Automated Test Generation from Vulnerability Signatures

Abdulbaki Aydin, Muath Alkhalaf, and Tevfik Bultan
Computer Science Department
University of California, Santa Barbara
Email: \{baki,muath,bultan\}@cs.ucsb.edu

Abstract—Web applications need to validate and sanitize user inputs in order to avoid attacks such as Cross Site Scripting (XSS) and SQL Injection. Writing string manipulation code for input validation and sanitization is an error-prone process leading to many vulnerabilities in real-world web applications. A vulnerability signature is a characterization of all user inputs that can be used to exploit a vulnerability. Recent research in automata-based static string analysis resulted in techniques for automated computation of vulnerability signatures represented as automata. However, there are several factors that limit the applicability of static string analysis in general: 1) undecidability of static string analysis requires the use of approximations leading to false positives, 2) static analysis tools do not handle all string operations, 3) dynamic nature of the web scripting languages makes static analysis difficult. In this paper, we show that vulnerability signatures computed for deliberately insecure web applications (developed for demonstrating different types of vulnerabilities) can be used to generate test cases for other applications. Given a vulnerability signature represented as an automaton, we present algorithms for test case generation based on state, transition, and path coverage. These automatically generated test cases can be used to test applications that are not analyzable statically, and to discover attack strings that demonstrate how the vulnerabilities can be exploited.

I. INTRODUCTION

Correctness of input validation and sanitization operations is a crucial problem for web applications. One of the main forms of interaction between a user and a web application is through text fields. The text entered by the user is parsed by the web application and used as the input parameter for the action that is executed in response to the user’s request. During action execution, user input can be passed as a parameter to security sensitive operations such as sending a query to the back-end database. If the input sent by the user inserts unintended commands to the generated database query (which is called SQL injection), then security of the application can be compromised resulting in unauthorized access to sensitive data or loss of data. In another attack scenario, a user can send an input that stores malicious code in the database that can later be used for attacking other users’ machines, which is called Cross Site Scripting (XSS). Even for input fields which are not entered as text fields (such as inputs that are entered using a drop box), a malicious user can change the input field and insert an attack by manipulating the http request that is generated by the browser.

In order to ensure the security of a web application, the user inputs that flow into security sensitive functions like databases queries must be correctly validated and sanitized. Due to global accessibility of web applications, malicious users all around the world can exploit a vulnerable web application and cause significant damage. Given the significance of this security threat, one would expect web application developers to be extremely careful about input validation and sanitization. Unfortunately, web applications are notorious for security vulnerabilities such as SQL injection and XSS that are due to lack of input validation and sanitization, or errors in string manipulation operations used for input validation and sanitization.

In this paper we present an automated testing framework that targets testing of input validation and sanitization operations in web applications for discovering vulnerabilities. Our framework combines automated testing techniques with recently developed static string analysis techniques for vulnerability analysis [1]. Although static string analysis techniques are powerful, they are not always feasible for analyzing real world applications due to various reasons such as cost of the analysis, missing models for library functions, and the difficulty of statically resolving dynamic behaviors of programs written in scripting languages. Moreover, since static string analysis is undecidable, these techniques use abstractions and approximations which lead to false positives.

In our approach we use static string analysis to obtain an over-approximation of all the input strings that can be used to exploit a certain type of vulnerability. This set of strings is called a vulnerability signature. Note that, this could be an infinite set and could include arbitrarily long strings. Since we use automata-based string analysis, the vulnerability signatures are characterized as automata.

For specification of different types of vulnerabilities we use attack patterns developed by security researchers. These are regular expressions that characterize the strings that would cause a vulnerability when sent to a security sensitive function. Given an attack pattern and a web application, we use automata-based string analysis techniques to generate an automaton that corresponds to the vulnerability signature for that application for the type of vulnerability characterized by the attack pattern. As input web applications, we use the deliberately insecure web applications that are developed by security researchers to demonstrate different types of programming practices that lead to vulnerabilities.

Using the vulnerability signature automata generated by analyzing the deliberately insecure web applications, we automatically generate test cases based on three coverage criteria: state, transition and path coverage. Each test case corresponds
to a string such that, when that string is given as a text field input to a web application, it may exploit the vulnerability that is characterized by the given vulnerability signature. Our automated test generation algorithm tries to minimize the number of test cases while achieving the given coverage criteria.

In order to demonstrate the effectiveness of our approach we experimented on several real-world web applications. As we report later in the paper, the automatically generated test sets were very effective in identifying vulnerabilities in these applications.

The rest of the paper is organized as follows. In Section II we give an overview of our approach. In Section III we review the vulnerability signature generation techniques we use. In Section V and VI we discuss the test generation algorithms we use. In Section VII we show the experimental results of our approach. In Section VIII we discuss the related work, and we conclude the paper in Section IX.

II. MOTIVATION AND OVERVIEW

The high-level flow of our automated testing framework for input validation and sanitization functions is shown in Figure 1. In this section we give an overview of different aspects of our approach, before explaining the technical details in the following sections.

A. Automata-based Static String Analysis

Our automated testing framework generates test cases from vulnerability signatures. A vulnerability signature is a characterization of all user inputs that can exploit a vulnerability. In our framework we use automata-based string analysis in which vulnerability signatures are represented as automata. Automata-based string analysis is a static program analysis technique. Given a set of input values represented as automata, it symbolically executes the program to compute the set of string values that can reach to each program point. Using a forward-analysis that propagates input values to sinks (i.e., security sensitive functions), it is possible to identify attack strings that can reach to a given sink. Then, a backward analysis that propagates the attack strings back to user input results in an automaton that corresponds to the vulnerability signature.

Automata-based static string analysis is challenging due to several reasons. Due to undecidability of string verification problem, string analysis techniques use conservative approximations that over-approximate the vulnerability signatures. Due to these approximations techniques use conservative approximations that over-approximate the vulnerability signatures. Due to these approximations techniques use conservative approximations that over-approximate the vulnerability signatures. Due to these approximations vulnerability signatures may contain strings that do not correspond to attacks, leading to false positives. Moreover, string analysis tools only model a subset of available string library functions, and when an unmodeled library function is encountered, the function has to be over-approximated to indicate that it can return all string values, which results in further loss of precision. Furthermore, forward and backward symbolic execution using automata can cause exponential blow-up in the size of the automata when complex string manipulation operations such as string-replace are used extensively. Finally, dynamic nature of scripting languages used in web application development makes static analysis very challenging and applicable to a restricted set of programs. Due to all these challenges it is not possible to have a push-button automata-based string analysis that works for all real-world applications.

In this paper we combine static vulnerability analysis techniques with automated test generation. The combined approach compensates for the weaknesses of the static vulnerability analysis techniques. In our approach static vulnerability analysis is applied to a small set of programs and the results from this analysis is used for testing other applications. Hence, programs with features that make static vulnerability analysis infeasible can still be checked using automated testing. Moreover, the approximations that are introduced by static vulnerability analysis that lead to false positives are eliminated during testing.

B. Generating Vulnerability Signatures from Deliberately Insecure Applications

Security researchers have developed applications that are deliberately insecure to demonstrate typical vulnerabilities. These applications are sometimes used to teach different pitfalls to avoid in developing secure applications, and sometimes they are used as benchmarks for evaluating different vulnerability analysis techniques. In our framework we use static string analysis techniques to analyze deliberately insecure applications and to compute a characterization of inputs that can exploit a given type of vulnerability.

In order to generate the vulnerability signature for an application, we need an attack pattern (specified as a regular expression) that characterizes a particular vulnerability. An attack pattern represents the set of attack strings that can exploit a particular vulnerability if they reach a sink (i.e., a security sensitive function). Attack patterns for different types
of vulnerabilities are publicly available and can be used for vulnerability analysis.

Given an attack pattern and a deliberately insecure web application, we use automata-based static string analysis techniques to generate a vulnerability signature that characterizes all the inputs for that application that can result in an exploit for the vulnerability characterized by the given attack pattern. Since we use automata-based string analysis techniques, at the end of the analysis we obtain an automaton that corresponds to the vulnerability signature. I.e., the vulnerability signature automaton only accepts the strings that are in the vulnerability signature. In the next phase of our approach we automatically generate test cases from the vulnerability signature automaton.

C. Automated Test Generation from Vulnerability Signatures

Given a vulnerability signature automaton, any string accepted by the automaton can be used as a test case. Hence, any path from the start state of the vulnerability signature automaton to an accepting state characterizes a string which can be used as a test case. However, a vulnerability signature automaton typically accepts an infinite number of strings since, typically, there are an infinite ways one can exploit a vulnerability. In order to use vulnerability signature automata for testing, we need to somehow prune this infinite search space. Our overall goal is to minimize the number of test cases while making sure that we cover all possible ways of exploiting a vulnerability.

The mechanism that allows an automaton to represent an infinite number of strings is the loops in the automaton. So, in order to minimize the number of test cases, we have to minimize the way the loops are traversed. We do this by identifying all the strongly-connected components (SCCs) in an automaton and then collapsing them to construct a directed acyclic graph (DAG) that only contains the transitions of the automaton that are not part of an SCC and represents each SCC as a single node. Using this DAG structure, we do test generation for three coverage criteria: 1) state coverage where the goal is to cover all states of the automaton (including the ones in an SCC), 2) transition coverage, where the goal is to cover all transitions of the automaton (including the ones in an SCC), 3) path coverage, where the goal is to cover all the paths in the DAG that is constructed from the automaton, while also covering all possible ways to enter and exit from an SCC.

We implement the state and transition coverage using the min-cover paths algorithm that we execute on the DAG representation followed by a phase where we ensure the coverage of the states and transitions inside the SCC nodes. We implement the path coverage using depth-first-traversal, where when an SCC node is encountered we ensure that all entry and exit combinations are covered in the generated test cases.

D. A Sanitization Example

One of the well-known XSS attack patterns is characterized by the following regular expression:

\( ^{/}.*<\text{script}\.*>.* \)

This attack pattern states that any string that contains the string <script> is potentially an attack. The script-tag indicates executable code and a malicious user might be trying to insert a malicious script that will later on be executed on another user's machine. Now, consider the example code in Figure 2 extracted from a deliberately insecure web application. This code is sanitizing the input provided by the user for the "name" field in line 7 by deleting all appearances of the string <script> (it deletes it by replacing each appearance of the string <script> with the empty string). Later on in the program the variable $html is used as an input for a security sensitive function, so if the sanitization is not done properly this application would have a vulnerability.

We can try to check if the application is vulnerable by generating a test string from the attack pattern. For example, we can test the application on the input <script> and as expected the sanitization code will correctly remove the script-tag and sanitize the input. So, based on this test input does not detect a vulnerability. However, this application has a vulnerability and the sanitization used in Figure 2 is incorrect.

When we run the automata-based string analysis on this example, we find out that the intersection of the set of strings that can reach the sink and the attack pattern is not empty, i.e., there are some inputs that will cause a string containing the script-tag reach the sink. So, we generate the vulnerability signature for this application which results in and automaton that contains 59 states and 8530 transitions. Note that, this vulnerability signature automaton captures the fact that the string-replace operation in line 7 will delete all appearances of the string <script> from the input. The reason that there are thousands of transitions is due to the fact that there is a transition for each ASCII character from each state.

We use our automated test generation technique to generate a test string from the vulnerability signature automaton and obtain the following test input:

<scriipt><script>

When we run the application with this input we discover an attack, i.e., the sink function receives an input that contains the string <script>. This is due to the fact that the incorrect sanitization function in Figure 2 deletes the substring <script> from the above test input and creates the attack string.

In our framework, we use the test strings generated from vulnerability signatures of deliberately insecure web applications to test other applications. If the applications we test contain sanitization errors similar to the errors in deliberately insecure web applications or if they do not use proper sanitization, then the generated test cases can discover their vulnerabilities without analyzing them statically. Note that the test inputs generated from vulnerability signatures can also be used for applications that are statically analyzable in order to eliminate false positives and construct exploits (i.e., to generate concrete inputs that demonstrate how a vulnerability can be exploited).

III. Vulnerability Signature Generation

We use an automata-based string analysis to generate the vulnerability signature from an application [2], [1]. This analysis takes as input a dependency graph for the input program. A dependency graph is a directed graph that specifies how the values of user inputs flow to the security sensitive functions (sinks). The analysis consists of two phases. In the first phase, we perform a forward symbolic reachability analysis starting
from nodes associated with input to compute all possible values that each node in the dependency graph can take. We use this information to collect vulnerable program points, as well as the reachable attack strings for those vulnerable program points. If the program is vulnerable, i.e., if there exists some vulnerable program points, we proceed to the second phase. In the second phase, we perform a backward symbolic reachability analysis from the vulnerable program points to compute all possible values of their predecessors that will result in attack strings at these vulnerable program points.

Figure 3 shows the algorithm used in our analysis. The algorithm takes three inputs: a dependency graph (denoted as $G$), a set of sink nodes (denoted as $\text{Sink}$), and an attack pattern (denoted as $\text{Attk}$). $G$ is a directed dependency graph that specifies how the values of user inputs flow to the security sensitive functions. $\text{Sink}$ denotes the nodes that are associated with security sensitive functions that might lead to vulnerabilities. $\text{Attk}$ is a regular expression represented as an automaton that accepts the set of attack strings. At each node, the set of reachable string values is approximated as a regular language and represented symbolically as an automaton that accepts the language. To associate each node with its automaton, we create two automata vectors $\text{POST}$ and $\text{PRE}$. The size of both is bounded by the number of nodes in $G$. $\text{POST}[n]$ is the automaton accepting all possible string values that can reach node $n$. $\text{PRE}[n]$ is the automaton accepting all possible string values that node $n$ can take to exploit the vulnerability. Initially, all these automata accept nothing, i.e., their language is empty. $\text{Vul} \subseteq \text{Sink}$ is the set of vulnerable program points, and initially it is set to an empty set.

At line 4, we first compute $\text{POST}$ by calling the forward analysis. At line 5, for each node $n \in \text{Sink}$, we generate an automaton $\text{tmp}$ by intersecting the attack pattern and the possible values of $n$. If the language of $\text{tmp}$, i.e., $L(\text{tmp})$, is not empty, we identify that $n$ is a vulnerable program point and add it to $\text{Vul}$ at line 8. In fact, $\text{tmp}$ accepts the set of reachable attack strings at node $n$ that can be used to exploit the vulnerability. Hence, we assign $\text{tmp}$ to $\text{PRE}[n]$ at line 9. If $\text{Vul}$ is not empty, we compute $\text{PRE}$ by calling our backward analysis at line 13. Note that for $n \in \text{Vul}$, $\text{PRE}[n]$ has been assigned. We report vulnerability signatures for each input node based on $\text{PRE}$ at line 14-16. If $\text{Vul}$ is an empty set, we report that the program is secure with respect to the attack pattern.

The forward symbolic reachability analysis is based on a standard work queue algorithm. We iteratively update the automata vector $\text{POST}$ until a fixpoint is reached [2]. Backward analysis uses the results of the forward analysis. Particularly, it computes all possible values of each node $n$ that can exploit the identified vulnerability. The challenge in both forward and backward analyses is computing pre and post-conditions of string manipulation functions such as concatenation, string-replace etc., where the inputs and outputs of the pre and post-condition operations are automata. We use the techniques described in [2] for pre and post-condition operations and the details of the symbolic automata-based forward and backward analyses can be found in [1].

The output of the vulnerability signature generation algorithm is a set of vulnerability signature automata. A vulnerability signature automaton is a tuple $V = (Q, \Sigma, \delta, q_0, F)$, where $Q$ is the set of states, $\Sigma$ is the input alphabet, $\delta \subseteq Q \times \Sigma \times Q$ is the transition relation, $q_0 \in Q$ is the initial state, and $F \subseteq Q$ is the set of final states. The alphabet $\Sigma$ is the set of ASCII characters. Each transition $t \in \delta$ is a tuple $t = (q, c, q')$ where $q = \text{source}(t)$, $q' = \text{target}(t)$ and $c \in \Sigma$. The vulnerability signature automata are deterministic, i.e., there is a single transition for each source state and alphabet symbol.

IV. CONVERTING VULNERABILITY SIGNATURE AUTOMATA TO DAGS

Some features of the vulnerability signature automata make test generation difficult. One feature is that there are large
number of transitions in $\delta$ where $source(t_0) = source(t_1) = source(t_2) = \ldots = source(t_n)\text{ and } target(t_0) = target(t_1) = target(t_2) = \ldots = target(t_n)$. Such transitions cause an exponential blow up in the number of accepting paths in the automaton, and this leads to a large search space for test generation. As an example consider state $q_2$ in Figure 4. For this relatively small automaton there are $128 \times 128$ accepting paths. Our solution to this problem is to collapse the transitions that have the same source and target states into one transition as shown in Figure 5. The label of the collapsed transition is a range of characters corresponding to each transition that it represents. During test generation we only pick one character from the range representing the all corresponding transitions. This allows us to avoid exponential blow up in the number of accepting paths. For the rest of the paper we assume that all transitions with the same source and target states are collapsed.

Another feature of vulnerability signature automata is that they can contain cycles which results in an infinite number of accepting paths, i.e., an infinite search space for test generation. As an example, in Figure 6, states $\{q_1, q_2, q_3\}$ and $\{q_4, q_5\}$ form cycles. In order to bound the number of accepting paths and, therefore the search space for test generation, we extract a high level representation of the given vulnerability signature automaton by identifying its strongly connected components (SCC). The high level representation we obtain is a directed acyclic graph $DAG = (N, E)$ where $N$ is the set of SCCs and $E$ is the set of edges between SCCs. At the automaton level each edge $e \in E$ is a transition such that $source(e) \in SCC_x$, $target(e) \in SCC_y$ and $SCC_x \neq SCC_y$. We use Tarjan’s strongly connected components algorithm to identify the cycles in the vulnerability signature automata [3]. The worst case time complexity of this algorithm is $O(|V| + |E|)$ for a given vulnerability signature automaton $V = (Q, \Sigma, \delta, q_0, F)$. High-level DAG representation for the automaton in Figure 6 is shown in Figure 7. It consists of four strongly connected components $N = \{SCC_0, SCC_1, SCC_2, SCC_3\}$, and six edges among them $E = \{e_a, e_b, e_k, e_n, e_f, e_h\}$.

In this section we discuss generating test cases from vulnerability signature automata based on state and transition coverage criteria. Given a vulnerability signature automaton $V = (Q, \Sigma, \delta, q_0, F)$, let $L(V)$ denote the set of strings accepted by $V$. Our aim is to find two sets of strings $S_{sc}, S_{tc} \subseteq L(V)$ that achieve state and transition coverage, respectively. The state and transition coverage definitions are as follows:

- For each state in $q \in Q$ there must be at least one string in $S_{sc}$ such that the accepting path for that strings visits $q$.
- For each (collapsed) transition $t \in \delta$ there must be at least on string in $S_{tc}$ such that the accepting path for that string includes $t$.

Finally, we want to generate the sets $S_{sc}$ and $S_{tc}$ in such a way that $|S_{sc}|$ and $|S_{tc}|$ are minimized.

The problem of finding minimum number of strings based state and transition coverage criteria is very similar to a well-known graph problem called minimum cover paths. Given a directed acyclic graph, minimum cover paths is the least number of paths that visits each edge of the graph at least once. Minimum cover paths problem has been studied in different research areas and there are well known solutions to this problem [4], [5], [6]. One known solution is to reduce minimum cover paths problem to the minimum flow problem [4], [6]. We follow this basic approach with some modifications. We can divide the state and transition coverage algorithms into five main steps: 1) Initialization of DAG, 2) Converting DAG into a flow graph, 3) Minimum flow algorithm, 4) Finding minimum covering paths, 5) Extending paths with SCC Coverage.

A. Initialization of DAG

Vulnerability signature automaton $V = (Q, \Sigma, \delta, q_0, F)$ has one start state $q_0$ and a set of final states $F$. In order to apply flow algorithms and minimum covering paths algorithm, one virtual final state $q_0'$ is added to $Q$, for each $q \in F$, a virtual transition $t_v = (q, \lambda, q')$ is added to the transition relation $\delta$ where $\lambda$ is a new symbol added to the alphabet $\Sigma$. The modified automaton has one start state $q_0$ and one final state.
1: procedure PREPROCESSRIGHT(node, queue)
2:   updated ← False
3:   for all edge ∈ outgoingEdges(node) do
4:     nextNode ← targetNode(edge)
5:     if flow(edge) = 0 then
6:       if #incomingEdges(nextNode) = 1 or
7:           #outgoingEdges(nextNode) = 1 then
8:         flow(edge) ← 1
9:       updated ← True
10:      end if
11:     end if
12:   end for
13:   if not updated or balanced(node) = 0 then
14:     return
15:   end if
16:   if updated and balanced(node) < 0 then
17:     queue.enqueue(node)
18:   end if
19:   if updated and balanced(node) > 0 then
20:     DISTRIBUTEFLOWS(EVENLY)(node)
21:   end if
22:   for all edge ∈ outgoingEdges(node) do
23:     nextNode ← targetNode(edge)
24:     PREPROCESSRIGHT(nextNode, queue)
25:   end for
26: end procedure

Figure 8. Phase 1 for Pre-Processing of State Coverage

Figure 9. Initial DAG for State Coverage

Phase 1 of the pre-processing algorithm for transition coverage is shown in Figure 10. The only modification compared to the algorithm shown in Figure 8 is inside the if block at line 5. The resulting flows for transition coverage are shown in Figure 11. Starting from the initial node, the algorithm first assigns a flow value of 1 to the edges ‘a’ and ‘b’. When it comes to SCC2 during depth first traversal, it first assigns a flow of 1 to the edges ‘k’ and ‘n’. As a result balance value of SCC2 becomes -1 and that SCC2 is queued for reverse pre-processing. Similarly when algorithm first visits the SCC1 using edges ‘a’ or ‘k’, balance value for SCC1 becomes negative and SCC1 is also queued for reverse pre-processing. However, when the algorithm visits SCC1 for the second time, balance value becomes 0 and reverse pre-processing of SCC1 does not have an effect.

B. Converting DAG into a Flow Graph

Given a DAG, an admissible flow assignment is needed for each edge in order to apply the min-flow algorithm. We use a pre-processing algorithm [4] to assign an initial flow to each edge based on the number of input and output edges for each node. This is a two phase algorithm that consists of a depth first traversal starting from start node (Phase 1) followed by a reverse depth first traversal (Phase 2) if necessary. The first phase of the initialization for state coverage is shown in Figure 8.

The statement at line 6 checks for the edges that can be removed safely. For example edges labeled with ‘f’ and ‘k’ can be safely removed from Figure 7. The resulting high level DAG is shown in Figure 9. Depending on the order that for loop retrieves the edges at line 3, algorithm may remove different edges at different runs. However, this does not affect the state coverage.

We can define the flow function flow(e) as number of visits for an edge e ∈ E. The balanced() function compares the total input flow and total output flow for a node n ∈ N based on flows for each incoming and outgoing edges. A positive balance means that the total input flow is larger than the total output flow. In that case line 20 distributes the input flows to the the output flows by updating the flow values of outgoing edges. For the case of a negative balance value, distribution is done in the reverse direction after Phase 1 finishes as described in [4]. Figure 9 also shows the initial flow values that are assigned to the example DAG. For the example shown in Figure 9, reverse pre-processing (Phase 2) is not necessary since in the first phase flows are already distributed correctly.

C. Minimum Flow Algorithm

After we have initial flows calculated, Ford-Fulkerson algorithm is applied to the flow graph [7]. Ford-Fulkerson algorithm computes the minimum flows to visit each transition at least once. The algorithm finds paths from the start node to the final node and removes the maximum amount of flow from each path without reaching 0. Assume that our initialization phase calculated the flow for the path "bkh" in Figure 11 as "b(4)k(3)h(3)" instead of "b(2)k(1)h(1)". We can take away 2 flows from all the edges in the path "bkh". The amount of flow that can be removed is called the residual value. Time complexity of the algorithm for a DAG is \(O(|p_{max}| \cdot (f_0 - f_{min}))\) where \(|p_{max}|\) is the maximum length path from start node to final node, \(f_0\) is initial flow set and \(f_{min}\) is the minimum flow [4].
for state coverage and
let for both coverage criteria is worst case time complexity for state and transition coverage.

maximum length path from start node to final node. Then, we have state and transition coverage (

as: “Figure 12 shows the general loop and the recursive path finding forms a path that ends at the final node (i.e., the virtual node).

flow looking for minimum covering paths. Minimum Covering among the sets E

D. Finding Minimum Covering Paths

After running Minimum Flow Algorithm we can start looking for minimum covering paths. Minimum Covering Paths algorithm finds the edges that have flow(e) > 0 and forms a path that ends at the final node (i.e., the virtual node). Figure 12 shows the general loop and the recursive path finding function. For example, given the DAG shown in Figure 11, the minimum covering paths for transition coverage are computed as: “afev”, “bkhev”, and “bnev” where ev is the virtual edge.

Let Nk be the set of nodes that are k edges away from the start node. Let Ek be the set of edges between Nk and Nk+1. Let Emax be the edge set with maximum size among the sets E0, E1, E2, ..., Es. Finally, let Pmax be the maximum length path from start node to final node. Then, worst case time complexity for state and transition coverage is O(|Pmax| × |Emax|) and the maximum size test set size for both coverage criteria is O(|Emax|) which is equal to the number of minimum covering paths. For the DAGs that are extracted from the same vulnerability signature automaton let Emaxsc denote the size of Emax for the DAG generated for state coverage and Emaxtc denote the size of Emax for the DAG generated for transition coverage. Then, we have |Emaxsc| ≤ |Emaxtc|. For the sets of test cases generated for state and transition coverage (Ssc and Stc, respectively) we have |Sc| ≤ |Stc|.

E. Extending Paths with SCC Coverage

Once we have the results for minimum covering paths we do a pass on each path and extend the SCC nodes n ∈ N that represent cycles. We can define a strongly connected component as SCC = (QSCC, Σ, δSCC) where QSCC ⊆ Q and δSCC ⊆ δ. Assume there is a state qx ∈ QSCC and a transition t ∈ δ. If q(x) = target(t) and source(t) /∈ QSCC, we say state qx is an entry point. Similarly, assume there is an edge qy ∈ QSCC and a transition t ∈ δ. If q(x) = source(t) and target(t) /∈ QSCC, we say state qx is an exit point.

There are two different strategies for SCC coverage based on DAG coverage algorithm in progress. Strategy for the state coverage algorithm is the following: Starting from an entry point visit all states q ∈ QSCC at least once and end up in an exit point. Similarly, for transition coverage starting from an entry point visit all transitions t ∈ δSCC at least once and end up at an exit point. If |δSCC| is greater than zero, then SCC must contain a cycle like SCC1, SCC2, and SCC3 in Figure 7. To terminate the algorithm we keep a queue for unvisited states or unvisited transitions and use depth first search whenever necessary. Figure 13 shows the algorithm we use for state coverage. DFS function at line 7 starts a depth first search from the state given as its first argument and searches for the state given as its second argument without being trapped in a cycle. Once it finds the state given as its second argument, it returns a path that includes all the states it visited. Algorithm for visiting all transitions t ∈ δSCC is the same except we keep a queue for unvisited transitions instead of unvisited states. Both algorithms have a worst case complexity of O(|δSCC|2) which depends on the overlapping cycles within a SCC. Worst case complexity of length of the returned path is also the same as the time complexity.
Consider the example vulnerability signature automaton shown in Figure 9. Based on state coverage algorithm it can produce a path \(a.h\), where each dot corresponds to a node in the DAG. Starting from the first dot which is actually \(SCC_0\) we extend the path. \(SCC_0\) returns an empty path and algorithm continues with next SCC in the path \(a.h\). \(SCC_1\) returns \(ce\) for entry point \(q_1\) and exit point \(q_3\) and algorithm extends the path as \(aceh..\) At the end the algorithm returns the extended path \(aceh\).

VI. PATH COVERAGE FOR FOR VULNERABILITY SIGNATURE AUTOMATA USING DEPTH FIRST TRAVERSAL

A straight forward definition of path coverage would result in an infinite set of test cases due to loops in automata. So, given a vulnerability signature automaton \(V\), we define \(Spc \subseteq L(V)\) as follows:

- For each path \(p\) in the \(DAG\) generated from \(V\) there must be a set of strings in \(Spc\) such that the accepting paths for those strings must correspond to \(p\) (i.e. they must visit the same set of SCCs in the same order), and there must be an accepting path for each combination of entry and exit nodes for all the SCCs in the path \(p\).

Path Coverage algorithm traverses DAG representation of vulnerability signature automata using a depth-first traversal (DFT). It does not have any initialization phase. It handles SCC entry-exit point coverage during path exploration. Assume current node in the DFT is \(n\) and \(n\) corresponds to a SCC. Again assume \(q_e\) is the entry point for the SCC corresponding to node \(n\). Path coverage algorithm calculates paths for all possible combinations of \(q_e\) with all exit points using the SCC coverage algorithm we have for transition coverage. Then, it continues to explore paths in the high level DAG representation by following exit points in a DFT manner. By doing so, path coverage algorithm calculates all possible combinations of all entry and exit points of a SCC. The path coverage algorithm generates 5 paths for the example shown in Figure 11.

Based on definitions we have in previous section the time complexity for path coverage is \(O(|Enax|^{1.5})\). Test size complexity is the same as the time complexity which is basically all paths from start node to final nodes. As a result we have the following test set size comparison for the three coverage criteria for the same vulnerability signature \(|Sc| \leq |Sc| \leq |Spc|\).

VII. IMPLEMENTATION AND EXPERIMENTS

In order to experiment with our automated testing framework, we used a deliberately insecure web application called Damn Vulnerable Web Application (DVWA) to generate vulnerability signatures. DVWA is listed in OWASP Broken Web Applications Project which lists deliberately insecure web applications. DVWA has several SQL injection, stored XSS and reflected XSS attacks with different security levels provided by the application. Security levels are no sanitization, custom sanitization, and incorrect use of built-in sanitization functions. We generated vulnerability signatures for each attack type considering different security levels. We used the Stranger static string analysis tool [8] to generate vulnerability signatures. We ran all the experiments on an Intel i5 machine with 2.5GHz X 4 processors and 32 GB of memory running Ubuntu 12.04.

Table I shows the properties of 5 vulnerability signatures generated from DVWA. We use the following well known attack patterns for vulnerability signature generation. Attack pattern \(\ldots<\script\ldots>\) or \(\ldots/\ldots\) is used for vulnerability signatures XSS 1, XSS 2, and XSS 3. Attack pattern \(\ldots or \ldots=\ldots\) is used for vulnerability signature SQLI 1 and attack pattern \(\ldots or \ldots=\ldots\) is used for vulnerability signature SQLI 2. The sizes of the vulnerability signature automata depend on the complexity and number of string operations that application has on user inputs. We can see that vulnerability signatures SQLI 1 and XSS 1 are larger than the other three vulnerability signature automata. That is because the corresponding application code has more sanitization on user input. The application code that corresponds to vulnerability signature SQLI 2 has no sanitization at all and the generated vulnerability signature is similar to the attack pattern. For each vulnerability signature, we can see that there is a big difference between actual number of transitions automata has and the number of collapsed transitions which allows us to reduce the sizes of the generated test sets. For a given vulnerability signature, the relation between the sizes of the test sets for different coverage criteria follow the ordering we expect where \(|Ssc| \leq |Sc| \leq |Spc|\). For larger vulnerability signatures, path coverage algorithm produces a large number of strings as expected. For a given vulnerability signature, average length of the strings generated for state coverage is the smallest. Since the number of states are smaller than the number of transitions this is not surprising. The SCC coverage algorithm for state coverage produces strings with smaller lengths for most of the cases.

We extended our test string generation framework to test the efficiencies of test suits generated.

In order to evaluate the effectiveness of our automated test generation techniques we experimented on five open-source applications 1) PHP-Fusion v7.02.05 2 (content management system), 2) RuubikCMS v1.1.1 (website content management tool), 3) UL Forum v1.1.7 (forum application), 4) Snipe Gallery v3.1.5 (image management system), 5) PHP Server Monitor v2.0.1 (server management script). We implemented a web application driver to test applications automatically with the automatically generated test strings. We execute each test string for selected fields from each application. We enable xdebug tool in our server to get the function call traces for each request our web application driver sends. After each
request, the web application driver checks the function calls in the trace files for any sink functions that are called. For the SQL injection attacks, the sinks we identify each call mysql_query function and for XSS attacks sinks we identify each call mysql_query function that executes INSERT or UPDATE statements. If the web application driver finds a sink function, it checks the value of the query parameter of the sink function to see if it contains any type of attack.

Table II shows the result for each application. The sum of the third column and fourth column shows the total number of strings generated from all vulnerability signatures for each coverage criteria. We can clearly say that path coverage and transition coverage have better detection rates than state coverage. The applications php_fusion and ruubik have lower detection rates compared to other three applications. That is because both applications have more string manipulation operations than the other three.

Table II. VULNERABILITY DETECTION PERFORMANCE PER APPLICATION

<table>
<thead>
<tr>
<th>Application</th>
<th>State</th>
<th>Transition</th>
<th>Path</th>
<th># Detected</th>
<th># Missed</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>php_fusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>snipe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phpservermon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III. VULNERABILITY DETECTION PERFORMANCE PER VULNERABILITY SIGNATURE

<table>
<thead>
<tr>
<th>Application</th>
<th>Coverage Type</th>
<th>State</th>
<th>Transition</th>
<th>Path</th>
<th># Detected</th>
<th># Missed</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQLI 1</td>
<td>State</td>
<td>8</td>
<td>11</td>
<td>42%</td>
<td>79</td>
<td>28</td>
<td>74%</td>
</tr>
<tr>
<td>SQLI 2</td>
<td>State</td>
<td>16</td>
<td>58</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>75%</td>
</tr>
<tr>
<td>XSS 1</td>
<td>State</td>
<td>100</td>
<td>481</td>
<td>19</td>
<td>3</td>
<td>1</td>
<td>21%</td>
</tr>
<tr>
<td>XSS 2</td>
<td>State</td>
<td>59</td>
<td>237</td>
<td>3</td>
<td>1</td>
<td>146</td>
<td>78%</td>
</tr>
<tr>
<td>XSS 3</td>
<td>State</td>
<td>11</td>
<td>37</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>73%</td>
</tr>
</tbody>
</table>

Overall we can say that path coverage has better detection rates for both tables as expected. Transition coverage detection rates are very close to path coverage detection rates with less number of strings in total. State coverage is not good enough to produce good attack strings for the vulnerability signatures we have.

VIII. RELATED WORK

Static analysis of strings has been an active research area, with the goal of finding and eliminating security vulnerabilities caused by misuse of string manipulation operations [9], [10], [11], [12], [2], [13] String analysis focuses on statically identifying all possible values of a string expression at a program point, and this knowledge can be leveraged to eliminate vulnerabilities such as SQL injection and XSS attacks. Due to undecidability of string analysis problem static string analysis approaches use conservative approximations such as widening [14], [15], [2], that can result in false positives.
Moreover static modeling of all string manipulation functions is challenging and typically limits the applicability of static string analysis techniques. We are not aware of any prior work that combines static string analysis and vulnerability signatures with automated test generation.

In [16], [17], [18] dynamic symbolic execution has been used for automatic testing of a web application. First, string constraints are generated using symbolic execution. Then, these constraints are solved to generate vulnerable input strings. In [17], [18] authors use a bounded string constraint solver that bounds the length of the strings before solving the constraint. In [16] string constraints are represented using finite state transducers. Unlike dynamic symbolic execution, which is a white box testing approach, our approach is a black-box specification-based testing approach. Dynamic symbolic execution tries to increase execution path coverage while in our case we try to increase coverage of the testing specification.

In XSS Analyzer [19] a black box testing approach is used where a very large database of attack strings is utilized to attack a web application. A learning algorithm is used to pick only a subset of this database. The authors do not discuss how they obtain the attack database. In our approach, we use static analysis to automatically generate vulnerability signatures from which the attack strings are generated. Also, since we generate attack strings from an automaton, the original size of the attack string database could be infinite whereas in XSS analyzer the size of the attack string database is finite.

In [20] a black box SQLI/XSS web vulnerability scanner is developed utilizing manually written attack strings with no specific criteria.

In [21] state machine based test generation using UML state charts is discussed. They define coverage criteria such as single UML transition coverage, full predicate coverage, transition-pair coverage, and complete sequence coverage. These coverage criteria are specific for UML diagrams. In [22] authors generate test cases from finite state machines that correspond to a software system specification. State machine based test generation has been used for different areas such as control systems, protocols, circuit design, data processing, navigation analyses.

Minimum cover paths algorithm has been studied for program testing [6]. It is used to generate minimum number of paths for certain features and generates test data for those paths. It is also used in the area of bioinformatics to convert the graphical data of complex biological experiments into tabular formats [4].

IX. CONCLUSION

We presented an automated testing framework for testing input validation and sanitization operations in web applications. In our framework the tests are generated from vulnerability signatures that are characterized as automata. Our experiments show that vulnerability signatures generated from deliberately insecure web applications can be used to generate effective tests for identifying vulnerabilities in other applications.