

## Research Article

# Experimental Evaluation of FSM-Based Testing Cost for Time-Delay Systems

Zheng Wei, Wu Xiaoxue, Cai Lu, and Hu Desheng

College of Software & Micro-Electronics, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Zheng Wei; zhengweizr@gmail.com

Received 28 April 2014; Accepted 26 June 2014; Published 22 July 2014

Academic Editor: Yuxin Zhao

Copyright © 2014 Zheng Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With time-delay systems arising, time-delay system testing has attracted much attention. Additionally, evaluating the cost and effectiveness is required to make a good test strategy in practice. In this paper, we take time-delay and other five factors (state number, input number, output number, completeness degree, and accessibility degree) into account and present a timer embedded FSM (TEFSM) model to design a comparative strategy for assessing the coverage criteria and test suites generation methods for time-delay systems. We explore the impact on the average length of test suites, in which the test suites generation methods, coverage criteria, and TEFSM model parameters are involved.

## 1. Introduction

In real world, time-delay exists in various engineering systems such as chemical processes and long transmission lines in pneumatic systems; the study of time-delay systems has gained considerable attention over the past years [1, 2].

The attendant increase in time-delay system complexity has made the timely development of reliable systems extremely challenging. To ensure the quality and reliability of time-delay systems, model based testing technique has been used widely. El-Fakih et al. present a method for deriving test suites with the guaranteed fault coverage for deterministic possibly partial timed finite state machines (TFSMs) [3]. Abramovici and Stroud present a delay-fault testing approach for Field Programmable Gate Arrays based on built-in self-test (BIST) [4]. And many conformance test derivation methods are based on a specification given in the form of a finite state machine (FSM), such as  $W$  [3, 5], partial  $W$  ( $W_p$ ) [6, 7], HIS [8], and  $H$  [9] test derivation methods.

However, the project schedule is always tight and it is impossible to test a system completely. Thus, the efficiency of testing method becomes a key factor for ensuring system reliability. There are some researchers studying the evaluation of the effectiveness of system testing. Briand [10] presents a critical analysis of empirical research in software testing and provides a structured overview and discussion of the validity

issues most commonly encountered while performing empirical studies of testing techniques. It gives us some guides to do experiments and analyze the validity of our research at last. Besides, there are some researchers exploring methods in the field of empirical research. For example, the literature [11] based on state charts proposes a precise simulation and analysis procedure and investigates the cost effectiveness of four adequacy criteria: all transitions (AT), all transition pairs (ATP), and all the paths in the transition tree (TT).

However, they do not have a further analysis on the effect of the parameters of FSMs to the derivation methods. Testers cannot determine derivation methods to use. Moreover, there are no papers to research these in combination (generation methods to coverage criteria, parameters to generation methods/coverage criteria).

Based on the previous discussion, aiming at decreasing the cost of time-delay systems, we present an experimental evaluation method for obtaining minimum test suite for time-delay systems. This method incorporates recent advances in THE literature [10–12]. We take time-delay and other five factors (state number, input number, output number, completeness degree, AND accessibility degree) into account and present a timer embedded FSM (TEFSM) model firstly. Moreover, by taking time-delay as an element of input, the TEFSM can be converted as normal FSM model, and

this makes the evaluation task much easier. Furthermore, the goal of our investigation is focused on minimizing the length of test suite since the minor the test sequence is, the less the system cost is. We present a novel method to evaluate the efficiency of FSM-based testing method and explore the situation that a test sequence generation method may have relatively good effect for a particular or some coverage criteria and how the parameters of a FSM model affect the FSM test suite. Namely, we investigate the relationship between generation methods, coverage criteria, FSM's parameters, and testing cost.

The rest of the paper is organized as follows. Section 2 describes some fundamental concepts such as FSM, TEFSM, and some classic coverage criteria of FSM-based systems. Section 3 presents the research method to covert TEFSM as FSM and decrease the size of test sequence for saving test cost. Section 4 presents experiments and analyzes the results. Finally, this paper makes a conclusion and presents directions for the further research.

## 2. Related Concepts Definition

In this section, some basic concepts are described which will be used in the proposed approach, and TEFSM model is especially described.

**2.1. FSM Model.** Formally, an FSM is a system  $M = (X, Y, S, s_0, \delta, \lambda)$  where  $X$  is a finite input character set,  $Y$  is a finite output character set,  $S$  is a finite state set,  $s_0 \in S$  is an initial state,  $\delta$  is a transition function, and  $\lambda$  is an output controlled function [3].

If each state of one FSM responds to all the inputs, tis is a complete FSM (CFSM); otherwise, it is a partial FSM. That is, if for all  $s_i \in S, x \in X, \exists s_j \in S, \delta(s_i, x) = s_j$ , it is CFSM.

If one state of a FSM reaches different states for the same input, this is a nondeterministic FSM; otherwise, it is a determined FSM. That is, if  $\exists s_i, s_j \in S, x \in X, i \neq j, \delta(s_i, x) = \delta(s_j, x)$ , it is nondeterministic FSM.

If the states of FSM are pair-wise distinguishable, this is a reduced FSM. Otherwise, it is nonreduced. That is, if  $\exists s_i, s_j \in S, \text{ for all } \gamma \in T, i \neq j, \lambda(s_i, \gamma) = \lambda(s_j, \gamma)$ , where  $T$  indicates a set of inputs sequences, it is nonreduced FSM.

**2.2. TEFSM.** TEFSM model is an extension of FSM; it is used for dealing with time-delay system. In this model, a timer is embedded as it is called timer embedded FSM model. Timer means delay time of a state, time-delay is considered as a factor during state transition, timers and entry actions are associated with states, conditions and actions are associated with transitions, and time is implicitly associated with inputs and outputs [12]. The state transition occurs automatically after timeout unless it receives further inputs before timeout. Given a TEFSM as  $M_t = (X_t, Y, S, S_0, \delta, \lambda)$ , in which  $X_t$  indicates input that contains a time-delay, a pair of  $(x, t)$ ,  $x \in X$  and  $t$  is a nonnegative rational and is a time-delay of state transaction, and input  $x$  is applied after time  $t$ .

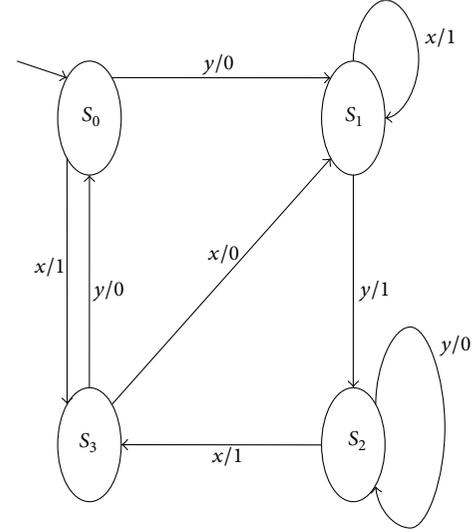


FIGURE 1: FSM.

**2.3. Coverage Criteria.** Coverage criteria rule the sufficient conditions for deciding when to end the testing process; that is, one test suite must satisfy one kind of testing requirements [13]. If  $K$  indicates one coverage criterion and  $T$  indicates a test suite,  $TR_K(M)$  is the testing requirements set of a given  $M$  for criterion  $K$ , and  $TS_K(M, T)$  is the requirement sets satisfied by a test suite  $T$ . Obviously, it is the subset of  $TR_K(M)$ .  $C_K(M, T)$  indicates the coverage rate of  $T$ ; that is  $C_K(M, T) = |TS_K(M, T)|/|TR_K(M)|$ . If  $TS_K(M, T) = TR_K(M)$  or  $C_K(M, T) = 1$ ,  $T$  is a test suite adequate for the criterion  $K$  of  $M$ .

Figure 1 is an example of FSM model which has four states  $S_0, S_1, S_2, S_3$ , respectively, and eight transactions; it will be used for the following coverage criteria explained in this section.

**(1) State Coverage (SC) Criterion.** Cover all states of  $M$ ; hence,  $TR_{SC}(M) = S$ , where  $S$  is the state set of  $M$  and  $TS_{SC}(M, T) = \{s \mid s \in S\}$ . For the FSM in Figure 1, there are four states  $S_0, S_1, S_2$ , and  $S_3$ , and the test suite  $\{x, yy\}$  is SC-adequate for all the four states that could be covered with this test suite. Note that the initial state is reachable with the empty transfer sequence, while the prefix  $y$  of the test  $yy$  is a transfer sequence to state  $S_1$ .

**(2) Transition Coverage (TC) Criterion.** A transition is a state change triggered by an input event. Such as in Figure 1, there are eight transitions between the four states [11]. To cover all transitions of  $M$ , conditions  $TR_{TC}(M) = D$  and  $TS_{TC}(M, T) = \{(s, x) \in D \mid \exists \pi \in T, ax \leq \pi, \delta(S_0, \alpha) = s\}$  need to be satisfied.

The test suite  $\{yx, yyy, yxx, xy, xx\}$  is TC-adequate for Figure 1 because this test suite could cover all the 8 transitions.

**(3) Initialization Fault (IF) Coverage Criterion.** Avoid confusing other states with the initial state, this is for checking whether the behaviors of other states are same as the initial

state, denotes as  $TR_{IF}(M) = \{s \in S \mid s \neq S_0\}$ . There might be no states that can be distinguished from the initial state [11]. However, for a minimum  $M$ , there are states which are different from the initial state. We can express  $TS_{IF}(M, T)$  formally as follows:

$$TS_{IF}(M, T) = \{s \in S \mid s \neq S_0, \exists \chi \in T, \beta\gamma \leq \chi, \\ \delta(S_0, \beta) = s, \lambda(S_0, \gamma) \neq \lambda(s, \gamma)\}. \quad (1)$$

It indicates that there are some sequences in  $T$  to distinguish the initial state from other states.

For the TEFSM in Figure 1, the test suite  $\{yyy, yyyy, xx\}$  is IF-adequate. We know that the input sequence  $y$  is a transfer sequence to state and is followed by  $y$ , which distinguishes state  $s_1$  from the initial state. Similarly, the sequences  $yyyy$  and  $xx$  also satisfy the requirements related to state  $S_3$  and  $S_4$ .

(4) *Transition Fault (TF) Coverage Criterion*. Detect whether the output or the reachable state of one transition is correct to certain input and whether the ending state could be distinguished from other states [14]. We can express  $TR_{IF}(M)$  formally in mathematics as follows:

$$TR_{IF}(M) = \{(s, x, s') \in D \times S \mid \delta(s, x) \neq s'\}. \quad (2)$$

$TS_{IF}(M, T)$  defines whether there are some sequences in  $T$  that could distinguish the ending states of all transitions of  $M$  from other states. We can express it formally as follows:

$$TS_{IF}(M, T) = \{(s, x, s') \in D \times S \mid \\ \delta(s, x) \neq s', \exists \pi, \chi \in T, \alpha\chi \leq \pi, \beta\gamma \leq \chi, \\ \delta(S_0, \alpha) = s', \lambda(\delta(S_0, \alpha\chi), \gamma) \neq \lambda(s', \gamma)\}. \quad (3)$$

For the example TEFSM in Figure 1, the test suite  $\{xxy, yyy, xxx\}$  is TF-adequate. Consider, for instance, the transition  $(S_3, x)$  whose tail state is  $S_1$ . Test  $xx$  covers this transition. States  $S_1$  and  $S_2$  are distinguished by  $y$ , which follows  $xx$  and  $yy$ ; states  $S_1$  and  $S_3$  are distinguished by  $x$ , which follows  $xx$  and  $x$ ; and states  $S_1$  and  $S_0$  are distinguished by  $y$ , which follows  $xx$  and empty sequence. Thus, all requirements related to transition  $(S_3, x)$  are satisfied. One can check the other requirements that are also satisfied.

2.4. *Test Sequences Generation Methods*. There are many methods reviewed in Section 5 of the literature [15]. This part describes some test sequences generation methods.

(1) *W-Method*. It uses a characterization set  $W$  to distinguish each state of the specification FSM [5]. For the FSM  $M$  of Figure 1,  $W = \{x, y, yy\}$  and  $M$  has the state cover set  $Q = \{q_0, q_1, q_2, q_3\} = \{\varepsilon, y, x, yy\}$  ( $\varepsilon$  means empty sequence), after

removing those sequences that are proper prefixes of other sequences; the test suite is

$$TS = Q \{x, y\} W = \{xxx, xxyy, xyx, xyyy, yxx, yxyy, \\ yyxx, yyxyyy, yyyy\}. \quad (4)$$

(2) *W<sub>p</sub>-Method*. It serves as an improvement of  $W$ -method [7]. In the testing phase of transition, while it does not use the characterization set  $W$ , it uses subset instead. According to the ending state of one transition, choose the corresponding state identifiers set  $W_i$ , which uses fewer test sequences than  $W$ -method. For the FSM in Figure 1,  $W_0 = \{yy\}$ ,  $W_1 = \{y\}$ ,  $W_2 = \{x, yy\}$ , and  $W_3 = \{x\}$ . In the first phase, test sequences that check the state are  $Q \cdot W$ .

(3) *HIS-Method*. It utilizes a separating family from the identifiers of each state to harmonize these identifiers [8]. For the example of Figure 1, we can obtain the separating family set  $F = \{H_0, H_1, H_2, H_3\}$ , where  $H_0 = \{x, yy\}$ ,  $H_1 = \{x, y\}$ ,  $H_2 = \{x, yy\}$ , and  $H_3 = \{x\}$ . So, in the first phase, we get  $q_j H_j$ ; in the second phase, we compute  $q_0 x H_3 + q_0 y H_1 + q_1 x H_1 + q_1 y H_2 + q_2 x H_3 + q_2 y H_2 + q_3 x H_2 + q_3 y H_0$ . The test suite is

$$\{xxx, xxy, xyx, xyyy, yxx, yxy, yyxx, yyyx, yyyy\}. \quad (5)$$

Besides, there are methods such as UIO presented in literature [8], DS, and  $H$  proposed in the literature [16]. And also some researchers think about the derivation methods of nondeterministic FSM.

### 3. Research Method

3.1. *Conversion between TEFSM and FSM*. According to the definition of TEFSM presented in Section 2.2, TEFSM diagram could be described as in Table 1. A number of symbols are used in order to increase readability: “?” symbol for input; “!” for output; “%” for probability; “~” for condition and probability; and “Δ” for time-delay [12].

Figure 2 is an example of TEFSM model, which has three states named  $S_0$  (initial state),  $S_1$ , and  $S_2$ .

There are transactions between each two states with different transaction conditions.  $\Delta t_1$  and  $\Delta t_2$  indicate the minimum delay time from state  $S_0$  to state  $S_1$  with transition conditions  $a$  and  $b$ , respectively. Set  $a' = a \cup (t > \Delta t_1)$  and  $b' = b \cup (t > \Delta t_2)$ ; then, the TEFSM in Figure 2 can be converted as Figure 3, which is now a normal FSM model.

In this way, theory of normal FSM could be applied for TEFSM model reasonably.

3.2. *Testing Cost*. Testing cost is a main aspect to evaluate testing technique performance. Briand presents that the testing cost involves two dimensions at least [10]: human effort and machine CPU time. In addition, Namin and Andrews describe the relationship among three properties of test suites: size, structural coverage, and fault location effectiveness [17].

We consider a technique more effective if it could generate minimum test suite and be highly active; hence, we make

TABLE 1: Conventions for constructing a TEFSM diagram.

Primitives	
	Notations
External transition edge	<div style="display: flex; align-items: center;"> <div style="border-bottom: 1px solid black; width: 100px; margin-right: 5px;"></div> <div style="margin-right: 5px;">Condition</div> <div style="border-bottom: 1px solid black; width: 100px; margin-right: 5px;"></div> <div style="margin-right: 5px;">%(probability) ?(input)</div> <div style="border-bottom: 1px solid black; width: 10px;"></div> </div> <div style="text-align: right; margin-right: 50px;"><math>T</math>-action</div>
Internal transition edge	<div style="display: flex; align-items: center;"> <div style="border-bottom: 1px dotted black; width: 100px; margin-right: 5px;"></div> <div style="margin-right: 5px;">Condition</div> <div style="border-bottom: 1px dotted black; width: 100px; margin-right: 5px;"></div> <div style="margin-right: 5px;">%(probability)</div> </div> <div style="text-align: right; margin-right: 50px;"><math>T</math>-action</div>
State node	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; border-radius: 50%; padding: 5px; text-align: center;">Initial state</div> <div style="border: 1px solid black; border-radius: 50%; padding: 5px; text-align: center;">State <math>\Delta t</math> <math>E</math>-action</div> <div style="border: 1px solid black; border-radius: 50%; padding: 5px; text-align: center;">Final state</div> </div>

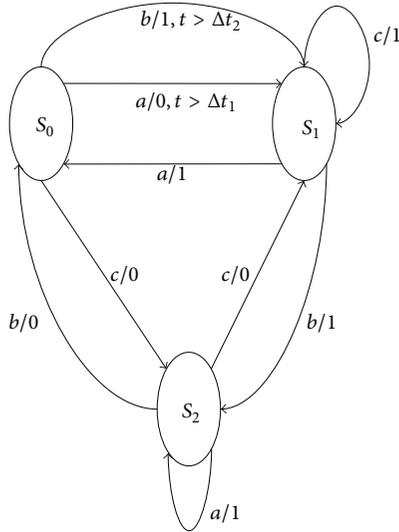


FIGURE 2: One example TEFSM.

use of the length of test suites as a measure of the cost. The length of test suites means the total length of test cases within the test suite. In the literature [4], it does not consider the accessibility degree and only changes one factor to observe the experiments each time. Here, we set five parameters of FSM model as the effect factors: the number of states, the number of inputs, the number of outputs, the completeness degree, and the accessibility.

**3.3. Evaluation Method.** This paper mainly focus on three test derivation methods ( $W$ ,  $W_p$ , HIS).  $W$ -method, as the most classic method, should be taken into consideration.  $W_p$ -method improves the  $W$ -method for reducing the number of test suites and thus should be considered. HIS-method is the method used in the literature [4, 14], so it is necessary to consider this method here for comparison.

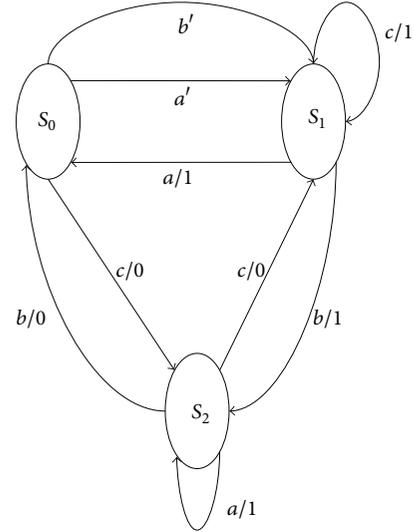


FIGURE 3: FSM converted from TEFSM of Figure 2.

Moreover, we use the coverage criteria as those considered in the literature [13]: state coverage (SC), transition coverage (TC), initialization fault (IF), and transition fault (TF) coverage criteria. When an FSM is the specification for testing, tests covering an FSM specification target one or several elements such as inputs, outputs, states, and fragments of its transition graph. Paths are typical fragments of the transition graph considered for coverage; therefore, path coverage has to be taken into account. We choose the traditional TC criterion. The latter two criteria (IF, TF) are proposed to direct at the fault models. IF criterion finds that the only possible faults in FSM implementation are related to a wrong initial state of a specification FSM, while TF criterion states that implementation faults occur in transitions. The four criteria could be representative of criteria for the FSM model.

For evaluation method, there are three ways.

- (1) Similar to the method proposed in the literature [1], generate a test set which is adequate for all coverage criteria as a test pool first; then, minimize it to get the minimum test set which is adequate for one specific coverage criteria.
- (2) Generate the test set which satisfies each kind of coverage criteria by generation method directly.
- (3) According to the relation among the coverage criteria shown in Table 3 of the literature [4], the test suite covering TC/IF/TF criteria almost covers the test requirements of SC (the coverage rate and the standard deviation in the first row are 1.000/0.000, 0.970/0.064 and 0.994/0.034). Since there may be some states that are not distinguishable from any other states, the latter two (IF/TF) do not cover all the states. However, we can generate the approximate minimum test set, which is adequate for the SC criterion, and then add some test sequences to the minimum test suite to satisfy other criteria. Contrast to (1), it resembles

**Inputs:**  
 FSM  
 $T$ : A set of initial test set  
 $K$ : A coverage criterion

**Output:**  
 $T'$ : The minimum test suite satisfy the criterion  $K$

**Process:**

- (1) Set  $T' = \emptyset$ ;
- (2) Initial  $TR_K(M)$ ; //test requirement set satisfy //criterion  $K$ .
- (3) Initial test coverage info TS by calculating covering weight  $(w(t))/|TR_K(M)|$  of  $T$ .
- (4) While  $TS_K(M, T') \neq TS_K(M, T)$  do
- (5)   If  $K$  in(SC||TC||IF) and NOT the 1st time
- (6)     Re-adjust covering weights of the remaining sequences;
- (7)     Select the minimum weight from the weights of all sequences in  $T$ ;
- (8)     For each  $t$  in  $T$
- (9)       If  $(TS_K(M, T' + \{t\}))/TS_K(M, T') \neq \emptyset$   
 $T' = T' + \{t\}$
- (10) End of While
- (11) return  $T'$

ALGORITHM 1: MinTSByWeight.

an incremental method. Because we can obtain the test suite adequate for some coverage criteria when utilizing the generation methods, so it is better to use this incremental method if the coverage criterion is very strict and one generation method is not able to generate the test set adequate for this coverage criterion. We use the first evaluation method in this paper, which we are coming to illustrate in Section 4 in detail.

After generating the final test set, we analyze the experimental data using the orthogonal testing method. The orthogonal test is useful when many factors influence one thing. Because at that situation the inputs would be many values, it would be effective through choosing a suitable orthogonal table that has advantage of equilibrium and dispersion. In this study, we choose the orthogonal table containing four levels and five factors which actually can be written as  $L_{16}(4^5)$ , which is shown in Section 4.2 in detail. The whole evaluation process is a general process for assessment as well. We can choose a way to solve the problem according to the situation. Figure 4 describes the process in detail.

How can we select a minimum test suite  $T'$  from  $T$  to satisfy the test coverage criteria? The literature [13] applies greedy algorithm to deal with this problem. Test suite selection for SC/TC criteria is still a weighted set-cover problem, while the selection for IF/TF criteria is defined as set-cover with pairs. However, for the definition of IF criterion, we can consider the selection of the minimum test suite for IF criterion as the weight set-cover problem as well.

We also define the cost  $w(\alpha)$  of a sequence  $\alpha$  in test suite  $T$  as  $|\alpha| + 1$ , that is, the length of the sequence plus the reset symbol used to bring the FSM back to the initial state.

In the literature [13], every time pick up the min ratio between the cost and coverage increments induced by  $t$  in  $T'$ , that is,  $(\omega(T' \cup \{t\}) - \omega(T'))/|TS_{SC}(M, T' \cup \{t\})/TS_{SC}(M, T')|$ . However, it may generate redundant test sequences. For example, in the case of SC criterion, the FSM model  $M$  shown in Figure 3 has three states  $S_0, S_1$  and  $S_2$ . The initial test set is:  $ba, bb$ . The first sequence “ $ba$ ” could cover state  $S_0$  and state  $S_1$  with min ratio =  $3/2 = 1.5$ . So, we can add it to  $T'$ , and then check the second sequence “ $bb$ ”, because the second input can make the system reach state  $S_2$ , which is not covered before, the ratio =  $3/1 = 3 > \text{min ratio}$ , so it should be ignored in this round. For the second loop, add “ $bb$ ” to it; as a result, there are “ $ba$ ” and “ $bb$ .” In fact, “ $bb$ ” is enough to satisfy the SC criterion.

To deal with this problem, we present a novel algorithm MinTSByWeight (minimum test suite by weight), which determines test sequence by weight. The algorithm is mainly based on the theory that the minor test suite it has, the less it costs. We do not need to consider time-delay in this algorithm since we already presented a method to covert TEFSM as FSM by taking time-delay as an element of input.

As shown in Algorithm 1, there are 3 inputs: FSM model, initial test set  $T$ , and test coverage criteria  $K$ ; and one output  $T'$ , which is the minimum test sequence of test object  $M$  and satisfies the coverage criteria  $K$ . At step (3), the weight of each test sequence or each pair test sequence is calculated based on the formula  $w(t)/|TR_K(M)|$ . In steps (4) to (7), select the minimum weight from the weights of all sequences or pairs and then add such sequence or pair “ $t$ ” or “ $t_1, t_2$ ” into the result set  $T'$  with steps (8)~(10). From the second round, the weights of the remaining sequences or pairs need to readjust in the initial test set by removing the cover points in  $t$  or  $t_1/t_2$  and then go to the select min weight as before. Cycle

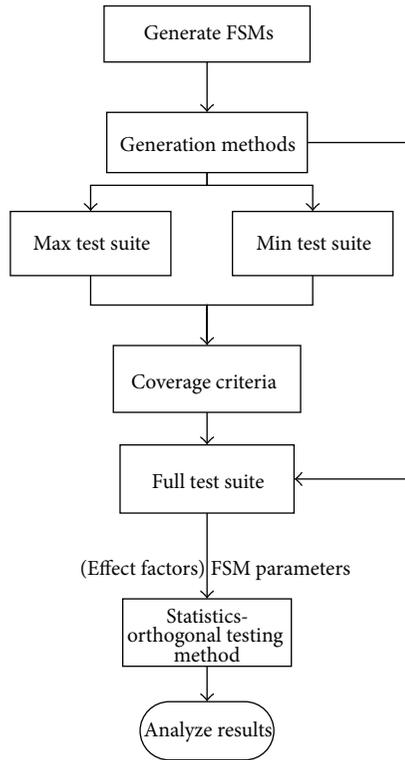


FIGURE 4: Assessment process.

this process until the selected sequences are adequate for the criterion or having checked all the sequences in the initial set  $T$ .

## 4. Experiments and Results Analysis

### 4.1. Coverage Criteria for Different Method

*Premise.* FSM is deterministic and completed finite state machine. When the generated FSM is nonreduced, the program does not proceed. In other words, although the automatic process of generating FSMs may produce nonreduced FSMs, this kind of model is not considered in the experiment. Therefore, we generate deterministic and completed models directly.

*Coverage Criteria and Generation Method.* This method includes average length of the test suite satisfying four coverage criteria (SC/TC/IF/TF) generated by  $W_p$ -method/ $W_p$ -method/HIS-method

*Experiment Application Scenario.* Considering that in practical applications, it is impossible to model the whole system, but some important parts, and the experiment contains eight experimental groups and sets the number of states: 3, 6, 8, 10, 12, 14, 16, and 20. Besides, we set the number of inputs to be 2, 5, and 7, respectively, and the number of outputs to be also 2, 5, and 7, respectively. Each group automatically generated 50 FSMs.

*Experiment Result.* The results of the three types under the three generation methods are shown in Tables 2, 3, and 4.

Figures 5(a)~5(e) show impact of kinds of methods to 4 criteria.

*Result Analysis from Test Suite Length.* Figures 5(a) and 5(c) show that the test suite length is, in low degree, affected by the generation methods. Almost all groups can obtain the sequence-suites with nearly the same length in average. Figure 5(b) shows that the gap between  $W_p$ -method and HIS-method under TC-adequacy is small, and  $W$ -method can generate a test set with shorter length. Figure 5(d) shows that the impact trend of these three methods on the TF-criterion is similar, but the average length of test suite with TF-adequacy generated by  $W$ -method and HIS-method is relatively longer. Figure 5(e) shows the difference of the initial test suite average length among the three generation methods. The  $W$ -method generates the longest test suite length, and the length gap of suite derived by HIS-method and  $W_p$ -method is insignificant, but  $W_p$ -method could generate less test suite because it does not need to harmonize the state identifiers. Moreover,  $W$ -method generates shorter TC-adequacy test suite and TF-adequacy test suite as well as shorter initial test suite.

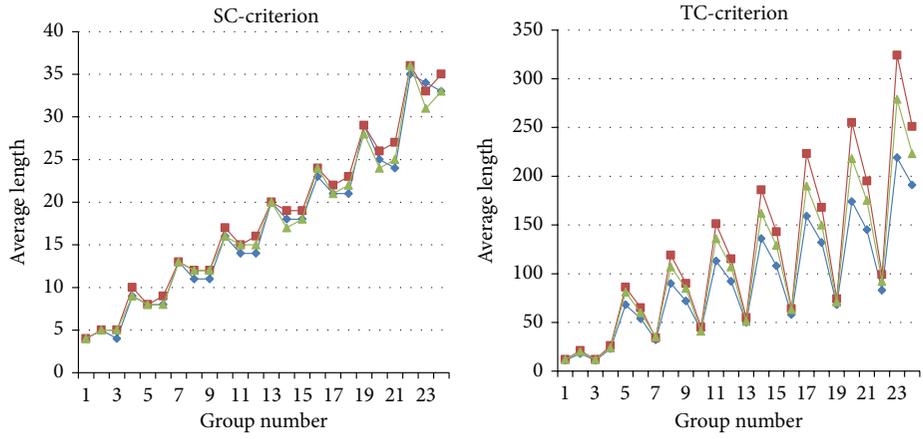
*Result Analysis from Effort.* When choosing the minimum set from a large pool suiting to each coverage criteria, the reduction rate is different. If the reduction rate is low, we could directly use the initial test suite instead of choosing the minimization set to improve the test efficiency, because the SC/IF criteria can be covered by fewer test sequences, while the TC/TF criteria need more test sequences to be covered. We only consider the latter two criteria under different derivation methods. The impact of each derivation method on the TC-criterion is shown in Figure 6(a), and the impact on TF-criterion is shown in Figure 6(b).

From Figure 6, we know that choosing the minimization test suite needs the largest effort. For the TC-criterion, Figures 5(b) and 5(e) show the reduction intensity to remove the useless sequences and Figure 6(a) reflects the high reduction rate, while there is little difference of the reduction rate between the other two generation methods ( $W_p$ -method and HIS-method) when choosing the minimum set to satisfy the TF-criterion. However, as for TC-criterion, when the number of inputs is small, the gap is small. But, when it is large, there is a clear gap. For example, if the groups of 7, 8 and 9 (the number of states, the number of inputs, and the number of 22 outputs) are 8/2/2, 8/4/7 and 8/5/5, respectively, in group NO. 7, the results selection rate is 0.62637 for  $W_p$ -method and 0.7059 for HIS-method, while in the NO. 8 the results selection rate is 0.31609 for  $W_p$ -method and 0.6409 for HIS-method. So, if the input number is large,  $W_p$ -method can be used to improve the quality of the test sequences.

### 4.2. The Impact of FSM Parameters on Coverage Criteria.

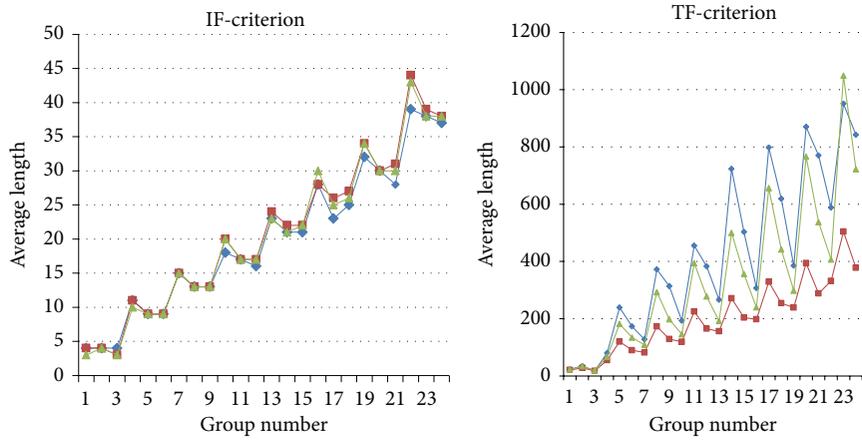
Consider five factors: the number of states, the number of inputs, the number of outputs, the completeness degree, and the accessibility degree. From the results above, for CFSMs, the minimization cost of  $W_p$ -method is the minimum. Hence, here, we choose  $W_p$ -method as generation method.

Using the orthogonal experimental method, we choose some representative samples from all level combinations for



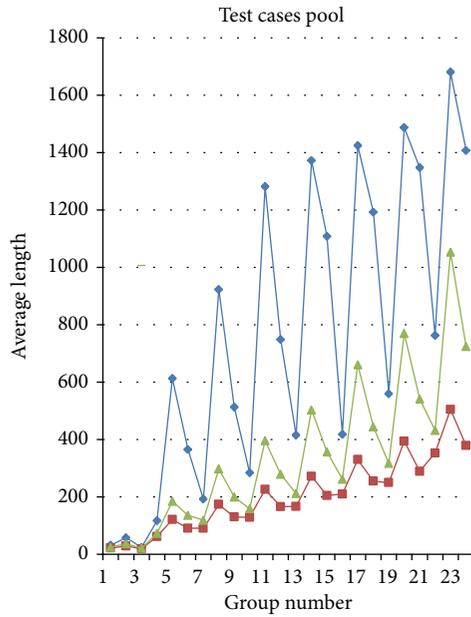
(a) Impact of test suites satisfy SC-Criterion

(b) Impact of test suites satisfy TC-Criterion



(c) Impact of test suites satisfy IF-Criterion

(d) Impact of test suites satisfy TF-Criterion



- W-method
- $W_p$ -method
- HIS-method

(e) Impact of original test suites

FIGURE 5: Impact of kinds of methods to criteria.

TABLE 2: The experimental data of different coverage criteria for  $W$ -method.

Group number	Number of states	Number of inputs	Number of outputs	Average length of initial test suites	$W$ -method/average length of test suites satisfying each coverage criteria			
					SC	TC	IF	TF
1	3	2	2	32	4	11	4	23
2	3	3	5	58	5	18	4	35
3	3	2	7	24	4	11	4	18
4	6	2	2	117	9	23	11	80
5	6	5	5	613	8	68	9	239
6	6	4	7	365	8	54	9	173
7	8	2	2	192	13	32	15	128
8	8	5	5	923	11	90	13	372
9	8	4	7	513	11	72	13	314
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
24	20	4	7	1407	33	191	37	842

TABLE 3: The experimental data of different coverage criteria for  $W_p$ -method.

Group number	Number of states	Number of inputs	Number of outputs	Average length of initial test suites	$W_p$ -method/average length of test suites satisfying each coverage criteria			
					SC	TC	IF	TF
1	3	2	2	22	4	12	4	21
2	3	3	5	29	5	21	4	28
3	3	2	7	19	5	12	3	18
4	6	2	2	62	10	26	11	55
5	6	5	5	121	8	86	9	120
6	6	4	7	91	9	65	9	90
7	8	2	2	91	13	34	15	82
8	8	5	5	174	12	119	13	173
9	8	4	7	130	12	90	13	129
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
24	20	4	7	379	35	251	38	378

TABLE 4: The experimental data of different coverage criteria for HIS-method.

Group number	Number of states	Number of inputs	Number of outputs	Average length of initial test suites	HIS-method/average length of test suites satisfying each coverage criteria			
					SC	TC	IF	TF
1	3	2	2	25	4	12	3	22
2	3	3	5	37	5	20	4	34
3	3	2	7	21	5	12	3	19
4	6	2	2	73	9	24	10	67
5	6	5	5	185	8	81	9	182
6	6	4	7	136	8	60	9	134
7	8	2	2	119	13	35	15	109
8	8	5	5	298	12	107	13	293
9	8	4	7	200	12	85	13	198
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
24	20	4	7	724	33	223	38	722

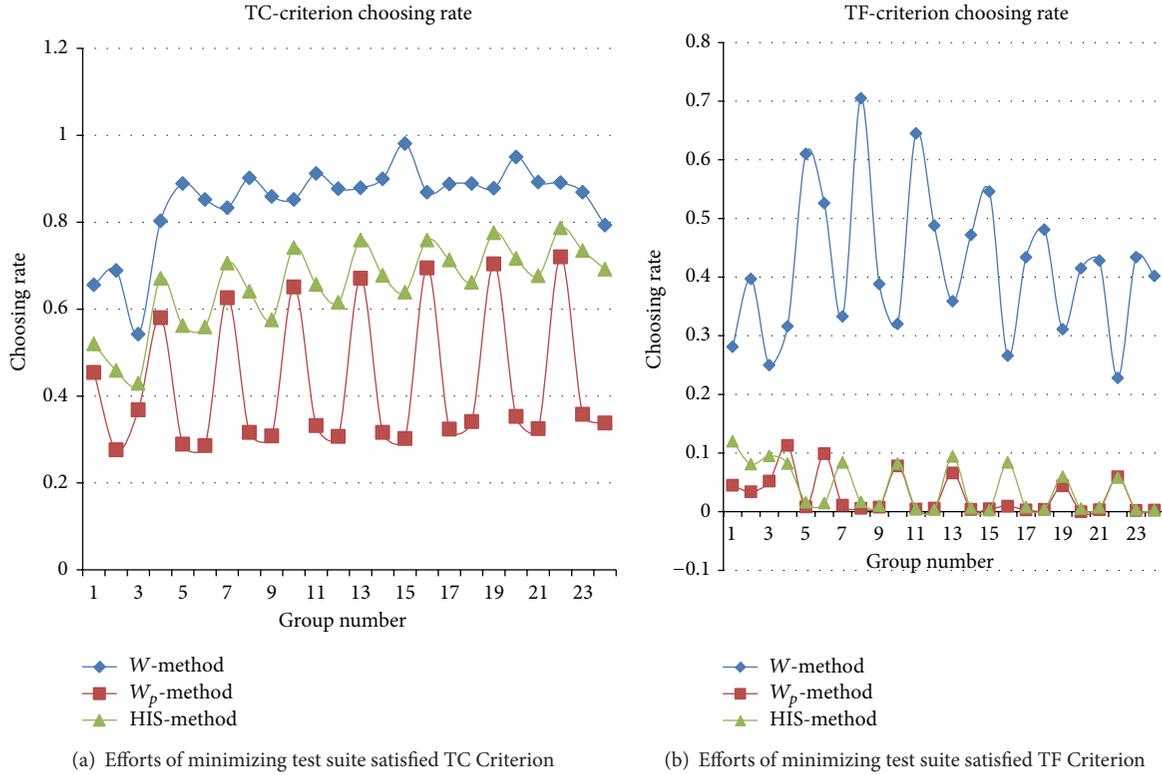


FIGURE 6: Screening ratio.

experiments and then find out the optimal level combination by analyzing the results of these samples to comprehend experiments completely.

In this experiment, there are 4 levels and 5 factors. Each group generates 500 FSMs automatically. We control the level of the number of inputs and accessibility degree related to the number of states to make the FSMs more similar to the reality. As one level among these factors is on the same situation, though the level values among each factor may be different, the control would not lead to the inequality among the factors.

Table 5 shows the levels for each factor and Table 6 shows the design scheme (number in brackets indicates level number).

*Experimental Results Analysis.* Applying the range-analysis method, we can obtain some values such as  $R1 = 1181.750$ ,  $R2 = 713.500$ ,  $R3 = 676$ ,  $R4 = 374.000$ , and  $R5 = 402.5$ ; from these range values, we can easily know  $R1 > R2 > R3 > R5 > R4$ . Also, the factor of the number of states has the greatest impact on the test sequences length, followed by the number of inputs and the number of outputs (outputs will affect the verification of transitions). By the orthogonal design tool, we can get the effect curve shown in Figure 7(a). The effect curve describes the impact for SC-criterion as shown in Figure 7(b).

Figure 7 shows that the number of states for the average length has a great impact, while the other three parameters have little effect on the average length of the test suites which satisfy the SC-criterion. The number of inputs has a positive

TABLE 5: Levels for each factor.

Factor	Levels	Level range
Number of states ( $n$ )	5, 8, 14, 20	3~20
Number of inputs ( $k$ )	$0.2n, 0.4n, 0.6n, 0.8n$	$0.2n \sim 0.8n$
Number of outputs ( $m$ )	2, 4, 7, 10	2~10
Completeness degree ( $d$ )	1.0	—
Accessibility degree ( $r$ )	$0.6 * (n - 1), 0.8 * (n - 1), 1.0 * (n - 1)$	$\sim(n - 1)$

influence on the average length as Figure 7(a) shows. And the number of outputs increased results in the growth of the length of the test sequences, because the more the inputs, the shorter the length of characterization sequences and the fewer the number of such sequences. As each derivation method is closely related to the sequences reaching each state, the impact of accessibility degree emerges as a wave form but is not positively correlated. Figure 7(a) shows that when the accessibility degree ( $r$ ) is 0.8 times as long as the maximum length (the number of states - 1), the test suite length nearly reaches the maximum. When it equals the 60 percent of the maximum length, the test suite length nearly reaches the minimum. After all,  $r$  represents the length of each sequence reaching one state.

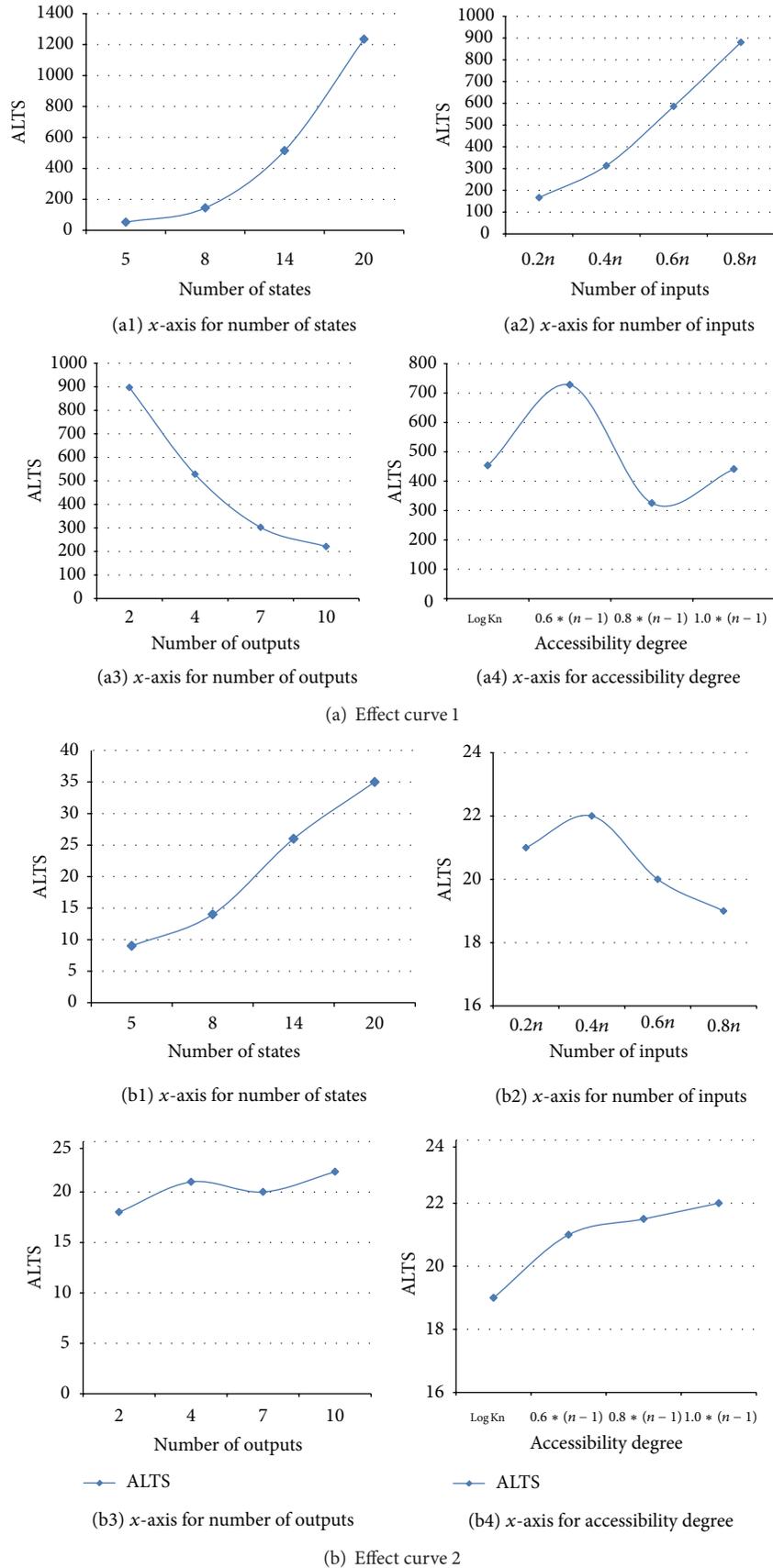


FIGURE 7: Effect curve.

TABLE 6: (a) Experimental design and data recording. (b) Parameter in different level.

(a)

Group number	Factors				Average length of test suites		
	Number of states ( $n$ )	Number of inputs ( $k$ )	Number of outputs ( $m$ )	Completeness degree ( $d$ )	Accessibility degree ( $r$ )	$W_p$ test suites	SC
1	5 (1)	2 (1)	2 (1)	1.0	2 (1)	48	9
2	5 (1)	2 (2)	4 (2)	1.0	2 (2)	37	10
3	5 (1)	3 (3)	7 (3)	1.0	3 (3)	55	8
4	5 (1)	4 (4)	10 (4)	1.0	4 (4)	71	8
5	8 (2)	2 (1)	4 (2)	1.0	7 (4)	69	14
6	8 (2)	3 (2)	2 (1)	1.0	5 (3)	157	13
7	8 (2)	5 (3)	10 (4)	1.0	4 (2)	158	13
8	8 (2)	6 (4)	7 (3)	1.0	2 (1)	195	15
9	14 (3)	3 (1)	7 (3)	1.0	7 (2)	188	26
10	14 (3)	5 (2)	10 (4)	1.0	2 (1)	291	32
11	14 (3)	8 (3)	2 (1)	1.0	13 (4)	853	22
12	14 (3)	10 (4)	4 (2)	1.0	10 (3)	727	22
13	20 (4)	4 (1)	10 (4)	1.0	15 (3)	362	35
14	20 (4)	8 (2)	7 (3)	1.0	19 (4)	770	32
15	20 (4)	12 (3)	4 (2)	1.0	3 (1)	1278	38
16	20 (4)	16 (4)	2 (1)	1.0	11 (2)	2528	30

(b)

Level number	Number of states ( $n$ )	Number of inputs ( $k$ )	Number of outputs ( $m$ )	Completeness degree ( $d$ )	Accessibility degree ( $r$ )
1	52.750	166.750	896.500	425.750	453.00
2	144.750	313.750	527.750	361.750	727.750
3	514.750	586.00	302.000	735.750	325.250
4	1234.500	880.250	220.500	423.500	440.750
Worst	1181.750	713.500	676.000	374.000	402.500

TABLE 7: The results of variance analysis.

Source of variance	Sum of squared deviations	Free degree	F ratio
Number of states ( $n$ )	$S_n = 3460936.188$	3	9.898
Number of inputs ( $k$ )	$S_k = 1188087.188$	3	3.398
Number of outputs ( $m$ )	$S_m = 1098390.688$	3	3.141
Accessibility degree ( $r$ )	$S_r = 349673.188$	3	0.976
error	$S_e = 349673.188$	3	1.000
Fa		$F_{0.05}(3, 3) = 9.28$	

Because the range analysis cannot distinguish the data changing caused by test conditions during the testing from that caused by data errors, without respect to whether the factor effect is significant, we cannot figure out the accuracy of analysis. Next, we perform the variance analysis of the impact of factors and errors, and let the column of accessibility degree be the column here that the experimental errors exist. Result is shown in Table 7.

We can see that the number of states ( $n$ ) has a most significant impact on the length of the test suites.

To sum up the above results, we draw the following conclusions. (a) For SC-criterion, IF-criterion, the length of the test suites has little difference among the derivation methods.  $W_p$ -method can produce a relatively shorter length of the test suites for TC coverage criterion adequacy, but it needs to pick out a relatively small part of the whole initial test suite. To satisfy the TF-criterion,  $W_p$ -method has some advantages. Relatively speaking, it generates the smallest test set length, and when choosing the minimum test suite, the number of removed sequences is few, so when testing, if the

coverage criterion is TF-criterion, we can use  $W_p$ -method directly and do not need to select the minimal criterion adequacy test suite. Moreover, if we only need to satisfy TC-criteria, we can make a trade-off and consider the HIS-method. Because  $W_p$ -method is influenced by the number of inputs of FSM in the process of selecting the minimum test set for TC-criterion, and in case the number of test sequences derived by HIS-method is more than that derived by  $W_p$ -method, the rate of the sequences needing to be removed is relatively stable. (b) One of the FSM model parameters, the number of states ( $n$ ), has a most significant impact on the length of the test suites. With the number of states increasing, the length increases much faster. In addition, the increase of the number of inputs will cause an increase in length, but the change rate is lower than that in the situation of increasing the number of states. And as the number of outputs increases, the length decreases. Because the increase of the outputs makes the length of the sequences distinct, the states are short and the number of distinguishable sequences is limited; and when the accessibility degree is up, the length changes as a wave form. So, when designing the complete FSM models, in the conditions of the same number of states, inputs, and outputs, if we try to make the accessibility degree be 0.6 times as long as the maximum length (the number of states - 1), the length of the generated test set will be fluctuating around the minimum.

*4.3. Threats to Validity.* The FSM considered in the experiments is generated randomly, and we only consider the complete and deterministic FSMs together with time-delay factor, so we could not make sure these would be close to the FSM specifications in the actual project. Therefore, the conclusions drawn from the random FSMs may not be well applied in the actual situation. In the future work, it needs further validation in real situation and to extend our study on the nondeterministic FSM and the partial FSM.

Measuring the cost of testing in such an experimental context is also a challenge. Briand presents that the testing cost involves at least two dimensions: human effort and machine CPU time, details of which are in Figure 3 of the literature [10]. It is rather clear that test suite size, regardless of how it is measured, is a very rough cost measure. Moreover, the literature [11] use the cumulative length of a test set to measure the cost for it's easy to count. Though many follow the simple way to measure the cost, it would be better to consider other factors.

We can assess the ability of the test criteria not only by the test suites average length but also through the malfunction versions generated by the use of mutation operators from a specification FSM together with time-delay. In this case, we compare their variation scores that measure the number of variants killed and use it to assess the fault detection capabilities of various test suites generation methods. However, it is only used for finding the faults. From here, it can be reflected that the former (the test suites average length) mainly evaluate the cost and the later (variation scores) mainly evaluate the effectiveness. Whether there are some

valuable suggestions on the problem of finding the balance point of cost-effectiveness is not probed here.

## 5. Conclusion

The main contribution of this paper is that we propose a comparative strategy for assessing the coverage criteria and test cost for time-delay systems test. A TEFSM model is adopted for time-delay system's test suite generation. In the future, if we want to assess the coverage criteria or the test suites generation methods that are not applied in this paper, the coverage criteria or the test suites generation methods can be added directly to the comparative architecture. Besides, we take many factors of the FSM model (FSM state number, input number, output number, completeness degree, accessibility degree, and time-delay which are not considered in the literature [13]) into account. Through empirical study, we explore the test suites generation methods, coverage criteria, and FSM model parameters in terms of the influence of the average length of test suites. Based on these relationships, test engineers can make a good test strategy on the part of the conformance testing based on the TEFSM model. For example, when testing, if the coverage criterion is TF-criterion, we can use that directly for its relatively minimum cost. When designing the complete FSM models, in the conditions of the same number of states, inputs, and outputs, and if we try to make the accessibility degree be 0.6 times as long as the maximum length (the number of states - 1), the length of the generated test set will be fluctuating around the minimum. In industrial practice, one may design FSM models for critical components and automatically obtain test sequences by suitable methods. If, based on the requirements, we could judge in which situation the test cost is lower than others' cost, one could know which method to choose or how to change the models to be better. However, due to the random generation of the FSMs for experiments, we only consider the complete and deterministic FSM; it may vary with the actual situation. Next, we will extend our work on the nondeterministic FSM and the partial FSM. It requires further validation in more practical situation. The literature [13] explores the correlation between the structure testing and the model testing. We will benefit from the efforts, if we solve these problems, such as, how to extract the commons of the testing phases, understand the kinds of errors detected by each of the various stages, and use the various test adequacy criteria and test suites generation methods reasonably to test efficiently.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to thank Professor Adenilso da Silva Simao for his help to their questions. The reviews are highly thanked for their useful comments.

## References

- [1] F. Zhang and M. Yeddanapudi, "Modeling and simulation of time-varying delays," in *Proceedings of the Symposium on Theory of Modeling and Simulation-DEVIS Integrative M&S Symposium*, Society for Computer Simulation International, March 2012.
- [2] J.-P. Richard, "Time-delay systems: an overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003.
- [3] K. El-Fakih, N. Yevtushenko, and H. Fouchal, "Testing timed finite state machines with guaranteed fault coverage," in *Testing of Software and Communication Systems*, pp. 66–80, Springer, Berlin, Germany, 2009.
- [4] M. Abramovici and C. E. Stroud, "BIST-based delay-fault testing in FPGAs," *Journal of Electronic Testing*, vol. 19, no. 5, pp. 549–558, 2003.
- [5] T. S. Chow, "Testing software design modeled by finite-state machines," *IEEE Transactions on Software Engineering*, vol. 4, no. 3, pp. 178–187, 1978.
- [6] G. Luo, G. V. Bochmann, and A. Petrenko, "Test selection based on communicating nondeterministic finite-state machines using a generalized  $W_p$ -method," *IEEE Transactions on Software Engineering*, vol. 20, no. 2, pp. 149–162, 1994.
- [7] X. Lu, Z. Li, W. Qiu, D. M. H. Walker, and W. Shi, "Longest-path selection for delay test under process variation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 12, pp. 1924–1929, 2005.
- [8] K. Sabnani and A. Dahbura, "A protocol test generation procedure," *Computer Networks and ISDN Systems*, vol. 15, no. 4, pp. 285–297, 1988.
- [9] R. Dorofeeva, K. El-Fakih, and N. Yevtushenko, "An improved conformance testing method," in *Formal Techniques for Networked and Distributed Systems—FORTE 2005*, F. Wang, Ed., vol. 3731 of *Lecture Notes in Computer Science*, pp. 204–218, Springer, 2005.
- [10] L. C. Briand, "A critical analysis of empirical research in software testing," in *Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement (ESEM '07)*, pp. 1–8, Madrid, Spain, September 2007.
- [11] L. C. Briand, Y. Labiche, and Y. Wang, "Using simulation to empirically investigate test coverage criteria based on state-chart," in *Proceedings of the 26th International Conference on Software Engineering (ICSE '04)*, pp. 86–95, May 2004.
- [12] D. W. Lee, B. K. Choi, and J. H. Kong, "Timer embedded finite state machine modeling and its application," in *Proceedings of the 24th European Conference on Modelling and Simulation (ECMS '10)*, Kuala Lumpur, Malaysia, June 2010.
- [13] A. Simão, A. Petrenko, and J. C. Maldonado, "Comparing finite state machine test coverage criteria," *IET Software*, vol. 3, no. 2, pp. 91–105, 2009.
- [14] L. Wu, X. Su, and P. Shi, "Sliding mode control with bounded  $\mathcal{L}_2$  gain performance of Markovian jump singular time-delay systems," *Automatica*, vol. 48, no. 8, pp. 1929–1933, 2012.
- [15] K. Karhu, T. Repo, O. Taipale, and K. Smolander, "Empirical observations on software testing automation," in *Proceedings of the 2nd International Conference on Software Testing, Verification, and Validation (ICST '09)*, pp. 201–209, Denver, Colo, USA, April 2009.
- [16] R. Dorofeeva, K. El-Fakih, and N. Yevtushenko, "An improved conformance testing method," in *Proceedings of the IFIP 25th International Conference on Formal Methods for Networked and Distributed Systems*, pp. 204–218, Taipei, Taiwan, October 2005.
- [17] A. S. Namin and J. H. Andrews, "The influence of size and coverage on test suite effectiveness," in *Proceedings of the 18th International Symposium on Software Testing and Analysis*, pp. 57–67, New York, NY, USA, 2009.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

