

**On Finite Covering of Infinite Spaces
for Protocol Test Selection**

by
Masaaki Mori
and
Son T. Vuong

Technical Report 93-7
April 1993

Department of Computer Science
University of British Columbia
Rm 333 - 6356 Agricultural Road
Vancouver, B.C.
CANADA V6T 1Z2

Telephone: (604) 822-3061
Fax: (604) 822-5485

On Finite Covering of Infinite Spaces for Protocol Test Selection

Masaaki Mori* and Son T. Vuong

Department of Computer Science
University of British Columbia
Vancouver, B.C., Canada V6T 1Z2
Email : mori@cs.ubc.ca, vuong@cs.ubc.ca

April 1993

Keywords: Conformance testing; Communication protocols; Metric topology; Formal language theory; Finite covering

Abstract

The core of the protocol test selection problem lies in how to derive a finite test suite from an infinite set of possible execution sequences (protocol behaviors). This paper presents two promising approaches to this problem : (i) the metric based topological approach, and (ii) the formal language theoretic approach ; both aim at producing finite coverings of an infinite set of execution sequences. The former approach makes use of the property of compactness of metric space, which guarantees the infinite metric space can be fully covered by a finite number of open "balls" (subspaces). The latter approach relies on the property that the Parikh mapping of a set of all execution sequences can be represented by a finite union of linear sets. Two simple protocol examples are given to elucidate the formal language theoretic approach.

1 Introduction

One of the fundamental problems in protocol conformance testing is how to generate and select test suites formally. Since communication protocols are generally complex and recursive, a test suite which covers all behaviors of such a protocol specification must be infinite. Thus, the main issue in the problem of test selection is how to characterize an infinite set of generated test cases (or behaviors) and how to select a finite number of test cases with some sound coverage measure from the infinite set. Such a finite selected subset of test cases

*Masaaki Mori is on sabbatical leave at the Univ. of British Columbia from Shiga Univ., Hikone, Japan.

must be representative of the infinite set in a sense and must "fully" cover the infinite set of the generated test cases. This problem of finite covering of the infinite set is fundamental and important for systematic test case selection.

There have been many research efforts on conformance test generation for protocols modeled as FSMs [1, 2, 3, 4], however only a few approaches [5, 6, 7] focused on the protocol specification structure from the point of view of coverage measure. In order to deal with the fundamental problem of generation and selection of test cases, it is essential to develop a method that characterizes an infinite set by means of a finite number of subsets which satisfies some test selection criteria defined on certain structures of the infinite set.

This paper aims at providing a theoretical framework for the derivation of viable formal methods for selecting finite test suites from infinite sets. This framework is comprised of two general approaches :

- Topological approach based on metric space : In this method, we consider a protocol as an infinite set of all possible execution sequences (i.e. protocol behaviors), where a metric (testing distance) is defined on pairs of execution sequences [8]. Here, the existence of finite covering for infinite metric space plays an crucial role in the problem of test suite selection and this finite covering is derived from the notion of compactness of metric space. In addition, the Lebesgue number is also an interesting notion in the sense that it gives the maximum diameter of the open subsets covering a given infinite metric space, and as such , it can be viewed as a measure of confidence in communication software reliability. In general, not every metric space is compact and moreover it is not trivial to prove a metric space is compact. In this paper, we show the compactness property holds for the metric space defined in the framework of [5].
- Formal language theoretic approach : This approach makes use of the notion of the Parikh mapping [9] of action sequences in dealing with the problem of test suite selection in a finite frameworks. In this method, a (potentially infinite) set of all execution sequences mapped by the Parikh's method can be represented by a *semilinear* set (i.e. a finite union of linear sets), provided the set (of execution sequences) can be generated by a context-free grammar. This is the case for most classes of execution sequences of protocols. For example, execution sequences with recursive actions taken from a set

$\{a_1, a_2, \dots, a_n\}$ in a protocol would in general take the form $(a_1^{m_1} a_2^{m_2} \dots a_n^{m_n})^p$ which can be shown to be a context-free language.

In this paper, the two approaches are presented in details and two simple protocol examples are provided to illustrate the novel idea of the formal language approach. Some conclusions are made regarding the potential ultimate solutions to the problem of finite test selection from an infinite space of execution sequences.

2 Topological Approach based on Metric Space

A topological property in a metric space, *finite covering*, is guaranteed by *compactness* property of the metric space. This property is important in that the finite covering makes it possible to address not only the problem of finite test selection from an infinite set of execution sequences, but also the issue of coverage measure by considering the radius of each open ball of the finite set of balls which covers the metric space. In this case, the radii of open balls can be the basis of the ε -dense notion defined in [5], and the largest radius for a given finite covering, i.e. the *Lebesgue number*, can serve as a sound coverage measure for confidence in conformance.

2.1 Basic Definitions

Definition 2.1 Let S be any nonempty set. A function $d : S \times S \rightarrow \mathbf{R}$ is called a *metric* on S if it satisfies the following conditions for all $x, y, z \in S$, where \mathbf{R} denote the set of all real numbers :

- (1) $d(x, y) \geq 0$
- (2) $d(x, y) = 0$ if and only if $x = y$.
- (3) $d(x, y) = d(y, x)$
- (4) $d(x, y) \leq d(x, z) + d(z, y)$

A set S together with a metric d is said to be a *metric space*, and is denoted by (S, d) .

Definition 2.2 A sequence $\{x_n\}$ from a metric space is called a *Cauchy sequence*, if for each $\varepsilon > 0$ there exists a number N such that for all n and m larger than N we have

$$d(x_n, x_m) < \varepsilon.$$

Definition 2.3 If a metric space has the property that every Cauchy sequence converges (to some point of the space), we say that the space is *complete*.

A sequence in \mathbf{R} is convergent if and only if it is a Cauchy sequence, however, for a general metric space, only one direction remains true in general. That is, the statement that every Cauchy sequence converges is not true in an arbitrary space, for example, in the space of rational numbers with the usual metric.

Definition 2.4 A metric space S is said to be *totally bounded* if, for each $\varepsilon > 0$, there is a finite collection of points $\{x_1, \dots, x_n\}$ such that each $x \in S$ is within a distance of ε of one of the x_i ($1 \leq i \leq n$). This is equivalent to saying that for each $\varepsilon > 0$ the space S is covered by a finite number of balls of radius ε .

Definition 2.5 Let S be a metric space and K be a subset of S . We say that a collection \mathcal{U} of open sets in S is an *open covering* for a subset K if K is contained in the union of the sets in \mathcal{U} . K is also said to be *compact* if every open covering \mathcal{U} of K has a finite number of elements, that is, if there is a finite collection $\{O_1, \dots, O_N\}$ of open sets such that $K \subset \bigcup_{i=1}^N O_i$.

If \mathcal{U} is an open covering of a metric space S , then each $x \in S$ is contained in some open set $O \in \mathcal{U}$, and hence for some $\delta > 0$ there is a ball $B_{x,\delta}$ about x that is contained in O .

We have the following proposition [10] : This proposition tells us that for a compact metric space this property holds uniformly in the sense that we can find δ independently of x , as illustrated in Figure 1. The largest number ε with the property stated in the proposition is called the *Lebesgue number* of the covering \mathcal{U} .

Proposition 2.1 Let S be a compact metric space, and let \mathcal{U} be an open covering of S . Then, there is a positive number $\varepsilon > 0$ such that for each $x \in S$ and each $\delta < \varepsilon$ the ball $B_{x,\delta}$ is contained in some open set $O \in \mathcal{U}$.

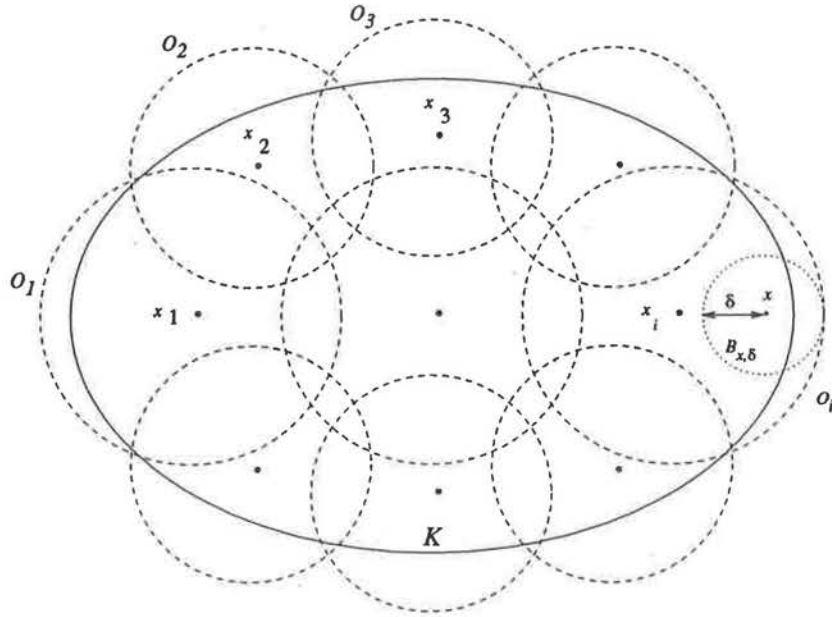


Figure 1: A finite covering of K by open balls with center points $\{x_1, \dots, x_n\}$

The following proposition [10] is useful for proving a metric space compact in the next subsection.

Proposition 2.2 A metric space S is compact if and only if it is complete and totally bounded.

Proposition 2.3 Let (X, d) be a metric space. If a function $d' : X \times X \rightarrow \mathbf{R}^+$ is defined by

$$d'(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$

then, d' becomes a bounded metric on X .

Proof It is trivial that the conditions (1), (2) and (3) in Definition 2.1 hold.

The condition (4) in Definition 2.1 is proved by using the following two inequalities :

$$\frac{a}{1+a} + \frac{b}{1+b} \geq \frac{a+b}{1+a+b} \quad (a, b \geq 0), \quad \frac{a}{1+a} \geq \frac{b}{1+b} \quad (a \geq b > 0)$$

From these inequalities, the following holds for $x, y, z \in X$:

$$d'(x, y) + d'(y, z) = \frac{d(x, y)}{1 + d(x, y)} + \frac{d(y, z)}{1 + d(y, z)} \geq \frac{d(x, y) + d(y, z)}{1 + d(x, y) + d(y, z)} \geq \frac{d(x, z)}{1 + d(x, z)} = d'(x, z)$$

This implies the condition (4) holds and d' becomes a metric on X .

On the other hand, since $d'(x, y) = \frac{d(x, y)}{1 + d(x, y)} < 1$ holds for any $x, y \in X$, d' is also bounded.

Proposition 2.4 Let (X_i, d_i) , $i \in N$ be a metric space. For two points $x = (x_n)$ and $y = (y_n)$ which belong to a direct product set $X = \prod_{i=1}^{\infty} X_i$, if a function $d : X \times X \rightarrow \mathbf{R}^+$ is defined by

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}$$

then, d becomes a bounded metric on X .

Proof The conditions (1), (2) and (3) in Definition 2.1 hold trivially.

We now show by using Proposition 2.3 the function d satisfies the triangle inequality as follows. Let $x = (x_n), y = (y_n), z = (z_n) \in X$. For any $n \in N$,

$$d(x, y) + d(y, z) \geq \sum_{i=1}^n \frac{1}{2^i} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)} + \sum_{i=1}^n \frac{1}{2^i} \frac{d_i(y_i, z_i)}{1 + d_i(y_i, z_i)} \geq \sum_{i=1}^n \frac{1}{2^i} \frac{d_i(x_i, z_i)}{1 + d_i(x_i, z_i)}$$

Therefore,

$$d(x, y) + d(y, z) \geq \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2^i} \frac{d_i(x_i, z_i)}{1 + d_i(x_i, z_i)} = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{d_i(x_i, z_i)}{1 + d_i(x_i, z_i)} = d(x, z)$$

This implies d is a metric on X . On the other hand, for any $x, y \in X$,

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)} \leq \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

Thus d is also bounded.

2.2 Examples of Metrics in Labeled Transition Systems

Whenever we examine an application problem by means of a metric system, it is important to ensure the metric space considered has the compactness property since this property guarantees the finite covering of the metric space, thus enabling the problem to be dealt

with in a finite context and providing a guideline for systematic handling of the problem. Here, we should note that in proving a set to be compact we have to show that any open cover (possibly containing uncountably many open sets) has a finite subcover. It is not sufficient to pick a particular open cover and extract a finite subcover. Therefore, it is generally difficult to apply the definition to show a set is compact.

Fortunately we can prove a metric space (D, dt) is compact by using the notion of totally boundedness as in the following examples. In these examples, we show that compactness is an important topological property : whether a given metric space is compact or not depends largely on the definition of the metric.

2.2.1 Example 2.2.1 - The case dt has rational numbers

Suppose that the functions and their constraints in the metric definition are given in Definitions 2.6 & 2.7, and Constraints 2.1 & 2.2. Then, (D, dt) becomes a complete metric space [5].

Definition 2.6 Let L be a set of observable actions and N be a set of recursion depth of their actions in a protocol specification, and let D be a set of all (finite and infinite) execution sequences derived from an initial term. Here, a set of all execution sequences is denoted by $Z = \{(a_i, \alpha_i) | a_i \in L, \alpha_i \in N, 0 \leq i \leq k\}$. $\delta_k(A, B)$, the measure of the difference in the level of recursion between two execution sequences $A = \{(a_k, \alpha_k)\}_{k=1}^K$ and $B = \{(b_k, \beta_k)\}_{k=1}^L$ of length K and L , is defined as follows :

$$\delta_k(A, B) = \begin{cases} 0 & \text{if } a_k = b_k \text{ and } \alpha_k = \beta_k \\ |\alpha_k - \beta_k| & \text{if } a_k = b_k \text{ and } \alpha_k \neq \beta_k \\ \infty & \text{if } a_k \neq b_k \end{cases}$$

for $k = 1, 2, \dots, \min\{K, L\}$, and we take $\delta_k(A, B) = \infty$, for $\min\{K, L\} < k \leq \max\{K, L\}$. Clearly, $\delta_k(A, B) \in N \cup \{\infty\}$.

Definition 2.7 For any two (finite or infinite) execution sequences $A = \{(a_k, \alpha_k)\}_{k=1}^K$ and $B = \{(b_k, \beta_k)\}_{k=1}^L$ in D , where $K, L \in N \cup \{\infty\}$, the testing distance dt between A and

B in D is defined as follows :

$$dt(A, B) = \sum_{k=1}^{\max\{K,L\}} p_k r(\delta_k(A, B))$$

Here, the functions p_k represents the weight of the individual calls reflecting the weight of the computational pattern, and $r(\delta_k(A, B))$ represents the weight of the difference in the level of recursion of a (recursive) call within the execution sequences. Both functions p and r are definable with the following constraints in order for dt to be a metric in D .

Constraint 2.1 $\{p_k\}_{k=1}^{\infty}$ is a sequence of positive numbers such that $\sum_{k=1}^{\infty} p_k$ converges.

Constraint 2.2 $\{r_k\}_{k=0}^{\infty}$ is an increasing sequence in $[0,1]$ such that $r(0) = 0$ and $\lim_{k \rightarrow \infty} r(k) = 1$. Furthermore, the sequence $\{\frac{r(k)}{k}\}_{k=1}^{\infty}$ is nondecreasing.

From these definitions and constraints, the following propositions hold [5].

Proposition 2.5 The dt defined in the previous definition is a distance in D .

Proposition 2.6 The metric space (D, dt) is *complete* (i.e. every Cauchy sequence in (D, dt) is convergent).

The definitions and constraints stated above (i.e. the definition of dt) are essentially similar to the definition of d_i as given in Proposition 2.4 if we take $d_i(x_i, y_i) = |x_i - y_i|$. However the metric dt (stated above) differs from the metric d (of Proposition 2.4) in that the range of d is \mathbf{R} (a set of real numbers), whereas the values of dt introduced here is a closed set $[0,1]$ in \mathbf{Q} (a set of rational numbers), and (D, dt) is not compact trivially. However, imposing a finiteness constraint to the recursion levels of each action, we can prove the compactness of the metric space (D, dt) as follows :

Theorem 2.1 The metric space (D, dt) is totally bounded.

Proof Since $\sum_{k=1}^{\infty} p_k$ is summable (i.e. converges) according to the definition of dt , for every

$\varepsilon > 0$ there exists some natural number N_0 such that $\sum_{k=N_0}^{\infty} p_k < \varepsilon$. Consider now all sequences $\{(a_i, \alpha_i)\} \in (D, dt)$ that have all different combinations of (a_i, α_i) for $i = 1, \dots, N_0$. Then, there is a finite number of such sequences, which we denote $X = (x_1, \dots, x_n)$, because the alphabet representing events a_i is finite (N_0) and the number of each α_i is also finite according to the constraint assumed. If we draw an ε -ball around each such finite sequence $\{(s_i, \sigma_i)\}$ in X , then all sequences $\{(a_i, \alpha_i)\}$ with initial sequences such that $i \leq N_0$, $a_i = s_i$ and $\alpha_i = \sigma_i$, will fall into the ε -ball as shown in Figure 2. In this way, an arbitrary sequence is within a distance of ε from one of the x_i . That is, (D, dt) is totally bounded.

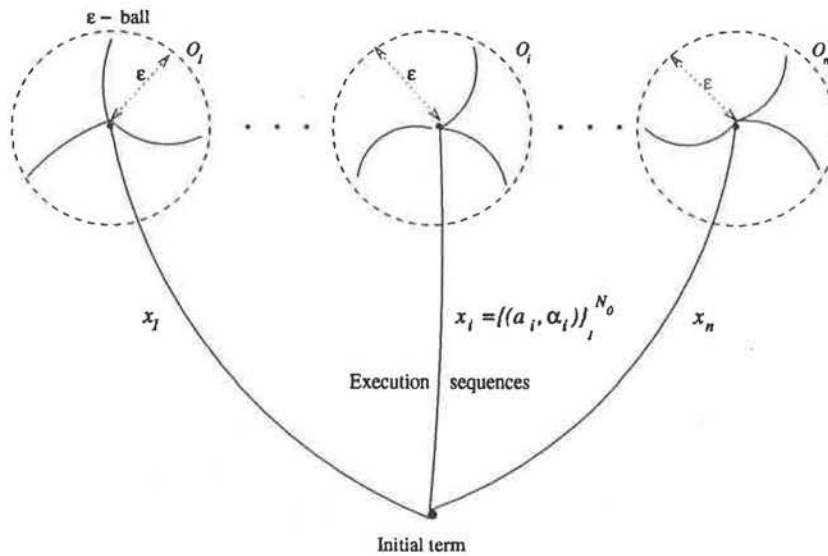


Figure 2: Sketch of the finite covering by open balls

This naturally leads to the next theorem.

Theorem 2.2 The metric space (D, dt) is compact.

Proof Since the metric space (D, dt) is both complete (Proposition 2.6) and totally bounded (Theorem 2.1), (D, dt) is compact, according to Proposition 2.2.

The proof of Theorem 2.1 is critical and depends on the assumption of finiteness on the recursion depths of each action in an execution sequence. This assumption may appear to

be too restrictive, but it is essential for the compactness property of the metric space (D, dt) to hold true. In fact, without this constraint each element of (D, dt) is a finite or infinite sequence of a tuple $\{(a_i, \alpha_i)\}$ with values of nonnegative integers, and it is well known that the closed interval of infinite set of rational numbers is not compact.

From a practical viewpoint, the depth of recursion should be restricted to a finite number in actual implementations. Thus, the finiteness assumption can be considered as plausible from the point of view of applying the compactness theory to conformance testing in the framework of the metric space (D, dt) proposed.

It is also worth while to note that the *Lebesgue number* is an important notion in the sense that the number gives a criteria on the largest finite covering of a given metric space, i.e. a finite covering with the smallest number of elements, i.e. a test selection with the smallest number of test cases to fully represent the original infinite set. This notion is closely related to the notion of ε -dense in [5]. Thus, with the Lebesgue number corresponding to a given finite covering of a metric space, we can define a sound coverage measure which clearly represents the limit of radius of the ε -ball with respect to the given finite covering.

2.2.2 Example 2.2.2 - The case dt has irrational numbers

As given in [7], the testing distance between two executions $\mathbf{s} = (a_1, \alpha_1) \cdots (a_n, \alpha_n)$ and $\mathbf{t} = (b_1, \beta_1) \cdots (b_n, \beta_n)$ (if \mathbf{s} and \mathbf{t} are of different lengths then the shortest sequence is padded to match the length of the longer sequence) is defined as :

$$dt(\mathbf{s}, \mathbf{t}) = \sum_{k=1}^n p_k r(\delta_k(\mathbf{s}, \mathbf{t}))$$

where the functions p and r are user configurable. If the functions p , r , and δ satisfy the following properties, then the distance function dt is a metric on the set of all control sequences starting at the initial protocol state (the proof is omitted as it is essentially the same as the one in [5]) :

- $\{p_k\}_{k=1}^{\infty}$ is a sequence of positive numbers such that $\sum_{k=1}^{\infty} p_k = p < \infty$.
- $\{r_k\}_{k=0}^{\infty}$ is an increasing sequence in $[0,1]$ such that $\lim_{k \rightarrow \infty} r_k = 1$. Put $r_{\infty} = 1$.

$$\delta_k(\mathbf{s}, \mathbf{t}) = \begin{cases} |r_{\alpha_k} - r_{\beta_k}| & \text{if } a_k = b_k \\ 1 & \text{if } a_k \neq b_k \end{cases}$$

These definitions are quite similar to those of Example 2.2.1, however, the critical difference between Examples 2.2.1 and 2.2.2 is in $\delta_k(\mathbf{s}, \mathbf{t})$. In this case, r_k is given as a continuous function whose range is $[0,1]$ in \mathbf{R}^+ (a set of positive real numbers), and the values of dt is a closed interval $[0, p]$. Therefore, the metric space (D, dt) becomes compact, as actually shown in [7].

In this section, the compactness property is discussed and also the theoretical feasibility of the metric based approach in test case selection environments is formally proven in terms of the existence of finite coverings of infinite metric spaces.

3 Formal Language Theoretic Approach

By examining the recursive structures of communication protocols, we note that most action (execution) sequences of protocols can be represented in the general form $(a_1^{m_1} a_2^{m_2} \dots a_n^{m_n})^p$, where the alphabets a_1 through a_n are primitives of a protocol representing actual actions, and m_1 through m_n denote the recursion depth of each primitive, and p denotes the (high) level of recursion of action sequences.

Fact $\{(a_1^{m_1} a_2^{m_2} \dots a_n^{m_n})^p\}$, ($p \geq 0$, $m_i (1 \leq i \leq n) \geq 1$) is a context-free language as it can be generated by a context-free grammar $G = (N, T, P, S)$ with the following production rules :

$$N = \{S, B, X_1, X_2, \dots, X_n\}, \quad T = \{a_1, a_2, \dots, a_n\},$$

$$P : S \rightarrow \lambda, \quad S \rightarrow SB$$

$$B \rightarrow X_1 X_2 \dots X_n$$

$$X_1 \rightarrow X_1 a_1, \quad X_1 \rightarrow a_1$$

$$X_2 \rightarrow X_2 a_2, \quad X_2 \rightarrow a_2$$

...

$$X_n \rightarrow X_n a_n, \quad X_n \rightarrow a_n$$

Definition 3.1 Let $\Sigma = \{a_1, a_2, \dots, a_{|\Sigma|}\}$ and L denote a set of actions and a language generated by grammar G , respectively. Moreover, let's consider the Parikh mapping $\psi(\omega)$ such that

$$\psi(\omega) = (\#_{a_1}(\omega), \#_{a_2}(\omega), \dots, \#_{a_{|\Sigma|}}(\omega))$$

where $\omega \in L$, and $\#_{a_i}(\omega)$ represents the number of occurrences of action a_i in the action sequence $\omega \in L$. That is, by $\psi(\omega)$, we represent a (potentially infinite) sequence of actions as a point in $|\Sigma|$ -dimensional vector space. We denote a set of $\psi(\omega)$ by $\Psi(L) = \{\psi(\omega) | \omega \in L\}$.

Definition 3.2 A set S of n-tuples is termed *linear* iff there is an integer $k \geq 0$ and n-tuples v_0, v_1, \dots, v_k such that S consists of all n-tuples of the form

$$v_0 + \sum_{i=1}^k x_i v_i$$

where $v_0, v_1, \dots, v_k \in \mathbb{N}^n$, $x_i \in \mathbb{N}$ and \mathbb{N} is a set of non-negative integers.

A set S of n-tuples is termed *semilinear* iff it is a *finite* union of linear sets.

$$S = \bigcup_{j=1}^h S_j, \quad S_j = \{v_{0j} + \sum_{i=1}^{k_j} x_{ij} v_{ij}\}$$

Theorem 3.1 [9] If a language L is context-free, then $\Psi(L)$ is semilinear. That is,

$$\Psi(L) = \bigcup_{j=1}^h \{v_{0j} + \sum_{i=1}^{k_j} x_{ij} v_{ij}\}$$

This theorem means that, if we map any execution sequence ω in a labeled transition systems [5] that belongs to context-free language into an ordered tuple by $\psi(\omega)$, there exists a canonical description for the set of all such tuples $\psi(\omega)$, which is a finite union of linear sets, as illustrated in Figure 3. It is interesting to note this is nothing more than the existence of a finite covering of an infinite set $\Psi(L) = \{\psi(\omega) | \omega \in L\}$.

Semilinear sets have desirable properties. It turns out that semilinear sets are exactly the sets describable within the first order logic of the natural numbers with addition, also known as *Presburger Arithmetic*, and the decision problems of Presburger formula is well

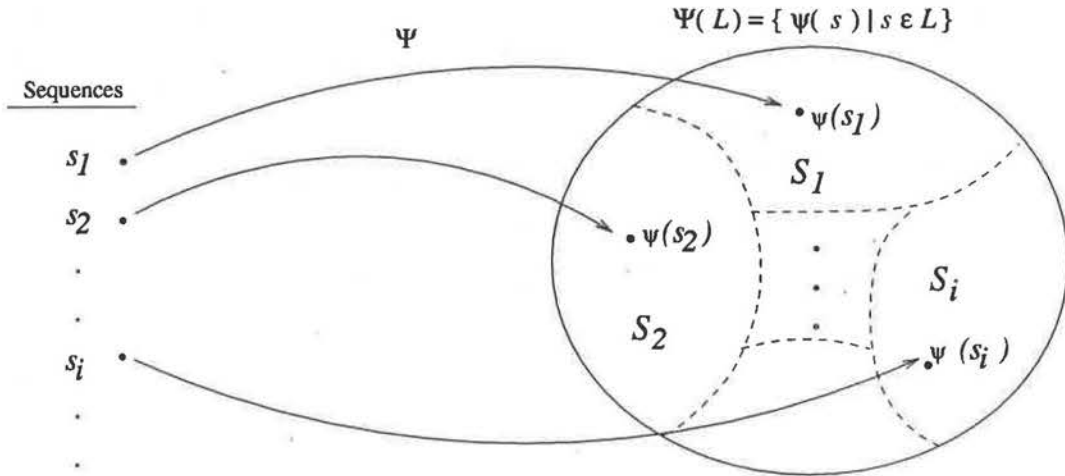


Figure 3: Finite covering of execution sequences by linear sets

known to be decidable [11]. That is, if a language L is context-free language, then $\Psi(L) = \{\psi(\omega) | \omega \in L\}$ is a semilinear set and there is a Presburger formula $P_L(a_1, a_2, \dots, a_{|\Sigma|})$ for $\Psi(L)$ such that $P_L(n_1, n_2, \dots, n_{|\Sigma|})$ is true if and only if $(n_1, n_2, \dots, n_{|\Sigma|}) \in \Psi(L)$ [9]. P_L is called Presburger formula for $\Psi(L)$. In the formalization by finite number of linear sets, we need not impose any assumptions related to a finite state machine but assume that the set of execution sequences of protocols must be context-free language. Such a weakness of assumptions on the protocol specification is worth noting. Moreover it is impossible to note the possibilities that the methods for protocol test sequence selection and the other similar decision problems for finding the input/output test cases satisfying given requirements, can be formulated in terms of decidable classes based on Presburger logic.

3.1 Application to protocols

The followings are examples to which the theory stated in this section can be applicable. One is a 3-state FSM example presented in [2] where the UIOV test cases selection algorithm is discussed for the FSM, and the other is a subset of the OSI Transport Service.

Example 3.1 A FSM specification example [2]

We denote a pair of input and output symbols attached to each transition in Figure 4 by one symbol a_i or b_j , where i and j denote the outputs for input a and b respectively; i is

0 or 1, and j is 1. Consider the sequences made up of symbols a_1 , a_0 , and b_1 . Let $State_i$ ($1 \leq i \leq 3$) denote a set of regular expressions of sequences of the symbols reachable from the initial state to state i in Figure 4. We have the following :

$$\begin{aligned} State_1 &= \{\lambda, a_1(b_1a_0)^*a_0, b_1a_0(b_1a_0)^*a_0, b_1(a_0b_1)^*b_1, a_1b_1(a_0b_1)^*b_1\} \\ State_2 &= \{a_1(b_1a_0)^*, b_1a_0(b_1a_0)^*\} \\ State_3 &= \{b_1(a_0b_1)^*, a_1b_1(a_0b_1)^*\} \end{aligned}$$

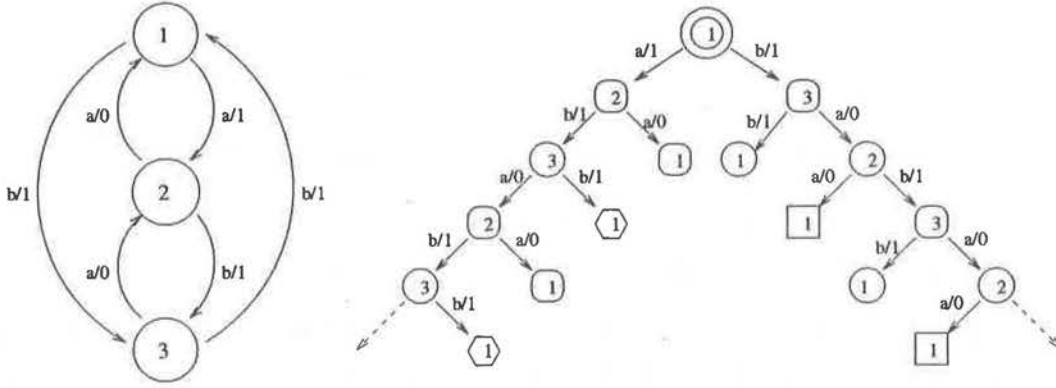


Figure 4: A FSM example and its transition tree

Trivially, a set $\{(\#_{a_1}(\omega), \#_{a_0}(\omega), \#_{b_1}(\omega)) | \omega \in (a_1, a_0, b_1)^*\}$ for all the input/output symbol sequences derived from the FSM model is semilinear. Thus this set can be represented as

$$\{\psi(\omega)\} = S_{State_1} \cup S_{State_2} \cup S_{State_3}$$

where each S_{State_i} is a finite union of linear sets as follows :

$$\begin{aligned} S_{State_1} &= \{(0, 0, 0) + \sum_{i=1}^N x_i(0, 0, 0)\} \cup \{(1, 1, 0) + \sum_{i=1}^N x_i(0, 1, 1)\} \cup \\ &\quad \{(0, 2, 1) + \sum_{i=1}^N x_i(0, 1, 1)\} \cup \{(0, 0, 2) + \sum_{i=1}^N x_i(0, 1, 1)\} \cup \\ &\quad \{(1, 0, 2) + \sum_{i=1}^N x_i(0, 1, 1)\} \\ S_{State_2} &= \{(1, 0, 0) + \sum_{i=1}^N x_i(0, 1, 1)\} \cup \{(0, 1, 1) + \sum_{i=1}^N x_i(0, 1, 1)\} \\ S_{State_3} &= \{(0, 0, 1) + \sum_{i=1}^N x_i(0, 1, 1)\} \cup \{(1, 0, 1) + \sum_{i=1}^N x_i(0, 1, 1)\} \end{aligned}$$

Here, if we set the parameter x_i to the minimal value 0, then each linear set stated above becomes a vector that is the Parikh mapping of a nonrecursive paths from an initial state to each state. That is, from each set $State_i$, we can get the sets of vectors by substituting 0 for x_i :

$$\begin{aligned} &\{(0, 0, 0), (1, 1, 0), (0, 2, 1), (0, 0, 2), (1, 0, 2)\} \text{ for the } State_1 \\ &\{(1, 0, 0), (0, 1, 1)\} \text{ for the } State_2 \\ &\{(0, 0, 1), (1, 0, 1)\} \text{ for the } State_3 \end{aligned}$$

This means that, neglecting the order of occurrence of input symbols, the candidates of the shortest paths from the initial state to $State_1$, $State_2$ and $State_3$, can be derived from the above vectors, respectively as follows :

$$\begin{aligned} &\{\lambda, a_1 a_0, a_0 a_0 b_1, b_1 b_1, a_1 b_1 b_1\} \text{ for the } State_1 \\ &\{a_1, a_0 b_1\} \text{ for the } State_2 \\ &\{b_1, a_1 b_1\} \text{ for the } State_3 \end{aligned}$$

Since these sets represent nonrecursive paths from the initial path to each state. The shortest paths from the initial state can be easily derived from these sets, as simply the shortest of those nonrecursive paths.

Example 3.2 OSI Transport Service [12]

The FSM model of the OSI Transport Service is given in Figure 5. Consider the global sequences of service primitives exchanged at both local and remote sides. The pairs of primitives, $ConReq$ and $ConInd$; $ConRsp$ and $ConCnf$; $DisReq$ and $DisInd$; and, $DataReq$ and $DataInd$; will be executed in sequence such that the first primitive (e.g. $ConReq$) executed at one side gives rise to the second primitive (e.g. $ConInd$) occurred at the remote side. Therefore, the set $\{\psi(\omega) | \omega \in \Sigma^*\}$ of Parikh mapping with respect to the valid global sequences of the OSI Transport service primitives (at both sides) can be represented in the general form :

$$\{(\#ConReq(\omega), \#ConInd(\omega), \#ConRsp(\omega), \#ConCnf(\omega), \#DisReq(\omega), \#DisInd(\omega), \#DataReq(\omega), \#DataInd(\omega))\} = \{(i, j, k, l, m, n, o, p) | i = j \geq 0 \text{ or } k = l \geq 0 \text{ or } m = n \geq 0 \text{ or } o = p \geq 0\}$$

In this case, the set $\{\psi(\omega)\}$ can be simply represented by a single linear set $\{v_0 + \sum_{i=1}^4 x_i v_i\}$, $v_0, v_i \in \mathbb{N}^8$, $x_i \in \mathbb{N}$, with the vectors v_0 through v_4 as follows :

$$\begin{aligned} v_0 &= (0, 0, 0, 0, 0, 0, 0, 0), \\ v_1 &= (1, 1, 0, 0, 0, 0, 0, 0), \quad v_2 = (0, 0, 1, 1, 0, 0, 0, 0), \\ v_3 &= (0, 0, 0, 0, 1, 1, 0, 0), \quad v_4 = (0, 0, 0, 0, 0, 0, 1, 1) \end{aligned}$$

The values of these vectors are uniquely decided by the conditions on the number of occur-

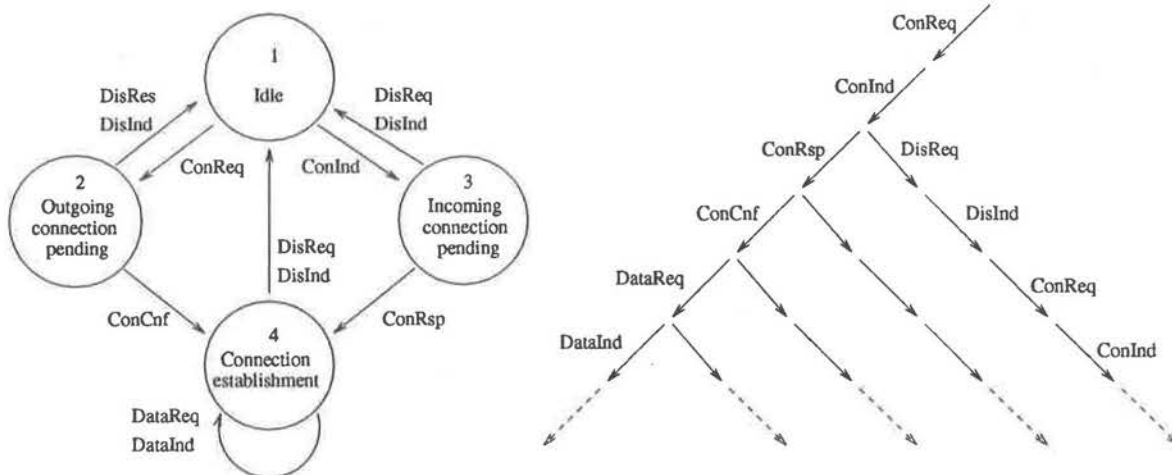


Figure 5: FSM model of the OSI Transport Service and the partial global transition tree

rences of each primitive. It is thus easy to determine these vectors. In general, we must determine $|\Sigma| + 1$ vectors.

4 Concluding remarks

In this paper, we present a framework, comprising two theoretical approaches to the problem of finite test selection from an infinite set of protocol behaviors (execution sequences). Both approaches are based on the notion of finite covering, which can be realized by the property of metric space compactness and the property of semilinear set in formal language, respectively. The metric based approach is abstract; although the ultimate theoretical solution is yet to be found, the theory developed so far has been demonstrated to be applicable to the practical problem of protocol test selection [6, 7]. The formal language based approach, on the other hand, is somewhat more intuitive, as illustrated in the simple protocol examples in the paper. The semilinear (i.e. "finite") representation of the Parikh mapping of execution sequences of protocols (which can be expressed as a context free language) opens door to the possibility of treating the problem of test selection from an infinite set of protocol behaviors in a finite context. The major research problem remains to define the coverage criteria and corresponding selection methods in the Parikh space so as to achieve finite coverings with optimal coverage measures.

It is hoped the framework presented in this paper provides enough insight to the theoret-

ical problem of finite covering of infinite set of protocol behaviors, which would encourage further work leading eventually to the complete solution to the problem of protocol test selection for conformance testing.

References

- [1] D.P.Sidhu and T.K.Leung, Formal Methods for Protocol Testing : A Detailed Study, *IEEE Trans. Software Eng.*, Vol.15, No.4, pp.413-426, April 1989.
- [2] S.T. Vuong, W.Y.L.Chan and M.R.Ito, The UIOv-method for protocol test sequence generation, *In Proceedings of the IWPTS II - Int. Workshop on Protocol Test Systems*, Berlin, 1989.
- [3] S.Fujiwara, G.V.Bochmann, F.Khendek, M.Amalou and A.Ghedamsi, Test Selection Based on Finite State Models, *IEEE Trans. Software Eng.*, Vol.17, No.6, pp.591-603, June 1991.
- [4] A.Petrenko and N.Yevtushenko, Test suite generation for a FSM with a given type of implementation errors, *Proc. 12th Int. Sympo. on Protocol Specification, Testing and Verification*, June 1992.
- [5] S.T. Vuong and J. Curgus, Test coverage metrics for communication protocols, *In Proceedings of the IWPTS IV - Int. Workshop on Protocol Test Systems*, Leinschendam, The Netherlands, 1991.
- [6] M.McAllister, S.T. Vuong and J. Curgus, Automated test case selection based on test coverage metrics, *In Proceedings of the IWPTS V - Int. Workshop on Protocol Test Systems*, Montreal, Canada, 1992.
- [7] J. Curgus and S.T. Vuong, A Metric Based Theory of Test Selection and Coverage, *In Proceedings of the PSTV XIII - Int. Symposium on Protocol Specification, Testing, and Verification*, Liege, Belgium, 1993.
- [8] I.Kaplansky, *Set Theory and Metric Spaces*, Allyn and Bacon, Inc., Boston, 1972.
- [9] S.Ginsburg, *The Mathematical Theory of Context Free Languages*, McGraw-hill, New York, 1966.
- [10] H.L.Royden, *Real Analysis*, Macmillan, New York, 1988.
- [11] M.Presburger, Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt, *Sprawozdaniez I Kongresu Matematykw Krajow Slowiaskich*, Warsaw, pp.92-101, 1930.
- [12] A.S.Tanenbaum, *Computer Networks*, Prentice Hall, Englewood Cliffs, 1989.