and the next generation is processed. This early backtracking forces the use of a new generation that will help the exploration of the reachable states. To achieve this speedup, we observe an overhead of 13% for the memory consumption. This overhead is the consequence of generating states that does not belong to the state space. The use of dedicated memory reduction techniques, such as partial order reduction technique, could help to reduce this footprint. Indeed, since POR only consider representative executions, some states will not be explored and thus will not be stored.

Variations of the number of gp threads. All the algorithms presented in this paper supposes that only half of the threads perform an exploration based on one or more artificial states. This restriction was chosen based on the experiments of Table 1 where half of the threads seems useless from the speedup point of view. Figure 7 describes possible variations on this approach. The main idea is to run Algorithm 5 with a variation on the percentage of threads using a genetic programming approach. This Figure displays three lines: *dfs* represents the state-of-the art while the other lines represent the fitness functions combined to the best parameters observed in Table 2. The first line is horizontal, since it represents the state-of-the-art, which does not use genetic programming, and thus insensitive to this variation. For the two other lines, it appears that the percentage of threads using a genetic approach has a strong impact on the results. Indeed, even with a few percentage (20%) of threads using genetic programming, we observe a 11% reduction of the total time to run the benchmark. Nonethe-

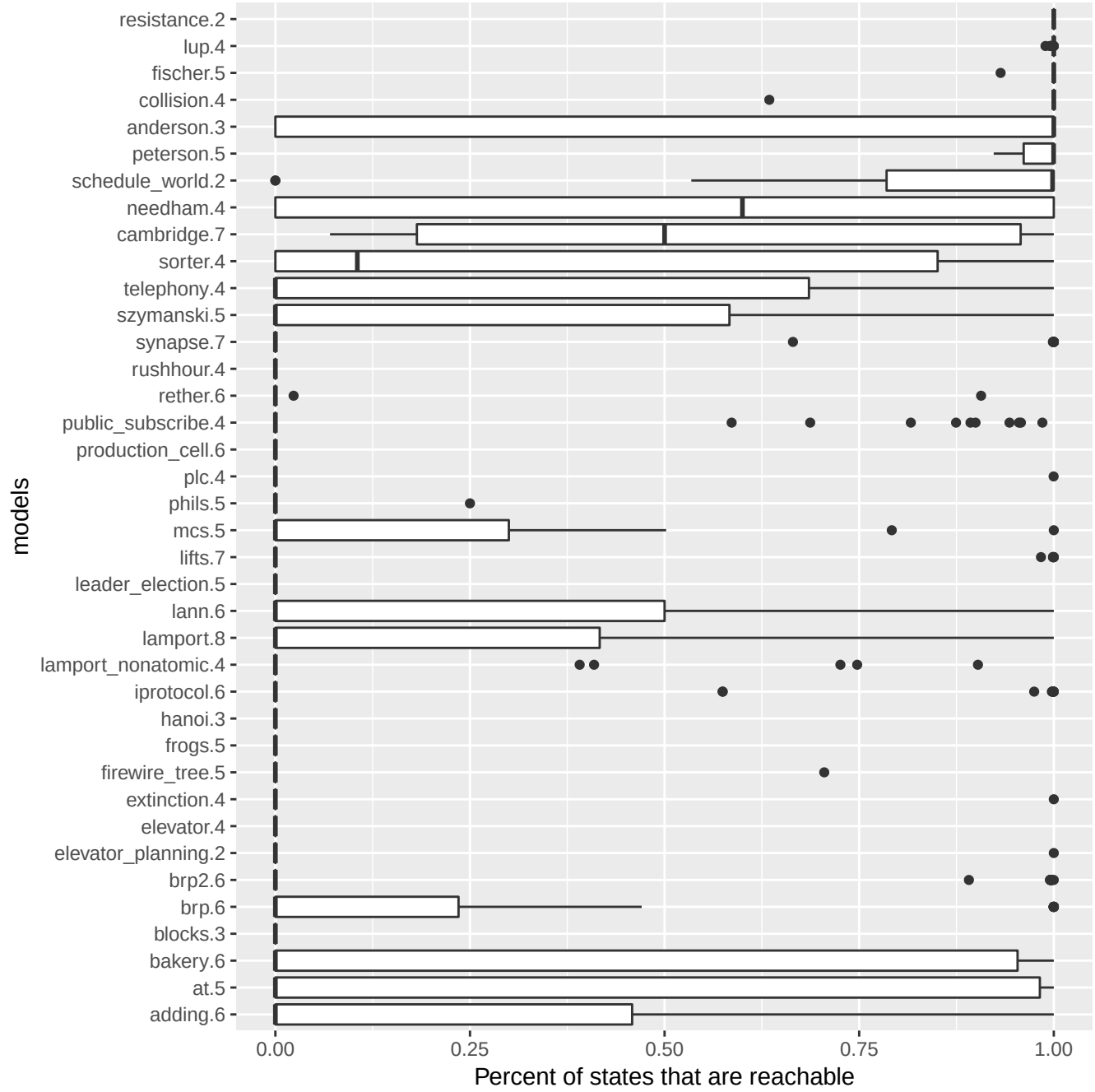

Fig. 6 Distributions of reachable states from a population of artificial initial states.

less we can observe that these performances are degraded while the percentage of *gp threads* augments. Indeed, *gp threads* may explore states that are not in the state space and using too much of these threads will not help *classical threads* to explore concurrently this state space. Notice that compared to results presented in Table 2, the variation of the percentage of the *gp threads* helps to obtain around 2.5% more reduction of the total time for this benchmark. Even if these variations are relatively small they tends to suggest that around 30% of threads are still useless. Portfolio algorithms, where 30% of threads run different exploration algorithms (e.g symbolic approaches), could help to decrease the global runtime.

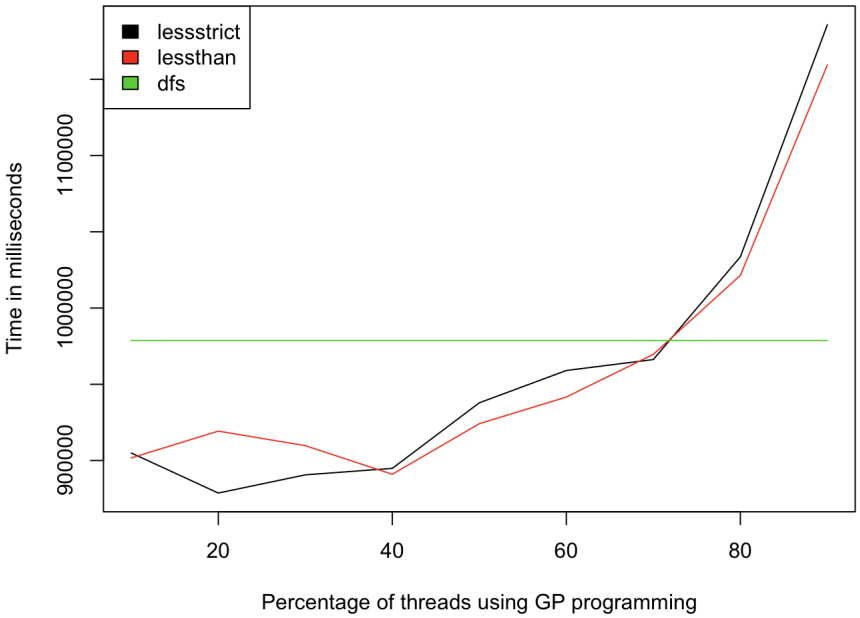

Fig. 7 Impact of the percentage of threads using genetic programming. Experiments were run with 12 threads. The *dfs* horizontal line represents state-of-the art while *lessthan* and *lessstrict* (the fastest for 20%) represents the fitness functions combined to the best parameters observed in Table 2.

6.2 Comparison with State-of-The-Art Algorithms

This section compares the techniques proposed in this paper with other approaches that are known to scale correctly up to 12 cores (the maximum we can test):

- `cndfs`[13]: the best parallel nested detpth-first search. It implements two swarmed nested-DFS with synchronizations to ensure the validity of the shared information.[12]
- `ufscc`[5]: the best parallel SCC-computation algorithm. It performs a swarmed DFS with a work-stealing mechanism integrated into a union-find data structure.[13]
- `symbolic` [36]: the best parallel symbolic reachability algorithm. It uses parallel BDD to compute the set of reachable states.[14]

Since the three previous algorithms were developed by the LTSmin team, we opted to run the benchmark of the previous section against their own tool[15]. Notice few details about this comparison:

1. The setup for running the benchmark were slightly modified: the 40 minutes runtime limitation is postponed to 3 hours. Nonetheless we still use the same machine.
2. All the algorithms load the same model, i.e., a shared library that only offers two functions: one for retrieving the initial state, and the other that compute a list of successors from a given state. This library respect the PINS interface [21] and it well suited for on-the-fly model-checking.

Figure 8 presents the comparison of our techniques against the aforementioned algorithms. This figure details, for each algorithm and for a given number of threads, the runtime required to process the whole benchmark (notice that the y-axis uses a logarithmic scale). The names of algorithms are straightforward: for instance `ltsmin/cndfs` refers the `cndfs` algorithm from the LTSmin tool. We opted to only benchmark the **lessthan** variation of our genetic approach (Algorithm 5), denoted `spot/gp-dfs-lessthan`, since it achieves the best results (empirically, according to the previous section).

First of all we observe that, for 12 threads, `cndfs` and `ufscc` obtain a speedup contained between six and eight, while our implementation of a single DFS (`spot/dfs`, see Algorithm 1) only achieves a speedup of three. Nonetheless, a closer look to sequential performances shows that `spot/dfs` is 5 time faster than `cndfs` and 3 time faster than `ufscc`. This difference introduces a bias when comparing the speedup of the different algorithms. To mitigate this effect one can compute the speedup relative to the fastest sequential algorithm, i.e. `spot/dfs`. When doing so, it appears

[12] *dve2lts-mc –strategy=cndfs –threads=... -s 29 –perm=shift*
[13] *dve2lts-mc –strategy=ufscc –threads=... -s 29 –perm=shift*
[14] *dve2lts–sym –saturation=sat –order=bfs-prev -rf –lace-workers=...*
[15] `https://ltsmin.utwente.nl`

that `cndfs` achieves a speedup of 1.27 and `ufscc` a speedup of 2.38. This confirms the results presented in Table 1 and highlights the need for new techniques to obtain better speedups in parallel model-checking.

Figure 8 also displays the result using parallel BDD (`ltsmin/symbolic`). It appears that one out of 38 models is not processed within the three hours limitation using less than 4 threads. Moreover, the results of the symbolic approach are really disappointing compared to explicit approaches even if a reasonable speedup is achieved with 12 threads. These poor results are the direct consequence of using the PINS interface which is well-suited for on-the-fly computation but not adapted for building the symbolic transition relation of the model. Indeed, LTSmin relies on two basic functions: INITIAL_STATE and NEXT_STATE: INITIAL_STATE returns the initial state s ; NEXT_STATE returns the successor states of state s. This interface allows explicit-state (enumerative) algorithms to be implemented for any language, but it has very limited use in combination with symbolic techniques due to the fact that it needs to call the next-state function for every state. Attempts have been made by LTSmin to optimize this using *event locality* and the implementation of PINS-2-PINS wrappers [7, 8, 26]. Our settings for the `ltsmin/symbolic` also uses *saturation* techniques well suited for multi-core on-the-fly symbolic algorithms [37]. During our experiments we observe differences that are quite modest compared to symbolic approach without saturation. To summarize, the symbolic approach cannot compete with explicit techniques on a benchmark (BEEM) dedicated to the comparison of explicit model-checkers.

Finally, Figure 8 shows that all (explicit) algorithms seem to reach a plateau (around 8 threads) where adding new threads does not improve significantly the results. If we observe the genetic programming approach of this paper, we note that it helps to reduce this "plateau effect".

6.3 Linear Topology

Few models in the previous benchmark have a linear topology (as the one of the Figure 1) which can be considered as the perfect one for the algorithms presented in this paper. In such model, using one or more threads will result in the same running time.

To evaluate the performances of our genetic programming algorithm, we decided to run Algorithm 5 (with the `lessthan` fitness function) against a linear (parametric) model with 150 processes and 300 variables. This model was built so that the genetic algorithm can exploit the bounds of the different variables (that are known a priori).

Figure 9 presents the results of this evaluation. For a given size of model, the running time is reported for the swarmed DFS with 12 threads, as well as for the **lessthan** variation. It should be noted that for all experiments, running `spot/dfs` with only one thread achieves 5% better. This is due to the contention over the hash table. Indeed, since the system is linear and since all threads have similar throughput, at any moment all threads try concurrently to insert the same states. This induces caches problems which slow down the algorithm. Moreover, each thread has its own stack to maintain, which implies a bigger memory footprint and thus a runtime overhead.

A closer look to the Figure 9, shows that Algorithm 5 is in average 40% faster than the swarmed DFS of Algorithm 1 regardless the size of the model considered.

These results are encouraging since it demonstrate that the generation of artificial initial states can bring significant speedup where classical parallel algorithm are stuck due to the topology of the model and their exploration order.

Discussion. Overall, we observe a global improvement of state-of-the-art algorithm over the various benchmarks we have tested. We believe that other fitness functions (based on interpolation or estimation of distribution) could help to generate better states, i.e. deep with respect to many DFS orders. Moreover, the generation of new states could be improved using other techniques like neural networks.

7 Conclusion

We have presented some first and new parallel graph exploration algorithms that rely on genetic algorithms. We suggested to see variables of the model as genes and states as chromosomes. With this definition we were able to build an algorithm that generates artificial initial states. To detect if such a state is relevant we proposed and evaluated various fitness functions. It appears that these seed states improve the swarming technique. This combination between swarming and genetic algorithms has never been proposed and the benchmark shows encouraging results: 10% faster than state-of-the-art algorithms on a general benchmark and 40% on a specialized benchmark. Since the performance of our algorithms highly relies on the generation of good artificial states we would like to see if other strategies could help to generate better states. We also observed that a small percentage of threads

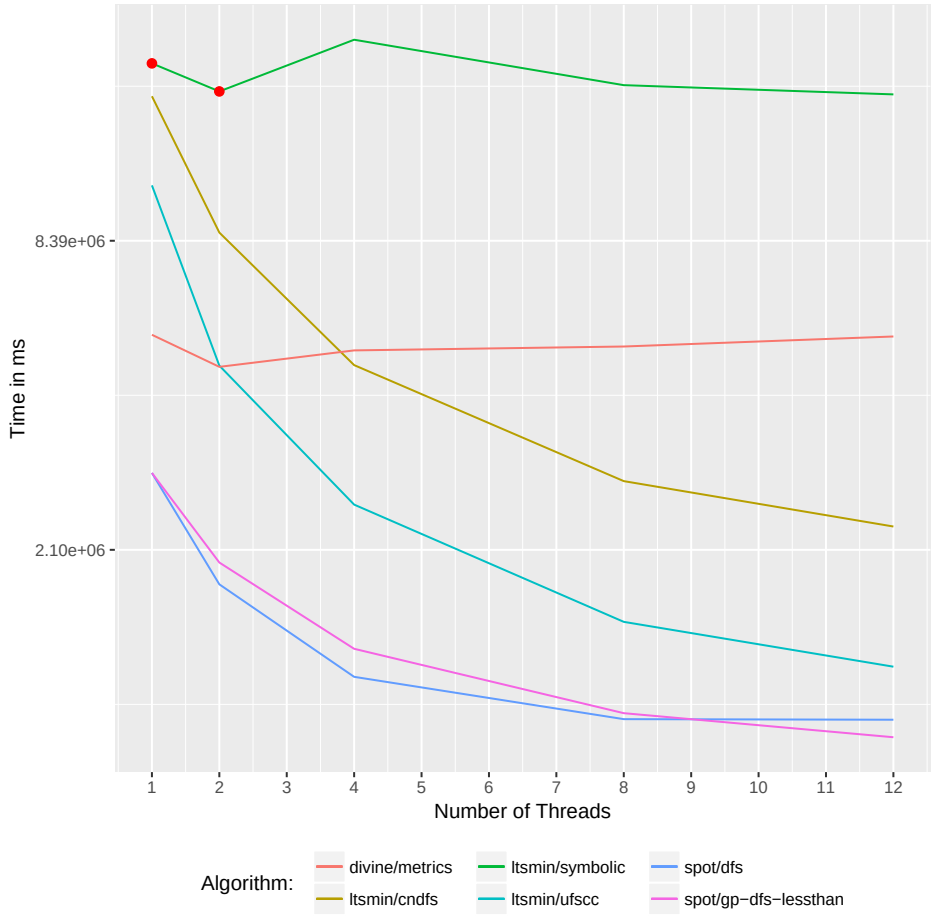

Fig. 8 Cumulated time for processing the whole benchmark w.r.t a number of threads. Red circles denote experiments where the runtime limitation was reached (only for **symbolic** approach). The other lines (from top to down in sequential) represent respectively: **cndfs**[13], **ufscc**[5], Algorithm 5 and Algorithm 1.

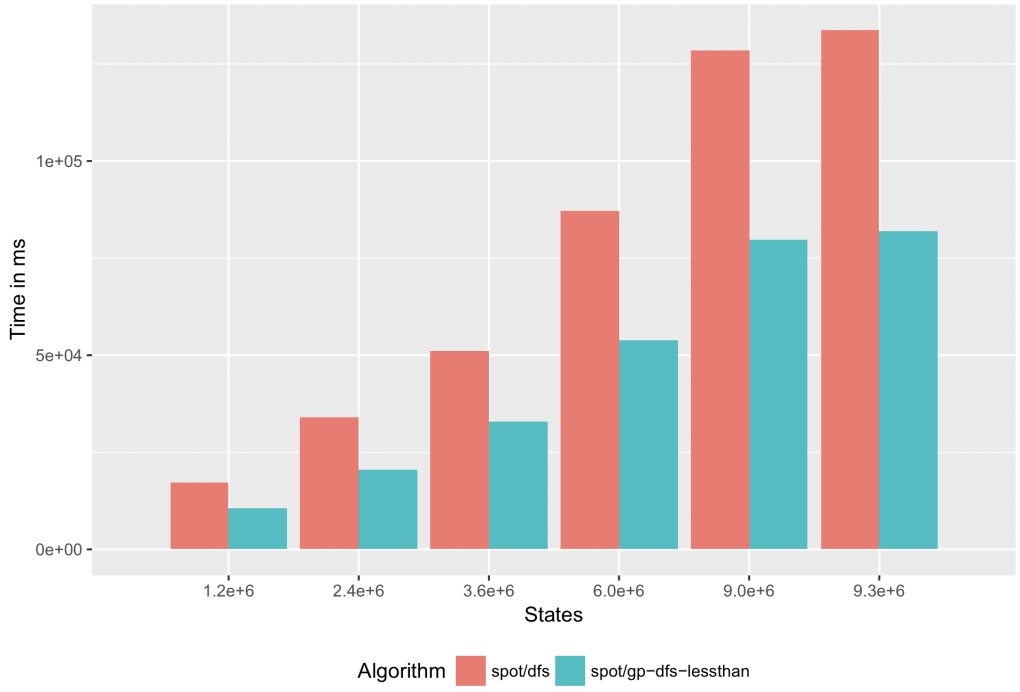

Fig. 9 Time, with 12 threads, for processing a linear parametric model of 300 bounded variables. The left bar represents Algorithm 1 and the right bar represents Algorithm 5.

using genetic programming is sufficient to obtain a good speedup.

We also demonstrated the correctness of our algorithms and described how they can be adapted to report counterexamples.

This work mainly focused on checking safety properties even if we proposed an adaptation for liveness properties. A future work would be to evaluate the performance of our algorithm in this latter case. We also want to investigate the relation between artificial state generation and POR, since both rely on the analysis of processes variables. Finally, we strongly believe that this paper could serve as a basis for combining parametric model-checking with neural network.

References

1. P. E. Ammann, P. E. Black, and W. Majurski. Using model checking to generate tests from specifications. In Oscar Nierstrasz and Michel Lemoine, editors, Proceedings of the 2nd International Conference on Formal Engineering Methods (ICFEM'98), volume 1687, pages 46–54. IEEE, december 1998.
2. Richard J. Anderson and Heather Woll. Wait-free parallel algorithms for the union-find problem. In Proceedings of the 23rd ACM Symposium on Theory of Computing (STOC'91), pages 370–380, 1994.
3. J. Barnat, P. Ročkai, V. Štill, and J. Weiser. Fast, dynamically-sized concurrent hash table. In Bernd Fischer and Jaco Geldenhuys, editors, Proceedings of the 22th international SPIN conference on Model Checking software (SPIN'15), volume 9232 of Lecture Notes in Computer Science, pages 49–65. Springer International Publishing, 2015.
4. Jiří Barnat, Luboš Brim, and Petr Ročkai. Scalable shared memory LTL model checking. International Journal on Software Tools for Technology Transfer (STTT), 12(2):139–153, 2010.
5. Vincent Bloemen and Jan Cornelis van de Pol. Multi-core scc-based LTL model checking. In Roderick Bloem and Eli Arbel, editors, Proceedings of the 12th International Haifa Verification Conference (HVC'16), Lecture Notes in Computer Science, pages 18–33. Springer International Publishing, november 2016.
6. Vincent Bloemen, Alfons Laarman, and Jaco van de Pol. Multi-core on-the-fly scc decomposition. In Proceedings of the 21st Symposium on Principles and Practice of Parallel Programming (PPOPP'16), volume 51, pages 8–20, march 2016.
7. Stefan Blom, Jaco van de Pol, and Michael Weber. Bridging the Gap between Enumerative and Symbolic Model Checkers. Technical Report TR-CTIT-09-30, University of Twente, Enschede, 2009.
8. Stefan Blom, Jaco van de Pol, and Michael Weber. LTSmin: Distributed and symbolic reachability. In Proceedings of the 22nd Computer Aided Verification International Conference (CAV'10) 2010, Edinburgh, UK, pages 354–359, July 2010.
9. Jerry R. Burch, Edmund M. Clarke, Kenneth L. McMillan, David L. Dill, and L.J. Hwang. Symbolic model checking: 10^{20} states and beyond. In Albert R. Meyer, editor, Proceedings of the Fifth Annual IEEE Symposium on Logic in Computer Science, pages 1–33, Washington, D.C., 1990. IEEE.
10. Edmund M. Clarke, Orna Grumberg, and Doron A. Peled. Model Checking. The MIT Press, 2000.
11. Costas Courcoubetis, Moshe Y. Vardi, Pierre Wolper, and Mihalis Yannakakis. Memory-efficient algorithm for the verification of temporal properties. In Edmund M. Clarke and Robert P. Kurshan, editors, Proceedings of the 2nd international workshop on Computer Aided Verification (CAV'90), volume 531 of Lecture Notes in Computer Science, pages 233–242. Springer-Verlag, 1991.
12. Alexandre Duret-Lutz, Alexandre Lewkowicz, Amaury Fauchille, Thibaud Michaud, Etienne Renault, and Laurent Xu. Spot 2.0 — a framework for LTL and ω-automata manipulation. In Proceedings of the 14th International Symposium on Automated Technology for Verification and Analysis (ATVA'16), volume 9938 of Lecture Notes in Computer Science, pages 122–129. Springer, October 2016.
13. Sami Evangelista, Alfons Laarman, Laure Petrucci, and Jaco van de Pol. Improved multi-core nested depth-first search. In Supratik Chakraborty, editor, Proceedings of the 10th international conference on Automated technology for verification and analysis (ATVA'12), volume 7561 of Lecture Notes in Computer Science, pages 269–283. Springer-Verlag, 2012.
14. Hubert Garavel, Radu Mateescu, and Irina Smarandache. Parallel State Space Construction for Model-Checking. Technical Report RR-4341, INRIA, 2001.
15. Patrice Godefroid and Sarfraz Khurshid. Exploring very large state spaces using genetic algorithms. In Joost-Pieter Katoen and Perdita Stevens, editors, Proceedings of the fourth International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'02), pages 266–280, Berlin, Heidelberg, april 2002. Springer Berlin Heidelberg.
16. Patrice Godefroid, Gerard J. Holzmann, and Didier Pirottin. State space caching revisited. In Gregor Bochmann and DavidKarl Probst, editors, Proceedings of the Fourth International Workshop on Computer Aided Verification (CAV'92), volume 663 of Lecture Notes in Computer Science, pages 178–191. Springer Berlin Heidelberg, 1992.
17. John H. Holland. Genetic algorithms. Scientific American, july 1992.
18. Gerard J. Holzmann. On limits and possibilities of automated protocol analysis. In Harry Rudin and Colin H. West, editors, Proceedings of the Seventh International Conference on Protocol Specification, Testing and Verification (PSTV'87), pages 339–344. North-Holland, May 1987.
19. Gerard J. Holzmann and Dragan Bosnacki. The design of a multicore extension of the SPIN model checker. IEEE Transaction on Software Engineering, 33(10): 659–674, 2007.
20. Gerard J. Holzmann, Rajeev Joshi, and Alex Groce. Swarm verification techniques. IEEE Transaction on Software Engineering, 37(6):845–857, 2011.
21. Gijs Kant, Alfons Laarman, Jeroen Meijer, Jaco van de Pol, Stefan Blom, and Tom van Dijk. LTSmin: High-performance language-independent model checking. In Christel Baier and Cesare Tinelli, editors, Proceedings of the 21th International Conference on Tools and Algorithms for the Construction and Analysis of Systems

(TACAS'15), volume 9035, pages 692–707, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.
22. Gal Katz and Doron A. Peled. Synthesis of parametric programs using genetic programming and model checking. In Lukás Holík and Lorenzo Clemente, editors, Proceedings 15th International Workshop on Verification of Infinite-State Systems (INFINITY'13), volume 140, pages 70–84, Hanoi, Vietnam, October 2013.
23. Alfons Laarman, Jaco van de Pol, and Michael Weber. Boosting multi-core reachability performance with shared hash tables. In Natasha Sharygina and Roderick Bloem, editors, Proceedings of the 2010 Conference on Formal Methods in Computer-Aided Design (FMCAD'10), pages 247–256. IEEE, 2010.
24. Alfons Laarman, Elwin Pater, Jaco Pol, and Henri Hansen. Guard-based partial-order reduction. International Journal on Software Tools for Technology Transfer (STTT), 7976:1–22, 2014.
25. L. Lamport. Proving the correctness of multiprocess programs. IEEE Transactions on Software Engineering, 2:125–143, 1977.
26. Jeroen Meijer, Gijs Kant, Stefan Blom, and Jaco van de Pol. Read, write and copy dependencies for symbolic model checking. In Hardware and Software: Verification and Testing - 10th International Haifa Verification Conference (HVC'14) 2014, Haifa, Israel, 18-20, 2014. Proceedings, volume 8855 of Lecture Notes in Computer Science, pages 204–219, november 2014.
27. R. Pelánek, T. Hanžl, I. Černá, and L. Brim. Enhancing random walk state space exploration. In Tiziana Margaria and Mieke Massink, editors, Proceedings of the 10th International Conference on Formal Methods for Industrial Critical Systems (FMICS'05), pages 98–105. ACM Press, 2005.
28. Radek Pelánek. BEEM: benchmarks for explicit model checkers. In Dragan Bošnački and Stefan Edelkamp, editors, Proceedings of the 14th international SPIN conference on Model checking software (SPIN'07), volume 4595 of Lecture Notes in Computer Science, pages 263–267. Springer-Verlag, 2007.
29. Radek Pelánek. Properties of state spaces and their applications. International Journal on Software Tools for Technology Transfer (STTT), 10:443–454, 2008.
30. Doron Peled. Combining partial order reductions with on-the-fly model-checking. In David L. Dill, editor, Proceedings of the 6th International Conference on Computer Aided Verification (CAV'94), volume 818 of Lecture Notes in Computer Science, pages 377–390. Springer-Verlag, 1994.
31. John H. Reif. Depth-first search is inherently sequential. Information Processing Letters, 20:229–234, 1985.
32. Etienne Renault. Improving parallel state-space exploration using genetic algorithms. In Mohamed Faouzi Atig, Saddek Bensalem, Simon Bliudze, and Bruno Monsuez, editors, Proceedings of the 12th International Conference on Verification and Evaluation of Computer and Communication Systems (VECOS'18), volume 11181 of Lecture Notes in Computer Science, pages 133–149, Grenoble, France, sept 2018. Springer.
33. Etienne Renault, Alexandre Duret-Lutz, Fabrice Kordon, and Denis Poitrenaud. Variations on parallel explicit model checking for generalized Büchi automata. International Journal on Software Tools for Technology Transfer (STTT), pages 1–21, April 2016.
34. Hemanthkumar Sivaraj and Ganesh Gopalakrishnan. Random walk based heuristic algorithms for distributed memory model checking. In Lubos Brim and Orna Grumberg, editors, Proceedings of the 2nd International Workshop on Parallel and Distributed Model Checking (PDMC'03), volume 89, pages 51–67, 2003.
35. Antti Valmari. Stubborn sets for reduced state space generation. In Grzegorz Rozenberg, editor, Proceedings of the 10th International Conference on Applications and Theory of Petri Nets (ICATPN'91), volume 618 of Lecture Notes in Computer Science, pages 491–515, London, UK, 1991. Springer-Verlag.
36. Tom van Dijk and Jaco van de Pol. Sylvan: multi-core framework for decision diagrams. International Journal on Software Tools for Technology Transfer, 19(6):675–696, Nov 2017.
37. Tom van Dijk, Jeroen Meijer, and Jaco van de Pol. Multi-core on-the-fly saturation. In Tomáš Vojnar and Lijun Zhang, editors, Proceeding of the 25th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'19), volume 11428 of Lecture Notes in Computer Science, pages 58–75. Springer International Publishing, 2019.
38. Moshe Y. Vardi. An automata-theoretic approach to linear temporal logic. In Faron Moller and Graham M. Birtwistle, editors, Proceedings of the 8th Banff Higher Order Workshop (Banff'94), volume 1043 of Lecture Notes in Computer Science, pages 238–266, Banff, Alberta, Canada, 1996. Springer-Verlag. ISBN 3-540-60915-6.