Interoperability Test Generation and Minimization for Communication Protocols Based on the Multiple Stimuli Principle

Soonuk Seol, Myungchul Kim, Member, IEEE, Samuel T. Chanson, Member, IEEE, and Sungwon Kang

Abstract—This paper presents an automatic test generation and minimization method for testing interoperability of communication protocols such as the ATM/B-ISDN signaling protocol and TCP. The method is based on a technique of composing finite state machines (FSMs). Traditionally, the generation of inputs utilizing FSM was mostly based on sequential test scripting languages and often done manually. Each input is processed on a one-at-a-time basis, which did not allow generating complex scenarios of simultaneous inputs on several interfaces of the system under test (SUT). In practice there exists the possibility that additional inputs have to be sent to the SUT while the previous input is still being processed, and/or multiple inputs need to be sent to different interfaces of the SUT at the same time. Our method generates interoperability test cases dealing with these more complex situations. Moreover, it minimizes the number of increased test cases due to the multiple stimuli without jeopardising transition coverage. Experimental results have shown that compared to the conventional methods, our test generation method generates more interoperability test cases resulting in higher transition coverage (26% higher for TCP and 12% higher for ATM signaling protocol). Also, our test minimization method reduces the interoperability test cases by about 94% in the case of the ATM signaling protocol with the same transition coverage.

Index Terms—conformance testing, interoperability testing, protocols, test case generation, test minimization, validation.

I. INTRODUCTION

To increase the confidence that protocol products conform to international standards, various protocol testing methodologies (see for example [5], [20], [22], and the bibliographies therein) have been developed and deployed. In particular, conformance testing that checks whether an implementation is correct with respect to the relevant standards has been standardized by ISO [1] and ITU-T [12] and applied to protocols such as N-ISDN and ATM/B-ISDN. Nevertheless, it is well known that conformance testing has limitations in ensuring interoperability. Thus, even two conforming implementations may fail to interoperate [2], [3]. The reasons include ambiguity of protocol standards, incompatible option settings and incomplete conformance testing [4]. Moreover, in interoperability testing, the degree of interaction between implementations depends not only on the implementations themselves but also on the environment, and hence some degree of direct testing of interoperability is considered indispensable [8], [21].

In the field of interoperability testing, most work has concentrated on generating interoperability test cases. A common approach is to select test cases from the composed finite state machine, which is constructed from component machines of the system via reachability analysis. Rafiq and Castanet [2] employed this approach to generate interoperability test cases. However, a rigorous definition of interoperability was not used and the work considered only the case where lower testers exist between two Implementations Under Test (IUTs). Arakawa et al. [3] derived a conformance test suite and an interoperability test suite separately and then manually combined them to reduce the number of conformance test cases. Kang and Kim [6], [8] developed a coherent framework of interoperability testing and a systematic interoperability test suite derivation method based on the framework. Griffith et al. [21] presented a general method for automatic generation of test cases with various redundancy criteria and applied it to VoIP systems. For testing the data part of protocols, Seol et al. [18] developed an automatic test generation method which assigns values to parameterized test cases based on experimental design techniques.

In fact, the approach used in interoperability testing can also be found in some other work in which the notion of interoperability testing is not explicitly mentioned. These include validation of interacting processes modeled by labelled transition systems [15], conformance testing based on communicating finite state machines [10], [19], and embedded testing [11].
In the previous studies, however, the single stimulus principle has always been explicitly or implicitly assumed either in the method, model, or architecture. According to the single stimulus principle, only a single stimulus is examined at each stable state when generating a reachability graph. A stable state for concurrent systems or communicating protocols is defined in [9] as the system state (i) that is reachable from the initial state adhering to the single stimulus principle, and (ii) from which no change can occur without another stimulus. Systems adhering to the single stimulus principle are said to be running in a slow environment [10]. In a slow environment, inputs can be sent from the environment to the system only when all the queues and all the channels are empty. Therefore, simultaneous stimuli and stimuli during transitions cannot be considered.

When protocol implementations interact with one another, it is possible that additional messages are sent to the implementation(s) while the previous message is still being processed, and/or simultaneous messages are sent to each implementation at the same time. This is known as multiple stimuli. For instance, TCP allows simultaneous open and simultaneous close of a TCP connection. Therefore, checking multiple stimuli in interoperability testing is important because multiple stimuli exist in many protocol behaviors and their implementation seems to be especially error-prone.

Reducing the size of the interoperability test cases is another important concern. In order to reduce the size of relevant state space, the notion of stable state has been introduced in the literature [9], [10]. Unfortunately, as mentioned above, these studies have assumed the single stimulus principle. An alternative is to use a guided random walk approach [19] to trim the state space at the cost of exploring less behavior space. Various redundancy criteria are discussed in the literature [21]. The existing work, however, is not capable of handling complex scenarios of simultaneous inputs on several interfaces of the system under test.

In this paper, we present a new method that supports concurrent protocol behavior in generating test cases for interoperability testing. Our scheme also minimizes the size of the test cases in order to reduce testing time and cost. The rest of this paper is organized as follows: Section II presents some related work regarding the limitation of the single stimulus principle. In Section III, we define formal models for communication protocols and interoperability test cases. Section IV describes our approach to interoperability testing and the procedure for generating interoperability test suites with applications to the connection establishment phase of TCP and the ATM/B-ISDN signaling protocol. Section V presents test case minimization algorithms and their applications to the interoperability test cases derived from the ATM signaling protocol. Finally, Section VI concludes the paper.

II. RELATED WORK AND MOTIVATION

In this section, we present some existing test generation methods that are closely related to our work and discuss their limitations due to the single stimulus assumption.

Luo et al. [10] proposed a method of generating test sequences for concurrent programs and communication protocols modeled as communicating nondeterministic finite state machines (CNFSMs). The method first reduces a system of CNFSMs into a single NFSM by reachability analysis. The NFSM is then transformed into a trace-equivalent observable NFSM from which test sequences are generated. Basically, a global state denotes contents of channels, input queues, and states in each machine. It is obvious that the number of states in a global state machine becomes infinite in case of unlimited channels and/or queues. For that reason, they made the assumptions that queues and channels are bounded, and that the system runs in a slow environment. The first assumption is reasonable since unbounded queues or unbounded channels cannot be implemented in practice. The second assumption, however, is not reasonable for those protocols that allow multiple and simultaneous messages.

Lee et al. [19] used a guided random walk procedure for conformance testing of protocols specified as communicating finite state machines. This procedure attempts to cover all edges of the component FSMs instead of all edges in the global FSM. It uses synchronous interprocess input/output operations to specify interactions between machines. For a message transmission, the input and output operations have to be matched and executed simultaneously. In other words, if a process attempts to send an output, it has to wait until the receiving process is ready to execute the matching input operation and vice versa. This means that the system adheres to the single stimulus principle and hence multiple stimuli during transitions between stable states and simultaneous inputs cannot be examined.

Lima and Cavalli [11] proposed a pragmatic approach to generating test sequences for embedded components of complex systems. Embedded testing and interoperability testing are similar in that internal actions are usually invisible and thus hard to control. Their method provides means to support multiple outputs (i.e., a signal from the environment brings about several simultaneous outputs) and nondeterminism of components. A drawback of their method is that it limits the number of external inputs between stable states to one. In other words, an input can be given to the system only if the previous input has been completely processed and the system is not undergoing any state change at the time.

In practice, multiple messages between communicating entities often occur. Fig. 1 shows three cases of multiple messages in MSC-like charts. In the example, two machines exchange messages with each other and with the environment. In this paper, we shall call the former type of messages internal messages and the latter external messages. Case (a) involves multiple stimuli from one entity to the peer entity for a given external input while cases (b) and (c) involve multiple stimuli from the environment to the system of machines. In (a), machine $M_a$ produces two internal messages $x$ and $y$ arising from the external input $a$. Note that multiple output messages

1 A preliminary version of this work appeared in TestCom 2002 [16].
can be produced by an external input as well as by an internal input. In (b), external input $b$ is issued while the previous input $a$ is still being processed. This is an example of multiple stimuli during a transition from one stable state to another. Note that the global states ($A1, B1$) and ($A4, B3$) in Fig. 1(b) are stable states. A similar case is shown in Fig. 1(c) where one input goes to $M_A$ and another input goes to $M_B$ at the same time. This is known as simultaneous stimuli which is a special case of multiple stimuli. Simultaneous stimuli can be seen in TCP's simultaneous open and close behaviors. These cases can often happen in the real world and may also occur in faulty implementations. Therefore, interoperability testing should test for these behaviors as well.

III. FORMAL MODELS

This section defines formal models for communicating entities and for interoperability test cases. The specifications and implementations of communicating entities can be modelled using a special type of FSMs known as input output state machine (IOSM). IOSM is defined as follows:

**Definition 1.** An IOSM is a 5-tuple ($S$, $S_0$, $L_{in}$, $L_{out}$, $Tr$) where:

1. $S = \{S_0, S_1, ..., S_n\}$ is a set of states;
2. $S_0 \in S$ is the initial state;
3. $L_{in} = \{v_1, v_2, ..., v_m\}$ is a set of input symbols;
4. $L_{out} = \{u_1, u_2, ..., u_k\}$ is a set of output symbols;
5. $Tr \subseteq \{S_i \longrightarrow v \in L_{in}, S_j \in S \wedge v \in L_{out} \}$ is a set of transitions where $L_{out}$ denotes the set of sequences of symbols in $L_{out}$ and $P(X)$ denotes the power set of the set $X$.

Bold letters in Definition 1 represent sets. $L_{in}$ is divided into the set of internal input symbols $L_{in,E}$ and the set of internal input symbols $L_{in,I}$. Similarly, $L_{out}$ is divided into two sets $L_{out,E}$ and $L_{out,I}$ that are, respectively, the set of external output symbols and the set of internal output symbols. For example, symbol $a$ in Fig. 1(a) is an external input symbol and $a'$ is an external output symbol. Symbols $x$ and $y$ are internal output symbols for machine $M_A$, and, at the same time, are internal input symbols for machine $M_B$. "$S_i \longrightarrow v \in L_{in}, S_j \in S"$ is a transition where $S_i$, $v$, and $U$ are, respectively, the starting state, the destination state, an input symbol and a sequence of output symbols. For a transition $tr$ in $Tr$, the $\cdot$ notation is used to refer to one of the elements of the following sets such as $S_i$, $S_j$, $v$, and $U$. For example, $tr.S_i$ represents the starting state of the transition $tr$. In case of a deterministic IOSM, there is at most one transition in $Tr$ for each input symbol while there is at least one transition in a completely defined IOSM.

**Definition 2.** For every transition $tr (\forall tr \in Tr)$ whose starting state is $S_i (tr.S_i = S_i)$, the state $S_i \in S$ is said to be:

1. **controllable** iff $tr.v \in L_{in,E}$;
2. **uncontrollable** iff $tr.v \in L_{in,I}$;
3. **semi-controllable** iff $\exists tr.v \in L_{in,E}$ and $\exists tr.v \in L_{in,I}$ where $j \neq k$.

We classify states in an IOSM into three types according to Definition 2. If a state can be changed only by external inputs, the state is said to be **controllable**. Such a state is controlled only by the testers’ inputs. If all possible inputs at a state are internal ones, the state is said to be **uncontrollable**. If some possible inputs at a state are internal and others are external, the state is **semi-controllable**. This classification is illustrated in Fig. 2. In IOSM $M_A$, a part of which is depicted in the box in Fig. 2, state $A1$ is controllable, $A3$ is uncontrollable, and $A2$ is semi-controllable. There are two transitions from state $A2$ as depicted in the box in Fig. 2. One transition goes to state $A3$ with external input $y$ and the other goes to state $A4$ with internal input $c$. Let us trace the behavior of multiple stimuli in this system. Suppose an external input $x$ is given when the system’s global state is ($A1, B1$) which is a stable state. Then, $M_A$ goes to state $A2$, sending an external output $x'$ and an internal output $a$, and $M_B$ replies with an internal message $c$. Before the message $c$ reaches $M_A$, another external message $y$ (multiple stimuli from the environment) is issued. This is because the state $A2$ of $M_A$ is a semi-controllable state. If $M_A$ receives the internal message $c$ in state $A2$, it will go to state $A4$, but it will go to state $A3$ if it receives the external input $y$. Although the system is in a transient global state ($A2, B2$), the next stable state of the system is unpredictable since the transient global state includes a semi-controllable state. In general, if a machine in a system is in a semi-controllable state and an external signal comes from a peer communication machine, the existing test generation methods will not be able to examine the external inputs since they adhere to the single stimulus principle.

Now we define a behavior model for two IOSMs communicating with each other and the environment. Let $[I]$ be
the composition of the two IOSMs, one, $M_A$, for the specification A and the other, $M_B$, for the specification B. Then $\Pi$ is defined as follows:

1. $S_{\Pi0} \in S_\Pi$ is the initial state. Namely, $S_{\Pi0} = (S_{A0}, S_{B0})$;
2. $L_{\Pi,\text{in}} = \{v_1, v_2, ..., v_{\text{in}}\}$ is a set of input symbols. These messages are external inputs of each machine. Namely, $L_{\Pi,\text{in}} = (L^A_{\text{in,E}} \cup L^B_{\text{in,E}})$;
3. $L_{\Pi,\text{out}} = \{u_1, u_2, ..., u_{\text{out}}\}$ is a set of output symbols. These messages are external outputs of each machine. Namely, $L_{\Pi,\text{out}} = (L^A_{\text{out,E}} \cup L^B_{\text{out,E}})$;
4. $T_{\Pi} \subseteq \{v_1, v_{\text{in}+1}, ..., v_{\text{in}+l-1}\}$ is a set of sequences of transitions, and a sequence of transitions $\{v_1\} \Leftrightarrow v_0, v_{i+1}, ..., v_{l-1}$, where $v_i \in (T_{\text{in}} \cup T_{\text{out}})$;
5. $Q_\Pi = (Q_A, Q_B)$ is a pair of directed channels where $Q_A$ is the channel from $M_B$ to $M_A$, and $Q_B$ from $M_A$ to $M_B$; both channels are represented by queues.

A global state and directed channels (modelled by queues) in a composite IOSM $\Pi$ determine the system state of the IOSM $\Pi$. Here, the system state is used to describe the state of the system as a whole while a global state defines each local state of the IOSMs in the system. Note that a global state is defined as one describing the entire system, including information for input queues and/or channels in [8] and [10].

An IOSM receives messages from either the environment or its channel queue. Then, the two IOSMs and the two directed channels constitute a composite IOSM $\Pi$ as shown in Fig. 3. Note that in this model IOSM $\Pi$ does not have any queue for inputs from the environment. This is the same as a single IOSM because its formal model does not define any queue. Since each state in an IOSM consumes one message at a time, messages following the current message do not affect the next state. We assume that internal outputs are instantly placed in the channel queue directed to the receiving IOSM. This implies first-come first-served (FCFS) message queuing and handling.

A global state is classified into one of two types depending on the channel contents.

**Definition 4.** The global states of two IOSMs are classified into two types.

\[ \forall S_\Pi = (S_A, S_B) \in S_\Pi \]

1. A global state $S_\Pi$ is said to be *stable* iff $Q_A$ and $Q_B$ are both empty.
2. $S_\Pi$ is in a *transient* state otherwise.

A system in a stable state will change state due to external inputs only. Both local states in a stable state cannot be uncontrollable states, otherwise the system is in a deadlock state because both machines will wait for messages from each other indefinitely. When a system is in a transient state, the system will change its state soon. However, the existing test generation methods do not allow for the system to receive additional external inputs. Our method, on the other hand, examines all possible external inputs. When one of the two local states in a transient global state is semi-controllable, we examine the cases where the IOSM receives an external input in the semi-controllable state as well as the cases where it consumes an internal message from its receiving channel.

Finally, we define an interoperability test suite in Definition 5. Each machine in a system consumes one input (either internal or external) at each state, that is, one transition at a time for each IOSM. Therefore, an interoperability test case can be represented by a sequence of transitions in either IOSMs in the system. An interoperability test suite is defined in terms of the interoperability test cases. Note that a test case whose length is 1 is simply a conformance test case.

**Definition 5.** An interoperability test suite $\text{IOPTS} \subseteq T_{\Pi}$ is defined as follows:

$\text{IOPTS} = \{i_{\text{opt}0}, i_{\text{opt}1}, ..., i_{\text{opt}n}\}$ is a set of interoperability test cases where an interoperability test case is defined as a sequence of transitions, $i_{\text{opt}} \Leftrightarrow t_{r0}, t_{r1}, ..., t_{r_{l-1}}$, where $t_{ri} \in (T_{\text{in}} \cup T_{\text{out}})$, $0 \leq i \leq l-1$, and $l \geq 2$.

The interoperability test case illustrated in Fig. 2 can be represented as follows. The length of this interoperability test case is 5. The starting stable state is $(A1, B1)$ and the destination stable state is $(A4, B3)$.


**IV. TEST CASE GENERATION**

This section describes our approach to generating interoperability test cases based on the formal models defined in Section III. First, we explain the overall procedure of our approach and the associated assumptions. Next, we select the test architecture that achieves the desired test coverage. Then, we discuss how to generate interoperability test cases based on the multiple stimulus principle. For the purpose of analysis, we divide the multiple stimuli into two sets as classified in Fig. 1 and consider them in Sections VI-C and VI-D, respectively. Lastly, our test generation algorithm is applied to two protocols, the ATM signaling protocol and TCP, with assessment of the results.

**A. An overview**

In order to derive test cases, we first need to derive IOSMs from the given protocol specifications. We assume that there is no autonomous transition, e.g., a timer. This is because even
though we are able to generate test cases involving time constraints, it is difficult to apply them to real test environments. The interoperability test derivation algorithm generates test cases from the IOSMs by constructing an interaction graph. The algorithm starts from the initial stable state consisting of each IOSM’s initial states and at each step examines all possible external inputs even in the transient states, until a stable state is reached. In the meantime, new paths are generated based on the multiple stimuli principle, and newly found stable states are added to the associated state space. This procedure is repeated for every new stable state until all the stable states are covered. Our algorithm can be used to derive interoperability test cases based on the single stimulus principle as well, especially when resources (such as available points of control and observation) are limited. The input-enumeration procedure used in our approach is similar to the concept of state-perturbation [15]. However, the goal of our procedure is to build a reachability tree and then generate interoperability test cases based on its stable states, while the goal of state-perturbation is to build a reachability tree for validating protocols between interacting processes without the concept of stable state. Moreover, in generating interoperability test cases, we mainly focus on inputs and outputs to and from the systems of machines which are regarded as black- or grey-boxes. However, the model (labelled transition systems) used in paper [15] is focused on internal interactions between systems of machines.

B. Test architecture

To cater to the multiple stimuli principle, we need a test architecture with a tester between the two IUTs as well as one at each end as shown in Fig. 4. The arrows in Fig. 4 represent points of control and observation (PCOs) with both observability and controllability functions. It is assumed that there is a communication channel through which the testers coordinate with each other. Tester C should be able to hold messages passing through it for a while. In this way, simulating multiple stimuli is feasible. In the case of TCP/IP, for example, tester C will be configured with two network interface cards for each IUT. The tester duplicates and drops all the messages between the two IUTs using libpcap [14], a library for system-independent packet capture, and then sends the messages according to the interoperability test cases, coordinating with the other two testers in doing so.

C. Multiple stimuli from one machine to another

Fig. 5 shows the channel status (modelled by queues) for every transition execution in deriving an interoperability test case in the given example. Starting with an external input $ex$, the algorithm pushes every internal output generated by the input into the channel $Q_b$. Next it moves to the peer machine $M_b$, processes all the messages in the channel, and pushes all internal outputs produced during this period of time into the channel $Q_c$. This procedure is repeated until both channels are empty. Note that the order of messages, $a$, $b$, $c$ (or, 0, 1, 2, 3) remains unchanged since first-come first-served scheduling is used. Moreover, protocol concurrency is still maintained, e.g., $tr_b | tr_a, tr_c$, or calling itself recursively. Lines 25 to 37 examine all possible transitions whose inputs are either external inputs or the current message from the queue. The program then calls itself recursively with appropriate variables for this test case. Finally, the rest of the function checks whether or not the current interoperability test case has encountered a specification error, namely, there is no transition for the current message (so-called implicit signal consumption).

This algorithm has been implemented in the C language with about 1000 lines of codes. The execution time of the program is negligible (80 milliseconds for generating 1012 test cases from...
Applications

We apply our test generation algorithm (Algorithm 1) to the connection establishment phase of TCP and the ATM/B-ISDN signaling protocol. Their IOSMs can be found in papers [7] and [8], respectively. These IOSMs may lack completeness compared to real specifications of the protocols. An example is illustrated later.

Given the IOSMs, the input data needed by our proposed method can be derived systematically. An IOSM is described in a file like that in Appendix I, in which each line represents a transition in the IOSM. For instance, “[-1]0:0 3:6 SETUPa CALL_PROC i_SETTUP” in Appendix I is the transition whose starting state is 0:0, destination state is 3:6, input message is SETUPa, and output messages are CALL_PROC and i_SETTUP. Each transition has a unique id. The minus sign is used to denote the transitions of one IOSM and the plus sign denotes the transitions of the other IOSM. The postfix is used to specify a specific IOSM, e.g., SETUPb is an external message being sent to IOSM M_b. The prefix ‘i_’ is used to denote an internal message.

The results of applying our method to the two protocols are summarized in Table I. In the case of ATM signaling protocol, a total of 1012 test cases were derived. 990 cases among them are for interoperability testing. An other 16 cases are for conformance testing because they involve one IUT only. The other 6 cases address specification errors for requesting the IUTs to receive unspecified input messages. Note that our method can also be used in detecting specification errors in the

Algorithm 1. Interoperability test suite derivation.

150 transitions).

E. Applications

We apply our test generation algorithm (Algorithm 1) to the connection establishment phase of TCP and the ATM/B-ISDN signaling protocol. Their IOSMs can be found in papers [7] and [8], respectively. These IOSMs may lack completeness compared to real specifications of the protocols. An example is illustrated later.

Given the IOSMs, the input data needed by our proposed method can be derived systematically. An IOSM is described in a file like that in Appendix I, in which each line represents a transition in the IOSM. For instance, “[-1]0:0 3:6 SETUPa CALL_PROC i_SETTUP” in Appendix I is the transition whose starting state is 0:0, destination state is 3:6, input message is SETUPa, and output messages are CALL_PROC and i_SETTUP. Each transition has a unique id. The minus sign is used to denote the transitions of one IOSM and the plus sign denotes the transitions of the other IOSM. The postfix is used to specify a specific IOSM, e.g., SETUPb is an external message being sent to IOSM M_b. The prefix ‘i_’ is used to denote an internal message.

The results of applying our method to the two protocols are summarized in Table I. In the case of ATM signaling protocol, a total of 1012 test cases were derived. 990 cases among them are for interoperability testing. An other 16 cases are for conformance testing because they involve one IUT only. The other 6 cases address specification errors for requesting the IUTs to receive unspecified input messages. Note that our method can also be used in detecting specification errors in the
Validation is based on the state space exploration technique. It interoperability test cases that are related to multiple stimuli. Our algorithm can be easily tailored to generate only those methods cannot detect faults due to multiple stimuli. Note that derived by the method in [7]. Therefore, it is clear the existing derived 990 interoperability test cases while only 54 were stimuli. In the case of the ATM signaling protocol, we have cases than existing methods [7], [8] in order to support multiple design phase. It is shown that our method generates more test cases than existing methods [7], [8] in order to support multiple stimuli. In the case of the ATM signaling protocol, we have derived 990 interoperability test cases while only 54 were derived by the method in [7]. Therefore, it is clear the existing methods cannot detect faults due to multiple stimuli. Note that our algorithm can be easily tailored to generate only those interoperability test cases that are related to multiple stimuli.

Fig. 6 shows two examples of the derived interoperability test cases. An interoperability test case consists of several transitions between the IOSMs. Illustrated in Fig. 6(a), the interoperability test case, $< -1, 4, -12, -61, 5, 16, 8, -60, -8 >$, involves multiple stimuli. This test case is based on the specification and represents a correct behavior of the protocol. Depicted in Fig. 6(b), the interoperability test case, $< -1, 75, -12, -9, -999 >$, reveals a specification error. This leads us to find incompleteness of the IOSM used in [8]. In fact, this case is covered in the error handling clause in ITU-T Q.2931 [23]. By adding a transition to the IOSM, in order to handle the unspecified input _i_REL’, we have derived 6 more interoperability test cases.

Table II gives the transition coverage for the IOSMs of TCP’s connection establishment phase and the ATM signaling protocol. Note that the traversed transitions include those covered by the conformance test cases. It is shown that our method has higher transition coverage than the existing methods that are based on the single stimulus principle only. Specifically, 5 more transitions of TCP and 9 more transitions of ATM are traversed than the existing methods in [7] and [8]. Our coverage is exactly the same as that covered by validation.

Validation is based on the state space exploration technique. It automatically checks a distributed system’s design for logical errors. Validation executes all possible combinations of events that can happen, and reports any indication that something has gone wrong.
techniques. A SDL commercial validation tool [13] produced the same result reported in Table II. In other words, our method can generate interoperability test suite whose transition coverage is equivalent to that of validation.

Table III shows the reasons why some transitions are not traversed. First, some transitions are incorrectly specified and will never be executed due to the absence of transitions that send the corresponding signals. For instance, the 38th transition in Appendix I needs an input signal \( \text{I\_RESTART} \), but there is no transition that sends the required signal. Note that a wrongly specified transition may give rise to subsequent untraversed transitions. Also, some transitions are over-specified. Over-specified transitions are superfluous and are unnecessary for our purpose. They include self-transitions whose start state and destination state are the same. Actually they are sometimes used for error handling. Since checking correct behavior is the objective of our method, such transitions are considered over-specified in this paper. Finally, some transitions are not traversed due to the limitation of the method used. As shown in the table, our method can traverse all the transitions that cannot be handled by the existing methods.

V. TEST CASE MINIMIZATION

It is observed that the multiple stimuli principle incurs a considerable number of interoperability test cases for the ATM signaling protocol. In this section, we shall minimize the interoperability test cases we have derived.

Our technique will maintain the same transition coverage, which is often measured by how many transitions in the given component IOSMs would be executed. Note that a component IOSM is used in this paper to mean a single machine in the composite IOSM. The same coverage criterion can be found in the guided random walk approach used for conformance testing [19]. As discussed in Section II, one main difference is that, in the guided random walk approach, the system adheres to the single stimulus principle and hence multiple simultaneous inputs cannot be examined. Another difference lies in when test minimization is carried out. The guided random walk approach is applied when the test cases are being generated. Since the approach always selects external transitions that have not been tested earlier, it may not be able to explore complex interaction scenarios of the composite machine and may generate more test cases than those minimized from the whole set of test cases.

More diverse redundancy criteria are discussed in the literature [21] with applications to interoperability testing of VoIP. A key idea is to ignore local transitions that are not relevant for interoperability testing such as obtaining information from an end-user in the VoIP system. Note that they did not consider multiple simultaneous inputs either. Our method, adopting multiple stimuli principle, uses a relatively restrictive criterion for minimization - it attempts to cover each transition in the component IOSMs. However, our test generation method already has some redundancy criteria applied, so that it could remove redundancy during test generation. The criteria (or assumptions) include ignoring interactions involving only one component machine and no cyclic reachability graph. Covering each transition of the component machines is the next applicable redundancy criterion in our test case minimization.

Throughout this section, the interoperability test cases derived from the ATM signaling protocol by Algorithm 1 are used for explanation and experiments.

A. Analysis of the given test cases

Let us define the transition execution rate for a finite state machine as the average number of executions for each transition in the IOSM. Given the test cases, the rate is defined for IOSM \( M \) as

\[
r = \frac{E(M)}{N}
\]

where \( E(M) \) is the total number of transition executions in IOSM \( M \) and \( N \) is the number of transitions of \( M \).

In the case of the ATM signaling protocol, 88 out of 150 transitions were covered by our derived interoperability test cases. The remaining 62 transitions were not traversed. We consider only those 88 transitions in the IOSM for minimization. Another factor is the total number of transition executions and it was 6308 times in this case. So the transition execution rate \( r \) is 71.68 (i.e., \( 6308 / 88 \)). This implies that each transition would be redundantly executed close to 72 times on average. The minimum cost (or most effective) test is achieved when the transition execution rate \( r \) is 1.

Our derived interoperability test cases originally constituted 990 test cases, which would execute 6308 transitions. In the following we provide methods to reduce the number of interoperability test cases while keeping the same transition coverage.

B. Definition of the minimization problem

Minimizing interoperability test cases with the same transition coverage is similar to the set-covering problem [17]. The set-covering problem is an optimization problem that models many resource-selection problems and is NP-hard.

An instance \((T, TC)\) of the set-covering problem consists of a finite set \( T \) and a family \( TC \) that are subsets of \( T \), such that every element of \( T \) belongs to at least one subset in \( TC \). Fig. 7 gives an example where \( T = \{t1, t2, t3, t4, t5, t6\} \) and \( TC = \{tc1, tc2, tc3, tc4\} \). The problem is to find a minimum-sized subset \( C \subseteq TC \) whose members cover all of \( T \). With regard to the problem of minimization of interoperability test cases, \( T \) represents a set of transitions in the individual machines of the composite IOSM and \( TC \) is a set of derived interoperability test cases. Each interoperability test case consists of more than one transition in \( T \) and every transition in \( T \) is covered by at least one interoperability test case (a subset) in \( TC \). Note that we do not consider transitions that are not covered by the originally derived interoperability test cases. The minimization of the interoperability test cases can be conducted with two policies;

Policy P1: to reduce the number of interoperability test cases, and
Policy **P2. to shorten the total length of interoperability test cases.**

The length of an interoperability test case is defined as the number of transitions contained in the test case. Note that a smaller number of test cases do not imply shorter total length. However, it is obvious that a long test case is costly in terms of testing time and a large number of test cases require effort in test preparation for each of them. Rigorous analysis on the cost of testing with the interaction effect between these two policies is not dealt with in this paper.

In essence, policy **P1** minimizes the number of subsets of C, i.e., \( n(C) \) while policy **P2** minimizes the total length of subsets of C, i.e., \( \sum n(c_i) \) where \( c_i \) is the \( i \)-th subset of C. In the example of Fig. 7, under policy **P1**, the minimum-sized set cover \( C = \{ tc1, tc2 \} \) or \( \{ t1, t2, t3, t4, t5, t6 \} \) while, under policy **P2**, \( C = \{ tc1, tc3, tc4 \} \) or \( \{ t1, t2, t3, t4, t2, t5, t3, t6 \} \). Although the latter solution has more interoperability test cases – three compared to the former’s two, the total length (= 8) is shorter than that of the former (= 9).

To solve the minimization problem, we consider a two-dimensional matrix. The minimization problem is modelled so as to find a subset of the columns of an \( M \times N \) zero-one matrix as shown in Fig. 8, and covers all rows. Here, M is the size of the finite set T and N is the number of subsets in T in an instance (T, TC) of minimization problem. More precisely, this problem is defined as follows:

\[
\text{Minimize } \sum_{i=1}^{M} s_i \text{ in policy } P1, \text{ or } \sum_{i=1}^{M} \sum_{k=1}^{N} a_{ik}s_i \text{ in policy } P2, \tag{1}
\]

\[
\text{subject to } \sum_{i=1}^{M} a_{ik}s_i \geq 1, \quad \text{for } k=1, ..., M, \tag{2}
\]

\[
\text{with } s_i \in \{0,1\}, \quad \text{for } i=1, ..., N. \tag{3}
\]

Equation (2) states that at least one column must cover a particular row. Fig. 8 illustrates this matrix model. With policy **P1**, we minimize the number of 1’s in \( s_i \) subject to the condition that every row must be set to 1 in at least one of the columns selected by \( s_i \). With policy **P2**, \( s_i \) is set such that the number of selected elements for all rows is minimized.

So far, we have defined the problem of minimizing interoperability test cases. In the following, we propose two approaches for this problem. One is a simple ad-hoc approach that does not consider either policy explicitly in selecting test cases. The other approach takes the policies into consideration and is a greedy heuristics that gives approximate solutions.

**C. Ad-hoc approach**

This approach selects an interoperability test case in an ad-hoc way until all transitions of the component IOSMs are covered. For every transition, it simply picks any test case that covers the transition. This scheme is presented in Algorithm 2.

The input is a \( M \times N \) zero-one matrix as shown in Fig. 8, and the output is a set of binary variables \( s_i \in \{0,1\} \) where \( i=1, ..., N \). Note that the input and output are the same in all the following algorithms we propose. For the ad-hoc approach, we introduce a temporary variable \( v \) to keep the visited (or covered) elements. From lines 2 through 6, for each element \( v_k \), the algorithm selects any \( i \) such that \( a_{ik}=1 \) and sets \( s_i=1 \). This means that the \( i \)-th column is selected as an element of the minimized set. Then, in line 5, it updates \( v \) for the other elements in the \( i \)-th column since the column covers them as well. The complexity of this algorithm is \( O(M) \), i.e., it is proportional to the number of given transitions to cover in the component IOSMs. The result obtained by using this ad-hoc algorithm can be costly because it does not consider redundancy or optimization. However, this algorithm can be embedded into our test generation method presented in Algorithm 1 so that the test cases are minimized at the same time as they are derived.

**D. Greedy Heuristics**

To get smaller size set, we examine greedy heuristics with a logarithmic ratio bound [17]. We propose two algorithms with policies **P1** and **P2**, respectively. With policy **P1**, which aims to reduce the number of interoperability test cases, our greedy method works by picking the set that covers the largest number of remaining uncovered elements at each stage. This algorithm is called quantity minimization and is shown in Algorithm 3. The difference from the ad-hoc minimization is that it selects, at each stage, a column that updates \( v_k \) at maximum. As a result, it
would cover all rows quicker and hence select relatively small number of columns. When there is more than one choice, it selects the column that has a minimum column sum, i.e., shortest-length test case.

With policy P2, which pursues the minimum total length of interoperability test cases, our greedy method works by picking, at each stage, the set that has the fewest number of elements that are already covered. In other words, it tries to keep redundancy as low as possible. This algorithm is called length minimization and is shown in Algorithm 4. Note that the difference between length minimization and quantity minimization is the priority they use. As shown in lines 8-9, it first finds the columns, which cover the already covered rows the least, and then chooses the column that updates $v_k$ the most. By doing so, the result approximates the optimal solution that minimizes the total length of test cases because at each stage it picks the column that increases redundancy least. When there is no redundancy in every stage, which means that each column always covers new rows, the total length of the interoperability test cases is the same as the number of rows (i.e., the number of transitions in the given component IOSMs).

E. Applications

By applying the three proposed algorithms, namely ad-hoc, quantity, and length minimizations, to the ATM signaling protocol, we obtained the results shown in appendixes II, III, and IV, respectively, and they are summarized in Table IV.

Basically, it is shown that we can reduce the number of interoperability test cases by about 94% even with the ad-hoc algorithm (from 990 to 60 test cases). Note that the quantity minimization algorithm results in the smallest number of test cases (35) with slightly more transition executions (i.e., total length of test cases) (182) than that of length minimization algorithm (172 transitions in 38 test cases). Both quantity and length minimizations have similar execution rate of about two, i.e., each transition in the given component IOSMs will be executed twice on average when our interoperability test cases are applied. It would be around five times in case of ad-hoc minimization. As stated earlier, ad-hoc minimization can be combined with our test generation algorithm shown in Algorithm 1 so that it is not necessary to explore all the patterns of interoperating behavior.

Although the results of quantity and length minimizations are similar, the cost required in actual testing may be different. For example, suppose the cost required in preparing and driving the machines to a state ready for applying each single interoperability test case is very high, it would be desirable to have a smaller number of test cases with slightly more transition executions. Rigorous analysis of this trade-off requires further research.

VI. CONCLUSION

We have proposed a new method for automatically generating and minimizing interoperability test suites that support multiple inputs to the system under test. This means that the method adopts the multiple stimuli principle rather than the traditional single stimulus principle used in existing methods. As a result, transition coverage obtained by our method covers the same behavior space that is managed by the
A well-designed protocol is supposed to be able to deal with the race condition, i.e., anomalous behavior caused by unexpected relative timing of events. In other words, the protocol designer has incorrectly assumed that a particular event would always happen before another. Our method can generate interoperability test cases that test for the ability of the implementation to handle possible race conditions, and can also be used in the design phase to check whether or not race conditions would occur in the protocol.

To do this, we have developed formal models on which the interoperability test suite derivation algorithm is based. We have applied our test generation method to the ATM signaling protocol and also to a part of the TCP protocol. Experimental results have shown that our method produces test suites with higher transition coverage than the existing methods: 26% higher for TCP and 12% higher for ATM. The additional transition coverage is for possible simultaneous interactions allowed by the protocol specifications.

Since executing test cases is expensive, we have also developed three automatic minimization algorithms that can reduce the test size. They are able to reduce the interoperability test cases considerably while maintaining the same transition coverage.

APPENDIX I. TRANSITIONS OF AN IOSM OF THE ATM SIGNAING PROTOCOL

APPENDIX II. TEST CASES MINIMIZED BY AD-HOC MINIMIZATION

APPENDIX III. TEST CASES MINIMIZED BY QUANTITY MINIMIZATION

APPENDIX IV. TEST CASES MINIMIZED BY LENGTH MINIMIZATION
REFERENCES


Soounk Seol received his B.S. in Information and Communication Engineering from the Korea University of Technology and Education in 1998, and M.S. and Ph.D. in Computer Science from Information and Communications University (ICU) in Korea in 2000 and 2004. From 1998 he was also a visiting researcher at the Protocol Engineering Center in Electronics and Telecommunications Research Institute (ETRI) until October 1999. Currently he is with the Technology Laboratory at KT. His research interests include protocol engineering, interoperability testing, QoS network, and mobile computing.

Myungchul Kim received B.A. in Electronics Engineering from Ajou University in 1982, M.S. in Computer Science from the Korea Advanced Institute of Science and Technology in 1984, and Ph. D in Computer Science from the University of British Columbia, Vancouver, Canada, in 1993. Currently he is with the faculty of the Information and Communications University, Taejon, Korea as the dean of School of Engineering. Before joining the university, he was a managing director in Korea Telecom Research and Development Group for 1984 - 1997 where he was in charge of research and development of protocol and QoS testing on ATM/B-ISDN, IN, PCS and Internet. He has also served as a member of Program Committees for IFIP International Workshop on Testing of Communicating Systems (IWTCs), IEEE International Conference on Distributed Computing Systems, and IFIP International Conference on Formal Description Technique / Protocol Specification, Testing and Verification, the chairman of Profile Test Specifications - Special Interest Group of Asia-Oceania Workshop, and co-chair of the IWTCs’97 and the FORTE’01. His research interests include Internet, mobile Internet, protocol engineering, and mobile computing.

Samuel T. Chanson received his BSc degree in Electrical Engineering from Hong Kong University in 1969, and his MSc and PhD degrees in Electrical Engineering and Computer Sciences from the University of California, Berkeley in 1971 and 1975 respectively. He was a faculty member in the School of Electrical Engineering at Purdue University for two years. Then he moved to the Department of Computer Science at the University of British Columbia, where he became a full Professor and Director of the Distributed Systems Research Laboratory. In 1993 he joined the Computer Science Department at the Hong Kong University of Science and Technology. Prof. Chanson has served as an Editor of IEEE/ACM Transactions on Networking, the Journal of Computing and Information, and also New Generation Computing. He is also a member of the IFIP TC6/WG6.1 Committee. He has chaired many international conferences on communication protocols and distributed systems and has published more than 160 papers in these areas. His research interests include web content delivery technologies, QoS routing, energy-aware scheduling, and Internet security.

Sungwon Kang received his B.A. from Seoul National University in Korea in 1982 and M.S. and Ph.D. in computer science from the University of Iowa in U.S.A. in 1989 and 1992. From 1993 he was a principal researcher at Korea Telecom R & D Group until October 2001. Throughout this period his main area of research was protocol testing and formal methods. Then he joined Information and Communications University where he holds the position of an assistant professor. Since 2003, he is an adjunct faculty of Carnegie-Mellon University, USA, for the Master of Software Engineering Program. Areas of his current research interest include software engineering, software architecture, software testing, and formal methods.