Heuristics for deriving adaptive homing and distinguishing sequences for nondeterministic finite state machines

Natalia Kushik, Husnu Yenigun

Heuristics for Deriving Adaptive Homing and Distinguishing Sequences for Nondeterministic Finite State Machines

Natalia Kushik[1,2], Husnu Yenigun[3]

[1] Tomsk State University, Tomsk, Russia
[2] Telecom SudParis, Evry, France
`ngkushik@gmail.com`
[3] Sabanci University, Istanbul, Turkey
`yenigun@sabanciuniv.edu`

Abstract. Distinguishing Sequence (DS) and Homing Sequences (HS) are used for state identification purposes in Finite State Machine (FSM) based testing. For deterministic FSMs, DS and HS related problems are well studied, for both preset and adaptive cases. There are also recent algorithms for checking the existence and constructing Adaptive DS and Adaptive HS for nondeterministic FSMs. However, most of the related problems are proven to be PSPACE-complete, while the worst case height of Adaptive DS and HS is known to be exponential. Therefore, novel heuristics and FSM classes where they can be applied need to be provided for effective derivation of such sequences. In this paper, we present a work in progress on the minimization of Adaptive DS and Adaptive HS for nondeterministic FSMs.

Keywords. Nondeterministic Finite State Machines, Adaptive Homing Sequence, Adaptive Distinguishing Sequence, Novel heuristics.

1   Introduction

Distinguishing Sequences (DS) and Homing Sequences (HS) are used for state identification purposes in Finite State Machine (FSM) based testing [1-3]. A DS identifies the initial state of the FSM under investigation, while an HS is used to identify the final state after the sequence has been applied. A sequence is *adaptive* if the next input to be applied to an FSM under investigation is chosen based on the previously observed outputs, and the sequence is *preset* if the outputs need to be observed only after the entire sequence applied. The methods to derive preset/adaptive HS/DS are well elaborated for complete and deterministic FSMs [3-5]. Even though the length of most of these sequences is polynomial w.r.t. the number of FSM states, the current complexity of digital systems and software makes it almost impossible to derive a complete deterministic behavior of the system integrated into the overall software and/or hardware environment. Moreover, current specifications of telecommunication protocols and other digital systems include an optionality of output responses under

the same queries. That is the reason why researchers turn their attention towards special FSM types, and in particular, nondeterministic FSMs.

A method for deriving a DS for two states of an observable nondeterministic FSM has been proposed in [6], where the length of this sequence is shown to be exponential for nondeterministic FSMs. The upper bound on the length of an HS for an observable nondeterministic FSM is shown to be exponential as well, and this upper bound is reachable [7]. As for adaptive experiments for nondeterministic FSMs, it has been shown that the length of a shortest adaptive DS for two states of an observable nondeterministic FSM with n states is at most $n(n-1)/2$ [8]. Whenever such an adaptive sequence is derived to distinguish $m > 2$ states of an observable nondeterministic FSM, the length of this sequence grows exponentially [9]. The problem of checking the existence of a preset DS is known to be PSPACE-complete even for complete *deterministic* FSMs [3]. The latter means one cannot directly apply these techniques to effectively derive test sequences against FSM specifications. That is the reason why a number of heuristics to decrease such complexity have been proposed. In particular, various heuristic methods to construct reduced adaptive DS for complete deterministic FSMs have been proposed in [10]. However, it has been also proven that constructing a minimal adaptive DS for complete deterministic FSMs is an NP-hard problem [11]. Therefore, novel heuristics need to be provided for effective derivation of adaptive HS and adaptive DS for nondeterministic FSMs.

In this paper, we focus on two techniques for deriving adaptive DS and adaptive HS for nondeterministic FSMs. In particular, we present a class of nondeterministic FSMs for which it is possible to construct an adaptive DS without using any nondeterministic transitions. We also argue how existing adaptive DS minimization approaches can be used for such nondeterministic FSMs. As for adaptive HS, we address a method proposed in [9] that constructs an adaptive HS for an observable nondeterministic FSM when each state pair is adaptively homing. We discuss how this method can be improved so that the resulting adaptive HS can be shorter. Therefore, the main contributions of this work in progress are two novel heuristics for effective derivation of adaptive DSs for nondeterministic FSMs and adaptive HSs for observable nondeterministic FSMs.

The paper is organized as follows. Section 2 contains preliminaries. A heuristic method for effective adaptive DS derivation for nondeterministic FSMs is given in Section 3 while Section 4 presents an approach for minimizing adaptive HS for observable nondeterministic FSMs. Section 5 concludes the paper.

2   Preliminaries

In this paper, we focus on minimizing adaptive HS and DS for nondeterministic FSMs. As usual, an *FSM* S is a 4-tuple (S, I, O, h_S), where S is a finite set of states; I and O are finite non-empty disjoint sets of inputs and outputs, respectively; $h_S \subseteq S \times I \times O \times S$ is a *transition relation*, where a 4-tuple $(s, i, o, s') \in h_S$ is a *transition*. An FSM is *complete* if for each pair $(s, i) \in S \times I$ there exists $(o, s') \in O \times S$ such that $(s, i, o, s') \in h_S$. Otherwise it is called *partially specified*. If for some pair $(s, i) \in S \times I$, there exist two transitions $(s, i, o_1, s_1), (s, i, o_2, s_2) \in h_S$, such that $o_1 \neq o_2$ or $s_1 \neq s_2$ then S is called *nondeterministic*. For a nondeterministic FSM the nondetermin-

ism is *observable* if each input/output pair *i/o* uniquely identifies the successor of each FSM state (if it exists). Given a state *s* of S and an input/output sequence α/β, the *α/β-successor* of state *s* is the set of all states that are reached from *s* via an application of α when an output reaction β is produced. Note that for an observable FSM S, the cardinality of the *α/β-successor* of state *s* is at most one for any input/output sequence α/β.

A sequence α is a *distinguishing* (*homing*) sequence (DS/HS) for the FSM S if after applying α and observing output reaction β one can uniquely conclude about the initial (final/current) state of S. The length of preset DS and HS is exponential for nondeterministic FSMs, however sometimes this length can be shorter when adaptivity is used. An adaptive DS/HS can be represented as a tree or as a specific acyclic single-input output-complete FSM that are called test cases [3, 12].

A number of methods for deriving adaptive DS and HS for nondeterministic FSMs have been proposed (see, for example [9]). However, the length of the corresponding sequence in general case remains exponential and thus, novel heuristics need to be provided for minimizing the length of such sequences.

3 Nondeterministic FSMs with a Deterministic Adaptive DS

For a complete deterministic FSM with *n* states, it is known that $n(n-1)/2$ is the tight upper bound for the height of an adaptive DS [3]. However, for nondeterministic FSMs, the height of an adaptive DS is exponential in general. In this section, we present a class of nondeterministic FSMs for which the height of an adaptive DS is at most $2n^2 - n - 1$ and existing algorithms of adaptive DS minimization can be readily applied.

A transition $(s, i, o', s') \in h_S$ is a *deterministic transition* if for any transition $(s, i, o'', s'') \in h_S$ we have $o'' = o'$ and $s'' = s'$. For a given FSM $S = (S, I, O, h_S)$, we define *the deterministic projection* $S^d = (S, I^d, O^d, h^d_S)$ of S as follows. S^d and S have the same set *S* of states. For a transition $(s, i, o', s') \in h_S$, $(s, i, o', s') \in h^d_S$, if and only if (s, i, o', s') is a deterministic transition in S. I^d and O^d consist of the inputs and outputs used in the transitions in h^d_S. Intuitively, S^d is the same FSM as S where the deterministic transitions are preserved but all other transitions are removed. Hence, S^d is a deterministic FSM by definition. It is easy to see that an adaptive DS for S^d can be directly used as an adaptive DS for S as well, and it will be a *deterministic adaptive DS* in the sense that only the deterministic transitions are used throughout the application of the adaptive DS in S. In the lucky case that S^d is a complete deterministic FSM, the existence check and the adaptive DS construction algorithms given in [3] can be directly applied. However, in general, S^d is a partially specified deterministic FSM. Moreover, in the worst case the transition relation of S^d can be empty, and in this case, the question of existence of an ADS for S remains open.

Although there usually is a complexity jump for the algorithms when one considers partially specified FSMs, this is not the case for the problems related to adaptive DS. In [13] a polynomial time algorithm is provided to check if a partially specified FSM has an adaptive DS or not. In this paper, we adapt this algorithm for efficient derivation of an ADS for a nondeterministic FSM that requires a partial deterministic projection. We also improve the upper bound on the length of such ADS compared to the

one given in [13]. In fact, the approach given in [13] is based on constructing a complete deterministic FSM $C(\mathsf{S}^d)$ from a given partially specified deterministic FSM S^d such that there exists an adaptive DS with the height H for $C(\mathsf{S}^d)$ iff there exists an adaptive DS with height $H - 1$ for S^d. The number of states in $C(\mathsf{S}^d)$ is $2n$, where n is the number of states in S^d. Therefore, it is possible to check if there exists an adaptive DS for S^d (and hence a deterministic ADS for S) in $O(pn\lg n)$ time [3], where p is the cardinality of I^d.

If there exists an adaptive DS for $C(\mathsf{S}^d)$, one can then use the LY algorithm (the adaptive DS construction algorithm given in [3]) to construct an adaptive DS. Although [13] provides an upper bound of $\pi^2 n^2/3$ for the height H of the adaptive DS constructed for $C(\mathsf{S}^d)$, using the result in [14], the LY algorithm actually constructs an adaptive DS with the height at most $H = 2n^2 - n$ for $C(\mathsf{S}^d)$. Therefore, the height of the adaptive DS for S^d is at most $H - 1 = 2n^2 - n - 1$.

Note that the LY algorithm does not aim for the minimization of the adaptive DS it constructs. However, there exist heuristics for the minimization of adaptive DSs with respect to different metrics (e.g. height, total external path length, etc.) for complete deterministic FSMs [10]. Since $C(\mathsf{S}^d)$ is a complete deterministic FSM, one can directly use these heuristics in order to construct minimized adaptive DSs for $C(\mathsf{S}^d)$, and hence for the nondeterministic FSM S. Note that S can be nonobservable as well.

4   Minimizing adaptive HS for Nondeterministic FSMs

In this section, we discuss how to optimize the procedure for deriving an adaptive HS for an observable FSM S. The procedure is taken from [9] while the complexity of the related problem is given in [15]. The homing test case derivation strategy is based on the condition that each state pair of S is adaptively homing. In other words, there exists a homing test case for S if there exists an adaptive homing sequence for each subset $\{s_i, s_j\} \subseteq S$ of states of the observable nondeterministic FSM S. A test case $\mathsf{P}_{i,j}$ is a *homing test case* for the subset $\{s_i, s_j\} \subseteq S$ of states if for every input/output sequence α/β defined in $\mathsf{P}_{i,j}$, the α/β-successor of the subset $\{s_i, s_j\} \subseteq S$ has at most one state.

If there exists a homing test case for the FSM S then the set S is a *homing set* and the test case P is a *homing test case for the set S* or the test case P *homes* states of the set S. Otherwise, the set S is not homing. The homing test case for the set $S = \{s_1, .., s_n\}$ is derived iteratively. As mentioned above, $\mathsf{P}_{i,j}$ is a homing test case for the subset (pair) $\{s_i, s_j\} \subseteq S$ of FSM states. The procedure starts with a homing test case $\mathsf{P}_{1,2}$ for the set $\{s_1, s_2\}$, then state s_3 is added to the subset $\{s_1, s_2\} \subseteq S$. Each input/output sequence α/β that is defined in the test case $\mathsf{P}_{1,2}$ is applied at state s_3. If the sequence α/β is defined at state s_3, then the deadlock state q in the test case $\mathsf{P}_{1,2}$ is replaced with a test case $\mathsf{P}_{q,z}$, where q is the α/β-successor of state pair $\{s_1, s_2\}$, while z is the α/β-successor of state s_3. All the input/output sequences α/β' that are defined at state s_3 but not defined in the test case $\mathsf{P}_{1,2}$ are also included into the test case $\mathsf{P}_{1,2,3}$. Proceeding in this way by iteratively adding the remaining states $s_4, s_5, ..., s_n$, the test case $\mathsf{P}_{1,2,...,n}$ is derived.

We note that there exist various homing test cases $\mathsf{P}_{i,j}$ for the same state pair $\{s_i, s_j\} \subseteq S$. Therefore, the first optimization step can be to consider a somehow opti-

mal test case $P_{i,j}$ that is chosen for the first state pair and for those test cases $P_{q,z}$ that are appended iteratively, when $P_{1,2,\ldots,n}$ is being derived. On the other hand, the length of the resulting test case can significantly depend on the choice of state s_j that is added to the test case $P_{1,2,\ldots,j-1}$. Thus, the second optimization criterion can be on an optimal choice of the next state s_j when the test case $P_{1,2,\ldots,j-1}$ is already constructed. We further discuss how these two ideas can be taken into account to effectively derive a test case $P_{1,2,\ldots,n}$. We mention that when optimizing the length of the corresponding adaptive HS, represented as a test case, we focus on optimizing not only the height of the corresponding tree but also the number of transitions in the acyclic FSM.

The first type of heuristics is related to making an optimal choice between possible initial state pairs $\{s_i, s_j\} \subseteq S$, $i < j$, as well as defining the best homing test case $P_{i,j}$ for this pair. As it is shown in [9], the length of the overall homing test case $P_{1,2,\ldots,n}$ significantly depends even on the first input to be included into the test case $P_{1,2}$. As the test case $P_{1,2,\ldots,j}$ is defined based on the test case $P_{1,2,\ldots,j-1}$, the height / the number of transitions of it significantly depends on the number of sequences that are defined in the test case $P_{1,2,\ldots,j-1}$ and in state s_j that will be added to the root of the corresponding successor tree. That is the reason why we suggest to choose a pair $\{s_i, s_j\} \subseteq S$ to start with in such a way that its homing test case $P_{i,j}$ is a homing test case for some other state pairs. It is naturally to assume that the more state pairs are 'covered' by a homing test case $P_{i,j}$, the better is the resulting test case $P_{1,2,\ldots,n}$. Therefore, we suggest to choose the first state pair $\{s_i, s_j\} \subseteq S$, $i < j$, such that there exists a homing test case $P_{i,j}$ for this pair that homes as much state pairs in S as possible.

The idea behind the second optimization step is the same as in the previous case. In particular, we suggest choosing the next state s_j to be added to a test case $P_{1,2,\ldots,j-1}$ in such a way that the number of input/output sequences that are defined in $P_{1,2,\ldots,j-1}$ and are not defined at state s_j would be minimal. This fact can help to reduce the number of homing test cases $P_{q,z}$ that have to be appended at each state z, where z is the α/β-successor of state s_j, and q is the α/β-successor of the set $\{s_1, \ldots, s_{j-1}\}$. Moreover, as for each pair of states $\{q, z\}$ there can exist various adaptive homing test cases, the best choice of the test case $P_{q,z}$ can affect the overall height / transition number of the test case $P_{1,2,\ldots,n}$. Therefore, it is also necessary to consider which test case $P_{q,z}$ should be chosen for an intermediate state pair $\{q, z\} \subseteq S$.

We mention that both optimization steps for deriving an adaptive homing test case $P_{1,2,\ldots,n}$ need to be thoroughly estimated. On one hand, theoretical investigation on FSM classes that have shorter adaptive HS need to be elaborated, and on the other, experimental evaluation on the efficiency of proposed heuristics needs to be performed.

5   Conclusion

In this paper, we proposed two heuristic methods for optimizing the size of adaptive DS and adaptive HS for nondeterministic FSMs. The topic of such optimization is always motivated by the fact that the length of such sequences is exponential for nondeterministic FSMs. Moreover, related decision problems on the existence of such sequences even for the preset case are known to be PSPACE-complete. We note that this current work in progress only presents the ideas behind optimization criteria and

optimization techniques to achieve shorter resulting trees for adaptive DS or HS. As a future work, we plan to perform experimental evaluation over machines of various types (random, protocol specifications, etc.) to estimate the efficiency of the proposed techniques.

Acknowledgments

References

1. Dorofeeva, R., El-Fakih, K., Maag, S., Cavalli, A. R., Yevtushenko, N.: FSM-based conformance testing methods: a survey annotated with experimental evaluation. Information and Software Technology, 52, pp. 1286-1297, Elsevier (2010)
2. Gill, A.: State-identification experiments in finite automata. Information and Control, pp. 132-154 (1961)
3. Lee, D., Yannakakis, M.: Testing finite-state machines: state identification and verification. IEEE Trans. on Computers, 43(3), pp. 306-320 (1994)
4. Kohavi, Z.: Switching and Finite Automata Theory. McGraw-Hill, New York (1978)
5. Moore, E.F.: Gedanken-experiments on sequential machines. In Automata Studies (Annals of Mathematical Studies no.1), Princeton University Press, pp. 129-153 (1956)
6. Spitsyna, N., El-Fakih, K., Yevtushenko, N.: Studying the separability relation between finite state machines. Software Testing, Verification and Reliability, 17(4), pp. 227-241 (2007)
7. Kushik, N., Yevtushenko, N.: On the Length of Homing Sequences for Nondeterministic Finite State Machines. In: Proc. of CIAA'2013, pp. 220-231 (2013)
8. Alur, R. Courcoubetis, C., Yannakakis, M.: Distinguishing tests for nondeterministic and probabilistic machines. In Proc. of the 27th ACM Symposium on Theory of Computing, pp. 363-372 (1995)
9. Kushik, N., El-Fakih, K., Yevtushenko, N., Cavalli A. R.: On Adaptive Experiments for Nondeterministic Finite State Machines. Software Tools for Technology Transfer, DOI 10.1007/s10009-014-0357-7 (2014)
10. Türker, U.C., Ünlüyurt, T., Yenigün, H.: Lookahead-Based Approaches for Minimizing Adaptive Distinguishing Sequences. In Proc. of International Conference on Testing Software and Systems (ICTSS 2014), pp. 32-47 (2014)
11. Türker, U.C., Yenigün, H.: Hardness and inapproximability of minimizing adaptive distinguishing sequences. Formal Methods in System Design, 44(3), pp. 264-294 (2014)
12. Petrenko, A., Yevtushenko, N.: Conformance Tests as Checking Experiments for Partial Nondeterministic FSM. In: Proc. of FATES'2005, pp. 118-133 (2005)
13. Hierons, R.M., Türker, U.C.: Distinguishing Sequences for Partially Specified FSMs. In: Proc. of NASA Formal Methods. Lecture Notes in Computer Science Volume 8430, pp. 62-76 (2014)
14. Sokolovskii, M.N.: Diagnostic experiments with automata. Cybernetics and Systems Analysis, Volume 7, pp. 988-994 (1971)
15. Kushik, N., Yevtushenko, N.: Adaptive Homing is in P. In: Proc. of MBT'2015, pp. 73-78 (2015)