Forward and Backward: Bounded Model Checking of Linear Hybrid Automata From Two Directions

Yang Yang*, Lei Bu†, and Xuandong Li∗

Abstract—Instead of encoding the bounded state space of a linear hybrid automaton (LHA) in given threshold \( k \) into SMT formulas then solving them by SMT solvers, the authors proposed a different approach to handle the bounded reachability verification (BMC) of LHA in the previous work. First, the reachability specification along one abstract path in LHA can be checked by linear programming (LP). Then, all the abstract paths under the threshold can be checked one by one by depth-first-search (DFS) traversing. This approach was implemented in a prototype tool BACH, and the experiment result shows it is efficient.

As BACH uses DFS to traverse the bounded state space, clearly, if DFS traverses more quickly, the BMC can be finished more efficiently. Nevertheless, in many cases, the path segments which make the system infeasible are hidden “deeply” in the model or have many entry points which makes the DFS difficult to find them or has to traverse them many times. This burdens the DFS-style BMC approach a lot.

To handle this problem, in this paper, the authors propose a backward-DFS BMC approach for LHA. First, reverse the graph structure of LHA. Then, conduct the DFS-style BMC on the reversed LHA. In this way, the “deep” path segments in the forward-DFS can be found very quickly to prune the DFS tree which is needed to be traversed. This backward-DFS approach is implemented into BACH. The experiment shows the performance of BACH is optimized significantly to handle large cases.

I. INTRODUCTION

Reachability verification of Linear Hybrid Automata (LHA) [1] is a very difficult and important problem. Currently, researchers always try to handle this problem in two ways:

Classical Model Checking (CMC)[4]: Compute the complete state space by methods like polyhedra computation, like HYTECH[6] and PHAVer[7]. First of all, the classical reachability verification problem of LHA is proven to be undecidable[2], [3]. Furthermore, these methods are very complex and sensitive to the number of variables. Thus, they do not scale well to the size of practical problems.

Bounded Model Checking (BMC)[5]: Encoding the bounded reachability problem into the satisfiability problem of a boolean combination of propositional variables and linear mathematical constraints, which can be solved by SMT solvers[8], [9]. As this technique requires to encode the bounded problem space firstly, when the system size or the given threshold is large, the object problem will be very huge, which restricts the size of the problem that can be solved.

Both classical and bounded model checking are feeding the (partly) complete state space to the underlying solver at one time which suffers from the state explosion problem and restricts the problem size that can be solved. Study[13] proposed a linear programming (LP)-based approach to check one abstract path at one time to find whether there exists a behavior of LHA along this path and satisfy the given reachability specification. Study[15] extended this approach by using forward depth-first-search (F-DFS) to traverse on the graph structure of LHA to enumerate and check all the abstract paths with length no longer than the threshold one by one to answer the bounded reachability problem. Furthermore, when the DFS is finished before touching the bound, this approach can prove the given specification is not satisfied in general, not only in the given bound. A prototype tool BACH[15], [16] was implemented based on this idea. The experiments show that BACH is efficient in many cases.

Clearly, if DFS traverses more quickly, the BMC can be finished more efficiently. In BACH, once a path is found to be infeasible, the DFS will ask the underlying LP solver to locate the path segment which makes this path infeasible and backtrack to the path segment to prune the behavior tree which is needed to be traversed [17]. Nevertheless, in many cases, the path segments which make the path infeasible are hidden “deeply” or have many entry points in the graph structure which makes the DFS difficult to find these path segments or has to traverse them time and time again. This asks the DFS to traverse a lot of obviously infeasible paths, which makes the backtracking strategy inefficient in many cases and burdens the DFS-style BMC a lot.

To handle this problem, in this paper, the authors propose a backward-DFS BMC (B-DFS)[10], [11] approach to complement the classical F-DFS. First, reverse the graph structure of LHA. Mark the original target location as initial location, and mark the original initial location as target. Then, conduct the F-DFS BMC on the reversed LHA. As these path segments are “deep” in the original graph structure, clearly, if the DFS is conducted in a backward way, the path segments will be “shallow” in the reversed graph and can be found more quickly. Furthermore, for those path segments which have many entry points, the abstract paths with these path segments as suffix will be pruned easily in the reversed LHA to shrink the size of the DFS tree which is needed to be traversed.

This B-DFS BMC idea is implemented into BACH as a complement to the F-DFS BMC. Once a LHA and a reachability specification is given, BACH will start two threads, one conducts the F-DFS BMC on the original LHA and the other conducts the B-DFS BMC on the reversed model. The procedure terminates when any of these two threads finish. We conduct a series of case studies on the new BACH, and
compare it with both state-of-the-art classical model checker PHAVer and SMT solver MathSAT. The experiment results show that:

- By B-DFS, the state space needed to search and verify is pruned substantially. Therefore, BACH outperforms the other competitors significantly.
- In many cases, the DFS is finished before touching the bound. Then, in this situation, BACH can prove the given specification is not satisfiable in general, not only in the given bound, which is incapable for other BMC checkers.

II. The Underlying Techniques

A. Forward-DFS Approach

Now, let’s recall the F-DFS BMC approach given in study [15]. The main idea is to traverse all the abstract paths with length shorter than or equal to the bound by DFS on the graph structure of the LHA. Whenever an abstract path $\rho$, which contains the target location, is found, a decision procedure will be called to check whether the LHA has a feasible behavior according with $\rho$. The decision procedure will encode all the syntax elements, i.e., invariants, guards and e.t.c., in $\rho$ into a linear constraint set $R$ as guided by [13] and [14]. In this manner, the reachability problem of $\rho$ is transformed into the feasibility of $R$ which can be answered by an LP solver very efficiently.

It is obvious that the performance of the F-DFS-based BMC depends on the performance of the DFS algorithm. If the DFS can be finished more quickly, the BMC can be finished more efficiently. Therefore, we gave an optimization technique to generate unsatisfiable core from the proven infeasible paths to tailor the state space needed to be traversed[17]. Once $\rho$ is proved to be infeasible by the underlying LP solver, the solver will provide a feasible irreducible set (IIS) [18] of $R$, which is a minimal set of constraints in $R$ that makes $R$ infeasible. Then the constraint set $R'$ that IIS located can be mapped back to a path segment $\rho'$ in the path $\rho$, which means $\rho'$ is the infeasible core in the path that makes $\rho$ infeasible. Since removing any constraint in $R'$ will make it feasible, so the DFS will backtrack to the location preceding the last location in $\rho$.

Take the sample LHA in Fig.1 for example. We want to check whether location $v_6$ is reachable within bound 20. Suppose the current visiting path is $\rho = \langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle \rightarrow \langle v_4 \rangle \rightarrow \langle v_5 \rangle \rightarrow \langle v_6 \rangle$ which is proved to be infeasible by an LP solver, and the infeasible path segment located by IIS is $\rho' = \langle v_2 \rangle \rightarrow \langle v_3 \rangle \rightarrow \langle v_4 \rangle$, so the location that DFS will backtrack to is $v_3$ instead of $v_5$. Then the DFS will begin to search the next branch under $\rho'' = \langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle$, which is $\langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle \rightarrow \langle v_4 \rangle \rightarrow \langle v_5 \rangle \rightarrow \langle v_6 \rangle$. As we can see, the subtree beneath $\rho'' = \langle v_2 \rangle \rightarrow \langle v_3 \rangle$ is pruned completely to speed up the DFS.

Nevertheless, let’s look at the following path $\rho_1 = \langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle \rightarrow \langle v_4 \rangle \rightarrow \langle v_5 \rangle \rightarrow \langle v_6 \rangle$. $\rho_1$ is a graphically correct path beneath $\rho''$ and contains target location $v_6$. Therefore the F-DFS BMC will ask the underlying decision procedure to check whether $\rho_1$ is feasible or not. Indeed, as $\rho_1$ contains $\rho'$, it can be proved to be infeasible at once. Then DFS will backtrack to the second $v_3$ in $\rho_1$. But, as the bound 20 is not reached yet, the DFS will traverse the loop $\langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle$ for the third time and then $\rho'$ again, i.e., traverse a new path $\langle v_1 \rangle \rightarrow \langle v_2 \rangle \rightarrow \langle v_3 \rangle \rightarrow \langle v_4 \rangle \rightarrow \langle v_5 \rangle \rightarrow \langle v_6 \rangle$. In this manner, lots of time will be wasted on DFS to traverse these paths which are doomed to be infeasible. Obviously, if there is a method to prune the DFS tree and avoid the visiting of such paths, the DFS-BMC will be much more efficient.

B. Backward-DFS Approach

As discussed above, in many times, the path segments which make the path infeasible are hidden quite “deeply” or could have many entries in the graph structure which makes the F-DFS difficult to find them or has to traverse them lots of times. To solve this problem, in this paper we propose to conduct the DFS search in a backward way(B-DFS BMC): First, reverse the transition relation of the LHA model. For each transition $\langle v_1 \rangle \rightarrow \langle v_2 \rangle$ in LHA $H$, there will be a transition $\langle v_2 \rangle \rightarrow \langle v_1 \rangle$ in the reversed LHA $\sim H$. Then mark the original target location as initial location, and mark the original initial location as target. For example, Fig.2 is the reversed model of the LHA given in Fig.1.

Now, we can conduct the F-DFS BMC on the reversed LHA, which means searching for a reversed path from the original target location to the initial location. Once a reversed path is found, we will ask the LP solver to check the accordingly path in the original model and locate the infeasible path segments as well. As many infeasible path segments are “deep” in the original graph structure, if the DFS is conducted in a backward way, these path segments will be “shallow” in the reversed structure and can be found more quickly. Furthermore, for those path segments which have many entry points, all the paths containing those path segments as “suffix” will be reversed, then those infeasible path segments will become “prefix” now. As a result these paths can be pruned easily.
in the reversed LHA to shrink the size of the DFS tree which is needed to be traversed.

![Fig. 2. Reversed Sample Automaton](image)

Now, let’s look at the previous example again (whether $v_6$ is reachable within bound 20 in Fig.1). On the reversed model, Fig.2, the first path that is traversed and solved is: $\gamma_1 = (v_6) \xrightarrow{e_1} (v_2) \xrightarrow{e_2} (v_1) \xrightarrow{e_3} (v_2) \xrightarrow{e_4} (v_1)$. The accordingly path in the original model is $\rho_1 = (v_1) \xrightarrow{e_4} (v_2) \xrightarrow{e_3} (v_5) \xrightarrow{e_2} (v_3) \xrightarrow{e_1} (v_2) \xrightarrow{e_5} (v_6)$. Therefore, $\rho_1$ is given to the underlying LP solver. The LP solver can tell that $\rho_1$ is infeasible and $\rho_1' = (v_2) \xrightarrow{e_2} (v_1) \xrightarrow{e_3} (v_2)$ is the infeasible core. The according infeasible core in $\gamma_1$ is $\gamma_1' = (v_6) \xrightarrow{e_6} (v_3) \xrightarrow{e_5} (v_2)$. So, the location which is backtracked to is $v_2$ in $\gamma_1$. As $v_3$ has no other successor locations in $\neg H$, the DFS keeps backtracking to $v_5$ and traverses the next path $\gamma_2 = (v_6) \xrightarrow{e_7} (v_3) \xrightarrow{e_8} (v_1)$. Similarly, LP solver locates the infeasible core path segment of $\gamma_2$ as $\gamma_2' = (v_6) \xrightarrow{e_7} (v_3) \xrightarrow{e_8} (v_1)$. In this case, the DFS backtracks to $v_5$ in $\neg H$ again and finds there is no more path to travel. This means there doesn’t exist a feasible path which can go from the initial location to the target. Therefore, the target location $v_5$ can be proved to be not reachable in general, not only within bound 20!

Clearly, in this example, the DFS tree that needed to be traversed are shrunk significantly by B-DFS BMC. Many paths like $(v_1) \xrightarrow{e_4} (v_2) \xrightarrow{e_3} (v_5) \xrightarrow{e_2} (v_3) \xrightarrow{e_1} (v_2)$ in F-DFS are pruned. The B-DFS only needs to check two paths at all to tell location $v_6$ is unreachable.

### C. Bidirectional-DFS Implementation

As discussed above, if the path segment which makes the whole path infeasible is “deep” or has many entry points, the B-DFS BMC method will be more efficient than the F-DFS BMC method. Nevertheless, in another word, if the path segment is “shallow” or has many exit points, then F-DFS BMC will outperform B-DFS BMC. So, none of these two techniques can take over the other one.

Therefore, we combine the F-DFS and B-DFS BMC together, which results in a bidirectional-DFS BMC. In our bidirectional-DFS, the algorithm will start two threads, one conducts the F-DFS BMC on the original LHA and the other conducts the B-DFS on the reversed model. Be different from classical bidirectional-DFS algorithms like[11], [12], where the two threads can cooperate with each other, in our algorithm, these two threads work in a competition nature, that these two threads will not communicate with each other during searching, and the algorithm terminates when any of these two threads finish the searching. Under this setting, no matter where the infeasible path segment is, our bidirectional-DFS BMC can traverse the state space efficiently. The pseudo code for our implementation is shown in Table. I.

### III. Case Studies

We implement the bidirectional-DFS BMC, (F-DFS plus B-DFS), into our LHA bounded model checker BACH [15], [16] as guided by the last section. The latest version of BACH (V4.0) is implemented in Java, and can be downloaded from http://seg.nju.edu.cn/BACH/. As the LP solver underlying the previous versions of BACH is OR-objects[19] which does not support the functionality of IIS analysis, BACH 4 calls the IBM CPLEX[20] instead, which gives a nice support of IIS analysis. The main functionality of BACH is provided by the following set of services:

- **Graphical LHA Editor:** This component allows users to construct, edit, and perform syntax analysis of LHA interactively. This Editor can also transform the graphical representation of LHA to a readable text file which is used as the input file for reachability checking.
- **Path-Oriented Reachability Checker:** The checker requires users to select a specific path in the model. Then, it can check whether the reachability specification is satisfied along with the given path.
- **Bounded Reachability Checker:** This checker uses the path-oriented checker as underlying solver. It traverses the behavior tree of the model under the threshold by our adapted DFS algorithm, and checks the related path for reachability to perform bounded reachability checking.

In order to evaluate the performance of the BACH 4.0, we conduct a series of case studies on a set of well-known cases, which includes the temperature control system in Fig.3 and water-level monitor system in Fig.4, the scalable automated highway system in Fig.5, and the sample automaton given in Fig.1. The target locations are all marked by double circle in the models and they are all unreachable.
We compare the BMC performance of BACH 4.0, marked as \textit{BACH}\textsubscript{F+B}, with the previous BACH, which only includes F-DFS and marked as \textit{BACH}\textsubscript{F}, and the state-of-the-art SMT solver MathSAT5\cite{9}. The experiments are conducted on a Think Center desktop machine (Intel Core2 Quad CPU 2.83GHz, 4GB RAM and Ubuntu 10.04). The time limit for experiment is set as one hour. The input models we used in experiments are all available from http://seg.nju.edu.cn/BACH/. 

The experiment data for the time (second) spent in each benchmark is shown below in Table.II-V respectively. If the checker fails to give a result in the time limit, the corresponding blank is marked as N/A. The number of bound means the largest number of discrete locations that a path can have in the state space under searching. Furthermore, as we mentioned that if the DFS finished before touching the bound, then BACH can prove the target is not reachable in general. In such cases, the time that BACH spent for solving the problem are marked with subscript G. For example, for the sample automaton, no matter how large the bound is, the new BACH only needs to check two paths to tell that the location \textit{v}6 is not reachable in general which only took 0.01 second. Therefore, in Table.II, all the blanks in column \textit{BACH}\textsubscript{F+B} are combined together. Besides of that, a special row \textit{∞} is added to emphasise this problem can be proved in general not only in given bound.

As we can see from these tables that, for sample, water and highway system, new BACH can all give a general proof of the unreachability of the targets. Therefore, we make a comparison of new BACH with PHAVer\cite{7} which is the state-of-the-art classical model checker for LHA. We find that for sample automaton and water automaton, both BACH and PHAVer can finish in 0.01 second. For highway system which size, number of locations and variables, is scalable by increasing the number of vehicles in the system, the experimental data is plotted in Fig.6. We can see that the largest highway system that new BACH solved in one hour has 150 vehicles included, which is a big model of 150 variables and 151 locations, while in one hour, PHAVer can only solve a system with 6 vehicles. Furthermore, the only model that new BACH can not give a general proof is the temperature control system, while the computation of PHAVer can not terminate on this model and it fails to give any result about this model.

From these data, we can see that:
- By the help of integrating backward-DFS into BACH, the performance of BACH is optimized significantly.
- As BACH only checks one path at a time, the complexity of the verification is well controlled. As a result, the scalability of BACH is much better than the SMT-style BMC solvers, like MathSAT.
- When the DFS terminates before touching the bound, DFS-style BMC can prove the unreachability of certain...
In the current setting, the F-DFS and B-DFS are running independently. In the future, we will try to let these two threads to cooperate with each other. Then the state space has been pruned by one thread can benefits the other one, and the DFS-based BMC shall be finished more efficiently.

**ACKNOWLEDGMENT**

We would like to thank the anonymous reviewers for their valuable comments and suggestions. This work is supported by the National Natural Science Foundation of China (No.61100036, No.90818022, and No.61170066), the National Grand Fundamental Research 973 Program of China (No.2009CB320702), the National 863 High-Tech Programme of China (No.2011AA010103, No.2012AA011205) and by the Jiangsu Province Research Foundation (No.BK2011558).

**References**


