

International Journal of Foundations of Computer Science
Vol. 24, No. 2 (2013) 165–185
© World Scientific Publishing Company
DOI: 10.1142/S0129054113400066



THE COMPLEXITY OF COVERAGE^{*,†}

KRISHNENDU CHATTERJEE

*IST Austria Institute of Science and Technology Austria
Krishnendu.Chatterjee@ist.ac.at*

LUCA DE ALFARO

*University of California, Santa Cruz
luca@ucsc.edu*

RUPAK MAJUMDAR

*Max Planck Institute for Software Systems, Germany
rupak@mpi-sws.org*

Received 30 November 2011

Accepted 27 July 2012

Communicated by Giorgio Delzanno and Igor Potapov

We study the problem of generating a test sequence that achieves maximal coverage for a reactive system under test. We formulate the problem as a repeated game between the tester and the system, where the system state space is partitioned according to some coverage criterion and the objective of the tester is to maximize the set of partitions (or coverage goals) visited during the game. We show the complexity of the maximal coverage problem for non-deterministic systems is PSPACE-complete, but is NP-complete for deterministic systems. For the special case of non-deterministic systems with a re-initializing “reset” action, which represent running a new test input on a re-initialized system, we show that the complexity is coNP-complete. Our proof technique for reset games uses randomized testing strategies that circumvent the exponentially large memory requirement of deterministic testing strategies. We also discuss the memory requirement for deterministic strategies and extensions of our results to other models, such as pushdown systems and timed systems.

Keywords: Author, please provide.

1. Introduction

Code coverage is a common metric in software and hardware testing that measures the degree to which an implementation has been tested with respect to some crite-

*This research was funded in part by the US NSF grants CCR-0132780 and CNS-0720884, CCF-0546170, CNS-0702881, DARPA grant HR0011-09-1-0037, Austrian Science Fund (FWF) NFN Grant S11407-N23 (RiSE), ERC Start Grant (279307: Graph Games) and a Microsoft faculty fellowship.

†A preliminary version of the paper appeared in APLAS 2008.

tion. In its simplest form, one starts with a model of the program, and a partition of the behaviors of the model into *coverage goals* [3]. A *test* is a sequence of inputs that determines a behavior of the program. The aim of testing is to explore as many coverage goals as possible, ideally as quickly as possible. In this paper, we give complexity results for several coverage problems. The problems are very basic in nature: they consist in deciding whether a certain level of coverage can be attained in a given system.

Finite-state directed graphs have been used as program models for test generation of reactive systems for a long time (see [8, 24] for surveys). A coverage goal is a partition of the states of the graph, and a test is a sequence of labels that determine a path in the graph. The maximal coverage test generation problem is to hit as many partitions as possible. In this paper, we show that the maximal coverage problem becomes NP-complete for graphs with partitions. We also distinguish between *system complexity* (the complexity of the problem in terms of the size of the graph) and the *coverage complexity* (the complexity of the problem in terms of the number of coverage goals). We show that the system complexity of the problem is NLOGSPACE (but that algorithm uses space polynomial in the number of partitions).

We consider the special case where the graph has a special “reset” action that takes it back to the initial state. This corresponds in a testing setting to the case where the system can be re-initialized before running a test (we refer to this special class of graphs as *re-initializable graphs* or *controllably recurrent graphs*). In this case, the maximal coverage problem can be solved in polynomial time for graphs with partitions.

Directed graphs form a convenient representation for deterministic systems, in which all the choices are under the control of the tester. Testing of non-deterministic systems in which certain actions are controllable (under the control of the tester) and other actions are uncontrollable lead to *game graphs* [23]. A game graph is a directed labeled graph where the nodes are partitioned into tester-nodes and system-nodes, and while the tester can choose the next input at a tester node, the system non-deterministically chooses the next state at a system node. Then, the test generation problem is to generate a test set that achieves maximal coverage no matter how the system moves. For general game graphs, we show the complexity of the maximal coverage problem is PSPACE-complete. However, there is an algorithm that runs in time linear in the size of the game graph but exponential in the number of coverage goals. Again, the re-initializability assumption reduces the complexity of coverage: in case there is a re-initialization strategy of the tester from any system state, the maximal coverage problem for games is coNP-complete. Dually, we show that the problem of whether it is possible to win a safety game while visiting fewer than a specified number of partitions is NP-complete.

Finally, we consider the coverage problem in bounded time, consisting in checking whether a specified number of partitions can be visited in a specified number of steps. We show that the problem is NP-complete for graphs and controllably recurrent graphs, and is PSPACE-complete for game graphs.

Table 1. The complexity of coverage.

	Graphs	Recurrent Graphs	Game Graphs	Recurrent Game Graphs
Maximal Coverage	NP-complete	PTime	PSPACE-complete	coNP-complete
Coverage in Bounded time	NP-complete	NP-complete	PSPACE-complete	NP-hard and coNP-hard in PSPACE

In summary, our main contributions can be enumerated as follows (Table 1 gives a summary of the complexity results).

- (1) *Graphs and controllably recurrent graphs.* We show the maximal coverage problem is NP-complete for graphs, and can be solved in polynomial time for controllably recurrent graphs. In contrast, the coverage problem in bounded time is NP-complete for both graphs and controllably recurrent graphs.
- (2) *Game graphs and controllably recurrent game graphs.* The maximal coverage problem is PSPACE-complete for game graphs, and for the special class of controllably recurrent game graphs the problem is coNP-complete. The coverage in bounded time problem is PSPACE-complete for game graphs, and for controllably recurrent game graphs the problem is both NP-hard and coNP-hard, and the problem can be solved in PSPACE.

Related results and extensions of our results. We also present results related to the memory requirements of strategies for all the cases. While we develop our theory for the finite-state, discrete case, we derive similar results for more general models from our results, such as those incorporating incomplete information (the tester can only observe part of the system state), or timing, or pushdown systems (modeling procedures). For timed systems modeled as timed automata, the maximal coverage problem is PSPACE-complete. For timed games as well as for (finite state) game graphs with incomplete information, the maximal coverage problem becomes EXPTIME-complete. For pushdown systems, the maximal coverage can be solved in polynomial time, and for pushdown games the maximal coverage problem is EXPTIME-complete.

Related works. Optimization problems arising out of test generation have been studied before in the context of both graphs and games [1, 7, 17, 23]. Constraint-solving for test generation is currently an active area of research [6, 9, 10, 15]. However, to the best of our knowledge, the complexities of the coverage problems studied here had not been studied before.

2. Definitions

In this section we define *labeled graphs* and *labeled games*, and then define the two decision problems of coverage, namely, *maximal coverage* problem and *coverage with bounded time* problem. We start with definition of graphs and games. To simplify the notation in subsequent arguments, it is convenient to define state-space partitions

via labeling functions from states to predicates; the (single) predicate associated with a state indicates the partition to which the state belongs.

Definition 1. (Labeled graphs) A labeled graph $\mathcal{G} = ((V, E), v_{in}, \text{AP}, \mathcal{L})$ consists of the following component:

- (1) a finite directed graph with vertex set V and edge set E ;
- (2) the initial vertex v_{in} ;
- (3) a finite set of atomic propositions AP ; and
- (4) a labeling (or a partition) function $\mathcal{L} : V \rightarrow \text{AP}$ that assigns to each vertex v , the atomic proposition $\mathcal{L}(v) \in \text{AP}$ true at v .

For technical convenience we will assume that for all vertices $v \in V$, there exists $u \in V$ such that $(v, u) \in E$, i.e., each vertex has at least one out-going edge.

Labeling and partition of vertex set. Given a labeled graph $\mathcal{G} = ((V, E), v_{in}, \text{AP}, \mathcal{L})$, the atomic propositions and the labeling (or the partition) function gives a partition of the vertex set V . Let $\text{AP} = \{p_1, p_2, \dots, p_\ell\}$, and for $1 \leq i \leq \ell$, let $V^i = \{v \in V \mid \mathcal{L}(v) = p_i\}$. Then $(V^1, V^2, \dots, V^\ell)$ gives a partition of the vertex set V .

Paths in graphs and reachability. Given a labeled graph \mathcal{G} , a path ω in \mathcal{G} is an infinite sequence of vertices $\langle v_0, v_1, v_2 \dots \rangle$ starting from the initial vertex v_{in} (i.e., $v_0 = v_{in}$) such that for all $i \geq 0$ we have $(v_i, v_{i+1}) \in E$. A vertex v_i is reachable from v_{in} if there is a path $\omega = \langle v_0, v_1, v_2 \dots \rangle$ in \mathcal{G} and $j \geq 0$ such that the vertex v_j in ω is the vertex v_i .

Definition 2. (Labeled game graphs) A labeled game graph $\mathcal{G} = ((V, E), (V_1, V_2), v_{in}, \text{AP}, \mathcal{L})$ consists of the components of a labeled graph along with a partition of the finite vertex set V into (V_1, V_2) . The vertices in V_1 are player 1 vertices where player 1 chooses outgoing edges, and analogously, the vertices in V_2 are player 2 vertices where player 2 chooses outgoing edges. Again for technical convenience we will assume that for all vertices $v \in V$, there exists $u \in V$ such that $(v, u) \in E$, i.e., each vertex has at least one out-going edge.

Plays and strategies in games. A play in a game graph is a path in the underlying graph of the game. A strategy for a player in a game is a recipe to specify how to extend the prefix of a play. Formally, a strategy π_1 for player 1 is a function $\pi_1 : V^* \cdot V_1 \rightarrow V$ that takes a finite sequence of vertices $w \cdot v \in V^* \cdot V_1$ ending in a player 1 vertex v ($w \cdot v$ represents the history of the play so far), and specifies the next vertex $\pi_1(w \cdot v)$ by choosing an out-going edge from v (i.e., $(v, \pi_1(w \cdot v)) \in E$). A strategy $\pi_2 : V^* \cdot V_2 \rightarrow V$ is defined analogously. We denote by Π_1 and Π_2 the set of all strategies for player 1 and player 2, respectively. Given strategies π_1 and π_2 for player 1 and player 2, there is a unique play (or a path)

$\omega(v_{in}, \pi_1, \pi_2) = \langle v_0, v_1, v_2, \dots \rangle$ such that (a) $v_0 = v_{in}$; (b) for all $i \geq 0$, if $v_i \in V_1$, then $\pi_1(v_0 \cdot v_1 \dots v_i) = v_{i+1}$; and if $v_i \in V_2$, then $\pi_2(v_0 \cdot v_1 \dots v_i) = v_{i+1}$.

Controllably recurrent graphs and games. Along with general labeled graphs and games, we will also consider graphs and games that are *controllably recurrent*. A labeled graph \mathcal{G} is *controllably recurrent* if for every vertex v_i that is reachable from v_{in} , there is a path starting from v_i that reaches v_{in} . A labeled game graph \mathcal{G} is *controllably recurrent* if for every vertex v_i that is reachable from v_{in} in the underlying graph, there is a strategy π_1 for player 1 such that against all player 2 strategies π_2 , the path starting from v_i given the strategies π_1 and π_2 reaches v_{in} . Controllable recurrence models the natural requirement that systems under test are *re-initializable*, that is, from any reachable state of the system, there is always a way to bring the system back to its initial state no matter how the system behaves.

The maximal coverage problem. The *maximal coverage problem* asks whether at least m different propositions can be visited, in other words, it asks whether at least m different partitions of the vertex set (given by the proposition labeling) can be visited. We now define the problem formally for graphs and games. Given a path $\omega = \langle v_0, v_1, v_2, \dots \rangle$, let $\mathcal{L}(\omega) = \bigcup_{i \geq 0} \mathcal{L}(v_i)$ be the set of propositions that appear in ω . Given a labeled graph \mathcal{G} and $0 \leq m \leq |\text{AP}|$, the maximal coverage problem asks whether there is path ω such that $|\mathcal{L}(\omega)| \geq m$. Given a labeled game graph \mathcal{G} and $0 \leq m \leq |\text{AP}|$, the maximal coverage problem asks whether player 1 can ensure that at least m propositions are visited, i.e., whether

$$\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2))| \geq m.$$

It may be noted that $\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2))| \geq m$ iff there exists a player 1 strategy π_1^* such that for all player 2 strategies π_2^* we have $|\mathcal{L}(\omega(v_{in}, \pi_1^*, \pi_2^*))| \geq m$.

The coverage with bounded time problem. The *coverage with bounded time problem* asks whether at least m different propositions can be visited within k -steps, that is, whether at least m different partitions of the vertex set can be visited within k -steps. We now define the problem formally for graphs and games. Given a path $\omega = \langle v_0, v_1, v_2, \dots \rangle$ and $k \geq 0$, we denote by $\omega \upharpoonright k$ the prefix of the path of length $k + 1$, i.e., $\omega \upharpoonright k = \langle v_0, v_1, \dots, v_k \rangle$. Given a path $\omega = \langle v_0, v_1, v_2, \dots \rangle$ and $k \geq 0$, we denote by $\mathcal{L}(\omega \upharpoonright k) = \bigcup_{0 \leq i \leq k} \mathcal{L}(v_i)$. Given a labeled graph \mathcal{G} and $0 \leq m \leq |\text{AP}|$ and $k \geq 0$, the coverage with bounded time problem asks whether there is path ω such that $|\mathcal{L}(\omega \upharpoonright k)| \geq m$. Given a labeled game graph \mathcal{G} and $0 \leq m \leq |\text{AP}|$, the maximal coverage problem asks whether player 1 can ensure that at least m propositions are visited within k -steps, i.e., whether

$$\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2) \upharpoonright k)| \geq m.$$

It may be noted that $\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2) \upharpoonright k)| \geq m$ iff there exists a player 1 strategy π_1^* such that for all player 2 strategies π_2^* we have $|\mathcal{L}(\omega(v_{in}, \pi_1^*, \pi_2^*) \upharpoonright k)| \geq m$.

2.1. Examples

System-tester game. A system $S = (Q, \Sigma, q_{in}, \Delta, AP, \mathcal{L})$ consists of the following components:

- a finite set Q of states with the starting state q_{in} ;
- a finite alphabet Σ of input letters;
- a transition relation $\Delta \subseteq Q \times \Sigma \times Q$; and
- a finite set of atomic propositions AP and a labeling function \mathcal{L} that assigns to each state q the atomic proposition $\mathcal{L}(q)$ true at q .

We consider *total* systems such that for all $q \in Q$ and $\sigma \in \Sigma$, there exists $q' \in Q$ such that $(q, \sigma, q') \in \Delta$. A system is *deterministic* if for all $q \in Q$ and $\sigma \in \Sigma$, there exists exactly one q' such that $(q, \sigma, q') \in \Delta$. The tester selects an input letter at every stage and the system resolves the non-determinism in transition to choose the successor state. The goal of the tester is to visit as many different propositions as possible. The interaction between the system and the tester can be reduced to a labeled game graph $\mathcal{G} = ((V, E), (V_1, V_2), v_{in}, AP, \mathcal{L}')$ as follows:

- *Vertices and partition.* $V = Q \cup (Q \times \Sigma)$; $V_1 = Q$ and $V_2 = Q \times \Sigma$; and $v_{in} = q_{in}$.
- *Edges.* $E = \{(q, (q, \sigma)) \mid q \in Q, \sigma \in \Sigma\} \cup \{((q, \sigma), q') \mid (q, \sigma, q') \in \Delta\}$.
- *Labeling.* $\mathcal{L}'(q) = \mathcal{L}(q)$ and $\mathcal{L}'((q, \sigma)) = \mathcal{L}(q)$.

The coverage question for game between tester and system can be answered by answering the question in the game graph. Also observe that if the system is deterministic, then for all player 2 vertices in the game graph, there is exactly one out-going edge, and hence the game can be reduced to a labeled graph. A system consists of a set of variables, and a state represents the valuation of variables. If a subset of variable valuations is of interest for the coverage criteria, then the partition of the state space of the system is obtained through the valuations of variables of interest, and the desired partition of the state space by the valuations of interest is captured by the labeling of the atomic propositions. In this paper we will present all the results for the labeled graph and game model. All the upper bounds we provide follow also for the game between tester and system. All the lower bounds we present can also be easily adapted to the model of the game between system and tester.

Graph coverage criteria. Graph-based coverage criteria, such as node or transition coverage on the control-flow graph or on a finite-state abstract model, are used commonly in software testing [3]. Such coverage criteria reduce to our notion of testing on labeled graphs. Intuitively, we define a *program state graph* where each node represents a program state (location as well as valuations to all variables) or an abstraction (e.g., a predicate abstraction [5]), and each edge (s, t) represents a program operation that takes a state s to the state t . For node coverage, we label a node with the value of the program location. The coverage goal is to maximize the number of visited locations. Other graph-based coverage criteria can be expressed

by adding auxiliary state. For example, to capture def-use coverage (for each pair of definition point d and use point u of a variable, find a test such that the definition d is used at u), one can add auxiliary state variables that track the point of current definition of a variable, and label each node of the program state graph with a definition-use pair for each variable such that the current definition of the variable is used by the incoming operation to the node. The coverage goal is again to maximize the number of visited labels.

3. The Complexity of Maximal Coverage Problems

In this section we study the complexity of the maximal coverage problem. In subsection 3.1 we study the complexity for graphs, and in subsection 3.2 we study the complexity for game graphs.

3.1. Graphs

We first show that the maximal coverage problem for labeled graphs is NP-complete.

Theorem 3. *The maximal coverage problem for labeled graphs is NP-complete.*

Proof. The proof consists of two parts.

- (1) *In NP.* The maximal coverage problem is in NP can be proved as follows. Given a labeled graph \mathcal{G} , let $n = |V|$. We show first that if there is a path ω in \mathcal{G} such that $|\mathcal{L}(\omega)| \geq m$, then there is a path ω' in \mathcal{G} such that $|\mathcal{L}(\omega' \upharpoonright (m \cdot n))| \geq m$, where $m \cdot n$ denotes the product of m and n . If ω visits at least m propositions, and there is a cycle in ω that does not visit a new proposition that is already visited in the prefix, then the cycle segment can be removed from ω and still the resulting path visits m propositions. Hence if the answer to the maximal coverage problem is “Yes”, then there is a path ω' of length at most $m \cdot n$ that is a witness to the “Yes” answer. Since $m \leq |\text{AP}|$, it follows that the problem is in NP.
- (2) *NP-hardness.* Now we show that the maximal coverage problem is NP-hard, and we present a reduction from the SAT-problem. Consider a SAT formula Φ , and let $X = \{x_1, x_2, \dots, x_n\}$ be the set of variables and C_1, C_2, \dots, C_m be the set of clauses. For a variable $x_j \in X$, let
 - (a) $\text{T}(x_j) = \{\ell \mid x_j \in C_\ell\}$ be the set of indices of the set of clauses C_ℓ that is satisfied if x_j is set to be true; and
 - (b) $\text{F}(x_j) = \{\ell \mid \bar{x}_j \in C_\ell\}$ be the set of indices of the set of clauses C_ℓ that is satisfied if x_j is set to be false.

Without loss of generality, we assume that $\text{T}(x_j)$ and $\text{F}(x_j)$ are non-empty for all $1 \leq j \leq n$ (this is because, for example, if $\text{F}(x_j) = \emptyset$, then we can set x_j to be true and reduce the problem where the variable x_j is not present). For a finite set $F \subseteq \mathbb{N}$ of natural numbers, let $\max(F)$ and $\min(F)$ denote the

maximum and minimum number of F , respectively. For an element $f \in F$ that is not the maximum element let $\text{next}(f, F)$ denote the next highest element to f that belongs to F ; i.e., (a) $\text{next}(f, F) \in F$; (b) $f < \text{next}(f, F)$; and (c) if $j \in F$ and $f < j$, then $\text{next}(f, F) \leq j$. We construct a labeled graph \mathcal{G}^Φ as follows. We first present an intuitive description: there are vertices named $x_1, x_2, \dots, x_n, x_{n+1}$, and all of them are labeled by a single proposition. The vertex x_{n+1} is an absorbing vertex (vertex with a self-loop only), and all other x_i vertex has two successors. The starting vertex is x_1 . In every vertex x_i given the right choice we visit in a line a set of vertices that are labeled by clauses that are true if x_i is true; and given the left choice we visit in a line a set of vertices that are labeled by clauses that are true if x_i is false; and then we move to vertex x_{i+1} . We now formally describe every component of the labeled graph $\mathcal{G}^\Phi = ((V^\Phi, E^\Phi), v_{in}^\Phi, \text{AP}^\Phi, \mathcal{L}^\Phi)$.

(a) The set of vertices is

$$V^\Phi = \{x_i \mid 1 \leq i \leq n+1\} \cup \{x_{j,i} \mid 1 \leq j \leq n, i \in \text{T}(x_j)\} \cup \{\bar{x}_{j,i} \mid 1 \leq j \leq n, i \in \text{F}(x_j)\}.$$

There is a vertex for every variable, and a vertex x_{n+1} . There is a vertex $x_{j,i}$ iff $C_i \in \text{T}(x_j)$, and there is a vertex $\bar{x}_{j,i}$ iff $C_i \in \text{F}(x_j)$,

(b) The set of edges is

$$E^\Phi = \{(x_{n+1}, x_{n+1})\} \cup \{(x_{j, \max(\text{T}(x_j))}, x_{j+1}), (\bar{x}_{j, \max(\text{F}(x_j))}, x_{j+1}) \mid 1 \leq j \leq n\} \cup \{(x_{j, \min(\text{T}(x_j))}, x_{j, \bar{x}_{j, \min(\text{F}(x_j))})} \mid 1 \leq j \leq n\} \cup \{(x_{j,i}, x_{j, \text{next}(i, \text{T}(x_j))}) \mid 1 \leq j \leq n, i < \max(\text{T}(x_j))\} \cup \{(\bar{x}_{j,i}, \bar{x}_{j, \text{next}(i, \text{F}(x_j))}) \mid 1 \leq j \leq n, i < \max(\text{F}(x_j))\}.$$

We now explain the role of each set of edges. The first edge is the self-loop at x_{n+1} . The second set of edges specifies that from $x_{j, \max(\text{T}(x_j))}$ the next vertex is x_{j+1} and similarly, from $\bar{x}_{j, \max(\text{F}(x_j))}$ the next vertex is again x_{j+1} . The third set of edges specifies that from x_j there are two successors that are $x_{j,i}$ and $\bar{x}_{j,i'}$, where $i = \min(\text{T}(x_j))$ and $i' = \min(\text{F}(x_j))$. The final sets of edges specifies (a) to move in a line from $x_{j, \min(\text{T}(x_j))}$ to visit the clauses that are satisfied by setting x_j as true, and (b) to move in a line from $\bar{x}_{j, \min(\text{F}(x_j))}$ to visit the clauses that are satisfied by setting x_j as false. Figure 1 gives a pictorial view of the reduction.

(c) The initial vertex is $v_{in}^\Phi = x_1$.

(d) $\text{AP}^\Phi = \{C_1, C_2, \dots, C_m, X\}$, i.e., there is a proposition C_i for each clause C_i and there is a proposition X for all variables;

(e) $\mathcal{L}^\Phi(x_j) = X$; i.e., every variable vertex is labeled by the proposition X ; and we have $\mathcal{L}^\Phi(x_{j,i}) = C_i$ and $\mathcal{L}^\Phi(\bar{x}_{j,i}) = C_i$, i.e., each vertex $x_{j,i}$ and $\bar{x}_{j,i}$ is labeled by the corresponding clause that is indexed.

The number of vertices in \mathcal{G}^Φ is $O(n \cdot m)$, and the reduction is polynomial in Φ . In this graph the maximal number of propositions visited is exactly equal to the

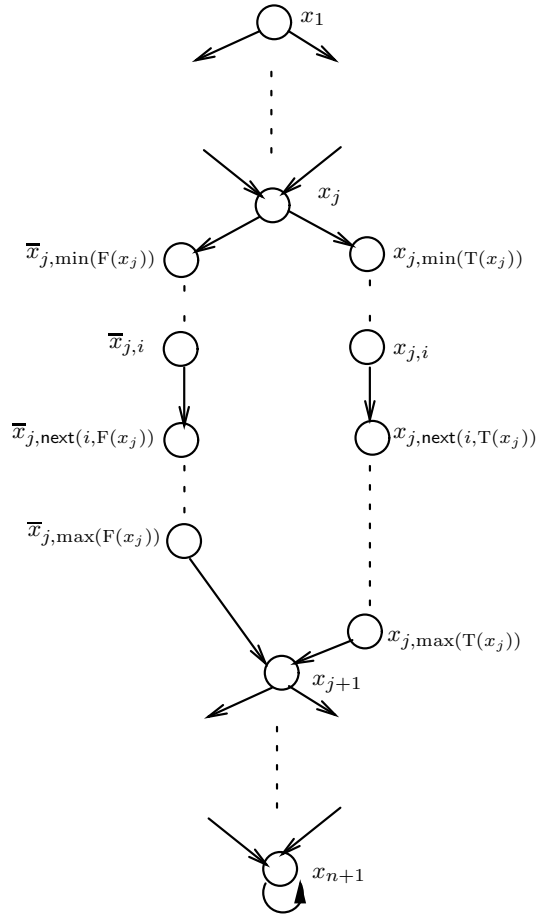


Fig. 1. The NP-hardness reduction in picture.

maximal number of satisfiable clauses plus 1 (since along with the propositions for clauses, the proposition X for all variables is always visited). The proof of the above claim is as follows. Given a path ω in \mathcal{G}^Φ we construct an assignment A for the variables as follows: if the choice at a vertex x_j is $x_{j,\min(T(x_j))}$, then we set x_j as true in A , else we set x_j as false. Hence if a path in \mathcal{G}^Φ visits a set $P \subseteq \text{AP}^\Phi$ of r propositions, then the assignment A satisfies $r - 1$ clauses (namely, $P \setminus \{X\}$). Conversely, given an assignment A of the variables, we construct a path ω^A in \mathcal{G}^Φ as follows: if x_j is true in the assignment A , then the path ω^A chooses $x_{j,\min(T(x_j))}$ at x_j , otherwise, it chooses $\bar{x}_{j,\min(F(x_j))}$ at x_j . If A satisfies a set Q of $r - 1$ clauses, then ω^A visits $r + 1$ propositions (namely, the set $Q \cup \{X\}$ of propositions). Hence Φ is satisfiable iff the answer to the maximal coverage problem with input \mathcal{G}^Φ and $m + 1$ is true.

The desired result follows. \square

We note that from the proof of Theorem 3 it follows that the MAX-SAT problem (i.e., computing the maximal number of clauses satisfiable for a SAT formula) can be reduced to the problem of computing the exact number for the maximal coverage problem. From hardness of approximation of the MAX-SAT problem [4], it follows that the maximal coverage problem for labeled graphs is hard to approximate.

Theorem 4. *The maximal coverage problem for labeled graphs that are controllably recurrent can be decided in PTIME.*

Proof. To solve the maximal coverage problem for labeled graphs that are controllably recurrent, we compute the maximal strongly connected component C that v_{in} belongs to. Since the graph is controllably recurrent, all vertices that are reachable from v_{in} belong to C . Hence the answer to the maximal coverage problem is “Yes” iff $|\bigcup_{v \in C} \mathcal{L}(v)| \geq m$. The result follows. \square

3.2. Game graphs

Theorem 5. *The maximal coverage problem for labeled game graphs is PSPACE-complete.*

Proof. The proof consists of two parts.

- (1) *In PSPACE.* We argue that the maximal coverage problem for labeled game graphs can be reduced to the coverage in bounded time problem. The reason is as follows: in a labeled game graph with n vertices, if player 1 can visit m propositions, then player 1 can visit m propositions within at most $m \cdot n$ steps; because player 1 can always play a strategy from the current position that visits a new proposition that is not visited and never needs to go through a cycle without visiting a new proposition unless the maximal coverage is achieved. Hence it follows that the maximal coverage problem for games reduces to the coverage in bounded time problem. The PSPACE inclusion will follow from the result of Theorem 10 where we show that the coverage in bounded time problem is in PSPACE.
- (2) *PSPACE-hardness.* The maximal coverage problem for game graphs is PSPACE-complete, even if the underlying graph is strongly connected. The proof is a reduction from QBF (truth of quantified boolean formulas) that is known to be PSPACE-complete [19], and it is a modification of the reduction of Theorem 3. Consider a QBF formula

$$\Phi = \exists x_1. \forall x_2. \exists x_3 \dots \exists x_n. C_1 \wedge C_2 \wedge \dots \wedge C_m;$$

defined on the set $X = \{x_1, x_2, \dots, x_n\}$ of variables, and C_1, C_2, \dots, C_m are the clauses of the formula. We apply the reduction of Theorem 3 with the following

modification to obtain the labeled game graph \mathcal{G}^Φ : the partition (V_1^Φ, V_2^Φ) of V^Φ is as follows. For a variable x_j if the quantifier before x_j is existential, then $x_j \in V_1^\Phi$ (i.e., for existentially quantified variable, player 1 chooses the out-going edges denoting whether to set the variable true or false); and for a variable x_j if the quantifier before x_j is universal, then $x_j \in V_2^\Phi$ (i.e., for universally quantified variable, the opposing player 2 chooses the out-going edges denoting whether to set the variable true or false). The vertex x_{n+1} is a player 2 vertex, and all other vertex has an single out-going edges and can be player 1 vertex. Given this game graph we have Φ is true iff player 1 can ensure that all the propositions can be visited in \mathcal{G}^Φ . Formally, let Π_1^Φ and Π_2^Φ denote the set of all strategies for player 1 and player 2, respectively, in \mathcal{G}^Φ . Then Φ is true iff $\max_{\pi_1 \in \Pi_1^\Phi} \min_{\pi_2 \in \Pi_2^\Phi} |\mathcal{L}^\Phi(\omega(x_1, \pi_1, \pi_2))| \geq m + 1$. Observe that since x_{n+1} is a player 2 vertex if we add an edge from x_{n+1} to x_1 , player 2 will never choose the edge x_{n+1} to x_1 (since the objective for player 2 is to minimize the coverage). However, adding the edge from x_{n+1} to x_1 makes the underlying graph strongly connected (i.e., the underlying graph of the game graph becomes controllably recurrent; but player 1 does not have a strategy to ensure that x_1 is reached, so the game is not controllably recurrent).

The desired result follows. \square

Complexity of maximal coverage in controllably recurrent games. We will now consider maximal coverage in controllably recurrent games. Our analysis will use fixing memoryless *randomized* strategy for player 1, and fixing a memoryless randomized strategy in labeled game graph we get a labeled Markov decision process (MDP). A labeled MDP consists of the same components as a labeled game graph, and for vertices in V_1 (which are randomized vertices in the MDP) the successors are chosen uniformly at random (i.e., player 1 does not have a proper choice of the successor but chooses all of them uniformly at random). Given a labeled game graph $\mathcal{G} = ((V, E), (V_1, V_2), v_{in}, \mathbf{AP}, \mathcal{L})$ we denote by $\text{Unif}(\mathcal{G})$ the MDP interpretation of \mathcal{G} where player 1 vertices chooses all successors uniformly at random. An *end-component* in $\text{Unif}(\mathcal{G})$ is a set U of vertices such that (i) U is strongly connected and (ii) U is player 1 *closed*, i.e., for all $u \in U \cap V_1$, for all u' such that $(u, u') \in E$ we have $u' \in U$ (in other words, for all player 1 vertices, all the out-going edges are contained in U).

Lemma 6. *Let \mathcal{G} be a labeled game graph that is controllably recurrent and let $\text{Unif}(\mathcal{G})$ be the MDP interpretation of \mathcal{G} . Then the following assertions hold.*

- (1) *Let U be an end-component in $\text{Unif}(\mathcal{G})$ with $v_{in} \in U$. Then $\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2))| \leq |\bigcup_{u \in U} \mathcal{L}(u)|$.*
- (2) *There exists an end-component $U \in \text{Unif}(\mathcal{G})$ with $v_{in} \in U$ such that $|\bigcup_{u \in U} \mathcal{L}(u)| \leq \max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2))|$.*

Proof. We prove both the claims below.

- (1) If U is an end-component in $\text{Unif}(\mathcal{G})$, then consider a memoryless strategy π_2^* for player 2, that for all vertices $u \in U \cap V_2$, chooses a successor $u' \in U$ (such a successor exists since U is strongly connected). Since U is player 1 closed (i.e., for all player 1 out-going edges from U , the end-point is in U), it follows that for all strategies of player 1, given the strategy π_2^* for player 2, the vertices visited in a play is contained in U . The desired result follows.
- (2) An optimal strategy π_1^* for player 1 in \mathcal{G} is as follows:
 - (a) Let $Z_0 = \{v \in V \mid \mathcal{L}(v) = \mathcal{L}(v_{in})\}$ and $i = 0$;
 - (b) At iteration i , let Z_i represent the vertices corresponding to the set of propositions already visited. At iteration i , player 1 plays a strategy to reach a vertex in $V \setminus Z_i$ (if such a strategy exists), and then reaches back v_{in} (a strategy to reach back v_{in} always exists since the game is controllably recurrent).
 - (c) If a new proposition p_i is visited at iteration i , then let $Z_{i+1} = Z_i \cup \{v \in V \mid \mathcal{L}(v) = p_i\}$. Goto step (b) for $i + 1$ iteration with Z_{i+1} . If no vertex in $V \setminus Z_i$ can be reached, then stop.

The strategy π_1^* is optimal since it iteratively visits all possible propositions that player 1 can ensure to visit from v_{in} . Let the above iteration stop with $Z_i = Z^*$. Let $X = V \setminus Z^*$, and let X^* be the set of vertices such that player 1 can reach X . Let $U^* = V \setminus X^*$. Then $v_{in} \in U^*$ and player 2 can ensure that from v_{in} the game can be confined to U^* . Hence the following conditions must hold: (a) for all $u \in U^* \cap V_2$, there exists $u' \in U^*$ such that $(u, u') \in E$; and (b) for all $u \in U^* \cap V_1$, for all $u' \in V$ such that $(u, u') \in E$ we have $u' \in U^*$. Consider the sub-graph \mathcal{G}' where player 2 restricts itself to edges only in U^* . A bottom maximal strongly connected component $U \subseteq U^*$ in the sub-graph is an end-component in $\text{Unif}(\mathcal{G})$, and we have

$$\left| \bigcup_{u \in U} \mathcal{L}(u) \right| \leq \left| \bigcup_{u \in U^*} \mathcal{L}(u) \right| \leq \left| \bigcup_{u \in Z^*} \mathcal{L}(u) \right|.$$

It follows that U is a witness end-component to prove the result.

The desired result follows. \square

Theorem 7. *The maximal coverage problem for labeled game graphs that are controllably recurrent is coNP-complete.*

Proof. We prove the following two claims to establish the result.

- (1) *In coNP.* The fact that the problem is in coNP can be proved using Lemma 6. Given a labeled game graph \mathcal{G} , if the answer to the maximal coverage problem (i.e., whether $\max_{\pi_1 \in \Pi_1} \min_{\pi_2 \in \Pi_2} |\mathcal{L}(\omega(v_{in}, \pi_1, \pi_2))| \geq m$) is NO, then by Lemma 6, there exists an end-component U in $\text{Unif}(\mathcal{G})$ such that $|\bigcup_{u \in U} \mathcal{L}(u)| < m$. The witness end-component U is a polynomial witness and

it can be guessed and verified in polynomial time. The verification that U is the correct witness is as follows: we check (a) U is strongly connected; (b) for all $u \in U \cap V_1$ and for all $u' \in V$ such that $(u, u') \in E$ we have $u' \in U$; (c) $v_{in} \in U$; and (d) $|\bigcup_{u \in U} \mathcal{L}(u)| < m$. Hence the result follows.

- (2) *coNP hardness.* We prove hardness using a reduction from the complement of the *Vertex Cover* problem. Given a graph $G = (V, E)$, a set $U \subseteq V$ is a *vertex cover* if for all edges $e = (u_1, u_2) \in E$ we have either $u_1 \in U$ or $u_2 \in U$. Given a graph G whether there is a vertex cover U of size at most m (i.e., $|U| \leq m$) is NP-complete [14]. We now present a reduction of the complement of the vertex cover problem to the maximal coverage problem in controllably recurrent games. Given a graph $G = (V, E)$ we construct a labeled game graph $\overline{\mathcal{G}}$ as follows. Let the set E of edges be enumerated as $\{e_1, e_2, \dots, e_\ell\}$, i.e., there are ℓ edges. The labeled game graph $\overline{\mathcal{G}} = ((\overline{V}, \overline{E}), (\overline{V}_1, \overline{V}_2), v_{in}, \text{AP}, \mathcal{L})$ is as follows.

- (a) *Vertex set and partition.* The vertex set \overline{V} is as follows:

$$\overline{V} = \{v_{in}\} \cup E \cup \{e_i^j \mid 1 \leq i \leq \ell, 1 \leq j \leq 2\}.$$

All vertices in E are player 2 vertices, and the other vertices are player 1 vertices, i.e., $\overline{V}_2 = E$, and $\overline{V}_1 = \overline{V} \setminus \overline{V}_2$.

- (b) *Edges.* The set \overline{E} of edges are as follows:

$$\begin{aligned} \overline{E} = & \{(v_{in}, e_j) \mid 1 \leq j \leq \ell\} \cup \{(e_i, e_i^j) \mid 1 \leq i \leq \ell, 1 \leq j \leq 2\} \\ & \cup \{(e_i^j, v_{in}) \mid 1 \leq i \leq \ell, 1 \leq j \leq 2\}. \end{aligned}$$

Intuitively, the edges in the game graph are as follows: from the initial vertex v_{in} , player 1 can choose any of the edges $e_i \in E$. For a vertex e_i in \overline{V} , player 2 can choose between two vertices e_i^1 and e_i^2 (which will eventually represent the two end-points of the edge e_i). From vertices of the form e_i^1 and e_i^2 , for $1 \leq i \leq \ell$, the next vertex is the initial vertex v_{in} . It follows that from all vertices, the game always comes back to v_{in} and hence we have a controllably recurrent game.

- (c) *Propositions and labelling.* $\text{AP} = V \cup \{\$ \mid \$ \notin V\}$, i.e., there is a proposition for every vertex in V and a special proposition $\$$. The vertex v_{in} and vertices in E are labeled by the special proposition $\$$, i.e., $\mathcal{L}(v_{in}) = \$$; and for all $e_i \in E$ we have $\mathcal{L}(e_i) = \$$. For a vertex e_i^j , let $e_i = (u_i^1, u_i^2)$, where u_i^1, u_i^2 are vertices in V , then $\mathcal{L}(e_i^1) = u_i^1$ and $\mathcal{L}(e_i^2) = u_i^2$. Note that the above proposition assignment ensures that at every vertex that represents an edge, player 2 has the choices of vertices that form the end-points of the edge.

The following case analysis completes the proof.

- Given a vertex cover U , consider a player 2 strategy, that at a vertex $e_i \in \overline{V}$, choose a successor e_i^j such that $\mathcal{L}(e_i^j) \in U$. The strategy for player 2 ensures that player 1 visits only propositions in $U \cup \{\$\}$, i.e., at most $|U| + 1$ propositions.

- Consider a strategy for player 1 that from v_{in} visits all vertices e_1, e_2, \dots, e_ℓ in order. Consider any counter-strategy for player 2 and let $U \subseteq V$ be the set of propositions other than $\$$ visited. Since all the edges are chosen, it follows that U is a vertex cover. Hence if all vertex cover in G is of size at least m , then player 1 can visit at least $m + 1$ propositions.

Hence there is a vertex cover in G of size at most m if and only if the answer to the maximal coverage problem in \mathcal{G} with $m + 1$ is NO. It follows that the maximal coverage problem in controllably recurrent games is coNP-hard.

The desired result follows. \square

Complexity of minimal safety games. As a corollary of the proof of Theorem 7 we obtain a complexity result about *minimal safety games*. Given a labeled game graph \mathcal{G} and m , the minimal safety game problem asks, whether there exists a set U such that a player can confine the game in U and U contains at most m propositions. An easy consequence of the hardness proof of Theorem 7 is minimal safety games are NP-hard, and also it is easy to argue that minimal safety games are in NP. Hence we obtain that the minimal safety game problem is NP-complete.

Randomized memoryless strategies. From our analysis it follows that for controllably recurrent game graphs the randomized memoryless strategy for player 1 that plays all edges uniformly at random is an optimal strategy with probability 1, since in an MDP for all strategies with probability 1 the set of states visited infinitely often is an end-component [11, 12].

4. The Complexity of Coverage in Bounded Time Problem

In this section we study the complexity of the coverage in bounded time problem. In subsection 4.1 we study the complexity for graphs, and in subsection 4.2 we study the complexity for game graphs.

4.1. Graphs

Theorem 8. *The coverage in bounded time problem for both labeled graphs and controllably recurrent labeled graphs is NP-complete.*

Proof. We prove the completeness result in two parts below.

- (1) *In NP.* Given a labeled graph with n vertices, if there is a path ω such that $|\mathcal{L}(\omega \upharpoonright k)| \geq m$, then there is path ω' such that $|\mathcal{L}(\omega' \upharpoonright (m \cdot n))| \geq m$. The above claim follows since any cycle that does not visit any new proposition can be omitted. Hence a path of length $j = \min(k, m \cdot n)$ can be guessed and it can be then checked in polynomial time if the path of length j visits at least m propositions.

- (2) *In NP-hard.* We reduce the *Hamiltonian-path (HAM-PATH)* [14] problem to the coverage in bounded time problem for labeled graphs. Given a directed graph $G = (V, E)$ and an initial vertex v , we consider the labeled graph \mathcal{G} with the directed graph G , with v as the initial vertex and $\text{AP} = V$ and $\mathcal{L}(u) = u$ for all $u \in V$, i.e., each vertex is labeled with a unique proposition. The answer to the coverage in bounded time with $k = n$ and $m = n$, for $n = |V|$ is “YES” iff there is a HAM-PATH in G starting from v .

The desired result follows. □

Complexity in size of the graph. We now argue that the maximal coverage and the coverage in bounded time problem on labeled graphs can be solved in non-deterministic log-space in the size of the graph, and polynomial space in the size of the atomic propositions. Given a labeled graph \mathcal{G} , with n vertices, we argued in Theorem 3 that if m propositions can be visited, then there is a path of length at most $m \cdot n$, that visits m propositions. The path of length $m \cdot n$, can be visited, storing the current vertex, and guessing the next vertex, and checking the set of propositions already visited. Hence this can be achieved in non-deterministic log-space in the size of the graph, and polynomial space in the size of the proposition set. A similar argument holds for the coverage in bounded time problem. This gives us the following result.

Theorem 9. *Given a labeled graph $\mathcal{G} = ((V, E), v_{in}, \text{AP}, \mathcal{L})$, the maximal coverage problem and the coverage in bounded time problem can be decided in NLOGSPACE in $|V| + |E|$, and in PSPACE in $|\text{AP}|$.*

4.2. Game graphs

Theorem 10. *The coverage in bounded time problem for labeled game graphs is PSPACE-complete.*

Proof. We prove the following two cases to prove the result.

- (1) *In PSPACE.* We say that an *exploration game tree* for a labeled game graph is a rooted, labeled tree which represents an unfolding of the graph. Every node α of the tree is labeled with a pair (v, b) , where v is a node of the game graph, and $b \subseteq \text{AP}$ is the set of propositions that have been visited in a branch leading from the root of the tree to α . The root of the tree is labeled with $(v_{in}, \mathcal{L}(v_{in}))$. A tree with label (v, b) has one descendant for each u with $(v, u) \in E$; the label of the descendant is $(u, b \cup \mathcal{L}(u))$.

In order to check if m different propositions can be visited within k -steps, the PSPACE algorithm traverses the game tree in depth first order. Each branch is explored up to one of the two following conditions is met: (i) depth k is reached, or (ii) a node is reached, which has the same label as an ancestor in the tree.

The bottom nodes, where conditions (i) or (ii) are met, are thus the leaves of the tree. In the course of the traversal, the algorithm computes in bottom-up fashion the *value* of the tree nodes. The value of a leaf node labeled (v, b) is $|b|$. For player-1 nodes, the value is the maximum of the values of the successors; for player-2 nodes, the value is the minimum of the value of the successors. Thus, the value of a tree node α represents the minimum number of propositions that player 1 can ensure are visited, in the course of a play of the game that has followed a path from the root of the tree to α , and that can last at most k steps. The algorithm returns Yes if the value at the root is at least m , and no otherwise.

To obtain the PSPACE bound, notice that if a node with label (v, b) is an ancestor of a node with label (v', b') in the tree, we have $b \subseteq b'$: thus, along a branch, the set of propositions appearing in the labels increases monotonically. Between two increases, there can be at most $|\mathcal{G}|$ nodes, due to the termination condition (ii). Thus, each branch needs to be traversed at most to depth $1 + |\mathcal{G}| \cdot (|\text{AP}| + 1)$, and the process requires only polynomial space.

- (2) *PSPACE-hardness.* It follows from the proof of Theorem 5 that the maximal coverage problem for labeled game graphs reduces to the coverage in bounded time problem for labeled game graphs. Since the maximal coverage problem for labeled game graphs is PSPACE-hard (Theorem 5), the result follows.

The result follows. □

Theorem 11. *The coverage in bounded time problem for labeled game graphs that are controllably recurrent is both NP-hard and coNP-hard, and can be decided in PSPACE.*

Proof. It follows from the (PSPACE-inclusion) argument of Theorem 5 that for labeled game graphs that are controllably recurrent the maximal coverage problem can be reduced to the coverage in bounded time problem (with bound $m \cdot n$). Hence the coNP-hardness follows from Theorem 7, and the NP-hardness follows from hardness in labeled graphs that are controllably recurrent (Theorem 8). The PSPACE-inclusion follows from the general case of labeled game graphs (Theorem 10). □

Theorem 11 shows that for controllably recurrent game graphs, the coverage in bounded time problem is both NP-hard and coNP-hard, and can be decided in PSPACE. A tight complexity bound remains an open problem.

Complexity in the size of the game. The maximal coverage problem can alternately be solved in time linear in the size of the game graph and exponential in the number of propositions. Given a game graph $\mathcal{G} = ((V, E), (V_1, V_2), v_{in}, \text{AP}, \mathcal{L})$, construct the game graph $\mathcal{G}' = ((V', E'), (V'_1, V'_2), v'_{in}, \text{AP}, \mathcal{L}')$ where $V' = V \times 2^{\text{AP}}$, $((v, b), (v', b')) \in E'$ iff $(v, v') \in E$ and $b' = b \cup \mathcal{L}(v')$, $V'_i = \{(v, b) \mid v \in V_i\}$ for

$i \in \{1, 2\}$, $v'_{in} = (v_{in}, \mathcal{L}(v_{in}))$, and $\mathcal{L}'(v, b) = \mathcal{L}'(v)$. Clearly, the size of the game graph \mathcal{G}' is linear in \mathcal{G} and exponential in AP. Now consider a reachability game on \mathcal{G}' with the goal $\{(v, b) \mid v \in V \text{ and } |b| \geq m\}$. Player-1 wins this game iff the maximal coverage problem is true for \mathcal{G} and m propositions. Since a reachability game can be solved in time linear in the game, the result follows. A similar construction, where we additionally track the length of the game so far, shows that the maximal coverage problem with bounded time can be solved in time linear in the size of the game graph and exponential in the number of propositions.

Theorem 12. *Given a labeled game graph $\mathcal{G} = ((V, E), (V_1, V_2), v_{in}, \text{AP}, \mathcal{L})$ the maximal coverage and the coverage in bounded time problem can be solved in time linear in $O(|V| + |E|)$ and in time exponential in $|\text{AP}|$.*

5. Related Results

Somewhat surprisingly, despite the central importance of graph coverage in system verification and testing, the basic complexity questions were not studied. In this paper we considered the complexity of the basic problems on finite graphs and game graphs. The basic setting of this paper on graphs and games can be extended in various directions, enabling the modeling of other system features. For example, all our results hold even if every state is labeled with possibly multiple atomic propositions, instead of a single atomic proposition. Below we mention results related to memory requirements of deterministic strategies for the problems considered in the paper, and few important directions in which our results are easily extended.

Memory requirements for deterministic strategies. We now discuss the memory requirement for deterministic strategies. For graphs (as well as recurrent graphs), both for the maximal coverage and the coverage in bounded time problems we have shown that witness polynomial size paths exist ($m \cdot n$ for maximal coverage and $\min(k, m \cdot n)$ for coverage in bounded time). Hence for graphs polynomial size memory is sufficient for deterministic strategies for the maximal coverage as well as the coverage in bounded time problem (the memory remembers the sequences of vertices in the witness path and chooses the next vertex according to the witness path). The requirement of polynomial size memory for recurrent graphs is easily shown as follows: consider the graph with starting vertex v_0 , and vertices v_1 to v_n such that v_0 has edges to all vertices, and all vertices have edges to v_0 . Each vertex is assigned a unique label. To visit all the vertices, a deterministic strategies need to remember the vertices visited. Hence it follows that memory of polynomial size is both necessary and sufficient for both maximal coverage as well as coverage in bounded time for graphs as well as recurrent graphs. It follows from the results of [13] for games with multiple reachability objectives (which is a special case of maximal coverage) that exponential size memory is required. Our PSPACE upper bound proof for the problem shows that exponential memory is sufficient even for the coverage in bounded time problem for game graphs. It follows that exponential

Table 2. Memory requirements for coverage.

	Graphs	Recurrent Graphs	Game Graphs	Recurrent Game Graphs
Maximal Coverage	Polynomial	Polynomial	Exponential	Polynomial
Coverage in Bounded time	Polynomial	Polynomial	Exponential	Polynomial lower bound Exponential upper bound

memory is both necessary and sufficient for game graphs for both maximal coverage and coverage in bounded time. For controllably recurrent game graphs, the polynomial memory requirement follows from the graph case. The polynomial memory upper bound for the maximal coverage problem is as follows: the memory is initially the set of labeled propositions to be visited, and the strategy at any point is to reach a set of labeled propositions not yet visited, then reach the initial state and update the memory with the proposition visited. It follows that polynomial memory is sufficient. For the coverage in bounded time in recurrent game graphs, the exponential memory upper bound follows from the more general case of general game graphs. The lower bound is polynomial, and an exponential lower bound is an open problem (note that a polynomial upper bound would imply NP upper bound as the strategy can be guessed and verified in polynomial time, and since the problem is coNP-hard, a polynomial upper bound is unlikely). We summarize the memory requirements in Table 2.

Incomplete information. So far, we have assumed that at each step, the tester has complete information about the state of the system under test. In practice, this may not be true, and the tester might be able to observe only a part of the state. This leads to graphs and games of *imperfect information* [20]. The maximal coverage and the coverage in bounded time problem for games of imperfect information can be solved in EXPTIME. The algorithm first constructs a perfect-information game graph by subset construction [20], and then run the algorithm of Theorem 12, that is linear in the size game graph and exponential in the number of propositions, on the perfect-information game graph. Thus, the complexity of this algorithm is EXPTIME. The reachability problem for imperfect-information games is already EXPTIME-hard [20], hence we obtain an optimal EXPTIME-complete complexity.

Timed systems. While we have studied the problem in the discrete, finite-state setting, similar questions can be studied for timed systems modeled as timed automata [2] or timed game graphs [18]. Such problems would arise in the testing of real-time systems. We omit the standard definitions of timed automata and timed games. The maximal coverage problem for timed automata (respectively, timed games) takes as input a timed automaton T (respectively, a timed game T), with the locations labeled by a set AP of propositions, and a number m , and asks whether m different propositions can be visited. An algorithm for the maximal coverage problem for timed automata constructs the region graph of the automaton [2] and runs

the algorithm of Theorem 9 on the labeled region graph. This gives us a PSPACE algorithm. Since the reachability problem for timed automata is PSPACE-hard, we obtain a PSPACE-complete complexity. Similarly, the maximal coverage problem for timed games can be solved in exponential time by running the algorithm of Theorem 12 on the region game graph. This gives an exponential time algorithm. Again, since reachability on timed games is EXPTIME-hard [16], we obtain that maximal coverage problem in timed games are EXPTIME-complete.

Pushdown systems. We now discuss the extension of our results to systems with stack or pushdown store to model systems with recursion. We omit the standard definitions of pushdown automata. Pushdown reachability is decidable in polynomial time (for example using the algorithm of [22]) and there is a polynomial witness of the existence of a reachable path (though the path itself can be exponential). The argument for the polynomial witness of a reachable path is as follows: it follows from the results of [22] that if there is reachable path, then there is a reachable path with stack height at most $(|S| \times |\Gamma|)^2$ (i.e., where $|S|$ is the size of the state space and $|\Gamma|$ is the stack alphabet), and the argument for bounding the stack height is essentially the pumping lemma style argument. The polynomial witness path is constructed by dynamic programming as follows: the iterative algorithm in iteration i already has the graph where there are edges such that there is a path with stack height at most $i - 1$ (i.e., reachability with stack height at most $i - 1$ is summarized in an edge), and then computes the reachability with one additional stack height (and the reachability information for stack height at most i is captured as an edge for the following iteration). For pushdown reachability, the algorithm only needs to consider i upto $(|S| \times |\Gamma|)^2$. Thus pushdown reachability and a polynomial witness of a (possibly of exponential length) path can be constructed in polynomial time. It follows that for the maximal coverage problem we have polynomial bounds on the witness of a path for maximal coverage. Hence the NP upper bound follows, and since the problem is NP-hard (as shown by our results even without pushdown store), it follows that the maximal coverage problem is NP-complete for pushdown systems. It was shown in [21] that pushdown game graphs with reachability objectives can be solved by solving the reachability problem on an exponentially large game graph. The maximal coverage problem for pushdown games can be solved in exponential time as well by running the algorithm of Theorem 12 on the same exponential size game graph for reachability objectives. This gives an exponential time algorithm. Again, since reachability on pushdown games is EXPTIME-hard [21], we obtain that maximal coverage problem in pushdown games is EXPTIME-complete.

6. Conclusion

In this work we considered the complexity of coverage problems for finite graphs, game graphs, and the special class of recurrent graphs and recurrent game graphs. In most cases we establish optimal complexity bounds, and also optimal memory bounds for the testing strategies. The exact complexity and memory bound for

the bounded coverage problem for recurrent game graphs remain open; we show the problem is in PSPACE and DP-hard (both NP-hard and coNP-hard). We also show that from our results for finite graphs and game graphs optimal complexity results can be established for several important extensions, such as incomplete-information games, timed systems for real-time systems and pushdown systems for systems with recursion.

References

- [1] R. Alur, C. Courcoubetis, and M. Yannakakis. Distinguishing tests for nondeterministic and probabilistic machines. In *Proc. 27th ACM Symp. Theory of Comp.*, 1995.
- [2] R. Alur and D. Dill. The theory of timed automata. In *Real-Time: Theory in Practice*, volume 600 of *Lect. Notes in Comp. Sci.*, pages 45–73. Springer-Verlag, 1991.
- [3] P. Ammann and J. Offutt. *Introduction to software testing*. Cambridge University Press, 2008.
- [4] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.
- [5] T. Ball. A theory of predicate-complete test coverage and generation. In *FMCO 04: Formal Methods for Components and Objects*, LNCS 3657, pages 1–22. Springer, 2004.
- [6] D. Beyer, A. Chlipala, T. A. Henzinger, R. Jhala, and R. Majumdar. Generating tests from counterexamples. In *ICSE 04*, pages 326–335. IEEE, 2004.
- [7] A. Blass, Y. Gurevich, L. Nachmanson, and M. Veanes. Play to test. In *FATES*, volume 3997 of *Lecture Notes in Computer Science*, pages 32–46. Springer, 2005.
- [8] E. Brinksma and J. Tretmans. Testing transition systems: An annotated bibliography. In *MOVEP 00: Modeling and Verification of Parallel Processes*, volume 2067 of *Lecture Notes in Computer Science*, pages 187–195. Springer, 2000.
- [9] C. Cadar, V. Ganesh, P. M. Pawlowski, D. L. Dill, and D. R. Engler. Exe: automatically generating inputs of death. In *ACM Conference on Computer and Communications Security*, pages 322–335, 2006.
- [10] E. M. Clarke, D. Kroening, and F. Lerda. A tool for checking ansi-c programs. In *TACAS 04*, volume 2988 of *Lecture Notes in Computer Science*, pages 168–176. Springer, 2004.
- [11] C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *J. ACM*, 42(4):857–907, 1995.
- [12] L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997. Technical Report STAN-CS-TR-98-1601.
- [13] N. Fijalkow and F. Horn. The surprising complexity of reachability games. *CoRR*, abs/1010.2420, 2010.
- [14] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman and Co., 1979.
- [15] P. Godefroid, N. Klarlund, and K. Sen. Dart: directed automated random testing. In *PLDI*, pages 213–223, 2005.
- [16] T. Henzinger and P. Kopke. Discrete-time control for rectangular hybrid automata. *Theoretical Computer Science*, 221:369–392, 1999.
- [17] D. Lee and M. Yannakakis. Optimization problems from feature testing of communication protocols. In *ICNP: International Conference on Network Protocols*, pages 66–75. IEEE Computer Society, 1996.
- [18] O. Maler, A. Pnueli, and J. Sifakis. On the synthesis of discrete controllers for timed systems. In *Proc. of 12th Annual Symp. on Theor. Asp. of Comp. Sci.*, volume 900 of *Lect. Notes in Comp. Sci.* Springer-Verlag, 1995.

- [19] C.H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1993.
- [20] J.H. Reif. The complexity of two-player games of incomplete information. *Journal of Computer and System Sciences*, 29:274–301, 1984.
- [21] I. Walukiewicz. Pushdown processes: Games and model checking. *Information and Computation*, 164:234–263, 2001.
- [22] M. Yannakakis. Graph-theoretic methods in database theory. In *PODS*, pages 230–242, 1990.
- [23] M. Yannakakis. Testing, optimization, and games. In *LICS*, pages 78–88. IEEE Computer Society, 2004.
- [24] M. Yannakakis and D. Lee. Testing for finite state systems. In *CSL 98: Computer Science Logic*, volume 1584 of *Lecture Notes in Computer Science*, pages 29–44. Springer, 1998.