

Layered and Staged Monte Carlo Tree Search for SMT Strategy Synthesis

Zhengyang Lu¹, Stefan Siemer², Piyush Jha³,
Joel Day⁴, Florin Manea² and Vijay Ganesh³

¹University of Waterloo

²University of Göttingen and CIDAS

³Georgia Institute of Technology

⁴Loughborough University

john.lu2@uwaterloo.ca, stefan.siemer@cs.uni-goettingen.de, piyush.jha@gatech.edu,
j.day@lboro.ac.uk, florin.manea@informatik.uni-goettingen.de, vganesh@gatech.edu

Abstract

Modern SMT solvers, such as Z3, offer user-controllable strategies that allow solver users the ability to tailor solving strategies for their unique set of instances, thus dramatically enhancing the solver performance for their specific use cases. However, this approach of strategy customization presents a significant challenge: handcrafting an optimized strategy for a class of SMT instances remains a complex and demanding task for both solver developers and users alike.

In this paper, we address the problem of automated SMT strategy synthesis via a novel Monte Carlo Tree Search (MCTS) based method. Our method treats strategy synthesis as a sequential decision-making process, whose search tree corresponds to the strategy space, and employs MCTS to navigate this vast search space. The key innovations that enable our method to identify effective strategies, while keeping costs low, are the ideas of layered and staged MCTS search. These novel heuristics allow for a deeper and more efficient exploration of the strategy space, enabling us to synthesize more effective strategies than the default ones in state-of-the-art (SOTA) SMT solvers. We implement our method, dubbed Z3alpha, as part of the Z3 SMT solver. Through extensive evaluations across six important SMT logics, Z3alpha demonstrates superior performance compared to the SOTA synthesis tool FastSMT, the default Z3 solver, and the CVC5 solver on most benchmarks. Remarkably, on a challenging QF.BV benchmark set, Z3alpha solves 42.7% more instances than the default strategy in Z3.

1 Introduction

Satisfiability Modulo Theories (SMT) solvers [De Moura and Bjørner, 2011] are key tools in diverse fields such as software engineering [Cadar *et al.*, 2008], verification [Gurfinkel

et al., 2015], security [Song *et al.*, 2008], and artificial intelligence [Pulina and Tacchella, 2012]. It has long been observed that no single solver excels across all instances from an SMT logic or a problem class. As a result, modern SMT solvers, such as Z3 [De Moura and Bjørner, 2008], offer *user-controllable strategies* [De Moura and Passmore, 2013], enabling users to customize a decision procedure for their problem class. A *strategy* can be considered an algorithmic recipe for selecting, sequencing, and parameterizing *tactics*. Each *tactic* is a well-defined algorithmic proof rule or symbolic reasoning step, provided by the solver. For example, **propagate-values** is a Z3 tactic that propagates equalities, while **sat** and **smt** are the tactic wrappers of the main SAT and SMT solver in Z3. A strategy builds a decision procedure by combining tactics, as shown in an exemplar strategy (**if is-pb (then propagate-values sat) smt**). This strategy specifies a solving algorithm that, given an input instance, applies **propagate-values** followed by **sat** if the instance is a pseudo-boolean problem (as checked using **is-pb**), or applies **smt** otherwise.

Default solver strategies are typically optimized for well-established benchmarks, such as those in the SMT-LIB library [Barrett *et al.*, 2016]. However, as the scope of SMT applications continues to grow rapidly, users frequently encounter specialized, evolving, and unprecedented classes of instances. In these scenarios, the default or the existing customized strategies might not be as effective. Consequently, there arises a need for novel customized strategies, specifically designed to efficiently address the unique challenges posed by users' specific problems. Traditionally, this task of strategy customization has been undertaken by human experts through extensive experimentation and benchmark analysis. However, even with their expertise and efforts, the task remains challenging due to the intricate interactions among tactics and the vast search space for potential strategies.

Early attempts have been made to synthesize SMT strategies automatically. For instance, StratEVO [Ramírez *et al.*, 2016] searches for an optimal strategy using evolutionary algorithms, while FastSMT [Balunovic *et al.*, 2018] synthesizes a tailored strategy using imitation learning and decision tree learning techniques. While these methods show promise

in automated strategy customization, they suffer from issues such as a lack of robustness, limited interpretability, and extensive training times.

To address these issues, we introduce a novel SMT strategy synthesis method that employs Monte Carlo Tree Search (MCTS). MCTS is a heuristic search algorithm, widely applied in computer board game players as a lookahead planning algorithm [Browne *et al.*, 2012]. Its prominence further escalated following its successful integration into the groundbreaking deep reinforcement learning systems AlphaGo [Silver *et al.*, 2016] and AlphaZero [Silver *et al.*, 2017], where MCTS was employed as a policy improvement operator. Recently, MCTS has shown remarkable success as a standalone framework in addressing complex symbolic, combinatorial, or optimization problems [Zhang *et al.*, 2018; Khalil *et al.*, 2022; Sun *et al.*, 2023]. Its key strengths, including its ability to effectively balance exploration and exploitation and its adaptiveness to the nuances of varied search problems, make it an excellent method for such challenging tasks. Our work is the first to apply MCTS to the SMT strategy synthesis problem.

An extended version of this paper is available at: <https://arxiv.org/abs/2401.17159>. Our code and data are available at: <https://github.com/JohnLyu2/z3alpha>.

1.1 The Strategy Synthesis Problem

The SMT strategy synthesis problem is defined as automatically identifying an optimal strategy that yields the best performance for a given benchmark set P . This performance is typically measured in terms of metrics such as the number of P -instances successfully solved within a specified wallclock timeout t . P is intended to be a representative subset of a broader benchmark set Q , which is of interest to the user, with the expectation that a strategy performing well on P generalizes effectively to Q . It is important to note that due to the infinite nature of the strategy search space and the empirical approach to strategy evaluation, finding a rigorously optimal solution is impractical. Consequently, our objective is to discover a near-optimal solution within a reasonable search time, based on empirical measurements. This research focuses on strategy synthesis for Z3, a solver that is widely regarded as one of the most prevalent SMT solvers in use today.

1.2 Our Contributions

1. **Z3alpha: An MCTS-based Strategy Synthesizing Solver:** We present a novel MCTS-based framework, dubbed Z3alpha, for the SMT strategy synthesis problem, which automatically constructs tailored solver strategies for a given class of problem instances. To the best of our knowledge, Z3alpha is the first MCTS-based method developed for the SMT strategy synthesis problem.
2. **Layered and Staged MCTS:** To address the unique challenges inherent to strategy synthesis, that cannot be solved by the conventional MCTS alone, we develop two innovative heuristics, namely, layered search and staged search on top of the MCTS framework. The layered search method effectively narrows the search space

by treating certain auxiliary tasks as independent search problems. On the other hand, the staged search technique segments the entire search problem into sequential sub-problems, enabling the use of results from early stages to expedite the search in later stages. Together, these two techniques work symbiotically to enhance the search efficiency in finding the optimal strategy, an essential improvement given the time-consuming nature of strategy evaluations.

3. **Extensive Experimental Evaluation:** We implemented our proposed method, dubbed Z3alpha, on top of the leading SMT solver Z3. To assess its performance, we conducted comprehensive experiments, comparing Z3alpha with the state-of-the-art (SOTA) synthesis tool FastSMT, as well as Z3’s own handcrafted default strategy and the CVC5 solver. These experiments spanned a broad spectrum, including instances from six different SMT logics (namely, QF_{BV, LIA, LRA, NIA, NRA, S}), representing a wide range of problem sizes and solver runtimes. Across all experiments, Z3alpha consistently demonstrated superior and robust performance. This impressive performance strongly highlights the benefits of automated strategy customization, promoting broader adoption of user-controllable strategies within the SMT community.

2 Related Work

2.1 MCTS for Symbolic/Combinatorial Problems

MCTS has long been viewed as a powerful planning algorithm for board games [Browne *et al.*, 2012]. Recently, there has been a noticeable trend towards its application in solving symbolic and combinatorial problems. For instance, BaMCTS [Khalil *et al.*, 2022] and the Symbolic Physics Learner [Sun *et al.*, 2023] uses the standalone MCTS method to identify backdoors in Mixed Integer Linear Programming (MIP) problems and to discover nonlinear mathematical formulas for symbolic regression problems, respectively. Cameron *et al.* [2022] proposed an extension to MCTS, Monte Carlo Forest Search (MCFS), which steers the SAT branching policy. AlphaMapleSAT [Jha *et al.*, 2024] is an MCTS-based cube-and-conquer SAT solver. AlphaDev [Mankowitz *et al.*, 2023] is an MCTS-guided deep reinforcement learning agent that synthesizes assembly programs. Remarkably, it has successfully discovered sorting algorithms that surpass the best previously known human-designed algorithms.

Our work is, to the best of our knowledge, the first application of MCTS to address the SMT strategy synthesis problem. Different from AlphaDev, which considers the synthesis of assembly programs as sequencing individual instructions, we present the strategy program as an expression tree. The increased complexity in program structure and the extended program evaluation time present unique challenges to the strategy synthesis problem.

2.2 MCTS Variants

It is a well-known problem that the basic MCTS method does not generalize well between related states and actions. It results in a notably inefficient search in scenarios where the

search space is extensive. To counter this issue, various techniques have been proposed.

One prevalent technique is the rapid action value estimation (RAVE) algorithm [Gelly and Silver, 2011], which incorporates an action-specific term Q_{RAVE} into the MCTS tree policy. Its basic assumption is that there is an intrinsic value associated with each action regardless of its position. Game abstraction [Johanson, 2013] is another common technique, especially used in Poker, to expedite the search. It abstracts similar actions or states into a single category to reduce the search space. Option Monte Carlo Tree Search (O-MCTS) [De Waard *et al.*, 2016] uses options [Sutton *et al.*, 1999] in the MCTS framework, to mimic the human behaviors of defining subgoals and subtasks in game playings. An option is a predefined method for reaching a specific subgoal, with its own policy and termination function. In O-MCTS, the agent selects among options instead of actions.

Our layered and staged search methods share similarities with these MCTS variants. The layered search method reduces action and state spaces, but, instead of abstracting, it separates certain auxiliary actions aside and optimizes them in parallel. This separation generalizes values among subtrees, but, unlike RAVE which shares values between actions in different positions, it merges subtrees. The staged search, which chooses among previously found rewarding sub-solutions, is similar to choosing among options, but without a separate predefined policy for each choice.

2.3 SMT Strategy Synthesis

StratEVO [Ramírez *et al.*, 2016] presents a pioneering effort in automated strategy generation, utilizing a genetic programming algorithm [Koza, 1994] to evolve strategies from a predefined strategy population. Unfortunately, the StratEVO tool is not publicly available, which precludes us from conducting an empirical comparison.

FastSMT [Balunovic *et al.*, 2018], recognized as the SOTA tool in SMT strategy synthesis, applies a dual-phase learning approach. First, it applies the DAGger [Ross *et al.*, 2011] algorithm to train a deep neural network (DNN), in order to discover a collection of branch-free strategies, each tailored for specific instances. These strategies are then synthesized into one single, unified strategy through an entropy-based decision-tree learning algorithm [Poole and Mackworth, 2010]. Chen *et al.* [2021] proposed a strategy synthesis method specifically for symbolic execution, where, similar to FastSMT, strategies are synthesized with decision tree techniques from a list of branch-free strategies found by an offline trained DNN. Our method also involves two steps. Different from the previous work, we utilize MCTS for both steps. Compared to FastSMT, our method shows superior and more robust empirical results across six SMT logics, as detailed in Section 5. Further, the strategies synthesized by FastSMT tend to be less interpretable, sometimes involving more than a thousand branches.

Our earlier exploratory work, AlphaSMT [Lu, 2023], employs the AlphaZero framework to adaptively select per-instance tactic sequences, achieving promising results on certain challenging benchmark sets. However, further development is needed for broader applicability.

$$\begin{aligned}
 \langle \text{Strategy} \rangle &\rightarrow \langle \text{Tactic} \rangle \\
 &| \textbf{(using-params} \langle \text{Strategy} \rangle \langle \text{ParamSets} \rangle) \\
 &| \textbf{(then} \langle \text{Tactic} \rangle \langle \text{Strategy} \rangle) \\
 &| \textbf{(or-else} \langle \text{Strategy} \rangle \langle \text{Strategy} \rangle) \\
 &| \textbf{(try-for} \langle \text{Strategy} \rangle \langle \text{Constant} \rangle) \\
 &| \textbf{(if} \langle \text{Predicate} \rangle \langle \text{Strategy} \rangle \langle \text{Strategy} \rangle) \\
 \langle \text{Tactic} \rangle &\rightarrow \textbf{simplify} | \textbf{smt} | \textbf{sat} | \textbf{solve-eqs} | \textbf{aig} | \dots \\
 \langle \text{ParamSets} \rangle &\rightarrow \langle \text{ParamSet} \rangle | \langle \text{ParamSet} \rangle \langle \text{ParamSets} \rangle \\
 \langle \text{ParamSet} \rangle &\rightarrow \textbf{:} \langle \text{Parameter} \rangle \langle \text{Constant} \rangle \\
 \langle \text{Parameter} \rangle &\rightarrow \textbf{som} | \textbf{flat} | \textbf{seed} | \textbf{elim_and} | \dots \\
 \langle \text{Predicate} \rangle &\rightarrow \langle \text{BProbe} \rangle | \langle \langle \text{Operator} \rangle \langle \text{NProbe} \rangle \langle \text{Constant} \rangle \rangle \\
 \langle \text{BProbe} \rangle &\rightarrow \textbf{is-unbounded} | \textbf{is-pb} | \textbf{is-qflia} | \dots \\
 \langle \text{Operator} \rangle &\rightarrow > | < | \geq | \leq | = | \neq \\
 \langle \text{NProbe} \rangle &\rightarrow \textbf{num-consts} | \textbf{num-exprs} | \textbf{size} | \dots \\
 \langle \text{Constant} \rangle &\rightarrow \textbf{true} | \textbf{false} | \textbf{0} | \textbf{-1} | \textbf{1} | \dots
 \end{aligned}$$

Figure 1: Context-free grammar G the Z3 strategy language

3 Preliminaries

SMT Solvers, SMT Logics, and SMT-LIB Satisfiability Modulo Theories (SMT) solvers determine the satisfiability of first-order logic formulas, with the interpretation of symbols constrained by specific theories [Kroening and Strichman, 2016]. SMT-LIB refers to an international initiative aimed at facilitating research and development in SMT solvers [Barrett *et al.*, 2016]. The SMT-LIB initiative maintains a large library of SMT benchmarks, grouped by various SMT logics. A logic consists of one or more theories with certain restrictions, and is named after such theories and restrictions. “QF” refers to the restriction to quantifier-free formulas, “BV” refers to the theory of fixed-size bit-vectors, “S” refers to the theory of strings and regular expressions, “IA” and “RA” refer to integer and real arithmetic. “N” before “IA” or “RA” means the non-linear fragment of these arithmetics. SMT-COMP [Bobot *et al.*, 2023] is an annually held competition that arose from the SMT-LIB initiative for SMT solvers.

The Z3 Strategy Language The Z3 SMT solver offers a user-controllable strategy language, allowing users to craft their customized decision procedure algorithm. A strategy selects, sequences, and parameterizes tactics, where each tactic is a built-in reasoning step in Z3. The context-free grammar (CFG) G for the strategy language we consider in this research is shown in Figure 1, where variables are enclosed in angle brackets and terminals are highlighted in bold.

The start symbol $\langle \text{Strategy} \rangle$ represents a strategy and is defined recursively. A strategy may consist of either a single tactic or a series of tactics linked in sequence by the tactic combinator **then**. Each tactic can be configured with a variety of parameters. The combinator **or-else** applies the second strategy if the first strategy fails, while the combinator **try-for** makes the strategy fail if it does not return within the specified timeout (millisecond). Z3 also provides built-in probes,

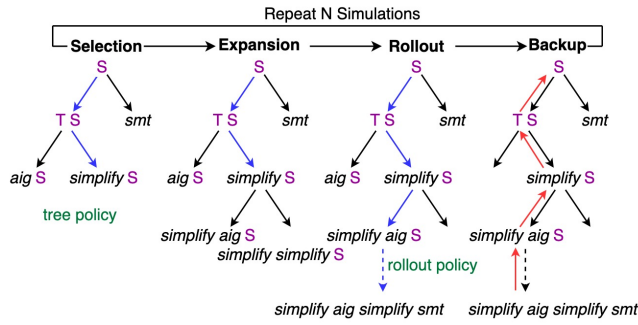


Figure 2: Illustration of the MCTS framework for strategy synthesis

which evaluate formula measures, e.g., the number of constants in the formula. Predicates over them can be built using relational operators. The **if** combinator constructs branching strategies based on these predicates. We refer readers to the official Z3 guide [Microsoft, 2023] for more information on the strategy language.

Monte Carlo Tree Search Monte Carlo Tree Search (MCTS) is a best-first search technique. It searches for the optimal decisions by estimating action values from numerous simulated trajectories. To search more efficiently, the method biases simulations towards previously rewarding trajectories, yet it maintains a balance by exploring less-visited paths as well. Alongside the simulations, an MCTS tree is progressively constructed to store the action value estimations. Each MCTS simulation consists of 4 steps:

1. *Selection*: Starting from the root node, a tree-search policy traverses the MCTS tree until a leaf node is selected. The Upper Confidence Bounds applied for Trees (UCT) [Kocsis and Szepesvári, 2006] is the most commonly used tree-search policy in MCTS. UCT balances exploiting the child node with the highest value estimation and exploring the less-visited children.
2. *Expansion*: The MCTS tree is expanded from the selected leaf node by adding child(ren) node(s) representing unexplored actions.
3. *Rollout*: If the selected leaf node is non-terminal, the simulation continues by subsequently choosing actions according to a rollout policy (usually a random policy) until reaching a terminal state.
4. *Backup*: After evaluation, the episode reward is backed up to update the action values alongside the traversed tree path.

4 z3alpha: MCTS for Strategy Synthesis

4.1 Modeling Strategy Synthesis as an MDP

If we view strategy synthesis as constructing a strategy string from G by sequentially applying production rules to the left-most variable, this process can be modeled as a deterministic Markov Decision Process (MDP). An MDP is a mathematical framework in which an agent makes action decisions in a series of states, with each action leading to a state transition. The agent seeks to maximize rewards over time through

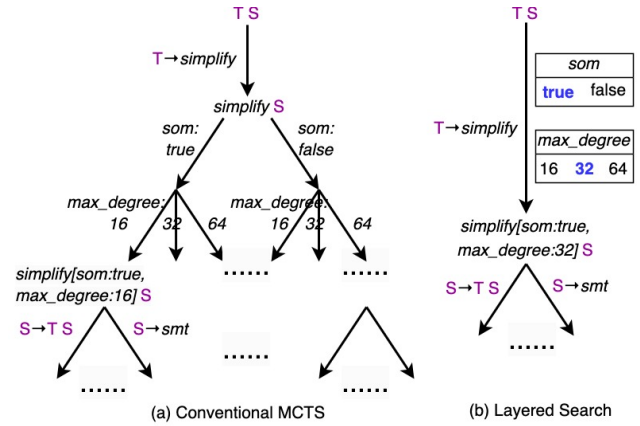


Figure 3: Comparison of the conventional MCTS and the layered search in treating tactic parameter tuning

choices of actions. In a deterministic MDP, each action results in a deterministic state transition.

In our formulation of the strategy synthesis problem, the *states* are the sentential strings derived from G , while the *actions* are the production-rule applications. The entire process of constructing a strategy is one single episode of the MDP. In an episode, the reward R_T is only received at the terminal step T . R_T is the performance measure of the synthesized strategy over a given benchmark set P .

Our reward system is designed to align with the evaluation criteria of SMT-COMP, prioritizing strategies that solve the highest number of instances. Simultaneously, when two strategies solve a similar number of P -instances, we want to steer the search towards the faster strategy. To embody these goals, we base our reward on the PAR-10 score over P . PAR-10 computes the average runtime for successfully solved instances and imposes a penalty for unsolved instances. The penalty is equal to the timeout value multiplied by a factor of 10.

With this modeling, the MDP search tree represents the strategy space, and the objective becomes identifying the path with the highest reward R_t in the search tree, which corresponds to the optimal strategy for P .

4.2 The MCTS Framework for Strategy Synthesis

We instantiate MCTS for this optimal strategy search problem. We use UCT as the tree policy in the *selection* phase and rollout randomly in the *rollout* phase. Notably, in the *backup* phase, we apply the max-backup rule [Sabharwal *et al.*, 2012; Sun *et al.*, 2023]. This approach updates the action values with the best observed return, rather than the average. It encourages more aggressive exploitation towards the previously best-performing strategy, aligning with our goal.

Therefore, the MCTS method continuously runs simulations, and in each simulation, the agent explores and assesses a single strategy, updating and retaining the best strategy seen so far. The MCTS stops when a simulation budget is reached. At the end of this process, the strategy with the highest reward R_T is selected and presented as the synthesized SMT strategy for the specified instance set P . Figure 2 illustrates

our basic MCTS framework, using a simplified CFG G' for illustrative purposes. G' is defined as $S \rightarrow T S \mid \mathbf{smt}$ and $T \rightarrow \mathbf{simplify} \mid \mathbf{aig}$, where S and T symbolize variables for strategy and tactic, respectively.

The primary challenge in strategy synthesis through this conventional MCTS method is the extensive time required to evaluate each strategy, which involves calling an SMT solver on all instances in P . This leads to a very limited exploration of potential paths, particularly given the immense search space created by the rich strategy language. To address this issue, we first add domain-knowledge rules restricting valid actions. For example, no tactic could be applied sequentially following a solver tactic such as **smt**. More importantly, we have introduced two heuristic methods, namely the layered search and the staged search, on top of the conventional MCTS, facilitating a deeper and more effective exploration of the strategy space.

4.3 Layered Search

To solve the above-mentioned challenge, we propose a layered search method to optimize the tactic parameters within strategy synthesis. As shown in our CFG G , each tactic can be paired with multiple parameters. Using the conventional MCTS with the grammar G , the selection of each candidate value for a parameter is represented by one production rule, and the agent needs to make sequential production-rule decisions to configure all parameters for a given tactic, leading to exponential growth in the problem search tree, as shown in Figure 3(a).

To address this issue, our layered search method approaches the tuning of each tactic parameter as a separate Multi-Armed Bandit (MAB) problem [Robbins, 1952]. As shown in Figure 3(b), the two parameters **som** and **max_degree** for one application of the tactic **simplify** are modeled as two MABs respectively. Each arm in the MAB represents one candidate value. For example in Figure 3(b), arm 16, 32, and 64 in the MAB **max_degree** are three pre-selected candidate values for this parameter. Note that the parameter MABs are associated with a tree edge, corresponding to one specific application of a tactic, not to this tactic in general. For instance, **simplify** may be applied several times in the strategy building. For each application, there will be two MABs representing the tuning of **som** and **max_degree** associated with it.

One key point is that these parameter-tuning MABs are not part of the main MCTS tree. They are engaged to select parameter values when their associated tree edge is traversed, and they are updated based on the episode reward during the *Backup* phase. However, such MABs do not expand the MCTS search tree after the parameter configuration, since they are separate components from the main search tree. This is in contrast to conventional MCTS, which also employs MAB principles to select among children nodes to explore, where these nodes constitute part of the search tree. For example, in Figure 3(a), the search tree is expanded sixfold to accommodate all possible combinations of these two parameters in the conventional MCTS framework. In contrast, in the layered search framework (Figure 3(b)), MABs for the two parameters are isolated from the search tree, creating no

additional branches in the tree.

The rationale behind the layer search is twofold. Firstly, tactics such as **simplify** may have dozens of parameters, and it is common for a tactic to be used multiple times within a strategy. Thus, navigating a search space that is fully expanded by all possible parameter combinations becomes impractical, especially given the time-intensive nature of strategy evaluation. Secondly, we argue that parameter tuning, although important, serves more as an auxiliary task in comparison to the tasks of tactic selection and sequencing. By employing the layered search method, we maintain the primary focus on the more important task. At the same time, the isolated MABs efficiently optimize the parameters without overwhelming the main search process.

4.4 Staged Search

While the layered search narrows down the search space, it does not alleviate the issue of significant time consumption required for each simulation evaluation. To discover a complex strategy, especially one involving nested branching, MCTS must expand over an enormous space and delve deeply, often resulting in an impractically lengthy search time. This is where the staged search comes into play.

The staged search method divides the entire search process into two stages. In the first stage, the MCTS focuses on finding high-performing *linear strategies* (actions introducing branches are not considered in the first stage). Here, a linear strategy is defined as a sequence of tactics without branching, where tactics are only connected with the combinator **then**. In the second stage, the MCTS looks for a single best strategy that combines these selected linear strategies. In other words, the second-stage MCTS works with the full grammar G but restricts the actions so that every explored strategy is a combination of the selected linear strategies through branching. By doing so, the key advantage is that the evaluation of the combined strategies can be done speedily based on the cached linear-strategy performances in the first stage, without costly SMT solver calls.

This is possible because, when executing any branched strategy S_c on a given instance f , there is always an equivalent sequence of branch-free strategy applications, $[S_0, S_1, \dots, S_N]$. For example, if we apply **((if is-pb (or-else (try-for (then simplify sat) 4000) smt) smt))** to a pseudo-boolean instance, the execution path is first to try **(then simplify sat)**, a linear strategy, for 4 seconds and then to execute **smt**, another linear strategy. Thus, when evaluating S_c for each input instance f , we first convert S_c to its equivalent linear strategy sequence $[S_0, S_1, \dots, S_N]$, where the performance of each linear strategy S_0, S_1, \dots, S_N on f is known and cached in the first stage. Then, the performance of S_c can be derived directly from these cached results without further calls to the SMT solver. This approach enables MCTS to traverse a significantly larger search space in the second stage, facilitating the discovery of effective complex strategies.

Another advantage of our staged search method is its adaptiveness to long timeouts. Strategy synthesis typically operates under short solver-instance timeouts, for example, 10 seconds in the `FastSMT` study. Increasing the timeout linearly raises the total synthesis time, making it prohibitively

Timeout(s)	Logic	Benchmark	Test Size	Z3alpha	Z3alpha0	FastSMT	Z3	CVC5	Z3str4
10	QF_BV	Sage2	6444	52.9	41.3	<u>51.3</u>	37.1	36.6	-
		core	270	99.6	<u>99.6</u>	100.0	75.6	82.6	-
	QF_NIA	AProVE	1712	94.8	<u>92.5</u>	90.3	90.0	69.9	-
		leipzig	68	91.2	89.1	88.2	<u>89.7</u>	25.0	-
	QF_NRA	hycomp	1982	91.4	<u>90.8</u>	89.1	84.7	85.5	-
	QF_LIA	entire logic	12476	76.2	-	31.2	<u>74.6</u>	64.8	-
	QF_LRA	entire logic	1003	<u>73.6</u>	-	74.0	71.0	63.6	-
	QF_S	entire logic	18173	99.0	-	-	98.2	<u>98.4</u>	97.2
60	QF_BV	Sage2	6444	69.8	-	67.7	57.7	77.8	-
300	QF_BV	Sage2	6444	<u>75.8</u>	-	72.6	69.8	93.3	-

Table 1: **Z3alpha vs. SOTA Solvers:** Percentage (%) of instances solved from the selected SMT-LIB benchmarks across six SMT logics (In each experiment, the result of the leading tool is highlighted in bold, while the second-best is underscored.)

costly to synthesize strategies with extended timeouts, like 5 minutes or more, through direct evaluation. Our staged search method provides a practical approach to synthesizing strategies under such extended timeouts. In the first stage, the linear strategy candidates are still selected based on their performance with a short timeout period (e.g., 10 seconds), ensuring broad exploration. Then, only the selected strategies are re-evaluated under the specified long timeout (e.g., 5 minutes). The second-stage MCTS leverages these re-evaluation caches, enabling the synthesis of customized strategies for these extended timeout scenarios.

5 Experiment Design, Results, and Analysis

5.1 Experimental Design

We evaluated Z3alpha across six logics, namely QF_{BV, NIA, NRA, LIA, LRA, S}. We benchmarked our performance against the SOTA strategy synthesis tool, FastSMT, as well as the default handcrafted strategy in Z3. CVC5 was also included as a baseline for comparison. To assess the robustness of Z3alpha, we designed a series of experiments tailored to different scenarios.

Experimental Design for Specific Classes of Benchmarks

Section 5.3 describes the evaluation of the solvers on five important SMT-LIB benchmark sets, namely, Sage2 (QF_BV), core (QF_BV), AProVE (QF_NIA), leipzig (QF_NIA), and hycomp (QF_NRA), covering three different logics. These specific benchmark sets were chosen as they were also utilized in the FastSMT study, following the same training-testing split. The size of these benchmark sets varies from 167 to 7436 instances. A 10-second timeout was chosen for evaluation, a common practice in the SMT strategy synthesis research [Ramírez *et al.*, 2016; Balunovic *et al.*, 2018]. This timeout period was also justified as Z3alpha solved over 90% of the testing instances in four out of the five benchmark sets within this time frame. Additionally, in this section, we included a version of our tool, Z3alpha0, that did not employ staged search, for ablation study purposes.

Experimental Design for Benchmarks from Diverse Applications

In Section 5.4, we extended our experiments to the entire SMT-LIB QF_LIA and QF_LRA benchmarks, which are obtained from diverse applications. These experiments were intended to evaluate the versatility and adaptabil-

ity of Z3alpha in handling diverse problem sets. For both QF_LIA and QF_LRA, we randomly selected 750 SMT-LIB instances for our training set, and utilized all remaining instances in the logic (12,476 instances and 1,003 instances for QF_LIA and QF_LRA, respectively) as our testing set.

Experimental Design for Long Timeout In Section 5.5, we expanded our experiments to include long timeouts, specifically 1-minute and 5-minute durations. These evaluations were conducted on the most challenging benchmark set, Sage2, from Section 5.3. Z3alpha synthesized new strategies for these scenarios, using the method described in Section 4.4. FastSMT used the identical strategy as in Section 5.3 for these extended timeout cases, adhering to the approach described in its original paper.

Experimental Design for User-defined Tactics

Z3 allows users to implement new tactics. Section 5.6 evaluated the capability of Z3alpha in synthesizing strategies that integrate both user-defined and built-in tactics. This experiment targeted the QF_S logic. In recent years, our team implemented new string tactics, such as the arrangement-based solver Z3str3 [Berzish *et al.*, 2017], the Length Abstraction Solver (LAS) [Mora *et al.*, 2021], and specialized rewrite rules for regular expressions [Berzish *et al.*, 2021] for string constraint problems. These tactics are in addition to the default built-in sequence solver in Z3. Z3str4 [Mora *et al.*, 2021] combines the aforementioned tactics using a meticulously handcrafted strategy. In this experiment, we leveraged Z3alpha to construct a strategy using the same set of tactics as in Z3str4 and then compared their performances. Z3alpha was trained on 750 randomly chosen QF_S instances from SMT-LIB. The testing was conducted on all the remaining 18,173 instances in the logic.

5.2 Experimental Setup

For every experiment, there were a training instance set and a testing instance set. Z3alpha, as well as FastSMT, synthesized strategies based on the training set while reporting experiment results on the testing set. The approach aligns with our problem statement, which aims to synthesize a strategy based on a representative set P , that is expected to generalize effectively on a larger scale.

In each experiment, Z3alpha first selected 20 linear strategies from 800 first-stage MCTS simulations. During the

second stage with 300,000 simulations, `Z3alpha` searched for the most effective strategy that combined these linear strategy candidates. The final synthesized strategy was the one with the lowest PAR-10 score during the second-stage search. To keep a similar time budget, the non-staged-search version `Z3alpha0` ran MCTS for 1,000 simulations with the full CFG G , in search of the best strategy.

Competing Solvers `Z3alpha` was implemented in Python 3.10 and was integrated with `Z3-4.12.2`. `FastSMT` was also integrated with the same version of `Z3`. Both tools constructed their strategies using the identical tactic and parameter set offered by `Z3`, and executed these strategies with `Z3`. Baseline solvers used in the experiment were `Z3-4.12.2` and `CVC5-1.0.5`. We compared our performance with `FastSMT` in all experiments other than the experiment described in Section 5.6, since `Z3str4` did not provide Python APIs for the user-defined tactics that were required by `FastSMT`.

Computational Environments Both our synthesis and testing were conducted on a high-performance CentOS 7 cluster equipped with Intel E5-2683 v4 (Broadwell) processors running at 2.10 GHz, accompanied by 75 gigabytes of memory.

Variability Both the `Z3alpha` and `FastSMT` algorithms make use of randomness, leading to the possibility of synthesizing different strategies in separate runs. To account for this variability, the results in Section 5.3 were average from five runs with different random seeds. Since little variability was found in Section 5.3 and the much more intense computational nature of the later experiments, we only report results from one run in later sections.

Metrics Consistent with the evaluation criteria used in the SMT-COMP, our performance metric was based on the number of correctly solved instances. For clearer comprehension, we present these results as a percentage, reflecting the proportion of solved instances out of the total tested.

5.3 Analysis of QF_BV, QF_NIA, QF_NRA Results

The first part of Table 1 summarizes the results on the five selected benchmark sets, namely `Sage2`, `core`, `AProVE`, `leipzig`, and `hycomp`, across the `QF_BV`, `QF_NIA`, and `QF_NRA` logics. Notably, `Z3alpha` surpassed the default `Z3` strategy, as well as `CVC5` in all of these benchmark sets, achieving the leading position in four of the five sets among all tested tools. In the particularly challenging `QF_BV` benchmark set `Sage2`, `Z3alpha` excelled by solving an impressive 42.7% more instances than the default strategy did. Furthermore, `Z3alpha` outperformed `Z3alpha0` across all benchmarks, underscoring the effectiveness of the staged search. The synthesis time for `Z3alpha` was on par with, and in most experiments, less than, the synthesis time for `FastSMT`. For instance, while the strategy synthesis for `AProVE` took 759.6 minutes for `FastSMT`, `Z3alpha` completed the task in 293.1 minutes, in which the stage-1 and stage-2 took 213.7 and 79.4 minutes respectively. One key distinction between the synthesized strategies from `Z3alpha` and `FastSMT` was that `Z3alpha` strategies are

more interpretable. For example, the `FastSMT` strategies for `AProVE` can have more than a thousand branches, while `Z3alpha` strategies usually have less than five branches.

5.4 Analysis of QF_LIA and QF_LRA Results

`Z3alpha` demonstrated consistent performance on the entire logics of `QF_LIA` and `QF_LRA`, as shown in Table 1. `Z3alpha` solved 2.2% and 3.7% more instances than `Z3` in `QF_LIA` and `QF_LRA`, respectively. While `FastSMT` solved four more instances than `Z3alpha` in `QF_LRA`, its performance suffered significantly in `QF_LIA`, solving 58.2% fewer instances than the default `Z3` strategy.

5.5 Analysis of QF_BV Results with Long Timeout

The results for experiments of 1-minute-timeout and 5-minute-timeout are shown in Table 1 as well. In every scenario, `Z3alpha` continued to maintain superior performance compared to both `FastSMT` and `Z3`. The performance advantage over `FastSMT` slightly increased when the timeouts were extended. However, an important shift was the significantly better performance of `CVC5` over the `Z3`-based methods for `Sage2` under long timeouts. This suggests a promising future research direction of extending the strategy synthesis method across different solvers.

5.6 Results with User-Defined Tactics for QF_S

The test results for `Z3alpha` with user-defined `QF_S` tactics are shown in the `QF_S` row of Table 1. `Z3alpha` demonstrated superior performance over all baseline solvers. Interestingly, the handcrafted `Z3str4` strategy, despite employing the same tactic portfolio as `Z3alpha`, performed worse than the default `Z3` strategy. The under-performance could be attributed to two factors: (1) the `Z3str4` strategy was optimized for logics including both `QF_S` and `QF_SLIA`, which could be sub-optimal for `QF_S` alone; (2) the tuning of the `Z3str4` strategy was carried out on an earlier version of `Z3`. These observations underscore the value of automated strategy customization, in terms of tailoring for specific problem contexts and adapting to solver versions.

6 Conclusions

In this work, we present `Z3alpha`, a novel MCTS-based method for SMT strategy synthesis. `Z3alpha` introduces layered and staged search heuristics upon the conventional MCTS framework, enabling a low-cost and effective search within the expansive strategy space. The superiority of `Z3alpha` was demonstrated by extensive experiments across six SMT logics. In all the experiments, `Z3alpha` consistently surpassed the default `Z3` solver and outperformed both `FastSMT` and `CVC5` in most testing cases.

Our method is currently implemented only on `Z3`, because other prominent solvers, like `CVC5`, to the best of our knowledge, do not offer an interface to group preprocessing and solving steps flexibly. We hope our strong empirical results will encourage a universal user-controllable strategy language in the SMT community. There is substantial potential to further enhance solver performance by applying our method across tactics from different solvers, thereby leveraging their complementary strengths.

Acknowledgements

We thank Kate Larson, Arie Gurfinkel, and Mark Crowley for their valuable feedback; Kaihang Jiang for his contribution to code testing and data collection; and the Digital Research Alliance of Canada for providing the computational resources and technical support. The work of Zhengyang Lu is supported by the Engineering Excellence Doctoral Fellowship (EEDF) at the University of Waterloo; the work of Florin Manea was supported by the DFG-Heisenberg grant no. 466789228.

References

- [Balunovic *et al.*, 2018] Mislav Balunovic, Pavol Bielik, and Martin Vechev. Learning to solve SMT formulas. In *Advances in Neural Information Processing Systems 31*, pages 10337–10348. Curran Associates, 2018.
- [Barrett *et al.*, 2016] Clark Barrett, Pascal Fontaine, and Cesare Tinelli. The satisfiability modulo theories library (SMT-LIB). <http://smtlib.cs.uiowa.edu/index.shtml>, 2016. Accessed: 2024-01-16.
- [Berzish *et al.*, 2017] Murphy Berzish, Vijay Ganesh, and Yunhui Zheng. Z3str3: A string solver with theory-aware heuristics. In *2017 Formal Methods in Computer Aided Design (FMCAD)*, pages 55–59. IEEE, 2017.
- [Berzish *et al.*, 2021] Murphy Berzish, Mitja Kulczynski, Federico Mora, Florin Manea, Joel D Day, Dirk Nowotka, and Vijay Ganesh. An SMT solver for regular expressions and linear arithmetic over string length. In *International Conference on Computer Aided Verification*, pages 289–312. Springer, 2021.
- [Bobot *et al.*, 2023] François Bobot, Martin Bromberger, and Jochen Hoenicke. SMT-COMP 2023. <https://smt-comp.github.io/2023/>, 2023. Accessed: 2024-01-16.
- [Browne *et al.*, 2012] Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1):1–43, 2012.
- [Cadaru *et al.*, 2008] Cristian Cadaru, Vijay Ganesh, Peter M Pawlowski, David L Dill, and Dawson R Engler. EXE: Automatically generating inputs of death. *ACM Transactions on Information and System Security (TISSEC)*, 12(2):1–38, 2008.
- [Cameron *et al.*, 2022] Chris Cameron, Jason Hartford, Taylor Lundy, Tuan Truong, Alan Milligan, Rex Chen, and Kevin Leyton-Brown. Monte carlo forest search: UNSAT solver synthesis via reinforcement learning. *arXiv preprint arXiv:2211.12581*, 2022.
- [Chen *et al.*, 2021] Zhenbang Chen, Zehua Chen, Ziqi Shuai, Guofeng Zhang, Weiyu Pan, Yufeng Zhang, and Ji Wang. Synthesize solving strategy for symbolic execution. In *Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis*, pages 348–360, 2021.
- [De Moura and Bjørner, 2008] Leonardo De Moura and Nikolaj Bjørner. Z3: An efficient SMT solver. In *International conference on Tools and Algorithms for the Construction and Analysis of Systems*, pages 337–340. Springer, 2008.
- [De Moura and Bjørner, 2011] Leonardo De Moura and Nikolaj Bjørner. Satisfiability modulo theories: introduction and applications. *Communications of the ACM*, 54(9):69–77, 2011.
- [De Moura and Passmore, 2013] Leonardo De Moura and Grant Olney Passmore. The strategy challenge in SMT solving. *Automated Reasoning and Mathematics: Essays in Memory of William W. McCune*, pages 15–44, 2013.
- [De Waard *et al.*, 2016] Maarten De Waard, Diederik M Roijers, and Sander CJ Bakkes. Monte carlo tree search with options for general video game playing. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8. IEEE, 2016.
- [Gelly and Silver, 2011] Sylvain Gelly and David Silver. Monte-carlo tree search and rapid action value estimation in computer go. *Artificial Intelligence*, 175(11):1856–1875, 2011.
- [Gurfinkel *et al.*, 2015] Arie Gurfinkel, Temesghen Kahsai, Anvesh Komuravelli, and Jorge A Navas. The SeaHorn verification framework. In *International Conference on Computer Aided Verification*, pages 343–361. Springer, 2015.
- [Jha *et al.*, 2024] Piyush Jha, Zhengyu Li, Zhengyang Lu, Curtis Bright, and Vijay Ganesh. AlphaMapleSAT: An MCTS-based cube-and-conquer SAT solver for hard combinatorial problems. *arXiv preprint arXiv:2401.13770*, 2024.
- [Johanson, 2013] Michael Johanson. Measuring the size of large no-limit poker games. *arXiv preprint arXiv:1302.7008*, 2013.
- [Khalil *et al.*, 2022] Elias B Khalil, Pashootan Vaezipoor, and Bistra Dilkina. Finding backdoors to integer programs: a monte carlo tree search framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3786–3795, 2022.
- [Kocsis and Szepesvári, 2006] Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *European conference on machine learning*, pages 282–293. Springer, 2006.
- [Koza, 1994] John R Koza. Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4:87–112, 1994.
- [Kroening and Strichman, 2016] Daniel Kroening and Ofer Strichman. *Decision Procedures: An Alogorithm Point of View*. Springer, 2016.
- [Lu, 2023] Zhengyang Lu. AlphaSMT: A reinforcement learning guided SMT solver. Master’s thesis, University of Waterloo, 2023.
- [Mankowitz *et al.*, 2023] Daniel J Mankowitz, Andrea Michi, Anton Zhernov, Marco Gelmi, Marco Selvi,

- Cosmin Paduraru, Edouard Leurent, Shariq Iqbal, Jean-Baptiste Lespiau, Alex Ahern, et al. Faster sorting algorithms discovered using deep reinforcement learning. *Nature*, 618(7964):257–263, 2023.
- [Microsoft, 2023] Microsoft. Online Z3 guide. <https://microsoft.github.io/z3guide/>, 2023. Accessed: 2024-01-16.
- [Mora et al., 2021] Federico Mora, Murphy Berzish, Mitja Kulczynski, Dirk Nowotka, and Vijay Ganesh. Z3str4: A multi-armed string solver. In *Formal Methods: 24th International Symposium, FM 2021, Virtual Event, November 20–26, 2021, Proceedings 24*, pages 389–406. Springer, 2021.
- [Poole and Mackworth, 2010] David L Poole and Alan K Mackworth. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2010.
- [Pulina and Tacchella, 2012] Luca Pulina and Armando Tacchella. Challenging SMT solvers to verify neural networks. *Ai Communications*, 25(2):117–135, 2012.
- [Ramírez et al., 2016] Nicolás Gálvez Ramírez, Youssef Hamadi, Eric Monfroy, and Frédéric Saubion. Evolving SMT strategies. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 247–254. IEEE, 2016.
- [Robbins, 1952] Herbert Robbins. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58(5):527–535, 1952.
- [Ross et al., 2011] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.
- [Sabharwal et al., 2012] Ashish Sabharwal, Horst Samulowitz, and Chandra Reddy. Guiding combinatorial optimization with UCT. In *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems: 9th International Conference, CPAIOR 2012, Nantes, France, May 28–June 1, 2012. Proceedings 9*, pages 356–361. Springer, 2012.
- [Silver et al., 2016] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.
- [Silver et al., 2017] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.
- [Song et al., 2008] Dawn Song, David Brumley, Heng Yin, Juan Caballero, Ivan Jager, Min Gyung Kang, Zhenkai Liang, James Newsome, Pongsin Poosankam, and Prateek Saxena. BitBlaze: A new approach to computer security via binary analysis. In *Information Systems Security: 4th International Conference, ICISS 2008, Hyderabad, India, December 16-20, 2008. Proceedings 4*, pages 1–25. Springer, 2008.
- [Sun et al., 2023] Fangzheng Sun, Yang Liu, Jian-Xun Wang, and Hao Sun. Symbolic physics learner: Discovering governing equations via monte carlo tree search. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.
- [Sutton et al., 1999] Richard S Sutton, Doina Precup, and Satinder Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.
- [Zhang et al., 2018] Zhenya Zhang, Gidon Ernst, Sean Sedwards, Paolo Arcaini, and Ichiro Hasuo. Two-layered falsification of hybrid systems guided by monte carlo tree search. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(11):2894–2905, 2018.